OXFORD MASTER SERIES IN CONDENSED MATTER PHYSICS

The Oxford Master Series in Condensed Matter Physics is designed for final year undergraduate and beginning graduate students in physics and related disciplines. It has been driven by a perceived gap in the literature today. While basic undergraduate condensed matter physics texts often show little or no connection with the huge explosion of research in condensed matter physics over the last two decades, more advanced and specialized texts tend to be rather daunting for students. In this series, all topics and their consequences are treated at a simple level, while pointers to recent developments are provided at various stages. The emphasis in on clear physical principles of symmetry, quantum mechanics, and electromagnetism which underlie the whole field. At the same time, the subjects are related to real measurements and to the experimental techniques and devices currently used by physicists in academe and industry.

Books in this series are written as course books, and include ample tutorial material, examples, illustrations, revision points, and problem sets. They can likewise be used as preparation for students starting a doctorate in condensed matter physics and related fields (e.g. in the fields of semiconductor devices, opto-electronic devices, or magnetic materials), or for recent graduates starting research in one of these fields in industry.

1. M. T. Dove: *Structure and dynamics: an atomic view of materials*
2. J. Singleton: *Band theory and electronic properties of solids*
3. A. M. Fox: *Optical properties of solids*
4. S. J. Blundell: *Magnetism in condensed matter*
5. J. F. Annett: *Superconductivity*
6. R. A. L. Jones: *Soft condensed matter*

Structure and Dynamics: An Atomic View of Materials

MARTIN T. DOVE

Department of Earth Sciences
University of Cambridge

OXFORD
UNIVERSITY PRESS

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

First published 2002

A catalogue record for this title
is available from the British Library

Library of Congress Cataloging in Publication Data

Dove, Martin T.
Structure and dynamics: an atomic view of materials/Martin T. Dove.
(Oxford master series in condensed matter physics)
Includes bibliographical references and index.
1. Solid–state physics. 2. Matter–Constitution. 3. Matter–Properties. I. Title.
II. Series.

QC176 .D69 2002 530.4'1–dc21 2002075712

ISBN 0 19 850677 5 (Hbk)
ISBN 0 19 850678 3 (Pbk)

10 9 8 7 6 5 4 3 2 1

Typeset using the author's TeX files by Cepha Imaging Pvt. Ltd.
Printed in Great Britain
on acid-free paper by The Bath Press, Avon

Preface

In the beginning you laid the foundations of the earth, and the heavens are the work of your hands.

Psalm 102:25

Objective of this book

This book is concerned with understanding how the structure of a material and the dynamics of its constituent atoms determine its properties and behaviour. From the outset I take a rather different approach from many standard textbooks in solid state physics. The weight of the content of most such books is concerned with the properties of the electrons, with topics such as electron distributions, electrical conductivity, semiconductors, superconductivity and ferromagnetism. When structure and dynamics are discussed, the examples are usually very simple, with one or two atoms in the unit cell. As vital as electrons are in understanding the behaviour of materials (as is made clear throughout this book) to simply focus on the electrons is to miss large areas of modern solid state physics than are both technologically important *and* intellectually stimulating. The point is that many important materials are more complex than the usual monatomic or diatomic crystals that find their way into solid state physics courses. The properties of these materials are mostly determined by the ways in which atoms are arranged, and through the ways in which these atoms respond to the forces between them. Of course, these forces arise from the electrons, but often in ways that can be understood using relatively simple mathematical representations. So this book is about *real* materials, the stuff of modern technology. And to understand the properties of these materials, we have to really understand what controls their structures and how the atoms move around inside them.

This book has been tailored for the purposes of the Oxford Master Series in Condensed Matter Physics in seeking to equip advanced level undergraduate or masters level students, doctoral students, and research level scientists in the study of materials. In writing this book I have taken the view that readers will be at a sufficiently advanced stage in their studies that they will want to understand enough to be able to use the techniques that are described, and this can only happen if the material is covered with a certain degree of rigour.

Use of this book

This book is arranged in two parts. The first (Chapters 1–7) is mostly concerned with the structural aspects. The opening chapter addresses the point that

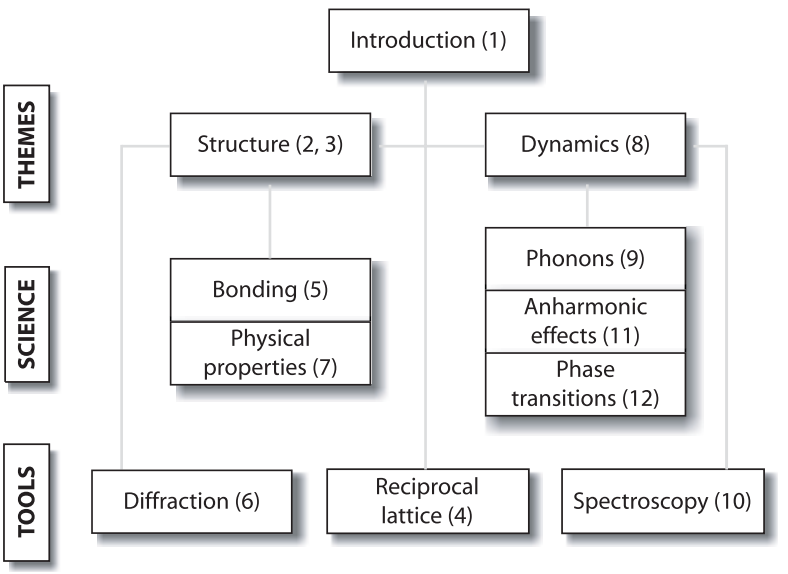
materials are made from atoms, and it shows that when armed with hindsight, and a little imagination, we can deduce properties about the atomic structure of materials from some common observations. This chapter is designed to be a light read. Chapter 2 then introduces a range of examples of **crystal structures**, both to show how wide the diversity of structures is, and also to highlight a number of important principles. At one level this chapter can be read as a guide book, picking some interesting examples to read in more detail, and to simply note the existence of other examples. Several important concepts are introduced through the examples without needing to develop the formalism in any detail. Chapter 3 picks up many of these concepts and draws them together in a basic formalism. One of the more important points in this regard is **symmetry**. Crystals are intrinsically symmetric, and their symmetry plays a very important role in determining their properties. To *properly* understand symmetry is a challenge that we should not shrink from, because it puts into our hands a tool of tremendous power and versatility. This chapter is likely to be one to which the reader will want to return later in the book. Chapter 4 delivers the second challenging concept, the **reciprocal lattice**, and again this gives us one of our more important tools with widespread applications. The reciprocal lattice is not a trivial concept, but neither is it gratuitously difficult, provided that it is tackled with an appropriate attitude. Some students feel that the reciprocal lattice is merely an invention of crystallographers to help interpret diffraction patterns, and if so it must be possible to ignore it! However, the fact that the reciprocal lattice is generated automatically, almost as a by-product, when tackling a diversity of issues such as lattice dynamics (Chapter 8) or electronic structure implies that the reciprocal lattice and reciprocal space are concepts that are as real as the space we inhabit. And therefore to achieve some familiarity with the reciprocal lattice is a worthy goal.

Chapters 3 and 4 equip us for the subjects of **bonding** (Chapter 5), **diffraction** (Chapter 6) and **physical properties of materials** (Chapter 7). Bonding is a vital topic because it unlocks many aspects of the properties of materials. Diffraction is covered in some depth because it is the key to being able to determine information about crystal structures from experiment. Physical properties are discussed in a wide sense because it is the fact that materials have a wide range of useful properties that gives the motivation for their study. In particular, it is in this chapter that we address the issue of anisotropy.

The second part of the book (Chapters 8–12) is primarily concerned with the dynamics of the atoms in a crystal. These are determined by the forces between atoms (hence Chapter 5), and we develop the basis of the **harmonic theory of lattice dynamics** in Chapter 8. One of the most important applications of this theory is to understand **thermodynamic properties**, which are covered in Chapter 9. Just as we covered the experimental methods to study structures in Chapter 6, we discuss the experimental methods used to study dynamics in Chapter 10. It might be said that the main emphasis of this chapter, namely **inelastic neutron scattering**, reflects the interests of the author (!), but it can be argued that the technique of **neutron scattering** requires a general understanding of general techniques, and that **Raman and infrared spectroscopy** can be understood as special cases of the general theory of spectroscopy. The theory of **harmonic lattice dynamics** gives an understanding of only some of the properties we might expect to be able to tackle, and for others (such as thermal

expansion) we need to develop the theory beyond the harmonic approximation into **anharmonic dynamics**. This is the area of Chapter 11. Among the most important aspects of material behaviour that can be understood by anharmonic lattice dynamics are **phase transitions**, which provide the subject area of Chapter 12. In fact, phase transitions link together many areas of this book: structure, symmetry, lattice dynamics, thermodynamic and physical properties, etc. One of the referees of an early draft thought that the whole book leads up to this chapter. This is not my intention, but since phase transitions have formed the core of my research since my graduate student days, it is probably inevitable that my excitement for this topic rather boils over at this point. I hope that an author should not need to apologise for such personal embellishments. Instead, I believe that the study of phase transitions can act as a fitting topic to bring together many of the various strands of this book.

The flow of the book is shown in the diagram below. A number of technical points and useful asides are covered in a series of appendices.



Any errors in this book which I discover after going to press, and any new information which may subsequently be of value, will be posted at http://www.esc.cam.ac.uk/astaff/dove/sd

Acknowledgements

I am very grateful to Sonke Adlung of Oxford University Press for inviting me to write this contribution to the series, and for his support and the support of Anja Tschoertner and Richard Lawrence. I am also grateful to a number of reviewers of my first draft, including Stephen Wells, Graeme Ackland and Artem Oganov, Mike Glazer, and two anonymous reviewers, for many useful comments and general encouragement.

I suspect that very few people can write a book without having to draw heavily on many influences. This book is surely a creation of my whole time in

the Mineral Physics group of the Department of Earth Sciences, University of Cambridge. I have enjoyed sharing in both research and teaching with many colleagues, amongst whom I include students as well as staff, and these interactions have shaped many aspects of my thinking. My debt to my colleagues in Cambridge is incalculable. Similarly, I was a student at both undergraduate and postgraduate level in Birmingham (Physics), then a post-doctoral worker in Edinburgh (Physics) and Cambridge (Theoretical Chemistry), and I am grateful for all that I have learned during my time in these different places.

There are, however, those who have helped in several specific ways in the production of this book, including providing data and pictures: Matt Tucker, Ming Zhang, Mark Calleja, Tony Abrahams, Archie Kirkland, Paul Midgely, David Keen, Mark Cooper, Mark Welch, Jeol Ltd, the Rutherford Appleton Laboratory and the Institute Laue-Langevin helped to provide me with data and figures for several parts of this book.

Finally, I am grateful to my wife Kate, and daughters Jennifer-Anne, Emma-Clare and Mary-Ellen, for all their support. I guess that by the time I came to write this book in earnest, they had long accepted the fact that family evenings could be spent with a laptop on my knee. I am extremely grateful for the combination of joy and encouragement that my four girls have given me.

Cambridge M.T.D.
December 2001

Contents

Table of symbols

We define the basic symbols used, giving the number of the equation where the symbol is first used or where it is defined.

a	one of the lattice parameters that gives the length of an edge of the unit cell
a	coefficient in Landau free energy function (12.8)
$\mathbf{a}$	one of the three basis lattice vectors (2.1)
$\mathbf{a}$	transformation matrix (7.5)
$\mathbf{a}^*$	one of the three reciprocal lattice vectors (4.4)
b	one of the lattice parameters that gives the length of an edge of the unit cell
b	coefficient in Landau free energy function (12.8)
b_i	neutron scattering length for atom of type i
$\mathbf{b}$	one of the three basis lattice vectors (2.1)
$\mathbf{b}^*$	one of the three reciprocal lattice vectors (4.5)
B	temperature factor (6.36)
B_{ij}	parameter in Born–Mayer potential (5.17)
c	one of the lattice parameters that gives the length of an edge of the unit cell
c	coefficient in Landau free energy function (12.19)
c	average velocity of sound (9.24)
c	heat capacity (1.6)
c_i	proportion of atom of type i
c_{ijkl}	component of elastic constant tensor (7.57)
c_n	parameter in linear combination of atomic orbitals (5.38)
c_V	heat capacity at constant volume
$\mathbf{c}$	one of the three basis lattice vectors (2.1)
$\mathbf{c}^*$	one of the three reciprocal lattice vectors (4.6)
C_{ij}	parameter in dispersive potential (5.23)
d	separation of core and shell in shell model potential (5.24)
d_{hkl}	spacing of hkl planes (6.2)
d_{ijk}	component of piezoectric tensor (7.38)
$\mathbf{d}$	piezoelectric tensor (7.38)
$\mathbf{d}^*$	vector between reciprocal lattice points (4.8)
$D(r)$	one of the formulations of the overall pair distribution function (J.13)
$\mathbf{D}$	dynamical matrix (8.57)
e_{ij}	component of strain tensor (7.15)
$e_{j,v}$	component of mode eigenvector (8.56)
$\mathbf{e}$	mode eigenvector (8.57)
E	modulus of the electric field (2.8)

E	energy (6.3)
E	energy transfer (10.9)
E	normalized structure factor (6.56)
E_{A}	activation energy (1.4)
$\mathbf{E}$	electric field vector (7.3)
f	x-ray atomic scattering factor (6.28)
F	Helmholtz free energy (5.3)
$F(\mathbf{Q})$	structure factor (6.21)
$g(k)$	distribution of wave vectors (9.26)
$g(\omega)$	frequency density of states (9.27)
$g_{ij}(r)$	pair distribution function for atoms of types i and j (2.9)
G	Gibbs free energy (5.1)
$G(r)$	one of the formulations of the overall pair distribution function (2.10)
$\mathbf{G}$	reciprocal lattice vector
h	one of the Miller indices (4.8)
$\mathbf{h}$	shorthand label for the reciprocal lattice vector (hkl) (6.46)
H	Enthalpy (5.2)
H	field conjugate to the order parameter (12.27)
$\mathcal{H}$	Hamiltonian (9.44)
$\hat{\mathcal{H}}$	Hamiltonian operator (5.33)
J	force constant (8.3)
J_{ph}	phonon flux (11.7)
k	one of the Miller indices (4.8)
k	parameter in the shell model potential (4.24)
k	shorthand notation for phonon wave vector and mode label (12.37)
k_{F}	Fermi wave vector
$\mathbf{k}$	wave vector
K	compressibility (11.13)
K	interatomic force constant (8.32)
l	one of the Miller indices (4.8)
$\mathcal{L}(\mathbf{r})$	Mathematical function to describe a lattice (3.8)
m	atomic mass
m	used to denote coordination number (5.19)
M	minimization residual (6.67)
$\mathbf{M}$	transformation matrix (4.13)
$n(\omega, T)$	Bose–Einstein distribution giving the number of phonons of angular frequency ω excited at a given temperature
N	number of atoms or unit cells in crystal
P	pressure
P	modulus of the dielectric polarization (2.8)
P_{c}	transition pressure
$\mathbf{P}$	dielectric polarization vector (7.3)
Q	modulus of the scattering vector (6.20)
Q_i	charge of atom of type i
$Q(\mathbf{k}, \nu)$	normal mode coordinate (9.33)
$\mathbf{Q}$	scattering vector (6.12)
r^+	radius of cation (2.3)
r^-	radius of anion (2.3)

r_{ij}	distance between atoms labelled i and j
$\mathbf{r}$	general vector, often used to denote position of atom or vector between two atoms (2.1)
R	agreement factor in crystal structure refinement (6.66)
$\mathbf{R}$	position of atomic nucleus (5.33)
$\mathcal{R}(\mathbf{k})$	Mathematical function to describe the reciprocal lattice (4.37)
s_{ijkl}	component of elastic compliance tensor (7.56)
S	entropy
$S(\mathbf{Q}, E)$	neutron scattering function (10.11)
t	time
T	temperature
T_c	Transition temperature (5.8)
$T(r)$	one of the formulations of the overall pair distribution function (J.16)
$\mathbf{t}_{[UVW]}$	lattice vector (3.1)
u	atomic displacement (1.10)
$u(\mathbf{r})$	periodic function in wave function (5.31)
$\langle u^2 \rangle$	mean square atomic displacement (1.10)
$\mathbf{u}$	vector atomic displacement (9.33)
U	one of the integer parameters to define a lattice vector (3.1)
U	internal energy (5.1)
U_Lattice	lattice energy (5.11)
V	one of the integer parameters to define a lattice vector (3.1)
V	volume
V_cell	volume of unit cell (3.2)
W	one of the integer parameters to define a lattice vector (3.1)
x	atomic fractional coordinate (2.1)
y	atomic fractional coordinate (2.1)
z	atomic fractional coordinate (2.1)
Z	partition function (9.4)
Z_i	integral charge of atomic nucleus (5.48)
α	coefficient of linear thermal expansion (1.5)
α	one of the lattice parameters, giving the angle between the **b** and **c** lattice vectors
α^*	angle between the $\mathbf{b}^*$ and $\mathbf{c}^*$ reciprocal lattice vectors
$\alpha^{(n)}$	n-th order anharmonic coefficient (11.25)
β	$1/k_\mathrm{B}T$ (9.3)
β	coefficient of volume thermal expansion (11.12)
β	one of the lattice parameters, giving the angle between the **a** and **c** lattice vectors
β^*	angle between the $\mathbf{a}^*$ and $\mathbf{c}^*$ reciprocal lattice vectors
γ	Grüneisen parameter (11.19)
γ	one of the lattice parameters, giving the angle between the **a** and **b** lattice vectors
γ^*	angle between the $\mathbf{a}^*$ and $\mathbf{b}^*$ reciprocal lattice vectors
$\delta(\mathbf{R})$	Dirac delta function
$\Delta(\mathbf{K})$	function that is only non-zero if **K** is a reciprocal lattice vector (11.3)

ϵ	energy of elementary excitation, usually used for $\hbar\omega$ (9.4)
ϵ_{F}	Fermi energy
$\boldsymbol{\epsilon}$	symmetric strain tensor (7.18)
η	order parameter (12.4)
θ_{hkl}	Bragg angle (6.2)
κ_2	cofficient in crystal potential energy function (12.39)
κ_4	cofficient in crystal potential energy function (12.39)
λ	wavelength
λ	mean free path length (11.4)
ν	used to label phonon branches
ρ	density
ρ	electron density (6.40)
ρ_{ij}	parameter in Born–Mayer potential (5.17)
σ	electrical conductivity (1.4)
σ	stress (7.38)
τ	lifetime (11.5)
ϕ	interatomic potential energy function, can be function of interatomic separation, core–shell separation, or bond angle (5.11)
ϕ	atomic wave function (5.38)
χ	dielectric susceptibility (2.8)
$\boldsymbol{\chi}$	susceptibility tensor (7.4)
ψ	wave function (5.31)
Ψ	overall wave function (5.54)
ω	angular frequency

Introduction

1

1.1 Observations

1.1.1 Microscopic and macroscopic properties of solids

One of the central themes of this book is the way that macroscopic properties of solids, whether physical or thermodynamic, are linked to the underlying arrangement of atoms and the forces between the atoms. Several of the *physics* aspects of these relationships can be appreciated from simple observations of crystals, which we consider in this opening chapter.

One of the most important physics principles that will emerge from our survey of simple observations is the role of symmetry. We will find throughout this book that symmetry underpins the form of many physical properties. The observations of thermodynamic properties will point to the important role of quantum mechanics, particularly as applied to the dynamics of the atoms within crystals. A third important feature is that of length scales. Simple observations allow us to appreciate that the forces between crystals operate over a much shorter length scale than the properties they determine. The interplay between different length scales is an important aspect of the physics of solids.

1.1.2 Physical characteristics of crystals

The shapes of crystals give an immediate clue to the underlying atomic arrangement within a crystal. One classic example is the crystal of calcite, shown in Fig. 1.1. Natural crystals of calcite have a characteristic rhomboid shape. When the crystal is cleaved, the shapes of the small chippings have an identical shape. The similarity of the shapes of crystals of calcite over all visible length scales suggests that there is an underlying microscopic object with a similar shape. The large crystal would then be composed of the underlying objects stuck together in a regular array. The same line of reasoning suggests that there must be forces between the underlying objects, which will be strong only over short distances since once we have chipped off a small crystal it cannot be stuck back onto the larger crystal.

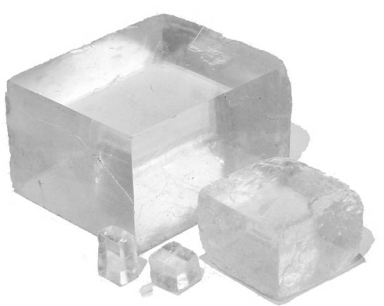


Fig. 1.1 Crystals of calcite, $CaCO_3$, showing that small bits chipped from the larger crystal have the same shape as the large crystal. This suggests that the basic shape is determined by an underlying microscopic building unit. The vertices pointing out of the photograph are symmetric with respect to rotations of 120°.

Careful inspection of the small crystals of calcite in Fig. 1.1 show some of the symmetry aspects of the crystal. The three edges at one of the apices subtend three equal angles. This suggests that the underlying atomic arrangement could be symmetrical with respect to rotations of 120°.

Studies of the shapes of crystals were among the first quantitative studies of the behaviour of materials. It was noted that there were many different

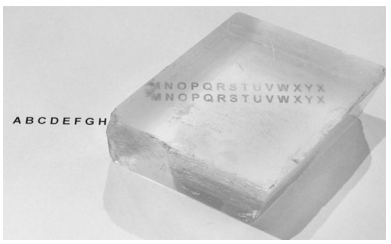


Fig. 1.2 Double refraction of calcite. The light passing through the crystal is split into two separate beams that are refracted differently, leading to two images transmitted through the crystal. The way that the double refraction changes as the crystal is rotated points to the presence of special axes within the crystal.

symmetry characteristics of the shapes of crystals. Some crystals form cube shapes, others hexagonal prisms. In fact the range of shapes is diverse, but by looking for particular elements of the shapes, such as angles between faces and relationships between the orientations of faces, it was possible to separate all the shapes into particular groups based on symmetry relations. Moreover, it was noted that some crystals can be cleaved in specific ways to give sets of faces with definite orientational relationships. These relationships could be quantified, and as a result it was possible to deduce the basic shape of the fundamental building blocks of many common materials.

Crystals of calcite show an interesting optical effect which is also likely to be linked to the underlying atomic structure. When an object is viewed through a natural crystal of calcite, two images are seen, which rotate around each other as the crystal is rotated. An example is shown in Fig. 1.2. This observation highlights one of the themes that will be developed later in this book, namely that in many crystals the underlying atomic structure leads to **anisotropy** in the properties of the macroscopic crystal (that is, the behaviour of a material depends on its orientation), and this anisotropy is crucial for many technological applications.

1.1.3 Physical properties of crystals

We are able to subject crystals to various types of stimuli and observe changes in their structures and hence behaviour and properties. How the crystals respond will give some indications of the underlying atomic structure of the crystal.

Optical properties

One of these issues is the response of the crystal to an incident light beam. We have just met a qualitative example in the case of calcite. To make the discussion more quantitative, consider a very simple experiment as shown in Fig. 1.3. A crystal is held between two crossed polars. A light beam is incident from one side and viewed from the other. This beam is polarized in one direction by the first polaroid sheet. If the crystal does nothing to the light beam, no light will pass through the second polaroid because the direction of polarization is at 90° to the second polaroid sheet. Indeed, many crystals have no effect. However, there are other crystals that do not behave in this way, and actually allow some of the light to pass through both polaroids. In these cases, the amount of light that is passed can be changed by rotating the crystal about the direction of the light beam. From experiment it is found that when rotating the crystal about the direction of the light beam, there are two orientations of the crystal for which no light will pass through the two polaroids, and rotation of the crystal from these orientations will give an increase in the intensity of the transmitted light up to some maximum.

If light can be transmitted, it suggests that the crystal is able to rotate the plane of polarization of the light. Furthermore, the amount of rotation depends on the orientation of the crystal, because for two orientations there is no rotation. It is known that light travels through solid matter with a velocity that is slower than in vacuum. We might suppose that in these crystals the velocity of light will depend on the orientation of its polarization. So we consider a light beam

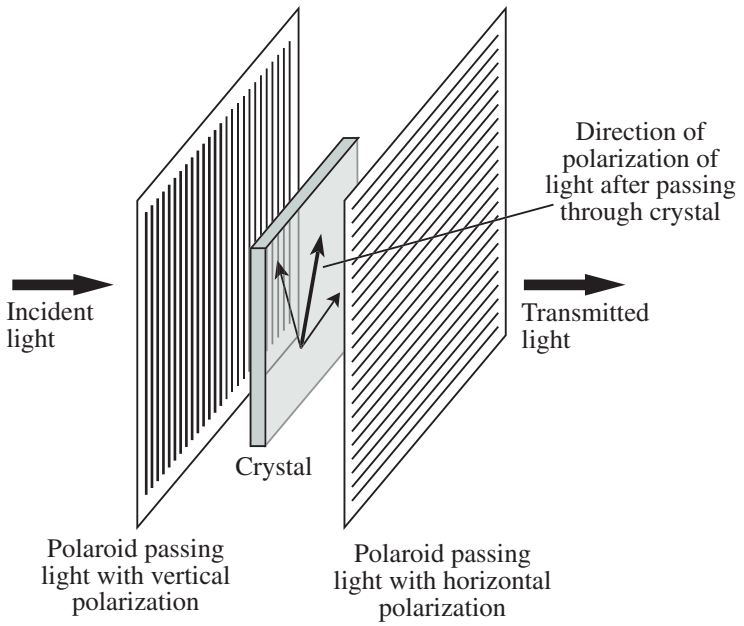


Fig. 1.3 Simple experiment in which a crystal is held between two polaroids oriented at 90° with respect to each other. The first polaroid transmits only the component of the light with polarization in the vertical direction. If none of the special axes of the crystal are aligned parallel to the polarization of the light, the crystal will effectively rotate the direction of the polarization of the transmitted light, allowing some intensity to pass through the second polaroid, which only transmits the component of horizontal polarization.

polarized in a direction that is at an angle θ with respect to one of the special axes. We can then split this beam into two components, each with polarization parallel to the two special axes. We assume that both components travel through the crystal with different velocities. The amplitude of the light beam in the two crystal directions is $A\cos\theta$ and $A\sin\theta$, where A is the amplitude of the incident beam. The velocity of light for the beam polarized along the two special directions is c_1 and c_2 respectively. Denoting the angular frequency as ω, the wavelengths of the two light beams will be $\lambda_1 = 2\pi c_1/\omega$ and $\lambda_2 = 2\pi c_2/\omega$ respectively. If the crystal has thickness d, the wave equations for the two beams after passing through the crystal, which we denote as i_1 and i_2, will be

$$i_1 = A\cos\theta\sin(2\pi d/\lambda_1 - \omega t)$$
$$i_2 = A\sin\theta\sin(2\pi d/\lambda_2 - \omega t)$$

Because of the different velocities along the two axes, the two components are now slightly out of phase. We can represent this phase difference by the angle δ:

$$\delta = 2\pi d\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right) \tag{1.1}$$

Thus we can write

$$i_2 = A\sin\theta\sin(2\pi d/\lambda_1 + \delta - \omega t)$$

The second polaroid selects and adds the components of i_1 and i_2 that lie at an angle of 90° to the first polaroid:

$$\begin{aligned} i' &= i_1\sin\theta - i_2\cos\theta \\ &= A\cos\theta\sin\theta[\sin(2\pi d/\lambda_1 - \omega t) - \sin(2\pi d/\lambda_1 + \delta - \omega t)] \\ &= A\sin 2\theta\cos(2\pi d/\lambda_1 + \delta/2 - \omega t)\sin(\delta/2) \end{aligned} \tag{1.2}$$

The overall intensity of the transmitted light beam is the time-averaged square:

$$I = A^2 \sin^2 2\theta \sin^2(\delta/2) \tag{1.3}$$

This depends on the orientation of the crystal axes with the polaroids through the $\sin^2 2\theta$ term. The important point is that the light can only be transmitted through the second polaroid because the crystal's optical properties depend on the orientation of the electric field with respect to the crystal axes.

The important principle illustrated here is that of crystal **anisotropy**. It is important because we have to understand why some crystals are anisotropic and some are not, but it is also important because anisotropy can be exploited for a number of technological applications (as we will explore in later chapters, particularly Chapter 7). Anisotropy is connected with symmetry, which we will study in Chapters 2 and 3. For example, in calcite, where we have a special axis at the meeting of three identical types of face, the crystal appears isotropic when viewed down this axis but is anisotropic when viewed at right angles to this axis. Many materials have this property, and are called **biaxial**. The existence of such materials implies that the underlying symmetry of the arrangement of atoms in the crystal has a direct effect on the symmetry of the physical properties of the crystal.

Dielectric and electrical properties

When an electric field is applied to a solid, several things can happen. For a metal, the electric field will generate an electric current, which is associated with the flow of nearly-free electrons. One of the successes of solid state physics, following from the development of quantum physics, is an understanding of how the electrical conductivity is determined by the distribution of electron energies. In fact, one can develop a general understanding of the electrical conductivity of metals without needing to know too much about the arrangement of atoms within the metal, since it is dominated by the contribution of the nearly-free electrons and the quantum mechanical distribution of electron energies and wave vectors. This topic is covered in many standard textbooks on condensed matter physics (see Further reading at the end of the chapter).

On the other hand, the response of a non-metallic solid (leaving aside the issue of semiconductors) to an applied electric field is rather different. In this case, an applied electric field will cause two main responses. The *first* response will be the generation of an electrical current, albeit one that is *much* weaker than if the material was a metal. This could imply that there are still some electrons to conduct electricity, but other experimental observations would seem to rule out this explanation. For example, the dependence on temperature T of the electrical conductivity, σ, of a material that is nominally an insulator often has the form

$$\sigma \propto \exp(-E_A/RT) \tag{1.4}$$

where E_A is an **activation energy**, and R is the gas constant ($R = 8.314\ \mathrm{J\,mol^{-1}\,K^{-1}}$). This temperature dependence implies that whatever is moving in the crystal is doing so in a way that depends on the thermal activation of the motion across some energy barrier characterized by E_A (see Appendix A).

In some respects the electrical conductivity of a non-metallic crystal is similar to that of an ionic solution. In a solution, application of an electric field moves cations to one electrode and anions to the other, which can clearly be seen through a coating of the electrodes that develops with time. This analogy leads to the conclusion that the ions have some degree of mobility in a crystal that can be thermally activated. Ionic mobility is another feature of crystalline materials that has important areas of application.

The *second* response of a non-metallic crystal to the application of an electric field is to generate a **dielectric polarization**. This has to be associated with the electric field causing slight displacements of the positive and negative charges in opposite directions.

The dielectric polarization can be pictured in a very simple way, Fig. 1.4. The equilibrium positions of the ions within the crystal are determined by the forces between them, such that the equilibrium positions are those for which all forces are perfectly balanced. When they are displaced by the application of an electric field, the balance is lost, and the interactions between ions in the crystal create new forces that eventually balance the force from the applied electric field. How large a dielectric polarization can be produced is determined by the size of forces acting on each ion.

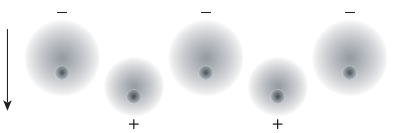


Fig. 1.4 Mechanisms of dielectric polarization in response to an applied electric filed indicated by the arrow. The cations and anions (marked with the + and − signs respectively) move in opposite directions. There is also a polarization of the ions, represented by the displacements of the electron clouds relative to the cores of the ions, the more so for the anions.

Both the ionic conductivity and dielectric polarization can be measured. Most experiments are carried out with electric fields that vary sinusoidally with time, with an angular frequency ω. For values of ω up to a certain point the amplitude of the induced dielectric polarization will be more-or-less constant for a constant amplitude of the applied field. However, at a certain value of ω, labelled ω_0, the amplitude of the induced polarization will fall, Fig. 1.5. It is as if something in the material cannot respond fast enough above a critical frequency. Some of the induced dielectric polarization remains, and this implies that there are two components to the polarization at low frequencies. A sensible explanation is that the two components arise from the structure of the atoms shown in Fig. 1.4. The component that only applies at the lower frequencies would be associated with the movement of the atomic nucleus and the tightly bound inner electrons, and the component that operates at both low and high frequencies would be associated with the outer electrons in the atom. Because they are so light, the outer electrons are able to relax instantaneously and therefore are not affected by the frequency of the applied electric field. On the other hand the nucleus has a much higher mass (by a factor of around 10^4), and therefore there is a more noticeable inertia associated with its motion. Once the frequency of the applied electric field is larger than ω_0, the mass component of the atom is unable to respond quickly enough and therefore does not play a role in the formation of a dielectric polarization. If this mechanism is correct, we might expect that the value of ω_0 for a material will be virtually independent of the direction of the applied electric field, which it is.

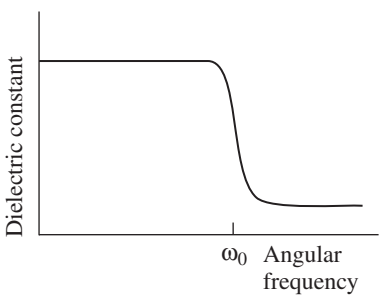


Fig. 1.5 Schematic plot of the dependence of the dielectric constant on the angular frequency of the applied field, ω. There is a marked decline at a particular frequency, ω_0, beyond which the atoms cannot move fast enough. In fact the behaviour for $\omega < \omega_0$ in this schematic plot is very oversimplified, and there may be other steps associated with other mechanisms. One such mechanism that is easily observed in frequency-dependent dielectric constant measurements arises from the motions of defects, as discussed in Appendix A.

For different crystals, the dielectric response and ionic conductivity will be either isotropic or anisotropic, as for the optical properties. The symmetry of the arrangements of ions has an important role to play in the response of a crystal to an electric field. It also has to be noted that some materials are electrically polarized even in the absence of an electric field. Since polarization is associated with the separation of the centres of the distribution of positive and negative charges, the existence of a dielectric polarization in the absence of an electric

field suggests that in these materials the atomic structure is such that there is a polarization of the charges within the basic atomic building block of the crystal.

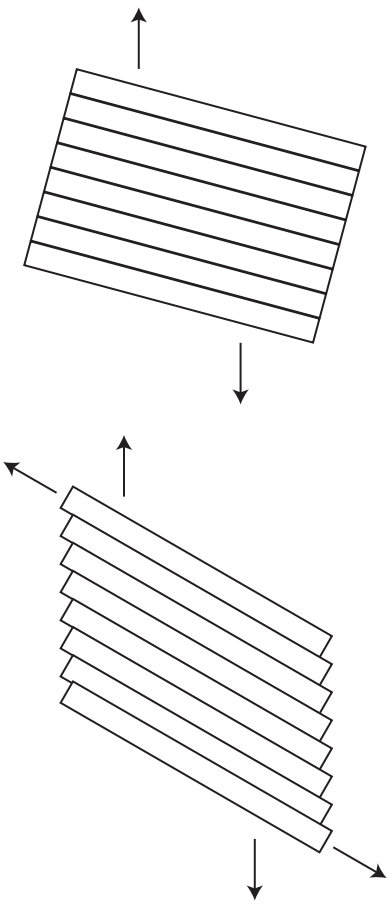


Fig. 1.6 Deformation of a single crystal by the slip of planes. The vertical arrows indicate the directions of the applied forces, and the arrows parallel to the crystal planes indicate the relative displacements of the planes in the process of slip.

Response of crystal to application of stress

Application of an external mechanical force to a material, which, when normalized against the area of the face on which it is applied, is called the **stress**, can produce several responses. The usual response is to change the size and shape of the material. Up to a certain size of force, this change will be **elastic**, by which we mean (in this context at least – we will see in this book how the word 'elastic' has several meanings in modern physics usage!) that on removal of the force the crystal will revert back to its initial size and shape. This implies the existence of a restoring force. Instantaneous deformation of the crystal at a macroscopic level must involve deformation at the atomic level. This will involve atoms moving closer together or further apart as they are displaced from their equilibrium positions. The macroscopic restoring force has to be related to the forces between the atoms. How much the size of the crystal can be changed on application of a mechanical stress gives an insight into the extent to which atoms can be squeezed together. Clearly, if atoms can be squeezed together, they cannot have the form of perfectly hard spheres (like pool or billiard balls appear to the players) permanently in contact with their neighbouring atoms, but they must be reasonably hard because scientists have to work very hard to achieve compressions of a few percent.

Forces normal to a crystal face can be applied in two senses, either in compression or in tension. When the size of a tension force applied along one direction (that is, a uniaxial tensile stress) reaches a certain level, the crystal will deform in a non-reversible way. One way appears to involve slippage of slivers of the crystal, as shown in Fig. 1.6. This produces terracing of the external faces of the crystal in a very characteristic manner. One can presume that this effect is related to the existence of discrete building blocks of the crystal, as illustrated in Fig. 1.6. This effect is found to be related to the orientation of the direction of the tensile force to the orientations of certain special directions, which will have some relationship (although not being the same as) the special directions we noted when discussing the optical effects of crystals.

Another non-reversible reaction of a crystal that can be produced by application of a uniaxial tensile stress is **fracture**. Like the process of slip, fracture has to be accompanied by the breaking of bonds between atoms. In principle, the force required to deform (whether by slip or fracture) a crystal might be directly related to the size of the forces between atoms. However, an early observation was made that measured fracture stresses are much too weak to be consistent with such a direct relationship. The solution to this problem, which then facilitated explanation of other phenomena, was found to require crystal imperfections – called **defects** – to be brought into the overall picture. These are outlined in Appendix A.

Deformation of the crystal is not the only possible response to stress. For some materials it is found that application of a stress will produce a dielectric polarization. This is called the **piezoelectric effect**, and is discussed in more detail in Chapter 7. This response is perhaps not surprising, since one might expect that the rearrangement of atoms due to an applied stress might produce a rearrangement of the charge distribution within the crystal. A little thought will

suggest that the generation of a dielectric polarization has to be related to the underlying symmetry of the atomic arrangement and the effect that the stress has on this symmetry. Not surprisingly, this symmetry property is related to the other symmetry properties already mentioned in this chapter.

Thermal expansion

The size of all materials will change on heating (aspects of thermal expansion are discussed in Chapters 7 and 11). Usually materials expand on heating, although we will note in later chapters that there are some examples of materials that actually shrink as the temperature is raised. Not surprisingly, the way that crystals change their size with temperature will depend on the same symmetry that determines the other physical properties. The rate of thermal expansion may be different along different directions, and these directions can be related to the other special directions noted above.

An example is calcite, whose optical properties were noted earlier. The coefficient of thermal expansion is related to the rate of change of a linear dimension, l, with temperature:

$$\alpha = \frac{1}{l}\frac{\partial l}{\partial T} \tag{1.5}$$

At room temperature, the coefficient of thermal expansion along the special symmetry direction of calcite, which we denote as α_3, is quite large, with value $3 \times 10^{-5}\,\mathrm{K}^{-1}$ (the reason why this value is relatively large has its origin in the subject matter of Chapter 12). On the other hand, the coefficient of thermal expansion along any direction normal to the special axis is small and negative, $\alpha_1 = -4.7 \times 10^{-6}\,\mathrm{K}^{-1}$.

The point that emerges from the observation of thermal expansion is that temperature has an effect on the properties of materials. In classical kinetic theories, temperature is related to atoms moving with a distribution of velocities. In a solid, where the atoms vibrate around fixed positions, the distances between neighbouring atoms will increase and decrease due to the thermal motions. If the size of the forces induced by the changing interatomic distance is the same for increasing or decreasing the distance, one might expect that the effects of the motion will not impinge on the crystal structure and hence on the physical properties. However, if the forces are stronger for atoms getting closer together than on moving apart, one might expect temperature to have an effect across all properties of the crystal. As temperature increases, the atoms will move faster and hence further apart, and if the energy favours increasing separation rather than decreasing separation, the distribution of atom separations will move towards increasing separation. This point is illustrated in Fig. 1.7, which shows the basic shape of the potential energy function between neighbouring atoms that is consistent with this simple analysis. This gives an explanation for the expansion of a material on heating, since it suggests an expansion of atomic separations on heating. The fact that there are some materials that have **negative thermal expansion**, not enough to blow a hole in the argument but enough to be interesting, suggests that the way in which the expansion of interatomic distances due to thermal motion is translated into the change in the size of the whole crystal has to involve some aspects of how the atoms are arranged with respect to each other.

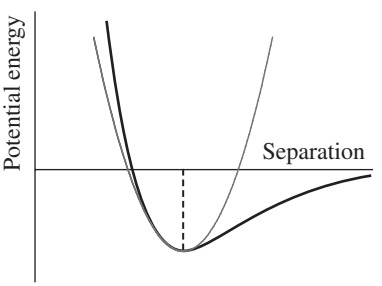


Fig. 1.7 Dependence of potential energy on the separation between a cation and anion (dark curve), compared with a symmetric potential (light curve). The steep rise at low separations arises from the repulsive interaction which prevents the ions from becoming too close. The slow tail at large separations is the residual attractive interaction. The two curves overlap around the minimum in the potential energy, denoted by the dashed line. The form of this plot will be explored in Chapter 5.

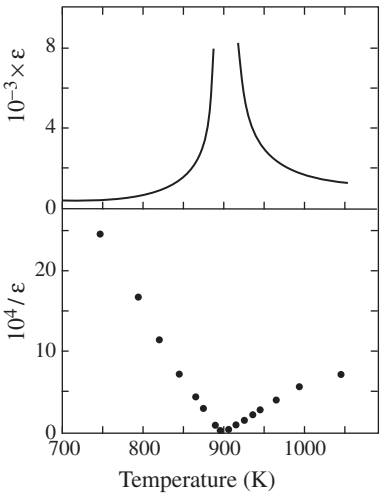


Fig. 1.8 Variation of the dielectric constant and its inverse in $LiTaO_3$ with temperature. The dramatic increase of the dielectric constant at temperatures around 900 K is due to the existence of a ferroelectric phase transition, see Section 12.1.2.

If changing temperature changes the atomic structure of a crystal, one might expect it to change other properties also. And indeed it does. For example, the dielectric response will vary with temperature, and in some cases dramatically so. This again reinforces the idea that there is an intrinsic link between the atomic structure of a material and its physical characteristics and properties.

It is worth exploring the idea that some physical properties can change dramatically with temperature. In some materials it is found that the dielectric polarization induced by an applied electric field gets larger and larger on heating or cooling towards a fixed temperature. Specifically, the coefficient linking polarization and electric field is the dielectric constant ϵ, and an example of the dramatic variation of the dielectric constant with temperature is shown in Fig. 1.8. Other measurements show that the optical properties change at this temperature, and that on cooling below this fixed temperature the crystal has a dielectric polarization even in the absence of an electric field. Since we have related physical properties to an underlying symmetry of the atomic arrangement, these observations suggest that changing temperature can also lead to a change in symmetry. Symmetry is not like a continuous variable; it cannot be changed gradually, only in a step-wise fashion. This is why the symmetry is changed at a fixed temperature. We call this process a **phase transition**. Because the change in symmetry is associated with large changes in the physical properties (Chapter 7), materials with phase transitions have many uniquely useful technological applications. Phase transitions are discussed in some detail in Chapter 12.

1.1.4 Thermodynamic properties of crystals

Bringing temperature into the picture opens up the whole realm of thermodynamics, and there are particular aspects of the thermodynamics of solids that give rise to new insights. The thermodynamic property of a solid that is most accessible to experiment is the heat capacity. This is defined as the amount of energy that is required to produce a given change in temperature, written mathematically as

$$c = \frac{\mathrm{d}E}{\mathrm{d}T} \tag{1.6}$$

From classical kinetic theory we have the concept of equipartition, by which all atoms have the same *average* kinetic energy, which is given as

$$\langle \mathrm{KE} \rangle = \tfrac{3}{2} k_\mathrm{B} T \tag{1.7}$$

We suppose that atoms in the crystal are held in place to some extent (were this not so we would have a liquid or gas), and that the forces are likely to be **harmonic**, by which we mean that the forces are proportional to the atomic displacements (in making this guess we are actually presupposing that even if the forces are more complex, the harmonic term will emerge as the first term in a Taylor expansion of the crystal energy, as we will investigate in Chapter 8). Harmonic forces give simple harmonic motion, in which there is a constant exchange of energy between kinetic and potential. In fact the bit of the potential

energy that can vary will have a mean value that is equal to the mean value of the kinetic energy, so the mean energy of one mole of atoms can be supposed to be equal to

$$\langle \mathrm{E} \rangle = \langle \mathrm{KE} \rangle + \langle \mathrm{PE} \rangle = 3RT \tag{1.8}$$

From this equation we can calculate the heat capacity *per mole of atoms*:

$$c = \frac{\mathrm{d}}{\mathrm{d}T} 3RT = 3R \tag{1.9}$$

This value, equal to $24.94\,\mathrm{J\,mol^{-1}\,K^{-1}}$ is found for many solids at high temperatures, but on cooling the heat capacity always falls, eventually dropping to zero at 0 K (Chapter 9). The temperature dependence of the heat capacity of NaCl is shown in Fig. 1.9. The observation of the classical equipartition value of the heat capacity at high temperatures reinforces our view that we have to take into account the behaviour of the atoms, and in particular their dynamics, in order to understand the thermodynamic properties of solids. However, the failure of the classical equipartition model to even hint at an explanation for the behaviour of the heat capacity at low temperatures points to the need for something new from physics. This, as you might guess, is the need to incorporate quantum mechanics into the atomic description of atomic motions. Similarly, the temperature dependence of many properties of a material (for example, crystal volume) is found to be different at low temperatures, where the rate of change usually falls to zero. The role of quantum mechanics in determining physical properties is discussed in Chapters 9 and 11.

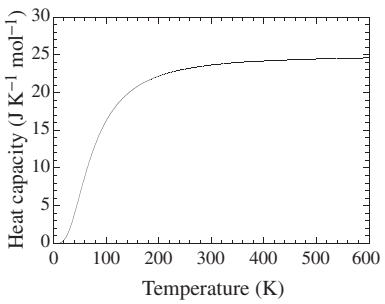


Fig. 1.9 Variation of the heat capacity of NaCl with temperature, in terms of per mole of atoms. The classical theory outlined in the text would predict a value of $3R = 24.9\,\mathrm{J\,mol^{-1}\,K^{-1}}$, to which the heat capacity tends at high temperatures. The effect of temperature on the heat capacity is discussed in Chapter 9.

In a fluid, the positions of the atoms are continuously changing with time (by a process called **diffusion**). However, in a crystal the average positions are more likely to be fixed, and the thermal motions will involve vibrations of the atoms about their mean positions rather than large-scale diffusion (not to disregard the fact that the atoms in a liquid vibrate within the cage formed by the instantaneous neighbouring atoms). The existence of vibrations suggests that there are associated frequencies. One might expect that since there are many atoms, there will also be many frequencies, so many in fact that one is forced into thinking of a distribution of frequencies. On the other hand, if there is symmetry and periodicity due to the stacking of some basic atomic building blocks, it may be that there are instead discrete frequencies. This latter viewpoint might be backed up by certain experiments. The easiest to picture is to measure the absorption of electromagnetic radiation (Section 10.6). Since electromagnetic radiation is oscillatory, one might expect that if the frequency of the radiation matches a particular vibration frequency it will induce vibrational motions of the atoms as a resonance. The most promising candidate for linking absorption of electromagnetic radiation to atomic vibrations occurs for infrared (IR) radiation. Experiments of IR absorption point to the existence of vibrational frequencies of order 10^{12}–10^{14} Hz (1 THz is 10^{12} Hz), as shown in Fig. 1.10 (data on vibrational frequencies for a range of materials are given in Chapter 8).

We can now match these frequencies to the observation of temperatures at which quantum mechanical properties come into play. We can equate an energy of a quantum of vibrational energy with $E_{\mathrm{quantum}} = \hbar\omega$, where $\hbar$ is Planck's constant divided by 2π (value $\hbar = 1.0 \times 10^{-34}\,\mathrm{J\,m^{-1}}$), and ω is the

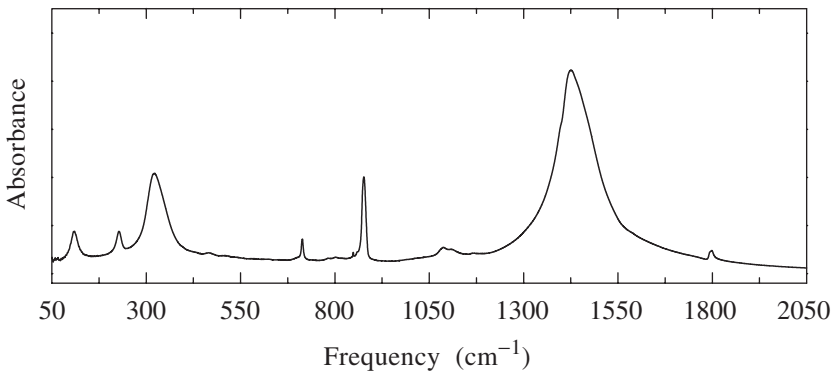


Fig. 1.10 Infrared absorption spectrum of a polycrystalline of calcite (data courtesy of Dr Ming Zhang, University of Cambridge). The technique of infrared absorption spectroscopy is discussed in Section 10.6.

angular frequency of the vibration. The classical energy is $k_B T$. One might expect quantum mechanics to take over from classical thermodynamics when $\hbar\omega \sim k_B T$ (as we will show in Chapter 9). Assuming an average frequency of $\omega = 2\pi \times 5 \times 10^{12}\,\text{rad}\,\text{s}^{-1}$ (which corresponds to $167\,\text{cm}^{-1}$; this is low on the scale of measurements for calcite shown in Fig. 1.10, but is typical of average frequency values in alkali halides for example), the cross-over temperature $T \sim \hbar\omega/k_B$ is calculated to be ~ 230 K. This is within the range of temperatures at which quantum mechanics effects are seen to come into play in experimental studies. This result can be compared with the temperature at which the heat capacity of NaCl falls fastest in Fig. 1.9.

One other important thermodynamic property of many solids is that they can change phase on changing temperature. Well, all materials will melt or sublime at a certain temperature, but other materials will also change from one solid phase to another in a way that can be detected through changes in physical characteristics or physical properties, as well as through appropriate microscopic experiments. Melting suggests that atoms eventually gain enough kinetic energy to be able to break away from the forces that constrain them to form a solid. However, the density of a liquid at a fixed temperature and pressure will be constant, and is often not significantly different from that of the crystal, and this suggests that the forces between atoms still play an important confining role, even if they are not able to preserve the rigidity of the solid. The existence of changes of phase whilst still within the solid state implies that atoms can gain enough kinetic energy to force a rearrangement of the atoms within the solid. This suggests that atoms have some degree of mobility within the solid state.

1.1.5 So what do some simple observations tell us?

We can now pull all these observations together to paint an impression of the fundamental nature of crystalline materials. Crystals appear to be made of small building blocks involving arrangements of atoms. Microscopic experiments, such as the diffraction experiments described in Chapter 6, point to the sizes of atoms being of order 10^{-10} m, which is given the unit 1 Å. These building blocks stack together in a periodic arrangement, like bricks in a wall or tiles on the floor. The form of this arrangement will have a particular symmetry that

is determined by the specific arrangement of atoms in the basic building block (discussed in Chapter 3), and the form of this symmetry will have an impact on the physical characteristic and properties of the crystal (Chapter 7).

The atoms are held in position by mutual forces (Chapter 5). These can be overcome at high temperature, as seen in melting. External forces can be applied in the form of mechanical stresses or applied electric fields, which cause small microscopic distortions of the atomic structure. Moreover, temperature will cause local vibrational motions of atoms (Chapter 8, Appendix N). We can estimate the size of the associated atomic motions by using an analogy to the vibrations of a mass held in place by springs. The energy associated with the displacement u of the mass is $m\omega^2\langle u^2\rangle$. Equating the mean displacement energy with $k_\mathrm{B}T$ gives

$$\langle u^2\rangle = \frac{k_\mathrm{B}T}{m\omega^2} \tag{1.10}$$

Using some typical numbers, $m = 0.05/N_\mathrm{A}$ kg, $\omega/2\pi = 5 \times 10^{12}$ THz, we have $\langle u^2\rangle^{1/2} \sim 0.07$ Å at a temperature of 300 K. If the typical spacing between atoms is 1.5 Å, the mean square displacement is around 5% of the mean atomic spacing. This is on a similar scale to the typical size of the distortion that can be achieved by application of mechanical stress on an electric field. For example, a typical size of the dielectric constant of an ionic crystal is 10 (a dimensionless quantity). In a crystal containing charges of ± 1 electron, an application of an electric field of $10^6\,\mathrm{V\,m^{-1}}$ will give displacements of ions of 1.4×10^{-4} Å. Whilst this is somewhat smaller than the amplitude of thermal motion, if the dielectric constant rose to 500 as happens at temperatures close to a ferroelectric phase transition (Fig. 1.8; see also Fig. 2.31 later) the induced displacement can become as large as the thermal motion.

Macroscopic measurements of sublimation energies point to binding energies of atoms in crystals of order of a few hundreds of $\mathrm{kJ\,mol^{-1}}$, or a few eV per atom ($1\,\mathrm{eV} = 96.484\,\mathrm{kJ\,mol^{-1}}$). For example, for NaCl the thermal energy of an atom at a temperature of 300 K is $2.5\,\mathrm{kJ\,mol^{-1}}$, which is substantially lower than the binding energy of $764.4\,\mathrm{mol^{-1}}$ (data from Kittel, 1996).

Without having to work too hard, but armed with a considerable degree of hindsight, we have been able to extract the basis of a plausible model for the atomic structure of a material, which links the physical properties of the crystal to the energetics of the distortions of this structure. The thermodynamic properties are associated with vibrations of the atoms with frequencies in the THz regime, which produce displacements of atoms of magnitude similar to static displacements that can be produced by external forces. Thus we have a picture that is making sense numerically, a picture that links structure to dynamics through the forces between atoms, and which provides the connection between macroscopic properties and the underlying atomic picture.

However, this is beginning to get a little dizzy, because nothing has been properly formalized or quantified. It is quite possible that one of the central assumptions we have used is simply wrong, in which case it is possible that the entire edifice might collapse. In fact, with hindsight we know that one of the points we have made is wrong because there are factors that cannot yet be introduced into the edifice. We assumed that the results of IR absorption experiments point to discrete vibrational frequencies rather than the

existence of a distribution of frequency values. It turns out that there really is a broad continuous distribution of discrete frequency values, and IR absorption spectroscopy can only detect a small subset of this distribution. This subset is so small that the spectra look as if there are only a few discrete frequency values. However, this does not destroy the overall picture we have put together. Implicit in eqn 1.10 is that there is one vibration for each degree of freedom of each atom. We automatically accommodated the fact that the total number of vibrations is three times the number of atoms in the crystal, and the relatively few vibrations seen in a spectroscopy experiment have frequencies that are representative of the complete distribution of frequency values.

So our task is to quantify the link between structure and dynamics, and also the links to the forces between atoms. Implicit in this task is the need to formalize these links. We need an appropriate language to link the mathematics and physics, and will need to develop particular tools for the job. These are the tasks of this book.

1.2 Length scales and time scales

We are now drawn to the question of length and time scales, and we start from the empirical issue of the length scale of the underlying object that causes a crystal to be cleaved to smaller and smaller pieces with a constantly repeating shape. We have to presume that we cannot keep chipping away at the crystal forever. That said, we are able to grind crystals of calcite down to a fine powder such that the shapes of the grains can only be determined using a microscope, and from this we have to conclude that the dimensions of the underlying periodic image is certainly less that the micron length scale. Thus we need a more powerful microscope.

This is the problem that faced scientists at the start of the twentieth century. There was some degree of consensus that matter appeared to be made from atoms, and that the atoms would form an internal ordered structure, but there was no way of telling whether or not this was so. Indeed, although there were compelling reasons for taking the existence of atoms seriously, the lack of any experimental evidence led some leading (and over-influential) scientists towards the view that atoms were little more than a mathematical construction designed to explain the empirical macroscopic observations, and that the idea of atoms might eventually be superseded by a better theory.

It was the development of new experimental methods acting as better microscopes that opened the box and tipped the atomic model onto the floor for scientists to assemble the theory of the structure and dynamics of solids from the many parts. The crucial microscope was X-ray diffraction. X-rays were discovered in 1895, and were shown by reflection from optical diffraction gratings to have wavelengths on the length scale of 1 Å. At the time there was evidence to suggest that this was the length scale of interatomic distances, and it was realized that X-rays might be diffracted by crystalline solids in the same way that light beams can be diffracted by ruled gratings. The year 1912 was when this idea was demonstrated experimentally, and within a short space of time the idea was turned into a quantitative tool for the investigation of the atomic structure of crystals. It was confirmed that the basic building blocks of

crystals are arrangements of atoms over length scales of a few Å, or longer in the case of more complex materials. The crystal structure of NaCl, Fig. 1.11, was one of the earliest structures solved by X-ray diffraction.

Later, once it had been established that beams of particles also have wave properties, it became possible to perform diffraction with other types of radiation. Beams of electrons provide a tool that most closely resembles a true microscope, giving a resolution that is truly on the atomic scale (the first electron microscope was developed in 1931). In some cases the images obtained in the electron microscope can be directly related to the atomic structure. A view of the arrangement of atoms in crystalline silicon is shown in Fig. 1.12. In recent years tools such as scanning tunnelling microscopes have given images that represent the structure of atoms on crystal surfaces.

We will therefore have to consider length scales that range from human length scales, of order 1 m, right down to the Å scale. In fact, when we consider diffraction of beams of neutrons (Section 6.2.4), we will also need to go down to the nuclear scale, namely 10^{-15} m. Thus the phenomena that we cover in this book will span 15 orders of magnitude!

Coupled with the range of length scales is the issue of time scales. We have already noted that the fastest time scales, namely those associated with atomic vibrations, involve frequencies in the 10 THz range, which correspond to time scales down to 10^{-13} s. Many phenomena of importance are associated with this time scale. Another time scale is shown by atomic diffusion, and this is around the 10^{-6} s ($= 1\,\mu$s) range. When compared with the human time scale, of the order of 1 s, the time scales span 13 orders of magnitude. The range of length and time scales is shown in Fig. 1.13, together with examples of significant distances found in crystals or in the probes used to study crystalline phenomena. This comparison of the different length and time scales is important to bear in mind throughout this book.

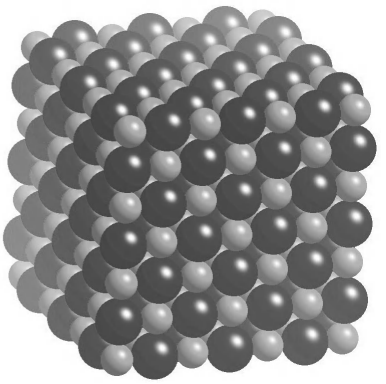


Fig. 1.11 Arrangement of the atoms in the crystal structure of NaCl, as revealed by X-ray diffraction, but having been previously predicted from arguments based on crystal shape.

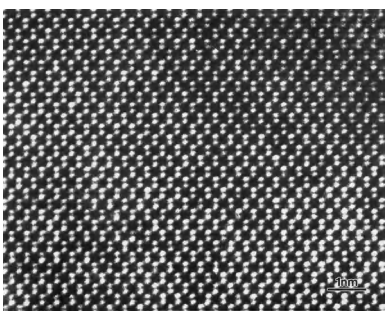


Fig. 1.12 Silicon atoms seen in a high-resolution transmission electron microscope (photograph courtesy of Jeol). Each spot can be identified with a single atom (although such an indentification is not always possible).

1.3 Tools of the trade

1.3.1 Theoretical and mathematical tools

In order to make sense of the empirical observations we have cited in this introduction, and to then develop a model for the structure and dynamics of crystalline materials that links the atoms and the forces between them to the important phenomena, we will need to make use of a wide diversity of theoretical tools. A number of these have already been highlighted in the preceding discussions.

Symmetry

Symmetry plays a critical role in many aspects of the study of the structure and dynamics of solids. Everyone has some idea of what symmetry is. Individual objects, such as gem crystals, have a type of symmetry in which the object will look the same when viewed from different directions or when the object is rotated through some angle. Nature constantly makes use of symmetry, as in the arrangement of petals in a flower (Fig. 1.14). Humankind particularly values symmetry. For example, symmetry is used in the design of many manufactured

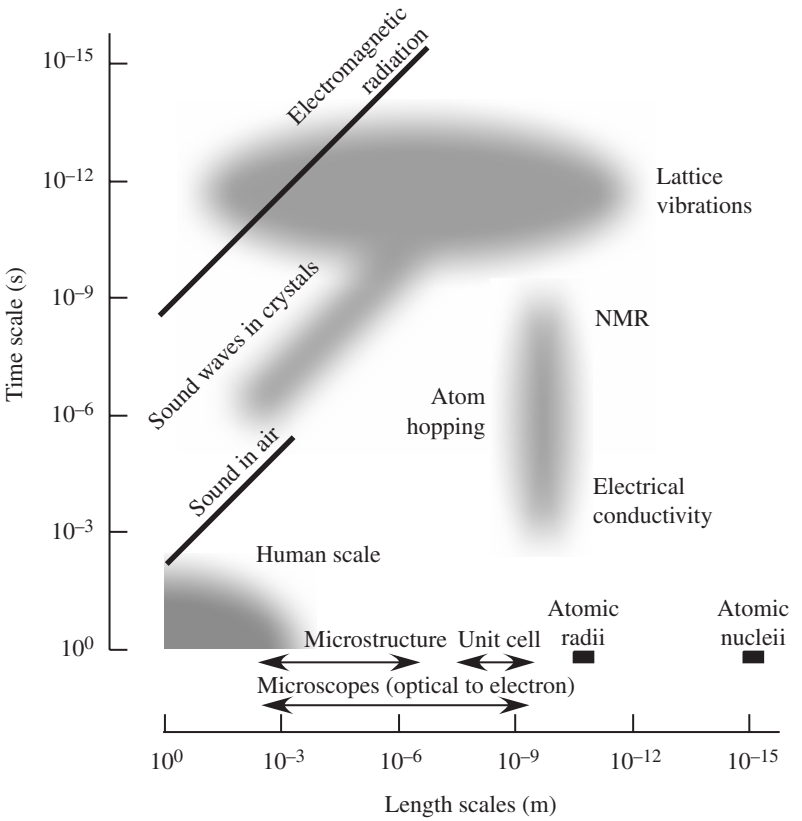


Fig. 1.13 Important length scales and time scales in the study of crystals.

Fig. 1.14 The structures of flowers provide an example of symmetry that everyone meets in daily life.

objects, both because of efficiency of production and because of aesthetics. Formal houses and gardens have symmetry at the heart of their design (Fig. 1.15). We also can appreciated the symmetry of repetition of a pattern across space, and again periodic symmetry is used for similar reasons in most areas of human activity. Wallpaper is based on this periodic symmetry, and tiles fill large areas of floors and walls by making use of periodic symmetry.

The symmetry of the arrangement of atoms in the crystal of silicon shown in Fig. 1.12 can readily be appreciated. Having an inherent appreciation of symmetry gives us a start in our study of the structure and dynamics of materials, but to properly exploit our understanding of symmetry will necessarily require a formalization of the concepts. This will lead into the ideas of group theory, which are introduced in Chapter 3.

Quantum mechanics

As we have anticipated in the discussion in this introduction, quantum mechanics will play a critical role in large parts of the model of the dynamics of atoms. The role of quantum mechanics is important in controlling the behaviour of the electrons that are involved in the forces that bind atoms together. We will give only a brief introduction to the quantum mechanics of electrons in solids with respect to the forces between atoms (Section 5.5). However, the important application of quantum mechanics for the structure and dynamics of crystals

is the quantization of atomic vibrations. This is a different form of quantum mechanics. Electrons are in a class of particles, called **fermions**, for which the **Pauli exclusion principle** is an important control on the form of the quantum description. On the other hand, the quantization of atomic vibrations, or indeed of any vibration, is not subject to the Pauli exclusion principle. The class of particles that includes these quanta are called **bosons**. Fermions and bosons obey different implementations of statistical mechanics, and we will make use of the boson statistics in Chapter 9.

Fig. 1.15 The house and garden of Wrest Park, Bedfordshire. In almost all styles of architecture and landscaping symmetry is the key component of design. (Photograph provided by English Heritage.)

Statistical mechanics and thermodynamics

The quantization of the vibrations of a crystal is not apparent from any experiment on the ground state of a crystal, but becomes apparent once temperature is introduced into the picture. As we have noted earlier, it is in linking the quantum mechanics of atomic vibrations to thermodynamics that gives us an accurate picture of the variation of the crystal structure and properties with temperature. Although quantum mechanics tells us the form of the tools of statistical thermodynamics we must use, it does not modify the central ideas of classical thermodynamics. We will find that one of the most powerful ideas given by classical thermodynamics is that the equilibrium state of matter is that for which the **free energy** is at a *minimum*. The free energy provides the link between energy and entropy, trading energy against the tendency for spontaneous creation of disorder. If it is possible to write down a formal expression for the free energy of a material (see Appendix L), we will be in a very powerful position, because minimization of the free energy function will give the full range of properties of the material. Our aim, therefore, will be to combine the quantum mechanics of the dynamics of atoms, and the energies of interatomic interactions, with the tools of statistical thermodynamics to produce the free energy of a crystal (Chapter 9).

Fourier transforms

One of the central mathematical tools we will use is the Fourier transform. This links frequency to time, and position to momentum and wave vector. The idea of the Fourier transform is that any function can be said to be broken down into an infinite number of components taken from a continuous distribution. For example, the time variation of a function can be represented by a set of component oscillations taken from a continuous distribution of frequencies. Fourier transforms are particularly useful for linking the behaviour of atoms to the quantities measured in experiments, and for matching the behaviour of individual atoms to that of the whole crystal. In many cases the Fourier transforms will emerge naturally in the development of the theory, diffraction (Chapter 6) being a particularly clear example. Fourier transforms are described in Appendix B.

1.3.2 Experimental tools

The dramatic growth of physics during the twentieth century was significantly aided by the development of sophisticated quantitative experimental methods. The development of X-ray diffraction was one of these, and the information

that it provided about the atomic structure of crystals was truly revolutionary. In fact most of the experimental tools we will use involve beams of radiation: X-rays, electrons, neutrons, infrared, and optical lasers (Chapters 6 and 10). Experiments with beams of radiation can provide information about the positions and dynamics of atoms, and as a result radiation beam experiments have provided many critical insights into the behaviour of atoms within solids.

Experiments with beams of radiation mostly probe the short length and time scales of Fig. 1.13. Intermediate length scales can be studied using microscopes, optical or electron. Intermediate time scales can be studied using nuclear magnetic resonance (NMR) spectroscopy or by measurements of dielectric properties. NMR tends to probe motions in the MHz to GHz frequency range, and can give information about atomic environments over nearest-neighbour distances. Dielectric constant measurements probe frequency scales from MHz down to below kHz, but do not provide direct information about spatial interactions.

In recent years advanced electronics and computing have revolutionized many experimental tools and have enabled the development of new tools that rely on high-precision movements and alignment. Examples of new tools are surface probes such as scanning tunnelling and atomic force microscopes. We will not have the scope to discuss these new tools in this book, but it is useful to be aware of the important impact they are making in our understanding of crystalline materials and their surfaces.

1.3.3 Special tools and concepts for the study of the structure and dynamics of crystals

To complete the survey of the theoretical, mathematical and experimental tools needed for understanding crystalline materials, we should note that we will need to develop some special tools. These are forced by the periodicity we encounter in crystals. Symmetry gives potential for gains in efficiency across the whole range of human activity, and we can expect that symmetry should be exploited in the study of crystalline materials. Periodicity, which is formally known as **translational symmetry**, is so important that it helps to have it built into the mathematical description of crystalline materials from the start. This will involve the concept of the **lattice**, which is an infinite array of points in which all points have an identical environment. The lattice provides the foundation of the mathematical description of translational symmetry. The formal tools are developed in Chapter 3, being anticipated by our review of the wide range of crystal structures in Chapter 2.

We will show that the Fourier transform of a lattice is a new lattice, not in the same space as that of the original crystal lattice, but in a completely new space which we will call the **reciprocal space**. The **reciprocal lattice** is discussed in some detail in Chapter 4, and is a concept that is essential for several other chapters. The variable of reciprocal space is the **wave vector**, which is given the symbol $\mathbf{k}$. This is linked to momentum $\mathbf{p}$ by $\mathbf{p} = \hbar\mathbf{k}$ (but we have to be careful in how we intepret this momentum – we will find that the periodicity of the reciprocal lattice means that the conservation laws on $\mathbf{k}$ are not as tight as the classical law of conservation of momentum). The reciprocal lattice will provide one of the main conceptual challenges in this book, but mastery of the

reciprocal lattice will give a tool that has a power similar in magnitude to that of the free energy.

Summary of chapter

- The **shapes of crystals** point to the existence of fundamental building blocks with particular **symmetry**. The existence of **slip deformation** points to the same building blocks. Optical properties of some crystals are **anisotropic**, with special directions being linked to the special directions of symmetry seen in their shapes. These building blocks are determined by **packing of atoms** whose arrangements have symmetry reflected on macroscopic length scales.
- Application of an electric field to an insulating crystal causes a **dielectric polarization** due to displacements and polarization of ions. These two effects can be seen separately using low-frequency and high-frequency fields.
- Many properties are affected by temperature, such as ionic conductivity, heat capacity, and crystal size (i.e. through **thermal expansion**). This factor points to the importance of the **dynamics** of atoms. In some cases, change of temperature produce a change of **phase** of the material, with dramatic changes in **physical properties**.
- **Quantum effects** are important in describing the thermodynamic properties of crystals at relatively high temperatures.

Further reading for Chapter 1

At the end of each chapter are references for further reading, which will either expand or deepen the discussion, or else will give an alternative presentation of the material. At this point it is worth recommending the following general textbooks on condensed matter, which touch to greater or lesser extent on the subject area of this book: Ashcroft and Mermin (1976), Blakemore (1985), Callaway (1991), Christman (1988), Hook and Hall (1991), Kittel (1986) and Rosenberg (1988). A more chemical perspective is given by Borg and Diennes (1992), Elliott (1998), Rao and Gopalakrishnan (1986) and Smart and Moore (1992). For the historically minded, Cruickshank et al. (1992), Hoddleson (1982) and Wilson and Weire (1987) provide some interesting insights. Finally, Hazen and Finger (1982) present many ideas from the perspective of high-pressure research.

2 Structure of materials

2.1 Introduction

In this chapter we will explore some of the principles that govern the way atoms can be packed together to form simple crystal structures. The main principles here are geometric in origin, concerned with packing atoms of different sizes or with specific bond geometries. The bond geometry will lead to the formation of simple structural units containing small groups of atoms, and the structures are then determined by the way that these small units can link together. The geometric viewpoint has some degree of simplification in it, but this can be seen as having some merit since the essence of physics is to look for simplifying principles that can get part way towards an understanding. To properly understand issues associated with structure we need to look in more detail at the nature of the interatomic bonding, which we will consider in Chapter 5.

One of the amazing things about crystalline solids is that just a few general principles and the scope of the periodic table of the elements lead to a great diversity of structures. Diversity is the flavour of our times (as highlighted by the growing concerns to protect our planet's biodiversity), and the world of crystal structures has an amazing diversity. On the one hand this diversity can be looked on as simply a large catalogue (structure #1, structure #2, . . .), but on the other hand the world of crystal structures can be viewed from the perspective as if entering an art gallery. Enjoy!

But we will have to take our gaze from the beauty of the crystal structures to analyse the details with a particular formalism. We cannot be merely observers of the art; we have to become art critics. The formalism has its own beauty, and to understand the formalism will help appreciate the diversity of crystal structures and facilitate the cataloguing and classification. The formalism also allows us to establish general principles that will make analysis of structures much easier. Moreover, these principles will allow us to interpret experimental data without the need for complex calculations, and will enable us to understand the link between structures at an atomic level and properties at a macroscopic level. The formalism also enables us to identify links between different structures, such as those between crystals and glasses. Finally, the formalism has its own elegance, enabling quite deep conclusions to be drawn from relatively simple applications of the formalism.

In this chapter we will chart our way through some of the simpler crystal structures, including materials with important technological applications. Our starting point is with the crystal structures of the elements. In many cases

these can be described in terms of simple models of packing of spheres, and at this level of simplicity we are able to identify some of the important general principles underlying the formation of crystal structures such as the role of symmetry. However, even the elements can form some quite complex crystal structures. By comparing these with some of the simpler structures, we will be able to develop ideas of bonding, which give new geometric principles, and allow for a much wider diversity of structure types than might have been envisaged. We will extend this discussion to diatomic compounds, and then to more complex materials. Several important concepts introduced in this chapter, such as symmetry (3), bonding (5), and phase stability (12), will be developed in later chapters.

2.2 Crystal structures of the elements

2.2.1 Close-packed metals

Close packing in two dimensions

For many elements, the overriding principle that governs the crystal structure is the need to have the highest possible density. How this can be achieved when chemical bonding does not throw up any constraints can be seen by looking at the packing of atoms in two dimensions, Fig. 2.1. Each atom has six neighbours, and there are groups of three atoms in close contact.

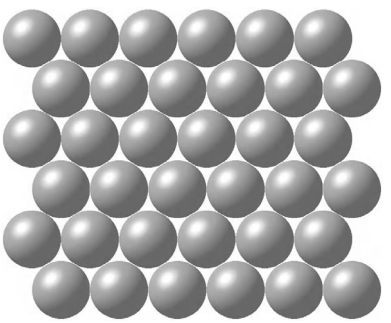


Fig. 2.1 Two-dimensional close-packed structure, with atoms in contact with six neighbours to give the maximum density.

Using the two-dimensional close-packed structure as an example, we can identify a number of important components involved in the description of any crystal structure. First we consider how the atoms are packed into a periodically repeating arrangement. The periodicity is obvious from Fig. 2.1, and to formalize this may seem to be a distraction at this point, but it will quickly become apparent that the formalism will help to make the description of crystal structures much easier. Our formalism involves separating the operation of the periodicity from the objects that are repeated. The process in two dimensions is illustrated in Fig. 2.2. We start with an infinite periodic array of points, which we will call the **lattice**.

Each lattice point represents a small region of space, which can be represented as a polygon bounded by four lattice points. Three examples are marked in Fig. 2.2. The whole of the space can be tiled by these parallelograms. One of the parallelograms (*a*) has edges of equal lengths and internal angles of 120° and 60°. This shape reflects the basic hexagonal symmetry of the close-packed structure, and is therefore the most natural parallelogram to consider as the basic tile. The region of space associated with each lattice point is called the **unit cell**.

In principle a unit cell can contain one or more atoms. The crystal is constructed by replacing each lattice point by the contents of the unit cell. The unit cell (*a*) of the two-dimensional structure in Fig. 2.1 is the trivial case of a unit cell containing just one atom, so in this case the process of replacing each lattice point by the contents of the unit cell is simply to place an atom on each lattice point. However, when the unit cell may contain several atoms, each lattice point will be replaced by a small arrangement of atoms. For future reference, the process of constructing an infinite crystal by replacing each lattice point

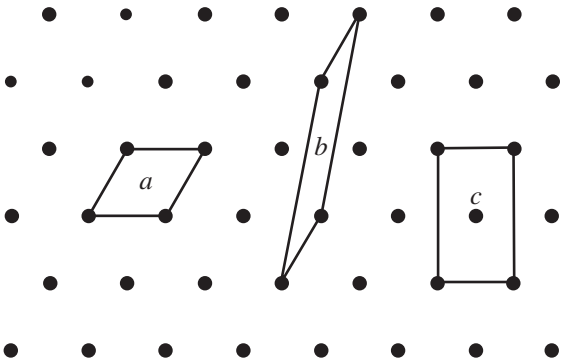


Fig. 2.2 Underlying lattice of the two-dimensional close-packed structure. The actual crystal structure is constructed by replacing the points on the lattice by atoms. The figure shows three possible **unit cells**. The unit cell marked *a* is a rhombus (internal angles 60° and 120°), whose shape reflects the underlying rotational symmetry. The shape of the unit cell marked *b* bears no obvious relationship to the symmetry and would not be a logical choice. The rectangular unit cell, *c*, has sides with lengths in the ratio $1:\sqrt{3}$. The orthogonal axes reflect some of the symmetry of the structure. This cell encompasses two lattice points. All three unit cells represent areas that can be repeated periodically and fill all the space.

by a small group of atoms is an example of the mathematical operation called **convolution** – this operation will be very important when we discuss diffraction methods in Chapter 6.

As we have noted, the natural unit cell of Figs 2.1 and 2.2 has edges of equal lengths, and angles of 120° and 60°, reflecting the hexagonal symmetry of the close-packed structure. We could also have defined a unit cell with orthogonal axes but with twice the area, as shown by *c* in Fig. 2.2. In this case, one cell edge would be $\sqrt{3}$ times as long as the other. This unit cell is not as perverse as the parallelogram marked *b*, because orthogonal axes are particularly easy to handle. We will see below that there are three-dimensional arrangements of atoms for which the most natural unit cell to reflect the symmetry has orthogonal axes, even though it may have a volume that is twice or four times larger than the volume associated with a single lattice point.

Before we move away from our two-dimensional example, we note that we can also formalize the **symmetry** of the structure. First, we note that from a point at the centre of each atom we could rotate the crystal structure by any multiple of 60° and get an exact copy of the starting structure. This is an example of **rotational symmetry**. In this case we can rotate the structure by multiples of $\frac{1}{6}$-th of a complete revolution, so the particular symmetry is called a **6-fold rotation symmetry**. The structure also has points with **3-fold rotational symmetry** (located in the middle of the small triangles described by three touching atoms) and **2-fold rotation symmetry** (located at the points where pairs of atoms touch). The axes associated with these rotational symmetries are shown in Fig. 2.3.

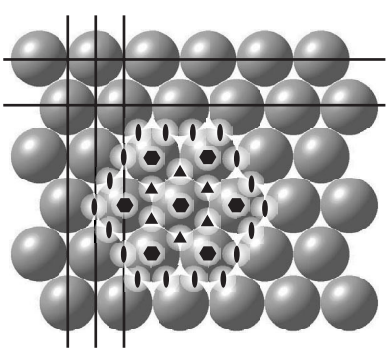


Fig. 2.3 Another view of the close-packed layer, showing some of the axes of the rotational symmetry (hexagons for 6-fold rotational symmetry, triangles for 3-fold symmetry, and ellipses for 2-fold symmetry), and the lines of the mirror planes.

The structure also has a type of symmetry associated with reflection rather than rotation. Consider a line through the centres of atoms in a horizontal row in Fig. 2.1. If we reflect the crystal through this line we get an exact copy of the structure. This line is called a **mirror plane** (because in three dimensions it is a plane, and one can imagine that the mirror is three-dimensional lying normal to the plane of the diagram). Because of the 6-fold rotation symmetry, there are identical mirror planes passing through the centre of each atom rotated by $\pm 60°$. There is another set of mirror planes passing through the centres of each atoms lying at 30° to the first set, and, again, these are reproduced by rotations of $\pm 60°$. Both sets of mirror planes are shown in Fig. 2.3.

To formally list all the rotation and mirror symmetries may seem to be a little gratuitous at this point, but later (Chapter 3) we will see that there is great power in being able to describe the complete symmetry of a crystal structure. For the moment, the important point to note is that we have introduced and defined both rotational symmetry and mirror planes.

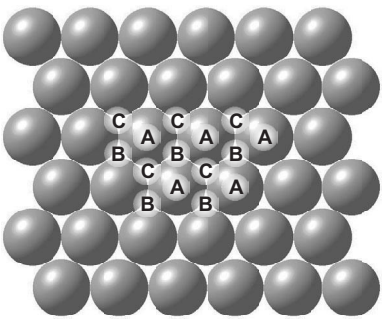


Fig. 2.4 Another view of the close-packed layer, showing the sites for the stacking of another close-packed layer. The A sites are occupied by the atoms in the layer, and the B and C sites fit into dimples on the top and bottom surfaces.

Three-dimensional close-packing: hexagonal close-packed structure

The close-packed layers of Fig. 2.1 can be stacked on top of each other in two ways that continue the close packing, as illustrated in Fig. 2.4. The atoms in adjacent layers are displaced with respect to each other in order to continue the close packing. If the first layer is denoted A in Fig. 2.4, the atoms in the next layer will be in either the B or C positions. With either of these stackings, an atom in one layer is able to touch three atoms in the layers above or below. When added to the six touching neighbours in the same layer, this means that each atom has 12 touching neighbours in a three-dimensional close-packed arrangement.

Because the layer on top of A can be in either B or C positions (Fig. 2.4), it is clear that there is not one unique way of building a three-dimensional crystal layer by layer. Given the tendency of nature to like to produce order, there are two main types of stacking favoured by the elements. The first is the case where the atoms in the third layer have the same relative positions as in the first layer, producing a repeat every other layer. In terms of the labels we have given the layers, the double-repeat can be described by the sequence –ABABAB–. This structure is called hexagonal close-packed, and is abbreviated as hcp.

We can define a three-dimensional unit cell. The base of the unit cell will be the parallelogram with equal sides and angles of 120° and 60° defined for the two-dimensional close-packed layer. The third axis is normal to the close-packed layer, and extends to the next-but-one layer up – recall that the stacking of layers in hcp repeats every other layer. The hcp unit cell is shown in Fig. 2.5a. We use the vectors **a** and **b** to represent the unit cell edges within the close-packed layer, and the vector **c** to represent the unit cell edge normal to the close-packed plane. The moduli of these three vectors are written as a, b and c respectively, noting that $a = b$ in this case.

A straightforward geometric analysis (see problem 2.2) can show that if the dimension of the unit cell in the horizontal plane, a, is equal to $2r$, where r is the radius of the atoms, the dimension of the unit cell in the third direction, c, would ideally be $4\sqrt{2/3}r$, so that $c/a = \sqrt{8/3} = 1.633$. In practice, some of the elements with hcp structure have values of a and c that come close to satisfying this ratio, such as Ce (1.63), La (1.62) and Mg (1.62), but because this ratio is not fixed by symmetry it will not be satisfied exactly. In some cases the ratio is broken to a much larger extent, being either larger or smaller than the ideal value. Examples are Cd (1.89), Zn (1.86), Be (1.56) and Ho (1.57). These examples are evidence that the bonding between atoms is rather more complex than the simple stacking of spheres, as will be discussed in Chapter 5.

The three-dimensional unit cell of the hcp structure contains two atoms, as shown in Fig. 2.5a. One of the atoms is at the origin of the unit cell, and the other is associated with the alternative close-packed layer. We can describe the

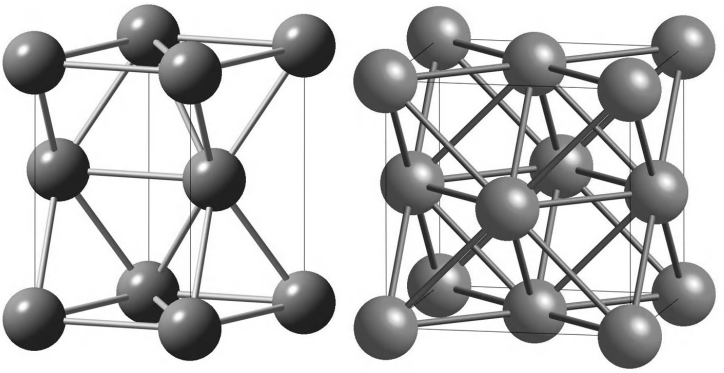


Fig. 2.5 The hexagonal (*left*) and cubic (*right*) close-packed structures showing the nearest-neighbour contacts.

position of the second atom using a convention called **fractional coordinates**. This provides a more general method than using absolute positions. We can describe the absolute position of an atom within a unit cell in terms of a combination of vectors parallel to each of the three lattice vectors. If the atom is at position **r**, it is described this way by

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} \tag{2.1}$$

where x, y, and z are numbers with values between ± 1. It may seem strange to want to use negative numbers, but in practice the use of negative numbers can sometimes help to highlight the coordination around the origin. The set of numbers x, y, z are the **fractional coordinates**. The lattice symmetry means that addition or subtraction of any integer to any one of the fractional coordinates simply gives an equivalent site in another unit cell. The use of fractional coordinates is valuable since it permits a description of the positions of atoms in the unit cell that is independent of the values of the lattice parameters. As a result, the set of fractional coordinates for all atoms in the unit cell can show the symmetry of a crystal structure. For hcp, the fractional coordinates of the two atoms are at 0, 0, 0 (the atom at the origin), and $\frac{1}{3}, \frac{2}{3}, \frac{1}{2}$.

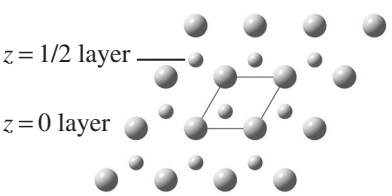


Fig. 2.6 Projection of the hcp structure showing the two close-packed layers, highlighting the different layers by using different size spheres. The point to note is that the atoms in the two layers have different environments, which are related to each other by a rotation of 180°.

The important point to note is that the vector from the origin to the second atom cannot be reduced to a lattice vector; doubling the length of the vector takes us to the point with fractional coordinates $\frac{2}{3}, \frac{4}{3}, 1 \equiv \frac{2}{3}, \frac{1}{3}, 0$, which is not another lattice point. In addition, the environments of the two atoms are not equivalent, but are related by 180° or a plane of reflection. This is seen in Fig. 2.6.

Three-dimensional close-packing: cubic close-packed structure

A second common three-dimensional arrangement of close-packed layers is where the repeat occurs every third layer. From Fig. 2.4, this arrangement of layers can be described by the sequence –ABCABC–. However, this structure can be described more easily than as just another version of hcp. In Fig. 2.5 we show a small part of this three-layer repeat structure, and draw in a new unit cell. By geometry it can be shown that the angles of this unit cell are 90°, and that the three lengths of the unit cell are equal (see problem 2.3 at the

end of the Chapter). The new unit cell is shown in a different orientation in Fig. 2.5. It contains four atoms, and is described as a cube with an atom on each corner and one at the centre of each face. This structure is called **cubic close-packed**, abbreviated as **ccp**, because it is the cubic version of the close-packed arrangement. This is the simplest version of a type of structure called **face-centred cubic** (**fcc**).

The ccp structure now introduces another aspect of the formalism of the description of the crystal structure. Each atom has an equivalent environment, unlike the case of hcp. Thus each atom can be said to be occupying a lattice point. The fcc unit cell, containing four lattice points, is therefore not the most fundamental unit cell that can be drawn. Quite how one then describes the fundamental unit cell is arbitrary, but the most sensible unit cell would be similar to that shown in Fig. 2.7. Denoting the three vectors describing the cubic unit cell by **a**, **b** and **c**, as shown in Fig. 2.7, the vectors describing the reduced unit cell can be written as

$$\begin{aligned} \mathbf{a}' &= \frac{1}{2}(\mathbf{a} + \mathbf{b}) \\ \mathbf{b}' &= \frac{1}{2}(\mathbf{b} + \mathbf{c}) \\ \mathbf{c}' &= \frac{1}{2}(\mathbf{c} + \mathbf{a}) \end{aligned} \tag{2.2}$$

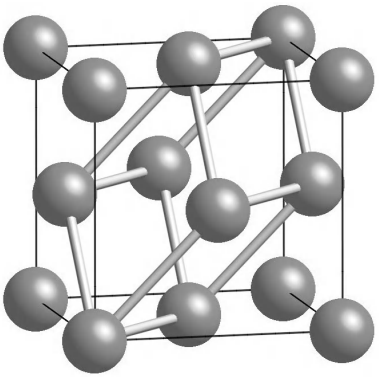


Fig. 2.7 One possible primitive unit cell of the ccp structure, shown as the large outline, with the outline of the conventional cubic unit cell.

If the edge of the fcc unit cell has length a, the new unit cell has all edges of length $a/\sqrt{2}$, and angles equal to 60°. This type of transformation will be discussed in more detail in Section 4.3.

This discussion concerning the two unit cells of the fcc structure illustrates the issue involved in deciding how to best describe a crystal structure. The unit cell with angles of 60° is the more fundamental since there is one lattice point per unit cell. It is called the **primitive unit cell**. However, it is much more complicated to think about than the fcc unit cell, and usually people stick to this one. As a result, it is called the **conventional unit cell**. The issue of having a conventional unit cell that differs from the primitive unit cell, and which contains two or four lattice points, is actually quite common.

It is worth comparing the symmetry of the ccp structure with that of the hcp structure. Two views of the ccp structure are shown in Fig. 2.8. The way that the close-packed layers stack together means that the 6-fold rotation symmetry is broken, preserving only the 3-fold symmetry. But we now also have 4-fold rotation symmetry down the cube axes, and 2-fold rotation symmetry axes lying between two orthogonal 4-fold axes. These rotation axes are shown in Fig. 2.8.

Inspection of the ccp structure in Fig. 2.8 also shows the presence of several different types of mirror planes. There is one other important symmetry operation that we should highlight. If you could view the environment from the centre of an atom in the ccp structure, you would find that the view along any direction is exactly the inverse of that seen from the opposite direction. Put more formally, anything at the position x, y, z would be identical to anything at the position $-x, -y, -z$. This is a very important symmetry operation that has profound implication for the physical properties of materials. The origin is called the **centre of symmetry**, and a crystal with a centre of symmetry is called **centrosymmetric**. The centre of symmetry is given the symbol $\bar{1}$.

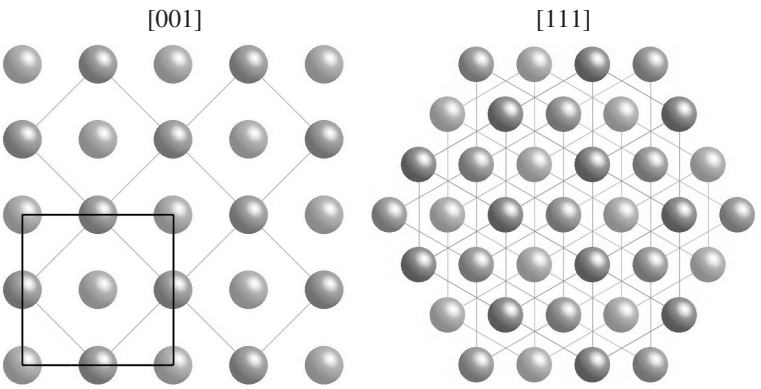


Fig. 2.8 Two views of the ccp structure, highlighting the symmetry operations active in the two directions. **Left** shows the view down one of the cube axes, labelled [001], showing the 4-fold rotation symmetry. **Right** shows the view down the cube diagonal and normal to the close-packed planes, labelled [111]. This view shows the 3-fold rotation symmetry. In both views atoms in the same horizontal layer are connected by the sticks drawn as guides.

Packing efficiency

To close the discussion of the two close-packed three-dimensional structures, we address the question of the efficiency of the packing of atoms in the close-packed structures. Clearly both the ccp and ideal hcp structures have the same efficiency of packing, so we will consider only the ccp structure because the geometry is a little easier. Let us consider each atom to be approximated by a hard sphere of radius R. In the ccp arrangement, the diagonal of the face of the cubic unit cell is equal to $\sqrt{2}a = 4R$. The total volume of the cubic unit cell is equal to a^3. There are four atoms in the cubic unit cell, each of volume $4\pi R^3/3$. Thus the fraction of space filled by the atoms is equal to $4 \times (4\pi R^3/3)/a^3 = \pi/3\sqrt{2} = 0.74$. This value will be compared with the packing efficiencies of other structures we will introduce later in this chapter.

Close-packed crystal structures

The number of elements that have either the hcp or ccp crystal structures can be seen by looking at the representation of the periodic table in Fig. 2.9. Many of the metals exist in one of either of these structures. There are around 30 elements that have the hcp structure, and a similar number with the ccp structure. Some elements can be found in either structure, sometimes because of a phase transition (as in Ca), sample preparation (Co), or by sample treatment (Ni). Because both structures are close-packed, it is to be expected that they will have similar energies. Moreover, since the hcp and ccp structures differ in how the close-packed layers are stacked on top of each other, it is possible to imagine that structures with other stacking sequences are possible. Some elements (e.g. Am, La) can be prepared with a hexagonal structure that has twice the repeat length and four atoms in the unit cell. From Fig. 2.4, this arrangement of layers can be described by the sequence –ABACABAC–, with the repetition occuring on the fourth layer. It is worth noting that we will briefly discuss other types of layer structures later in this chapter, and that layer structures can adopt quite complex periodic stacking sequences, such as –ABCACBABCACB– (sixth-neighbour repeat). Phases with long-period repeats are called **polytypes**. Some of these long-period repeats, particular those extending over tens of layers, can arise as non-equilibrium states due to crystal growth conditions, but there is evidence that other polytypes exist as equilibrium states, arising from the

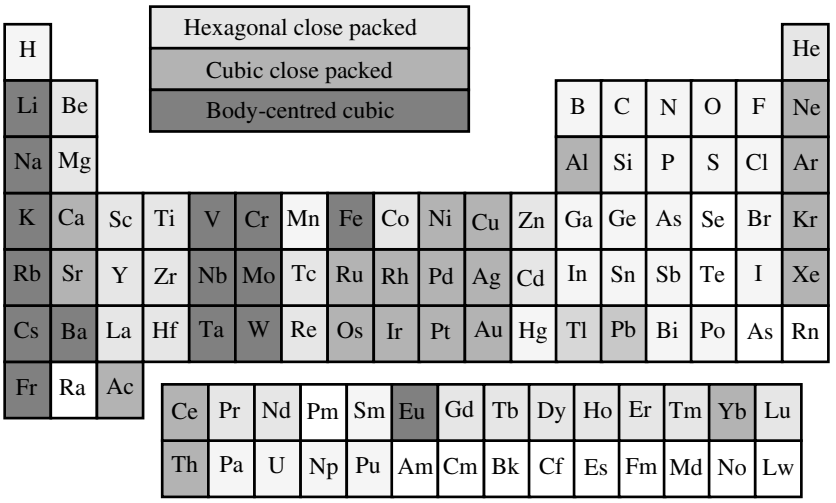


Fig. 2.9 The periodic table of the elements, showing the different types of crystal structures formed by the elements.

presence of competing interactions between different sets of neighbours. We will meet examples later in this chapter. It is also easy for a layer structure to have stacking faults. Thus any fault that might be described by stacking sequences, such as –**ABABABCBCBC**– (hcp with stacking fault) or –**ABCABCBACBA**– (ccp with stacking fault), will not have a high energy.

2.2.2 Body-centred cubic packing

Several metallic elements, including some of the alkali metals and iron, have a packing that is a little less efficient than close packing. This packing is easily described as a cubic cell with a second atom in the middle, with each atom having eight touching neighbouring atoms. This structure is called **body-centred cubic**, abbreviated as **bcc**, and is shown in Fig. 2.10. Each atom in the bcc structure actually has an equivalent environment, meaning that the cubic unit cell is not the primitive unit cell.

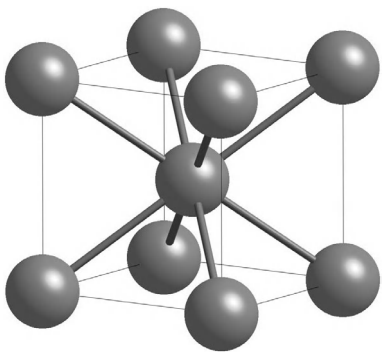


Fig. 2.10 Body-centred cubic structure. Each atom has eight neighbours, as indicated by the bonds drawn between atoms.

The efficiency of the packing of atoms in the bcc structure is actually not much lower than the close-packed structures described above. The body diagonal of the cubic unit cell is equal to $\sqrt{3}a = 4R$. With two atoms in the cubic unit cell, the efficiency of packing is equal to $2 \times (4\pi R^3/3)/a^3 = \sqrt{3}\pi/8 = 0.68$.

Just under 30 elements crystallise in the bcc structure. The is highlighted in Fig. 2.9. These include the Group I (alkali metal) elements, and some of the transition metals. Moreover, several elements can exist in both close-packed and bcc structures as equilibrium phases under different conditions of temperature and pressure, or as non-equilibrium states depending on the crystallisation process. For example, iron exists in both fcc and bcc structures as stable phases at different temperatures under ambient pressure, and it is believed that the stable phase under the temperature and pressure conditions of the inner Earth has the hcp structure.

2.2.3 Simple cubic packing

It might be thought that the simplest cubic packing would simply have a cubic unit cell with only one lattice point in the unit cell. This structure is shown in

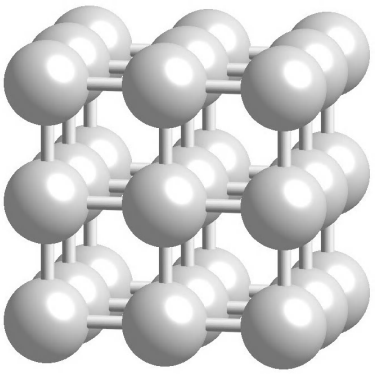


Fig. 2.11 Simple cubic structure. Each atom has six neighbours, as indicated by the bonds drawn between atoms.

Fig. 2.11. Each atom has six nearest neighbours, touching along the cube edges. The packing efficiency of this structure is $\pi/6 = 0.52$ (problem 2.4). In spite of its apparent simplicity, there is only one element that crystallises with the simple cubic structure, namely polonium. If atoms only had simple nearest-neighbour bonds, the simple cubic structure would be inherently unstable against any shear deformation, and it therefore needs to have strong directional or angle-dependent bonds to make it stable. Thus, although the simple cubic structure is nice because it does not have the complexity of the other structures we have discussed so far, it is in practice more-or-less irrelevant as a potential crystal structure!

2.2.4 Crystal structures of elements with covalent bonding

Diamond structure

The discussion above was primarily concerned with the geometric effects of packing of atoms. The observation that many metals have bcc structure instead of one of the close-packed structures suggests that there is more to the packing of atoms than merely trying to obtain the most efficient packing possible. When covalent bonding becomes an issue, not only does it allow for many more structure types, but it also considerably reduces the importance of packing efficiency. In such cases, satisfying the natural bond geometries of the atoms, such as the tetrahedral bond coordination of atoms such as carbon and silicon, is more important than packing efficiency. With a reduction in packing efficiency comes a reduction in density.

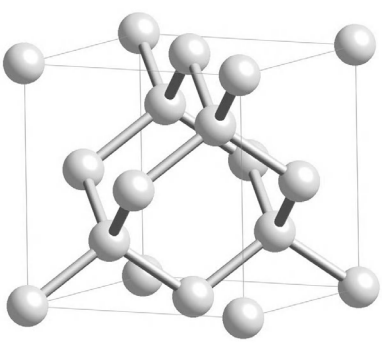


Fig. 2.12 Diamond structure. Each atom has four neighbours, as indicated by the bonds drawn between atoms. The lattice of the diamond structure is fcc, with eight atoms in the conventional cubic unit cell.

These points are highlighted by the crystal structures of the Group IV elements from carbon to tin, including the technologically important silicon. These have the **diamond structure**. The simplest way to describe the diamond structure is to start from the ccp structure. Take the group of atoms around the origin, with fractional coordinates 0, 0, 0; $\frac{1}{2}$, $\frac{1}{2}$, 0; $\frac{1}{2}$, 0, $\frac{1}{2}$; and 0, $\frac{1}{2}$, $\frac{1}{2}$. The midpoint has coordinates $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, and the coordination of this site with its neighbours forms a perfect tetrahedron. There are eight sites of this type in the ccp unit cell, and in the diamond structure half of these sites are occupied by atoms preserving the face-centre symmetry. The resultant structure is shown in Fig. 2.12 (the calculation of the packing efficiency is left to Problem 2.4).

The origin of the diamond structure arises from the electronic structure of these elements, which favours tetrahedral bonding. In forming compounds, carbon atoms favour tetrahedral bonding to other carbon atoms or hydrogen in organic molecules, and silicon usually bonds to four oxygen atoms in silicate oxides. This bonding gives both biology and geology!

Tin: an example of polymorphism

Tin, like iron mentioned earlier, is an example of a material that can exist in more than one structure. Different phases are also called **polymorphs**, and the ability of a structure to exist with different polymorphs is known as **polymorphism**. In the case of tin, the transformation between the two structures is produced by changing temperature, since the energy balance between the two structures is changed by heating or cooling through 260 K. The low-temperature phase,

known as **grey tin**, has the diamond structure ($a = 6.49$ Å). The high-temperature phase, known as **white tin**, has the structure shown in Fig. 2.13. This structure has orthogonal axes, with the two axes in the plane of the figure having equal lengths ($a = b = 5.82$ Å), but with the third axis being somewhat shorter ($c = 3.17$ Å). The structure has a 4-fold rotation axis parallel to the c-axis. This is an example of a **tetragonal** lattice (actually defined by the presence of the 4-fold rotation axis). In this specific case, there is a lattice point at the centre of the unit cell, so this is an example of a **body-centred tetragonal** lattice.

The two phases of tin have different electronic properties. Like silicon, the diamond form of tin is a semiconductor, whereas white tin is a metal. The change in electronic structure is likely to be an important factor in the energetics of the polymorphism of tin.

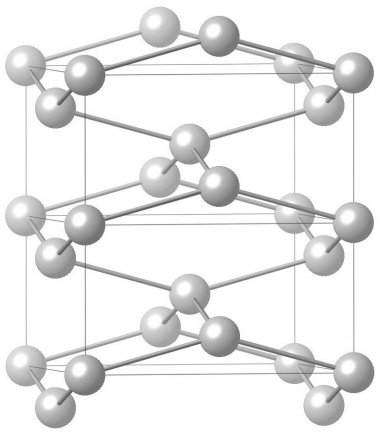


Fig. 2.13 Tetragonal structure of white tin. Each atom has four neighbours, but the coordination is more distorted than in the diamond structure. This structure can be viewed as a heavily distorted cubic close-packed structure.

Layered carbon and other forms of carbon

Carbon is another element to show polymorphism. The most familiar form of carbon is not diamond, but graphite. Diamond is only stable as a high-pressure phase. In the graphite structure there are layers of covalently bonded atoms. These form hexagonal layers, and the full crystal structure consists of these layers stacked on top of each other with a two-layer repeat. The graphite structure is shown in Fig. 2.14 ($a = 2.46$ Å, $c = 6.80$ Å). In this structure the carbon atoms form the hexagonal rings of the benzene molecule, with resonating double bonds.

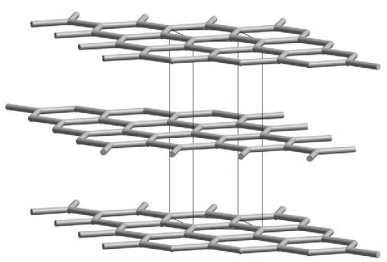


Fig. 2.14 Representation of the graphite structure, showing the bonds between atoms, with the atom positions corresponding to the points where three bonds meet. The layer structure is particularly clear in this representation.

The ability of carbon to form ring structures goes far beyond the graphite structure. Carbon is able to form molecular structures containing hexagonal and pentagonal rings of atoms, which are linked to form curved surfaces as ball-shaped molecules or long tubes. The most famous molecule of carbon is the fullerene C_{60}, Fig. 2.15, where the molecule looks like a football. These molecules crystallise in a fcc structure, where each molecule lies on a lattice point in exactly the same way that individual atoms do in the ccp structures seen earlier. We now encounter another aspect of the importance of symmetry. The sites in the ccp structure have the intersection of 2-fold, 3-fold and 4-fold rotation axes. However, the fullerene molecule has a different symmetry, as highlighted by the existence of the rings of five atoms (there are no 4-fold rotation axes in the molecule). The differences in symmetry can only be reconciled if the fullerene molecule is aligned in several different ways, flipping between different orientations, so that the average of the atom positions over all orientations has the same symmetry of the site. This situation is more common than might have been guessed, and we will encounter other examples below.

2.2.5 Diatomic molecular structures and other molecular elements

Diatomic molecules

The example of C_{60} showed that some elements can form crystal structures that involve packing of discrete molecules. The simplest examples are H_2, N_2 and O_2. Although the molecules appear to have shapes like dumbells, with atoms at either ends of bonds, in practice the outer electrons give a nearly spherical shape

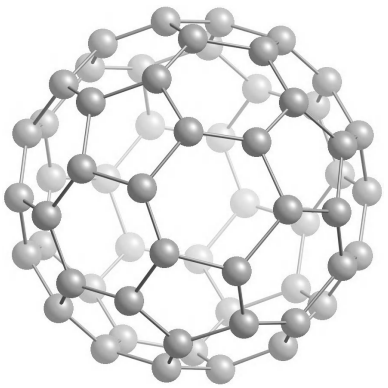


Fig. 2.15 Fullerene molecule, C_{60}, showing bonds between atoms that create rings of five and six atoms.

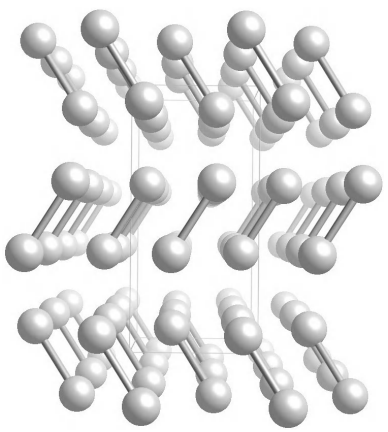


Fig. 2.16 Crystal structure of molecular iodine, I_2, viewed down the **a** axis.

to the molecules. Like C_{60}, the molecules pack into close-packed structures with orientational disorder of the molecules. For example, N_2 molecules crystallise in a hcp structure on cooling below the freezing temperature of 77 K. There is a phase transition at 35 K to a cubic structure in which the orientations of the N_2 molecules are ordered.

The molecules of the halogen atoms have a clearer dumbell shape, and the crystal structures of these elements reflect this by using a **herring bone** type packing. The example of crystalline iodine, I_2, is shown in Fig. 2.16, and other halogen molecules form similar crystal structures. The axes of this structure are orthogonal, but the three lengths of the unit cell are not equal ($a = 7.27$ Å, $b = 9.79$ Å, $c = 4.79$ Å). This structure type is called **orthorhombic** (we will meet the symmetry characteristics of the orthorhombic system in Chapter 3). In the case of the iodine structure, there are lattice points at the origin, 0, 0, 0, and at $\frac{1}{2}$, 0, $\frac{1}{2}$. Thus there is a lattice point at the centre of the *B*-face, namely the face of the unit cell in the **a**–**c** plane. Thus the unit cell is not primitive, and the lattice is called *B*-**centred orthorhombic**.

There is a lot of interest in what happens to these crystal structures at high pressures. On increasing pressure, the molecules get closer together until the distances between atoms from neighbouring molecules become close to the distances between the atoms in a molecule. When this happens the definition of the molecule becomes less distinct, and the description of the crystal becomes more of a case of packing of individual atoms. When this happens, it is likely that the crystal becomes a metal. Metallization of some of the halogen crystals has been observed, and one of the key objectives of current high-pressure research is to observe the metallization of hydrogen.

Other elements

A number of elements crystallize with structures that are rather more complicated. One example is sulphur, which can exist in two distinct structures, one **monoclinic** (where the three edge lengths of the unit cell are unequal, and two of the lattice vectors are not orthogonal), and the other having an orthorhombic unit cell containing distinct S_8 molecules. This is another example of polymorphism. The structure of orthorhombic molecular sulphur is shown in Fig. 2.17.

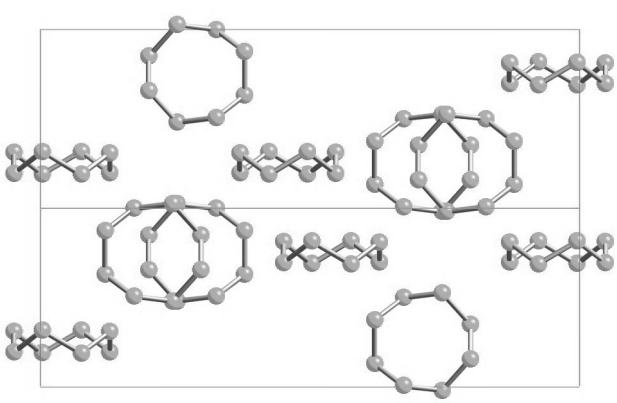


Fig. 2.17 Crystal structure of molecular sulphur, showing S_8 molecules. The view is along the diagonal between the **a** and **b** axes – this viewing direction gives a particularly clear view of the molecular structure.

Selenium and tellurium form hexagonal structures containing spiral chains of atoms that are bonded to two others. This structure is shown in Fig. 2.18. These are not the only phases of these two elements.

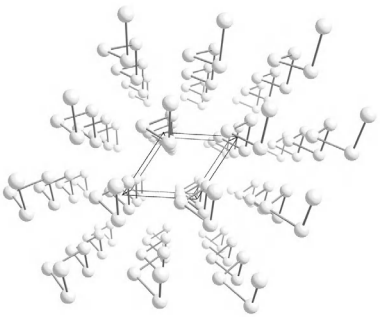


Fig. 2.18 Crystal structure of selenium as viewed down the hexagonal **c** axis. The bonds between atoms are drawn to highlight the spiral chain structure.

2.2.6 Summary of the structures of the elements

The crystal structures of the elements have shown a number of important principles that hold for the crystal structures of the compounds. One is the issue of close packing to maximize the density, as in ccp and hcp, and to a lesser extent in bcc. We have also seen the competing role of covalent bonds in creating structures with different bond geometries, such as in the diamond and graphite structures, and as also seen in selenium and the molecular crystals. These will have a lower density than the possible close-packed structures that could have been formed from these elements, but in crystals such as the diamond structure the covalent bonds can make the structures mechanically stiffer than the close-packed structures.

The elements show a wide diversity of possible structure types. We have begun to explore the role of symmetry in these structures, and have seen the possibility of phase transitions between different structure types. One feature that may have been unexpected is the possibility of the crystals composed of discrete molecules to have structural disorder (such as C_{60} and N_2) in order to achieve higher-symmetry molecular packing. The wide diversity of structural possibilities is even more pronounced in the crystal structures of the compounds, which we now move on to discuss.

2.3 Crystal structures of some simple inorganic compounds

2.3.1 Diatomic compounds

We recall the point made earlier in this chapter that a crystal structure has two main components, namely the lattice that gives the periodicity and the contents of the unit cell associated with each lattice point. For the ccp and bcc structures, the primitive unit cell contained just one atom. However, for hcp and the diamond structures, the unit cell contains two atoms, and for other structures of the elements there are even more atoms in the unit cell. All the structures we now consider have more than one *type* of atom in the unit cell. All the principles we have discussed with regard to the crystal structures of the elements, such as close packing, atomic coordination and symmetry, will continue to be important in our discussion of the crystal structures of more complex materials.

The most common structures of the diatomic compounds are all cubic and named after representative examples, namely the NaCl, CsCl and ZnS structures. These three structures are shown in Fig. 2.19. We will see that it is possible to describe the crystal structures in several ways. The formal way is simply to list the positions of atoms in the unit cell, using fractional coordinates. Whilst this is useful, it doesn't flag any special insights. Instead, we can also think about the coordination around the atoms, and the types of sites occupied by the atoms.

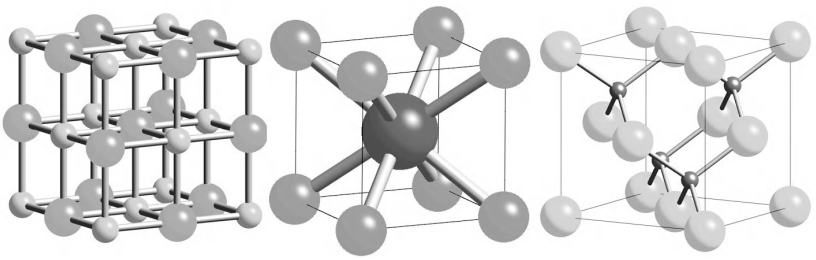


Fig. 2.19 Common crystal structures of diatomic compounds: *left*, NaCl structure; *centre*, CsCl structure; *right*, cubic ZnS structure.

NaCl structure

The **NaCl structure** has a fcc lattice. There is one type of atom at position 0, 0, 0 and another at position $\frac{1}{2}, 0, 0$. The other atoms in the conventional cubic unit cell are generated by addition of the three vectors to the face centres, namely by addition of $\frac{1}{2}, \frac{1}{2}, 0$; $\frac{1}{2}, 0, \frac{1}{2}$; and $0, \frac{1}{2}, \frac{1}{2}$ to the two initial atomic positions. The NaCl structure is shown in Fig. 2.19. An easy way to describe this structure is to begin with one atom type in a ccp arrangement. There is a set of equivalent empty sites with coordinates $\frac{1}{2}, 0, 0$; $0, \frac{1}{2}, 0$; $0, 0, \frac{1}{2}$; and $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$. These sites are at the centres of groups of six ions with perfect octahedral arrangement. In the NaCl structure these sites are occupied by the second set of atoms. As a result of the symmetry, both types of atom have six neighbours of the other type with perfect octahedral arrangement.

CsCl structure

The **CsCl structure** appears to related to the bcc packing we met earlier. The unit cell is a cube, with an atom of one type at the origin, and another type of atom in the centre of the cube. Because the two atoms are different, the site at the centre of the unit cell is no longer another lattice point, and the CsCl structure must not be confused with the bcc lattice. The lattice type is therefore primitive cubic. The CsCl structure is shown in Fig. 2.19. Each atom has eight neighbours of the opposite type.

Cubic ZnS structure

The **cubic ZnS structure** is related to the diamond structure. The lattice is fcc. In the diamond structure we started from the ccp structure, and added a second set of atoms on one of the two sets of tetrahedral sites. In the ZnS structure, one type of atom occupies the initial ccp arrangement, with coordinates 0, 0, 0 etc., and the other type of atom occupies one of the set of tetrahedral sites at $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ etc. The ZnS structure is shown in Fig. 2.19.

There is also a hexagonal polymorph of ZnS, called **wurtzite**. The origin of this polymorph can be considered as an **ABAB** stacking of ZnS double layers, compared with the **ABCABC** stacking along the [111] direction in the cubic polymorph. This is similar to the cubic close-packed metals. SiC is a compound that adopts the cubic ZnS structure but which can exist in many different polytypes arising from complex stacking sequences of the double layers.

2.3.2 Ionic packing

The concept of the ionic radius

Some aspects of the structures of diatomic compounds can be rationalized using the concept of the **ionic radius**. The central idea is that the distance between two nearest-neighbour ions, such as between the Na^+ and Cl^- ions in NaCl, is approximately equal to the sum of the nominal radii of the two ions. These radii can be obtained by systematic comparisons of interionic distances in a wide range of different materials. A comprehensive tabulation of rules of ionic radii can be obtained from http://www.webelements.com/, from which all values used in this book have been taken.

The idea that there is an effective ionic radius suggests that we can think of ions as hard spheres in contact with neighbouring hard-sphere ions. However, the same types of systematic comparisons suggest that the effective ionic radius will depend on the coordination of the ion, particularly for tetrahedral and octahedral coordination. For example the Mg^{2+} cation has an effective radius with values 0.71 Å, 0.86 Å and 1.04 Å for four, six and eight coordination. We therefore have to be careful not to overinterpret the idea of ionic radius. The main value of the concept of the ionic radius is that when packing atoms together in a crystal, the relative size of atoms is an important factor, and will give some understanding of why some particular structures may be stable over alternatives.

Ionic radii and ionic packing

One of the most important applications of the concept of ionic radius is in predicting the stabilities of crystal structures. To illustrate this we consider a simple crystal containing two ions where the ionic radii of the cation and anion are r^+ and r^- respectively. The basic principle for structure stability is that the ions in the crystal structures should pack together in such a way that the ions of the same type should not be 'touching'. We illustrate this with reference to the NaCl structure. The distances between neighbouring cations or neighbouring anions is $a/\sqrt{2}$, and the distance between neighbouring cations and anions is $a/2$. The geometry is shown in Fig. 2.19. In the extreme case where both the cation–cation and cation–anion distances correspond to the sum of the ionic radii, we have

$$\begin{aligned} a/\sqrt{2} = 2r^+ \qquad a/2 = r^+ + r^- \\ \Rightarrow r^+/r^- = \sqrt{2} - 1 \end{aligned} \tag{2.3}$$

A similar relation holds for the other extreme where the anion–anion distance is equal to $2r^-$. Taking both extremes into account, the NaCl structure is stable (in the sense that like ions are not touching) when the following condition holds:

$$\text{NaCl structure:} \quad \frac{1}{\sqrt{2}-1} < r^+/r^- < \sqrt{2} - 1 \tag{2.4}$$

A similar analysis for the CsCl structure gives

$$\text{CsCl structure:} \quad \frac{1}{\sqrt{3}-1} < r^+/r^- < \sqrt{3} - 1 \tag{2.5}$$

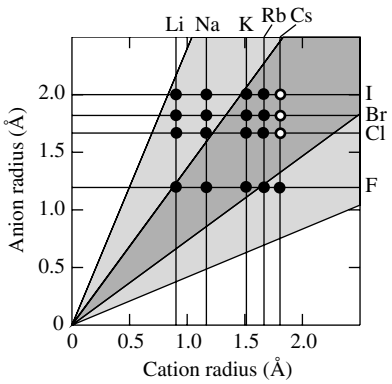


Fig. 2.20 Plot of ionic radii of family of alkali halide structures, showing the region of the stability of the CsCl structure, inner shaded area, and the region of stability of the NaCl structure, outer shaded area. Experimental structures are shown as closed circles for the NaCl structure and open circles for the CsCl structure.

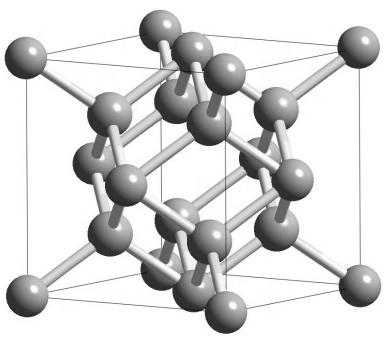


Fig. 2.21 Crystal structure of fluorite, CaF_2.

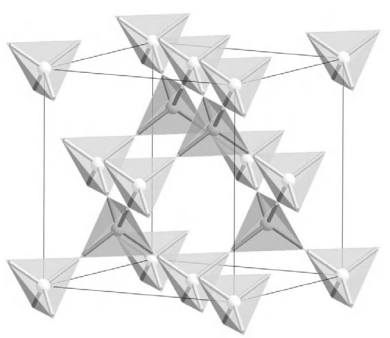


Fig. 2.22 Crystal structure of cristobalite, a polymorph of silica with a cubic structure at high temperature. This representation emphasizes the network of corner-linked SiO_4 tetrahedra: the Si cations lie at the centres of the tetrahedra, and the oxygen atoms lie at the vertices.

These results are illustrated for the alkali halides in Fig. 2.20. It can be seen that the Cs halides with the CsCl structure lie within the stability regime for this structure, and all other alkali halides fall within the stability condition for the NaCl structure. Some structures with the NaCl structure are within the stability condition for the CsCl structure – the two conditions are not exclusive of each other.

2.3.3 Crystals with general formula AX_m, and the general idea of coordination polyhedra

Fluorite structure

The general principles explored in our discussion of diatomic crystals can readily be extended to more complex cases. Here we develop the discussion to encompass compounds of general formula AX_m. The crystal structures of these compounds follow the trend in giving an even wider diversity of crystal structures. The simplest structure to understand is the **fluorite structure**, named after the mineral with formula CaF_2. We start with the A ion in a ccp arrangement. When discussing the diamond and ZnS structures, we noted that there are two sets of tetrahedral sites in the ccp arrangement. In the diamond and ZnS structures only one set is occupied, but in the fluorite structure both sets of tetrahedral sites are occupied by the X ion, as shown in Fig. 2.21. The fluorite structure retains the fcc lattice. The conventional cubic unit cell contains 12 atoms, with three atoms in the primitive unit cell. CaF_2 and ZrO_2 are two examples of crystals with the fluorite structure.

The crystal structure of the cristobalite polymorph of SiO_2

The crystal structure of the cristobalite polymorph of silica, SiO_2, is the next easiest one of this form to understand. This structure is based on that of elemental silicon with the diamond structure. The cristobalite structure is then formed by placing an oxygen atom half-way between each neighbouring pair of silicon atoms. The lattice type remains as fcc, and there are 24 atoms in the cubic unit cell (six atoms in the primitive unit cell). Although each oxygen atom has two neighbouring silicon atoms, each silicon atom has four neighbouring oxygen atoms in perfect tetrahedral coordination. The crystal structure of cristobalite is shown in Fig. 2.22 with a representation that emphasizes the existence of the SiO_4 tetrahedra as basic structural units.

The description of the crystal structure of cristobalite given above implies that the connection Si–O–Si is a straight line. Indeed, the simplest analysis of diffraction data (using the methods described in Chapter 6) gives the average positions of the atoms in this coordination. However, because silicon and oxygen are such common elements, there is a large variety of natural materials whose crystal structures contain the Si–O–Si connectivity, and invariably the connection is *not* straight but bent with an angle of around 145° subtended at the silicon atom. The natural tendency of the Si–O–Si connection to be bent can be reconciled with the ideal cristobalite structure if it is supposed that the linear Si–O–Si connection is merely an illusion obtained by the oxygen atom moving

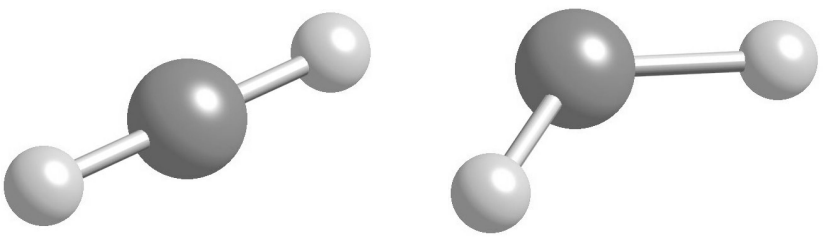


Fig. 2.23 Si–O–Si connections, showing the potential disorder of the position of the oxygen atom. The diagram on the *left* shows the configuration of atoms in their average positions, with apparently linear Si–O–Si bonds. The diagram on the *right* shows the distortion of this linkage to give an Si–O–Si angle with the usual value of 145°. In the cubic phase of cristobalite, it is probable that the oxygen atoms are constantly moving out of alignment to give an instantaneously bond angle close to the energetically preferable angle, but the average over all positions is still half-way between the two silicon atoms.

around so as to give instantaneous Si–O–Si connections that are bent, but with the oxygen atom having an average position exactly half-way between the two silicon atoms. This interpretation is shown in Fig. 2.23. We will develop the discussion of this disorder later in this chapter.

Coordination polyhedra

The example of cristobalite shows that it can often be useful to highlight the atomic coordination in describing the structure of a polyatomic crystal. In this case we can consider the SiO_4 tetrahedron to be a structural unit in its own right, and to then think of the whole crystal structure in terms of a three-dimensional network of corner-linked SiO_4 tetrahedra. In the case of cristobalite, each oxygen atom is shared by two tetrahedra, but this is not a necessary condition. The idea of representing structures in terms of component **polyhedra** is very powerful. It can be applied to a great variety of structures, and actually allows us to understand the relationship between different polymorphs of compounds.

The importance of the whole family of silica structures cannot be over-stated. One of its polymorphs, quartz, is the most common piezoelectric material, often used as the oscillator in electronic clocks. Silica is very important for geology, being the principle component of sand and an important phase in many rock types. The crystal structures of all the lower-pressure phases can be described as networks of corner-linked SiO_4 tetrahedra, there being an unlimited number of ways of forming such networks. Common to many of these structures are rings of tetrahedra, often as four-membered or six-membered rings (cristobalite contains such rings), or spirals of tetrahedra as in quartz. Silica glass (which we discuss later in this chapter) also has a network of SiO_4 tetrahedra, and in fact the structure of silica glass has many features in common with the crystalline phases.

At higher pressures, silica forms a structure, called **stishovite**, which has a higher density that is achieved by having the silicon atoms in 6-fold coordination with respect to the oxygen atoms. This structure is similar to that of the **rutile** phase of TiO_2, after which the structure type is generally known. The rutile structure is shown in Fig. 2.24. It consists of chains of TiO_6 (or SiO_6) octahedra linked along the edges, and with the chains linked by the sharing of the corners of octahedra.

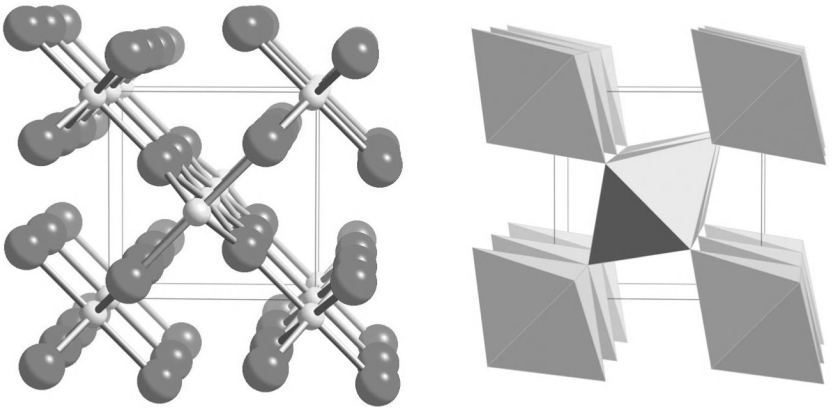


Fig. 2.24 Rutile structure of TiO_2, showing atoms (*left*) and TiO_6 octahedra (*right*).

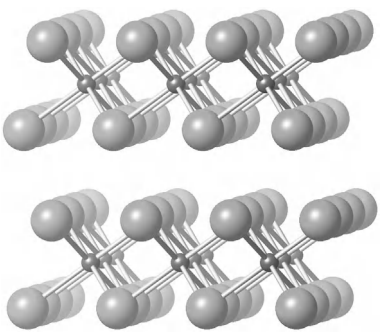


Fig. 2.25 Crystal structure of CdI_2. The bonds link the atoms within CdI_6 octahedra, which are connected within layers.

TiO_2 also has polymorphism, since in addition to the rutile phase there is also a phase called **anatase**, which has a tetragonal structure, and a phase called **brookite**, which has an orthorhombic structure. Both these phases also consist of TiO_6 octahedra as the principal structural building block.

Many other examples of AX_2 compounds form crystal structures, most of which can be described in terms of connected coordination polyhedra. There are many ways in which octahedra can pack together. Neighbouring coordination octahedra can share corners, edges (as in the rutile structure), or faces, and crystal structures containing coordination octahedra can use any combination of these. This is an important structural principle, and enables us to rationalize the wide diversity of crystal structure types in more complex compound materials. Figure 2.25 shows the example of CdI_2. This structure is shared by several dihalide crystals. It consists of layers of edge-sharing CdI_6 octahedra in a hexagonal lattice. There are several ways that the layers can stack above each other, and polytypism is very common in this family of structures.

Generalization to more complex compounds

The way of visualizing crystal structures in terms of coordination polyhedra can be applied to more complex structures. The examples of Al_2O_3 and ZrP_2O_7 are shown in Fig. 2.26. Al_2O_3 (called corrundum, but also known as ruby when containing the Cr^{3+} impurities which give rise to its red colour), consists of AlO_6 octahedra that are linked on faces and edges to form a very rigid network. ZrP_2O_7 is a representative of a class of materials with low or negative values of thermal expansion. Its structure consists of ZrO_6 octahedra and PO_4 tetrahedra linked in a network of corner-sharing polyhedra.

Among the most amazing structures are the **zeolite** phases. These consist of networks of corner-linked SiO_4, AlO_4 and PO_4 tetrahedra, as in the silica phases cristobalite and quartz, but the zeolite structures are surprisingly open with a low density. These structures have large cavities with connecting channels. An example is shown in Fig. 2.27. Various cations and molecular groups can be held within the cavities, with mobility between the cavities through the channels. These structural features enable zeolites to be used for a wide range of unique applications, including uses as molecular sieves or as catalysts for organic reactions. Zeolites are also used in washing detergents

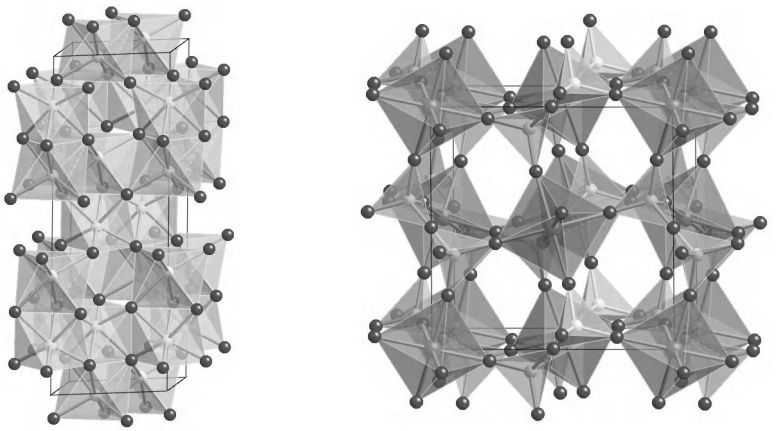


Fig. 2.26 Crystal structures of Al_2O_3 (*left*) and ZrP_2O_7 (*right*), highlighting the way in which the structures can be described in terms of a network of linked polyhedra: face- and edge-linked AlO_6 octahedra in the case of Al_2O_3, and corner-linked ZrO_6 octahedra and PO_4 tetrahedra in the case of ZrP_2O_7.

because their structures can absorb molecules of dirt. The framework can be relatively flexible, and this allows many zeolites to have negative thermal expansion.

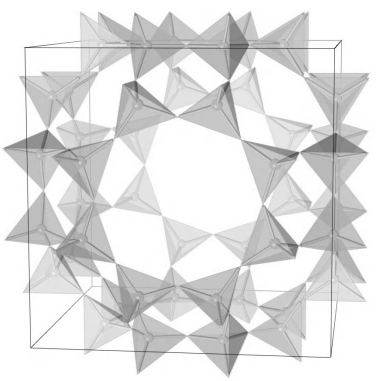


Fig. 2.27 Crystal structure of a zeolite, constructed as a three-dimensional network of corner-linked SiO_4 tetrahedra with large cavities.

2.4 The perovskite family of crystal structures

2.4.1 The ideal perovskite structure

We have already met a number of technologically important materials, and in this section we will introduce one very special group of materials that have had far-reaching applications across many aspects of modern life. This group is the **perovskite** family, which at first sight appears to be based on a deceptively simple crystal structure. The *ideal* perovskite structure is shown in Fig. 2.28. The chemical formula is represented by ABO_3, where A and B are cations. The A cation has fractional coordinates $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$, the B cation is on the origin, and the oxygen atoms have coordinates $\frac{1}{2}, 0, 0; 0, \frac{1}{2}, 0$; and $0, 0, \frac{1}{2}$ (the structure can also be drawn with the origin on the A site instead). From Fig. 2.28 it can be seen that the B site is in octahedral coordination with the neighbouring oxygen anions, and neighbouring BO_6 octahedra are joined at corners to form a network that runs throughout the entire crystal. The A sites have 12 oxygen neighbours. Common examples have Ti and Zr on the B site, with Ca, Ba, Sr or Pb on the A sites. Examples also exist with A and B cations having the same ionic charge (e.g. $LaAlO_3$), or with ionic charges 1+ and 5+ (e.g. $KMnO_3$). There are a few examples where the A site is empty (e.g. WO_3). The perovskite structure is very flexible in terms of the chemical variations it can tolerate.

The important feature of the perovskite structure is that there are many structural phase transitions, and these are controlled by the details of the chemical composition. To see why phase transitions are so common in the perovskite family of structures, it is useful to consider the ideal structure in terms of the ionic radii, which we label r_A, r_B and r_O. The cubic structure has only one degree of freedom, namely the unit cell parameter a. The nearest-neighbour

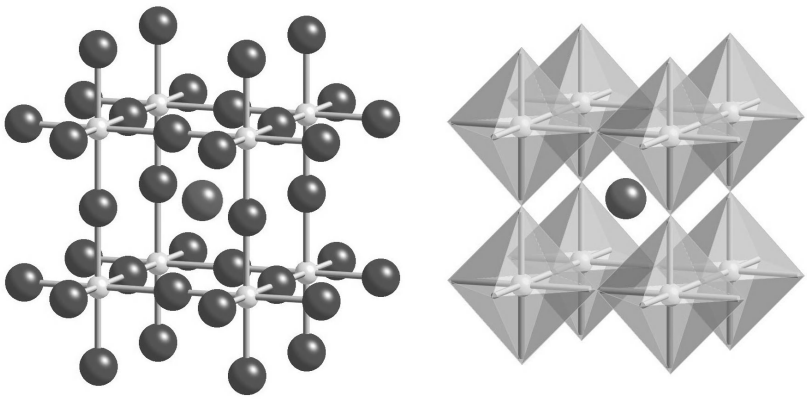


Fig. 2.28 The ideal perovskite structure, shown as atoms with bonds (*left*) or as a network of corner-linked octahedra (*right*).

Table 2.1 Ionic radii of cations of valence 2 and 4 that readily form perovskites. The radii are combined with the ionic radius of the oxygen anion (1.25 Å) in a way that gives comparison with the lattice parameter a. Data are from http://www.webelements.com/.

A cation	r_A (Å)	$\sqrt{2}(r_A + r_O)$	B cation	r_B (Å)	$2(r_B + r_O)$
Ca^{2+}	1.14	3.38	Zr^{4+}	0.86	4.22
Sr^{2+}	1.32	3.63	Ti^{4+}	0.75	4.00
Ba^{2+}	1.49	3.88			
Pb^{2+}	0.92	3.08			

distances are

$$d_{A–O} = a/\sqrt{2}$$
$$d_{B–O} = a/2$$
$$d_{O–O} = a/\sqrt{2}$$

All three distances are determined by the value of the lattice parameter a. If there was perfect matching between the ionic radii, we would have $d_{A–O} = r_A + r_O$, $d_{B–O} = r_B + r_O$, and perhaps even $d_{O–O} = 2r_O$. Clearly nature is not going to provide the perfect size matching across the wide range of pairs of A and B cations that can occur within the perovskite structure. Thus there will usually be a size mismatch, and this is what allows the structure to distort at a structural phase transition. And it is the ability of the structure to distort that in turns allows the perovskite structure to be formed with so many pairs of cations, and extended to include mixtures of more than two types of cations.

The point about the ionic radii is illustrated by the comparison of a group of perovskites in Table 2.1. The table gives ionic radii with the distances $\sqrt{2}(r_A + r_O)$ and $2(r_B + r_O)$ assuming an oxygen ionic radius of 1.25 Å. By comparison, $2\sqrt{2}r_O = 3.53$ Å. For perfect fitting of the ions, these three distances will be equal. The largest distance will determine the size of the unit cell, and the smaller of the distances involving the A and B cations will partly determine possible distortions of the structure. It can be seen from Table 2.1 that the most significant distance in determining the lattice parameter is that due to the B–O distance. Indeed, for $PbTiO_3$ the cubic lattice parameter (3.97 Å) is almost

exactly equal to twice the sum of the Ti^{4+} and O^{2-} ionic radii. The A cation in each case is smaller than the site it occupies, and this will allow for distortions of the structure involving rotations of the BO_6 octahedra.

The flexibility in the structure given by the conditions of ionic radii allow a number of different **displacive phase transitions** to occur. Displacive phase transitions are phase transitions that involve small displacements of atoms in such a way that the symmetry of the crystal is changed. These transitions are not like many of the polymorphic transitions we have already encountered, such as these between the phases of iron or tin. In those cases the changes in the structures are very distinct, involving changes in coordination and bonding. In displacive phase transitions the local environment of any atom only changes slightly, and there is a very clear relationship between the structures of the two phases.

2.4.2 Ferroelectric phase transitions in perovskites

The most important phase transition in the perovskite structure for modern technology is the ferroelectric phase transition. This type of phase transition is exemplified by $PbTiO_3$. This has the ideal cubic perovskite structure for temperatures above 763 K. Below this temperature, the structure is tetragonal. The cations and anions are displaced in opposite directions along the tetragonal **c** axis, as illustrated in Fig. 2.29. As a result of these displacements, a number of symmetry elements of crystal structure are lost. These include the 3-fold rotation axes, the 4-fold rotation axes perpendicular to **c**, the mirror plane normal to **c**, and the centre of symmetry. As a result of these changes in symmetry, the unit cell of the tetragonal phase has a small electric dipole moment, given in terms of the ionic displacements as

$$p = \sum_j q_j u_j^{(z)} \tag{2.6}$$

where q_j is the charge of ion j, and $u_j^{(z)}$ is the displacement of ion j along the direction of the tetragonal **c** axis. The sum is over all atoms in the unit cell. The dipole moments of all the unit cells are combined to give a net electrical polarization of the crystal along the tetragonal **c** axis. In $PbTiO_3$ the polarization P at room temperature is around 0.75 C m^{-2} (see Fig. 2.30). If this is caused by equal displacements of the two cations relative to the oxygen anions (a gross oversimplification which also neglects the role of the polarization of the ions), the polarization can be calculated as

$$P = \frac{6eu}{a^3} \tag{2.7}$$

where u is the displacement, a^3 is the volume of the unit cell (for the cubic phase $a = 3.97$ Å), and $6e$ is the total charge of the two cations. The volume is included because the definition of P is *dipole moment per unit volume*, with units of *charge per unit area*. Using these values, we obtain $u = 0.5$ Å. Because we have not included the effects of polarization, this is likely to be an overestimate, but it is reasonably close to experimental values ($u_{Pb} = 0.46$ Å, $u_{Ti} = 0.30$ Å).

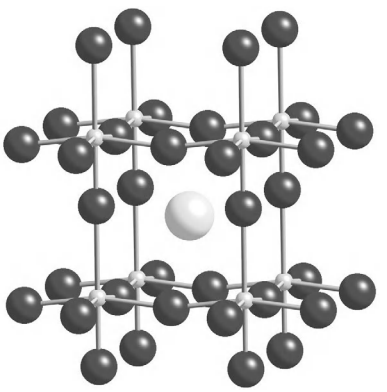


Fig. 2.29 Ferroelectric distortions of $PbTiO_3$, showing counter-displacements of the cations and anions along the direction of the *c*-axis.

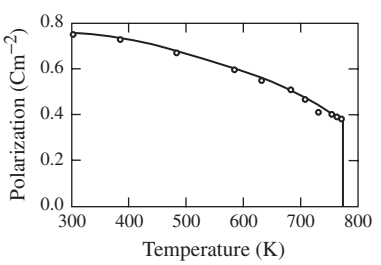


Fig. 2.30 Temperature dependence of ferroelectric polarisation of $PbTiO_3$.

Pyroelectricity in ferroelectric materials

The polarization induced by the ferroelectric phase transition varies with temperature. For $PbTiO_3$, the temperature dependence of the polarization is shown in Fig. 2.30. Any crystal in which the electrical polarization varies with temperature is called **pyroelectric**. There are symmetry restrictions that determine whether a material is pyroelectric. First, the crystal must not have a centre of symmetry, but this is not a sufficient condition. It is also essential that the combination of mirror planes and rotation axes does not cancel out the overall polarization. Any material that fulfils the symmetry conditions will have a macroscopic polarization, even if very small, and invariably this polarization will change slightly with temperature. In the case of ferroelectric materials, the variation of the polarization with temperature is particularly strong because of the existence of the phase transition.

The pyroelectric effect is not hard to measure in ferroelectric materials, and as a result it leads to important applications. The key point is that a small temperature change will lead to a measurable change in the electrical polarization, so that a ferroelectric pyroelectric material would make an excellent sensor of changes in temperature. In particular, ferroelectric materials can be used as sensors of body heat through the infrared radiation emitted by the body and detected through the subsequent heating of the ferroeletric pyroelectric. For $PbTiO_3$, a change in temperature of 1 K at an ambient temperature of 300 K will give rise to a change in polarization of 400 $\mu C\,m^{-2}$. This is easily measured.

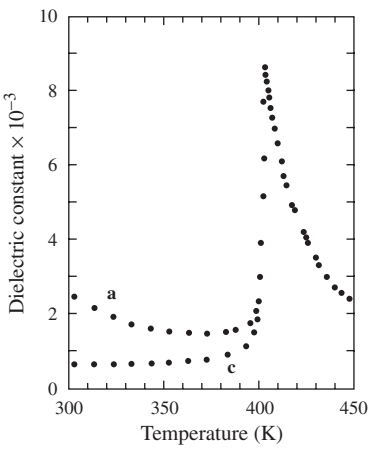


Fig. 2.31 Dielectric constant of $BaTiO_3$ in the vicinity of the cubic–tetragonal ferroelectric phase transition. In the ferroelectric phase there are differences along the **a** and **c** axes.

Dielectric properties

The polarization induced by an applied electric field is written as

$$P = \epsilon_0 \chi E = \epsilon_0(\epsilon - 1)E \tag{2.8}$$

where χ is the **dielectric susceptibility**, and $\epsilon = 1 + \chi$ is the **dielectric constant**. The susceptibility is the relevant coefficient for detailed analysis (as in Chapter 7), but the dielectric constant is more relevant for technical applications since capacitance is proportional to ϵ.

The important point is that the dielectric constant becomes very large for temperatures in the vicinity of a ferroelectric phase transition. This is illustrated in Fig. 2.31 (see also Fig. 1.8).

2.4.3 Rotational phase transitions

The second important type of phase transition in the perovskite family of structures involves small rotations of the BO_6 tetrahedra. The simplest example is $SrTiO_3$. In this case the TiO_6 octahedra rotate about one of the axes to give a transformation from the cubic structure to a tetragonal structure. Alternate layers of octahedra normal to the **c**-axis rotate in opposite directions. The structures of the two phases of $SrTiO_3$ are illustrated in Fig. 2.32. More complex rotations are possible in other perovskites. For example, in $CaTiO_3$ the low-temperature structure has rotations of the TiO_6 octahedra by different angles about all three crystal axes. The more complex rotations are correlated with the decreasing size of the A cation, as seen in Table 2.1.

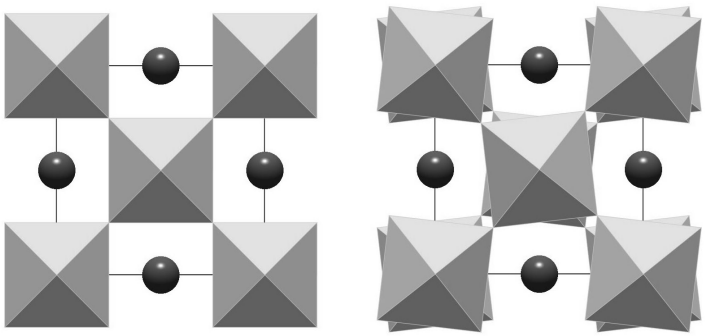


Fig. 2.32 Rotations of the TiO_6 octahedra associated with the displacive phase transition in $SrTiO_3$. The left-hand figure shows the structure of the cubic phase, and the right-hand picture shows the structure of the tetragonal phase, with the octahedra in the two layers down the direction of view having opposite rotations.

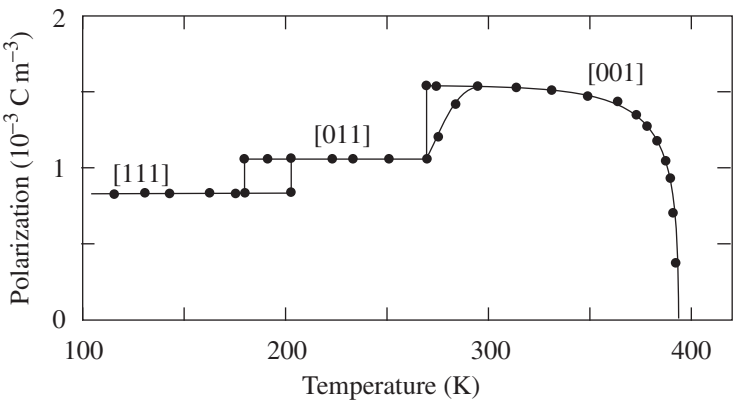


Fig. 2.33 Ferroelectric polarization in $BaTiO_3$ along the direction of the **c** axis of the cubic and tetragonal phases. The changes are due to the succession of phase transitions, which involve switching of the polarization directions. [001] denotes polarization along the **c** axis, [011] denotes the direction along the **b**–**c** cube face diagonal, and [111] denotes the direction along the body diagonal (these symbols will be explained in Chapter 3). The transition temperatures differ on heating or cooling, a process known as **hysteresis**.

2.4.4 Effects of chemical variation

The effects of chemical composition are quite dramatic. For example, $PbTiO_3$ has a single ferroelectric phase transition, whereas $BaTiO_3$ has a complex sequence of ferroelectric phase transitions. The first ferroelectric phase transition on cooling from the cubic phase is tetragonal, with the polarization along the **c** axis as in $PbTiO_3$. But on further cooling the subsequent phase transitions involve switching of the direction of the polarization, as shown in Fig. 2.33. On the other hand, $SrTiO_3$ would have a ferroelectric phase transition at 30 K, except that it is suppressed by quantum fluctuations (as we will discuss further in Chapter 12). As noted above, $SrTiO_3$ has a rotational phase transition involving rotation about a single axis, whereas $CaTiO_3$ has a series of rotational phase transitions involving three rotational axes. The fact that the phase transitions are so sensitive to chemical composition suggests that the properties of perovskite materials might be tailored by careful control of the chemical composition. This is exemplified by the solid solution $Pb(Zr,Ti)O_3$, known as PZT. The temperature–composition phase diagram of the PZT solid solution is shown in Fig. 2.34. The two main phases are the ferroelectric phases with electrical polarization along different directions, namely along the cube axis, labelled [001] (see Chapter 3), in the Ti-rich phase, and along the cube diagonal, labelled [111], in the Zr-rich phase. There are additional rotational phase transitions at

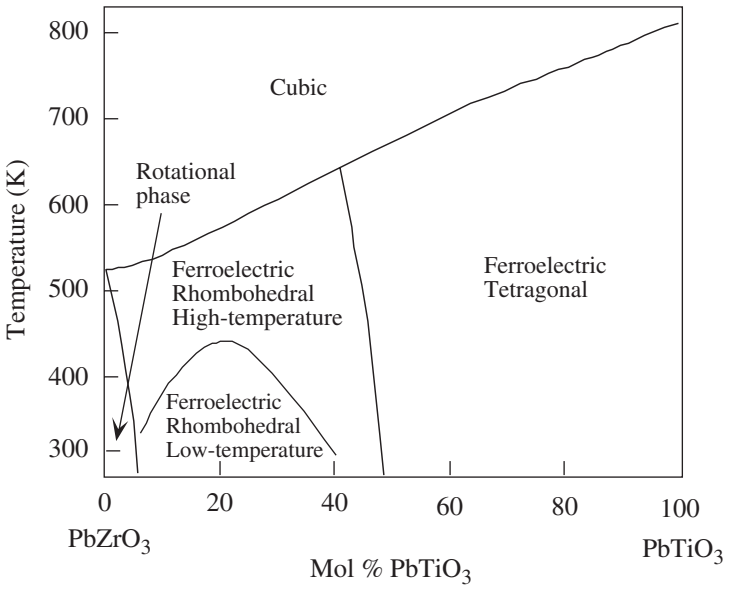


Fig. 2.34 Phase diagram for the solid solution $Pb(Zr,Ti)O_3$. The rotational phases on the Zr side of the phase diagram are too complicated to show all details.

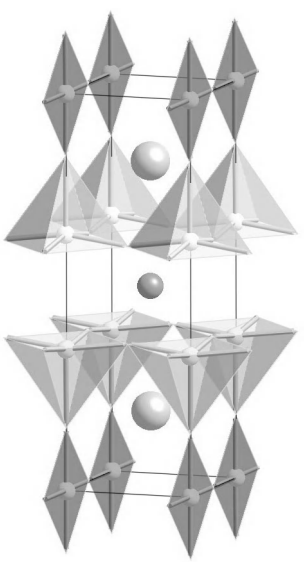


Fig. 2.35 Perovskite-based structure of the high-temperature superconducting ceramic $YBa_2Cu_3O_7$. The Y atom lies at the centre of the unit cell, the Ba atoms lie at heights $\pm 1/4$, and the Cu atoms occupy the sites with pyramidal and square coordination with the oxygen atoms.

the $PbZrO_3$ end of the phase diagram. The most important aspect of the phase diagram is the phase boundary between the two ferroelectric phases. This phase boundary is almost vertical, and is important because there is enhancement of the dielectric constant and piezoelectric constant at this composition. PZT ceramics with compositions around the phase boundary are amongst the most common ferroelectric ceramics for technological applications.

2.4.5 Variations on the perovskite theme

Perovskite structures are the simplest structures of a large group of related structures. For example, there are related structure with formulae of the form ABO_4, which have layers of corner-linked BO_6 octahedra but without connections between the layers. There are more complicated variants of this theme, for example with some of the oxygen sites unoccupied. Materials like this form the basis of the high-temperature superconducting cuprates. The example of $YBa_2Cu_3O_7$, the parent phase for one of the most important high-temperature superconducting ceramic materials, is shown in Fig. 2.35.

2.5 Organic crystals

We have already met a number of crystals that can be described as a packing of discrete molecules. The examples given earlier have been for molecules formed from one element only, such as C_{60} and N_2. However, it is not surprising that any molecule could be the basis of a crystal structure, whether a simple molecule such as methane, or a complicated molecule such as a protein or DNA. Close-packing principles are important, but because the molecules have complicated

shapes these principles lead to similarly complicated structures. Moreover, if the molecules do not have a centre of symmetry, the structures may pair molecules in opposition orientations to cancel the molecular dipole moments. Directional hydrogen bonds (see Chapter 5) may also have a significant impact on the form of the crystal structure. The crystal structure of one organic molecule, benzoic acid, $(C_6H_7)C(OH)_2$, is shown in Fig. 2.36. Molecules pair in opposite directions and link with hydrogen bonds, and pairs of molecules then pack together in an efficient arrangement.

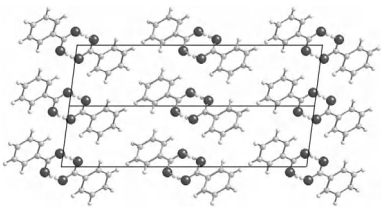


Fig. 2.36 Crystal structure of benzoic acid.

2.6 Disordered materials

2.6.1 The importance of structural disorder

Most of this book is concerned with crystalline materials, and the easier parts of the subject concern well-ordered crystalline materials. However, disordered solids are equally as important for intellectual as well as practical reasons. We have already met several possibilities for crystals to have structural disorder, some of which we will discuss below.

There is a wide range of types of disorder in materials, and any classification will be neither complete nor unambiguous. Some types of disorder can occur as an equilibrium condition, and others will occur through sample preparation or treatment. Some types of disorder will be effectively static, and some will involve dynamics over time scales that are close to the time scales for atomic vibrations. One could argue that atomic vibrations, which increase with temperature and are present in all materials, are a form of disorder. Certainly atomic vibrations are the dominant form of entropy in many materials — we will treat atomic vibrations as a separate topic in Chapters 8–10. However, there are materials where the amplitudes of atomic motions are so large that they can no longer be easily described as simple vibrations, and these are called **dynamically disordered** materials. We have already met some examples, and will give further examples below.

When we consider disorder, it is useful to take account of the length scale over which any order exists. In particular, it is useful to distinguish between **long-range order**, which effectively is the order that exists over length scales approaching infinity, and **short-range order**, which exists over length scales down to nearest-neighbour distances. Ordered materials are those for which the short-range order closely matches the long-range order.

We group some types of disorder under the label **defects**, and discuss these in Appendix A. Our rationale is based on the fact that the short-range order away from the defect sites closely matches the long-range order of a defect-free material. On the other hand, in the examples we discuss below, there are clear differences between long-range and short-range order.

2.6.2 Orientational disorder in molecular crystals

We met examples of orientational disorder when we discussed the crystal structures of C_{60} and N_2 above. In many cases, the molecules have higher symmetry than the average symmetry of the sites they occupy. This implies

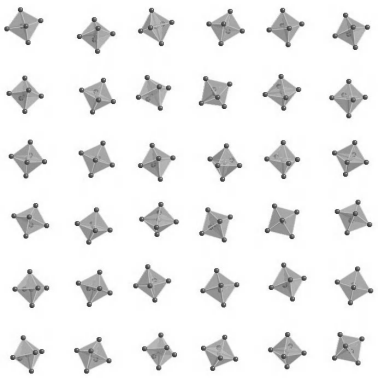


Fig. 2.37 Configuration of octahedral SF_6 molecules in the orientationally disordered phase. The crystal structure is bcc, and the configuration includes molecules lying on the plane that contains two lattice vectors. The average orientations will have the S–F bonds lying along the directions between the neighbours shown in this projection. However, short-range repulsions between closest fluorine atoms force the molecules to tumble, and several molecules can be seen with disordered orientations. (Data for figure from Matthew Tucker and the author.)

that the molecules must be rotating. The time scales for these rotations can be very fast, down to the order of 1 ps.

As an example, a configuration of orientationally disordered SF_6 molecules is shown in Fig. 2.37. The molecules have an octahedral shape, and are located on the lattice points of a bcc structure. In this case there is no mismatch between the molecular and site symmetries, but the molecules are forced to tumble because the F. . .F distances along the axes of the unit cell are too short to allow the orientations of the molecules to be ordered. The disordered phase exists between the melting point at 230 K down to 96 K, at which point there is a phase transition to an ordered structure.

It is interesting to view orientationally disordered materials as being intermediate states between ordered crystalline and liquid phases. The positions of the molecules are ordered as in a crystal, but the orientations are disordered as in a liquid.

2.6.3 Orientational disorder in framework structures

When we discussed the crystal structure of cristobalite earlier (Section 2.3.3), we noted that there must be short-range disorder of the bending of the Si–O–Si bonds in order to avoid forming linear bonds (Fig. 2.23). Since the SiO_4 tetrahedra are relatively stiff units, it can be imagined that any disorder of the alignment of atoms along the Si–O–Si bonds that link pairs of tetrahedra must involve rotations of the tetrahedra. Since the tetrahedra are linked, they cannot tumble as in orientationally disordered molecular crystals. However, they are able to rotate by large angles, and can cause considerable flexing of the network of linked tetrahedra. An example of a configuration of SiO_4 tetrahedra in the disordered phase of cristobalite is shown in Fig. 2.38. The disorder obviously creates a considerable degree of entropy, and this is reduced when the structure undergoes a phase transition to an ordered phase (this phase transition is briefly discussed in Chapter 12).

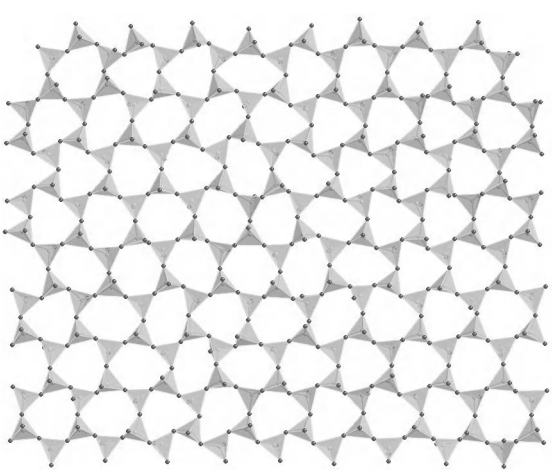


Fig. 2.38 Configuration of SiO_4 tetrahedra in the disordered phase of cristobalite. In this projection, the tetrahedra are arranged in a hexagonal network with neighbouring tetrahedra pointing up or down in opposite directions. In the 'average' structure, the space between the tetrahedra within the rings is a perfect hexagon. The flexing of the network through the rotations of the SiO_4 tetrahehdra to avoid the formation of linear Si–O–Si bonds is clearly seen through the large distortions of the hexagonal rings. (Data for the figure from Matthew Tucker, David Keen and the author.)

2.6.4 Fast-ion conductors

An important class of materials are those called **fast-ion conductors** or **superionic conductors**. These materials contain some ions that are extraordinarily mobile. In some cases this mobility means that the ions are able to hop between crystallographic sites, and in other cases it appears that the mobile ions almost form a fluid phase within the matrix of the crystalline order of the other ions.

The simplest example is AgI. At room temperature AgI has the cubic ZnS structure. At 420 K there is a phase transition to the fast-ion conducting phase with bcc structure. The I anions are located on the lattice points of the bcc structure, but the Ag cations have no fixed positions. There are a number of possible sites that can be occupied by the cations, and diffraction evidence (Chapter 6) indicates that the most important sites are the 12 with fractional coordinates with permutations of $\frac{1}{4}, 0, \frac{1}{2}$. Other sites are also occupied to a lesser extent, and these lie on the pathways for the movement of the Ag cations. The main candidate sites of the Ag cations are shown in a picture of the crystal structure of bcc AgI in Fig. 2.39.

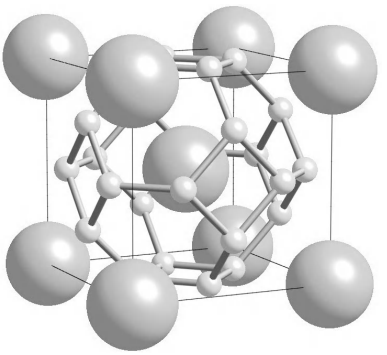
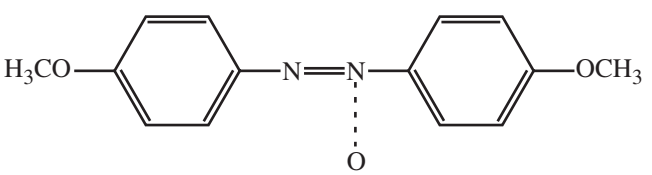


Fig. 2.39 Crystal structure of the bcc fast-ion phase of AgI, showing the main crystallographic sites occupied by the Ag cations and some of the routes for cation movement.

Fast-ion conductors are likely to be important as solid state electrolytes, particularly for battery technology. There is a wide range of materials with fast-ion conducting phases, some as simple as AgI. Most fast-ion conducting phases only exist at high temperatures, and the challenge is to discover or tailor chemical compositions that give fast-ion conduction at ambient temperature.

2.6.5 Liquid crystals

We noted above that orientationally disordered molecular crystals are like an intermediate state between a crystal and a liquid, with long-range order of the positions of the molecules but disorder of their orientations. Liquid crystals are the complementary intermediate state. Typically liquid crystals are made of organic molecules that have rod-like shape. Like real sticks or rods, the orientations of these molecules will align with respect to each other, but the positions of the molecules are disordered to varying degrees. An example molecule is



There are a number of different types of liquid crystal phases, with different degrees of positional order. In some liquid crystal phases, the positions are completely disordered, whereas in other liquid crystal phases the molecules may line up in layers but with positions being disordered within each layer. Some of the different phases are represented schematically in Fig. 2.40. In other phases there is also a variation of the angle of molecule orientation. In this case, the wavelength of the angle of orientation can be close to the wavelengths of light, giving colour interference effects.

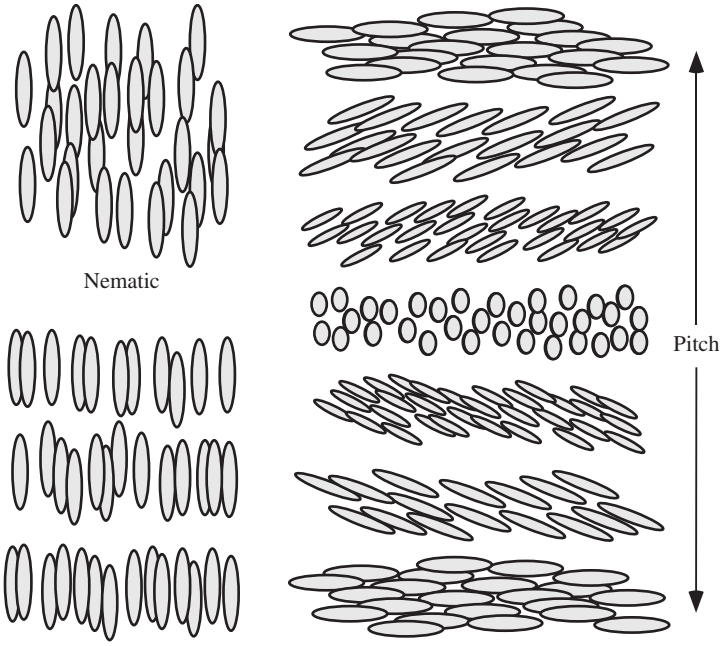


Fig. 2.40 Liquid crystal phases.

2.7 Glasses and amorphous phases

2.7.1 Glasses and structural disorder

Glasses are the exemplar of a disordered solid. The positions of the atoms are frozen into a structure that has no long-range order. However, the local environment may have some degree of short-range order, which may be very similar to the short-range structure in related crystalline phases. For example, diffraction experiments (discussed in Chapter 6) show that in silicate glasses the local structure consists of SiO_4 tetrahedra linked together in ways that are similar to the linkages in crystalline phases. This point is illustrated in two dimensions in Fig. 2.41.

The usual way to produce glass phases is to quickly quench a liquid. With fast cooling rates the atoms do not have time to organize themselves into a crystal lattice before they lose the kinetic energy associated with mobility. The most common glasses produced by cooling the fluid are based on silica – these are the glasses most often used in modern technology. They are relatively easy to work with because the fluid phases have relatively high viscosities. Similar to the silicate glasses are those based on boron or phosphorous, also with structural units linked together. In borate glasses, the boron atoms have 3-fold coordination with the oxygen atoms.

Another important family of glasses are the chalcogenide glasses, which are glasses based on Group VIb elements, namely S, Se, and Te, but excluding oxygen. Examples are As_2S_3 and $GeSe_2$. The family of chalcogenide glasses is important because of the electronic (semiconducting) properties of many of these materials.

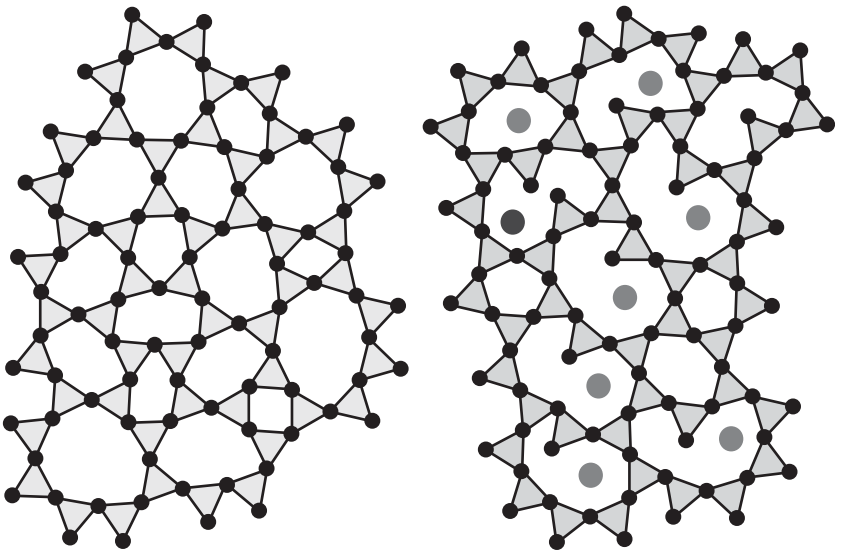


Fig. 2.41 Two-dimensional glass structures, showing linkages of fundamental building blocks. On the left the network is fully linked, and on the right the structure has non-bridging bonds and charge-balancing cations.

Some organic materials can be produced in glass phases, examples being ethanol at low temperatures, and glycerol. It is possible also to produce metallic glasses by very fast quenching. To produce fast quench rates, liquid metal is skirted onto a cold substrate and removed immediately in the form of a ribbon.

One of the most important amorphous phases is silicon. This has to be produced by deposition of the atoms into an amorphous state rather than by quenching the liquid. Amorphous Si has similar semiconducting properties to the crystalline phase. Because it is possible to produce large-area sheets of amorphous Si more economically than crystalline Si, amorphous Si is useful for applications such as solar panels. A model of amorphous silicon is shown in Fig. 2.42. In this model, all silicon atoms are bonded to four others, as suggested by diffraction data discussed below.

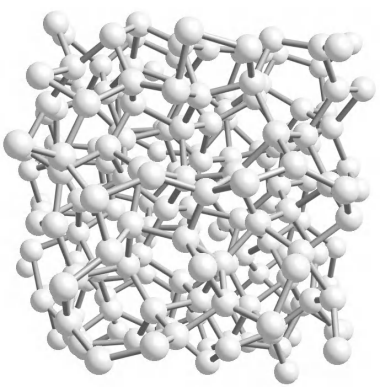


Fig. 2.42 Three-dimensional model of amorphous silicon, showing a network of atoms linked to four neighbours. (Data for figure kindly provided by Michael Thorpe, University of Michigan.)

2.7.2 Quantifying short-range order

Although there is no long-range order in a glass of the form that we have discussed for crystalline solids, there is considerable short-range order. When we introduced the basic structure types, we introduced a number of formal concepts that enable us to describe the crystal structure, concepts such as lattice, unit cell and lattice parameters, atomic coordinates and symmetry. These concepts will be developed in more detail in Chapter 3. None of these concepts are relevant for understanding the atomic structure of glasses. For crystalline materials, the long-range order is described using vectors in a frame of reference (lattice vectors, positions of atoms). There is no such frame of reference in a glass. Instead of thinking about absolute positions of atoms with respect to a reference point, we need to think about the distances between atoms. We could of course describe a crystal in the same way, and indeed we do so when we think about a cation being bonded to, say, four or six oxygen atoms. However, there is a significant difference between saying that an atom is bonded to four neighbours

and saying that an atom is in a tetrahedral coordination. The difference is that a statement of the shape of the coordination polyhedron implies that we have information about the directions of bonds, and this is precisely the information that we may lack in describing a glass. Certainly, it is not possible to obtain information about coordination polyhedra *directly* from experiments, and this information has to be inferred indirectly, as will be discussed in Chapter 6 (see also Appendix J).

If we are limited to describing the atomic structure of a glass in terms of the number of neighbours about different types of atoms, we need a new formalism to quantify this description. We introduce the **pair distribution function** (PDF), $g_{kl}(r)$. Consider at atom of type k. The number of atoms of type l that lie within a spherical shell of thickness dr at a distance r is given by

$$n_{kl}(r) = 4\pi r^2 \rho_l g_{kl}(r)\,\mathrm{d}r \tag{2.9}$$

where ρ_l is the number of atoms of type l per unit volume (known as the **number density**). For distances lower than the shortest interatomic distance, $g_{kl}(r) = 0$. For very large distances, $g_{kl}(r) = 1$ since the number of atoms within the spherical shell depends only the number density and not on the arrangement of atoms in the glass. In between these two limits, $g_{kl}(r)$ varies with r according to the distribution of atoms around the atoms of type k. $g_{kl}(r)$ will have peaks for values of r corresponding to well-defined interatomic distances, particularly nearest-neighbour bonds. If a peak is found to lie between two distances, r_1 and r_2, the number of bonds can be obtained from

$$\int_{r_1}^{r_2} n_{kl}(r)\,\mathrm{d}r$$

For a glass with two atomic species, there will be three PDFs. Figure 2.43 shows these functions for silica glass, namely $g_{\mathrm{Si,O}}(r)$, $g_{\mathrm{O,O}}(r)$ and $g_{\mathrm{Si,Si}}(r)$,

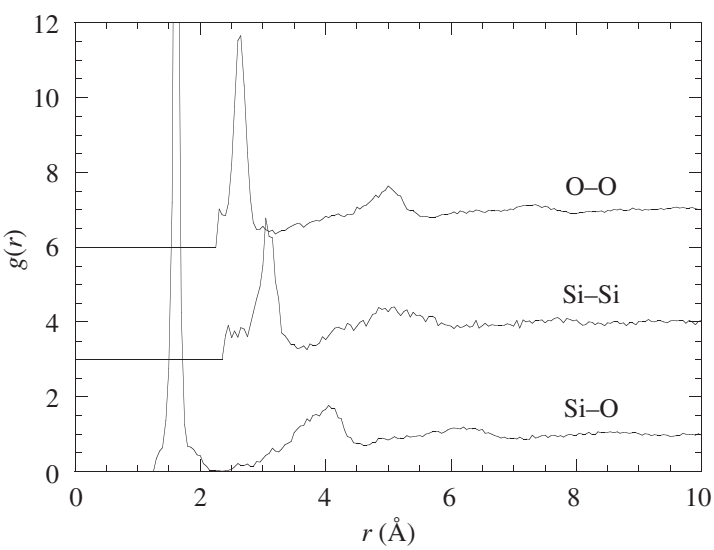


Fig. 2.43 The three $g(r)$ functions for amorphous silica, obtained from a model constructed to best match experimental diffraction data. Construction of perfect models of amorphous silica is extremely difficult, and small defects account for difficulties in obtaining symmetric peaks. (Data for figure kindly provided by David Keen, Rutherford Appleton Laboratory.)

which have been constructed from modelling studies based on experimental diffraction data obtained from silica glass (see Chapter 6). The Si–O PDF shows a large peak at $r = 1.6$ Å. This corresponds to the length of the Si–O bond in crystalline silicates. Moreover, the integral of this peak is close to 4, suggesting the existence of SiO_4 polyhedra within the glass structure. The O–O PDF has a strong peak at $r = 2.3$ Å, with integral of 6. This is what would be expected if the SiO_4 polyhedra are in fact well-defined tetrahedra, as in the crystalline phases. The Si–Si PDF has its first peak at $r = 3.2$ Å. This can be reconciled with the positions of the first peaks in the Si–O and O–O PDFs if the tetrahedra are linked at corners, with an Si–O–Si angle of around 145°. This analysis of the local environment in silica glass gives a picture that is remarkably similar to the local environment in the crystalline phases of silica, a point that will be developed below.

Based on the information contained in the PDFs, it is possible to build models of silica glass, or else to test models constructed by other means. Figure 2.44 shows a model of silica glass that is consistent with the PDF analysis, namely being based on a network of corner-linked SiO_4 tetrahedra with a distribution of Si–O–Si angles centred around 145°. This model was constructed without reference to experimental data, but experimental measurements of the PDFs have been used to verify the accuracy of the model.

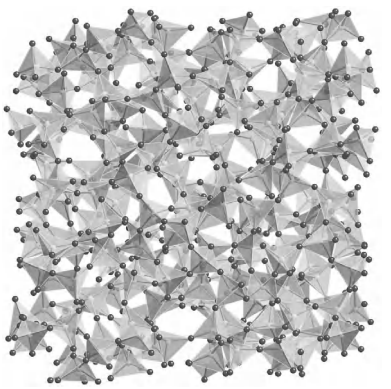


Fig. 2.44 Three-dimensional model of amorphous silica, showing a network of connected SiO_4 tetrahedra. (Data for figure kindly provided by Kostya Trachenko, University of Cambridge.)

2.7.3 Amorphous and crystalline phases of silica

The picture that emerges for amorphous silica is of a local structure that is very similar to that found in crystalline silicates. This similarity can be explored further by comparing the RDF for amorphous silica with the RDFs obtained experimentally from powdered samples of other crystalline silica phases. To link with experiment, we will work with the overall PDF formed by combining the separate PDFs, weighted according to the scattering power of each atom in neutron diffraction experiments (Chapter 6, Appendix J), which is called $G(r)$:

$$G(r) = \sum_{k,l} c_k c_l b_k b_l (g_{kl}(r) - 1) \tag{2.10}$$

where the sum is over pairs of atom types k and l, c_k is the proportion of atom type k, and b_k is the scattering power of the same atom type (this will be discussed in detail in Chapter 6). To aid clarity, we will actually plot the function $D(r) = rG(r)$, which emphasizes the structure in the PDFs at higher values of r (as will be seen in Chapter 6 and Appendix J, this is the actual quantity that is measured in a diffraction experiment). The $D(r)$ function for silica glass is compared with the $D(r)$ functions for the high-temperature phases of the crystalline polymorphs of silica, cristobalite, tridymite and quartz in Fig. 2.45. As we noted above, this analysis ignores the full three-dimensional relationship between the positions of atoms, and treats the crystal as an isotropic material. Over the range of r to 10 Å, the closest similarity is with tridymite, with quartz having the least similarity. However, in all cases the peaks in the functions are in the same places. The main difference between the $D(r)$ functions is that the peaks in $D(r)$ for amorphous silica decay on increasing r as compared to the crystalline phases, which reflects the fact that the structures cease to be similar

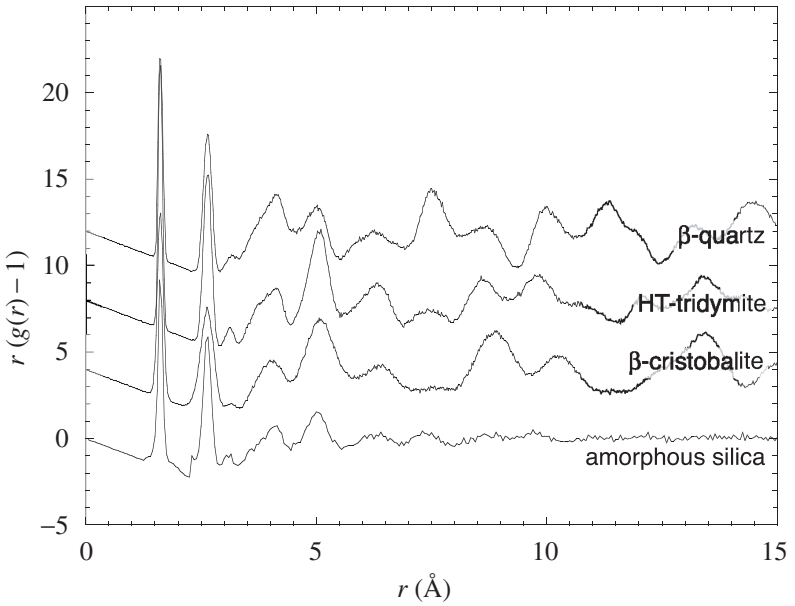


Fig. 2.45 The functions $r(g(r) - 1)$ for amorphous silica and three high-temperature polymorphs, obtained directly from experimental data. (Data for figure taken from Keen and Dove, J. Phys.: Cond. Matter *11*, 9263, 1999.)

once the periodicity of the crystal lattices has an influence on the PDFs. The important point of the comparison is that the structures of the amorphous and crystalline phases of silica have similar groups of linked SiO_4 tetrahdra. Both cristobalite and tridymite contain six-membered rings of tetrahedra, as shown in Fig. 2.38, with slightly different stacking of layers. Quartz has a spiral structure not found in the other polymorphs. It should also be noted that the densities of cristobalite and tridymite are very close to the usual density of silica glass, which supports the idea from the PDF analysis that the structure of silica glass uses building blocks that are very similar to these crystalline phases.

We are forced to conclude that the main difference between the structures of the amorphous and crystalline phases of silica only arise on length scales that are sensitive to the periodicity of the lattice. For phenomena that only depend on the short-range structure, glasses may be remarkably similar to related crystals. We will discuss a comparison of the dynamic properties of the amorphous and crystalline phases of silica in Chapter 10.

Silicate glasses, like crystalline silicates, can also contain Al incorporated into the basic network of AlO_4 tetrahedra. Other metal cations (such as the alkali metals, Mg and Ca) will act as charge balancing cations. These cations may occupy cavities between the tetrahedra, or else can modify the network to give octahedral coordination with oxygen atoms as in crystalline phases.

2.8 Conclusions

This chapter has been mostly concerned with examples of the range of crystal structures. We started from the simplest close-packed structures of the elements, and then developed the discussion to include more and more complex structures. We then looked at crystal structures of compounds, again starting from the

simplest close-packed structures and extending the discussion to include more complex structures.

The structures we discussed brought out a number of important principles. Some of these concerned the formal description of crystal structures, anticipating the more detailed treatment of the next chapter. The points highlighted in this chapter have included the definition of the **lattice**, the idea of **lattice vectors** and the **unit cell**, and the role of **symmetry**. The structures also highlighted a number of factors that are important in determining crystal structures, factors such as **close-packing**, **ionic radii**, and **structural polyhedra**. These factors will be properly quantified in Chapter 5.

Two other important features have been highlighted by the examples discussed in this chapter. The first is the link between structure at an atomic level and macroscopic physical properties. The second is the existence of **polymorphism** and **phase transitions**. One link between these two features is that different polymorphs may have such different structures that they have different physical properties. Moreover, the existence of a phase transition between two phases with similar structures, perhaps occuring as a result of small symmetry-breaking displacements of some of the atoms, often leads to an enhancement of the physical properties. A detailed discussion of physical properties will be given in Chapter 7, and phase transitions will be discussed in Chapter 12.

Summary of chapter

- The simplest two structures of the elements are the **cubic and hexagonal close packed structures**, in which close-packed planes of atoms are stacked above each other with repeats every three and two layers respectively.
- Two other important simple structures for elements are the **body-centred cubic** structure and the **diamond** structure. The **simple cubic structure** is of little practical relevance.
- The **packing fraction** for single-element structures gives the fraction of space within the crystal that is encompassed within the atoms. The close-packed, body-centred cubic and diamond structures have packing fractions 0.75, 0.68 and 0.34 respectively.
- There is, in fact, a wide range of structures available to the pure elements. Carbon and selenium are two examples of structures that form networks of covalent bonds, and carbon, nitrogen, halogens and sulphur form molecules.
- Simple crystal structures of diatomic crystals are arranged with unlike ions forming nearest-neighbour pairs, with a range of coordination numbers.
- Some aspects of the crystal structures of compounds can be rationalized in terms of the effective radii of the constituent ions.
- Structures of more complex materials can be described in terms of networks of **coordination polyhedra**. One example is the family of silica structures, which are all formed as networks of corner-sharing SiO_4 tetrahedra. A second example is the family of **perovskite** structures, which are formed as networks of corner-sharing octahedra with a simple-cubic parent structure.

- Crystals of the perovskite family of structures undergo a variety of **phase transitions**, which give rise to changes in physical properties that can be exploited for technological applications.
- Some crystals can accommodate a certain degree of disorder. Examples include crystals with molecules that have disordered orientations, fast-ion conductors in which mobile atoms can hop between vacant interstitial sites, and liquid crystals in which molecules have ordered orientations but only partial crystalline order.
- Glasses do not have the long-range periodicity of crystals, but can have features of the short-range structure that are also found in corresponding crystal structures. A prime example is silica, whose glass phase has a network of corner-sharing SiO_4 tetrahedra as in the crystalline phases. The short-range structural order can be characterized using the pair distribution function (PDF), and there can be close similarities between the PDFs of corresponding glass and crystal phases.

Further reading

The structures of crystals are discussed in many books. Kittel (1996) and Ashcroft and Mermin (1976) give an introduction to the basic structure types. More detailed discussions are given by Alcock (1990) and Borg and Dienes (1992). More general discussions of crystal structure are given by Hazen and Finger (1982), Ladd (1999), and Mak (1997). Glasses are described by Elliott (1990). Liquid crystals are discussed by Collings (1997) and Collings and Patel (1997).

Exercises

(2.1) Many crystals have atoms in tetrahedral coordination with other atoms. If l is the length of an edge of a tetrahedron, calculate (a) the distance of the centre of the tetrahedron to an apex, (b) the distance from the centre of a face to one of the three closest apices, and (c) the distance from a centre of a face to the opposite apex, each in terms of l. (*Hint*: it is helpful to describe the tetrahedron with reference to a cube of edges $l/\sqrt{2}$, with apices on four of the corners of the cube and edges along face diagonals.)

(2.2) Use the geometry of the tetrahedon obtained above to show that the ideal value of c for a hcp structure with atomic radius r is $4\sqrt{2/3}r$. Hence show that the ideal c/a ratio for the hcp structure is equal to $\sqrt{8/3}$.

(2.3) Show, using the geometric constructions of the previous questions, that the ideal **ABCABC** stacking of close-packed planes leads to a cubic structure.

(2.4) Show that the packing efficiency of the simple cubic structure is $\pi/6$, and calculate the packing efficiency for the diamond structure.

(2.5) One of the phases of tungsten is cubic, $a = 5.1$ Å, with atomic coordinates:

W(1): $0, 0, 0; \frac{1}{2}, \frac{1}{2}, \frac{1}{2}$;

W(2): $0, \frac{1}{2}, \pm\frac{1}{4}; \frac{1}{2}, \pm\frac{1}{4}, 0; \pm\frac{1}{4}, 0, \frac{1}{2}$.

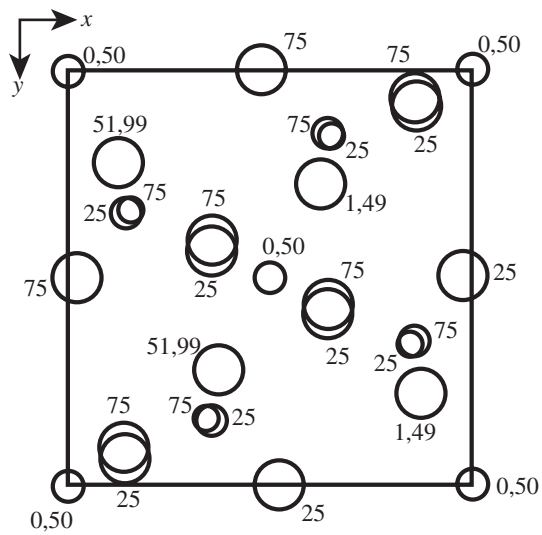
Draw a 2×2 plan of the structure down one of the cube axes. Identify the symmetry operations. Identify the coordination of one of the W(1) atoms. (For one of the types of symmetry you will need to back up your reading of this chapter with the next.)

(2.6) Assuming that like ions do not touch in either the NaCl or CsCl structures, calculate the relative packing efficiences

of both structures. Verify the result for the limit where the sizes of the cations and anions become the same.

(2.7) The ionic radii for Zn^{2+} and S^{2-} are 0.74 and 0.26 Å respectively. Calculate the lattice parameter for the cubic structure of ZnS.

(2.8) The figure below shows the crystal structure of Al_2SiO_5 sillimanite. The large circles represent oxygen atoms, the small circles represent silicon atoms, and the intermediate size circles represent aluminium atoms.



Identify the coordination polyhedra and the way in which they are connected.

(2.9) Considering only the electrostatic energy, which varies as $1/r$, show that the potential energy of a cation at the centre of an octahedral arrangement of anions is at a maximum with respect to displacements towards the anions. (*Hint*: express the distance between the cation and any anion in terms of the displacement of the cation from the centre.)

(2.10) Oxides formed from the singly charged alkali cations, A^+ and 5-valent cations, B^{5+} have the formula $A^+B^{5+}O_3$, and many crystallise in the perovskite structure. The ionic radii of the alkali cations and some 5-valent cations are

Cation	Li^+	Na^+	K^+	Rb^+	Cs^+	V^{5+}	Nb^{5+}	Ta^{5+}	
Radius	1.06	1.32	1.65	1.75	1.88	0.68	0.78	0.78	Å

The ionic radius of oxygen is 1.25 Å. Predict the stabilities of each of the possible perovskite structures.

(2.11) The rutile structure, TiO_2, is tetragonal, $a = 4.59$ Å, $c = 2.96$ Å. The fractional coordinates are Ti at $0, 0, 0$ and $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$, and O at $u, u, 0$; $-u, -u, 0$; $\frac{1}{2}+u, \frac{1}{2}-u, \frac{1}{2}$; and $\frac{1}{2}-u, \frac{1}{2}+u, \frac{1}{2}$, where $u \simeq 0.3$. Sketch the structure viewed down [001], and identify the orientations of the TiO_6 octahedra in this projection. If the length of the Ti–O bond is denoted by l, write equations for the lattice parameters a and c in terms of l assuming that the octahedra have ideal shape, and show that the ideal a/c ratio is equal to $1 + 1/\sqrt{2}$. Compare this value with the actual a/c ratio, and rationalize any difference. Show that the ideal value of u is $1/(2+\sqrt{2})$.

(2.12) Calculate the packing efficiencies of the following AX_2 structures: fluorite (fcc, fractional coordinates of cation at $0, 0, 0$ and anion at $\pm(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$), ideal rutile (see previous structure) and cristobalite (fcc, fractional coordinates of cation at $0, 0, 0$ and anion at $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ and $-\frac{1}{4}, -\frac{1}{4}, \frac{1}{4}$). Hence show that the cristobalite and rutile structures are 4 and 1.338 times less dense than the fluorite structure. Account for these differences in density.

3 Formal description of crystal structures

3.1 Introduction

In earlier chapters we met some of the fundamental concepts involved in describing crystal structures, but only in a case-by-case way. The purpose of this chapter is to properly formalize the description of crystal structures. Formalism can be like reading the instructions that go with a new gadget – it is nice to avoid reading them when starting out, but eventually there are many good reasons for studying them in detail. We can get quite far in our study of crystalline materials without worrying too much about the details of the formalism, but eventually the power behind the formalism makes it sensible to come to terms with it. So the reader should feel free to skip over this chapter when first approaching the study of crystalline materials, but bookmark it for future reference when there is a need to grasp some of the finer details needed to appreciate the arguments used in the deeper applications of the subject.

The formalism that we develop in this chapter is mostly concerned with **symmetry**, to a greater or lesser extent, including the symmetry operations we have already encountered (rotation axes, mirror planes, etc.), and the symmetry associated with the periodicity of the crystal lattice. We will take the ideas beyond the simple description of symmetry to show how symmetry can be converted into a quantitative tool. This tool has a number of uses, including the means to perform detailed mathematical calculations on both microscopic and macroscopic length scales, and intepretation of diffraction data.

To set the scene, Figs 3.1 and 3.2 show aspects of the symmetry we attempt to highlight. Figure 3.1 shows a number of tiled panels from the Moorish Alhambra Palace in Granada. These pack shapes together in periodic patterns that reflect different possible symmetries. These patterns use exactly the same symmetries (or at least the two-dimensional representations) as nature uses in packing atoms together. In each of the three patterns it is possible to see

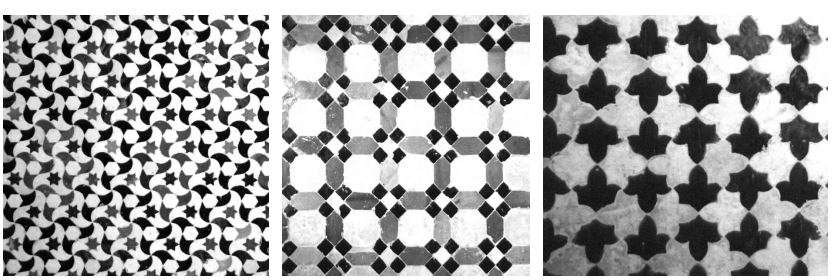


Fig. 3.1 Examples of some of the two-dimensional symmetry patterns used in tiled panels in the Alhambra Palace in Granada. The panels show uses of rotational and mirror symmetry together with periodic repeats of a basic pattern.

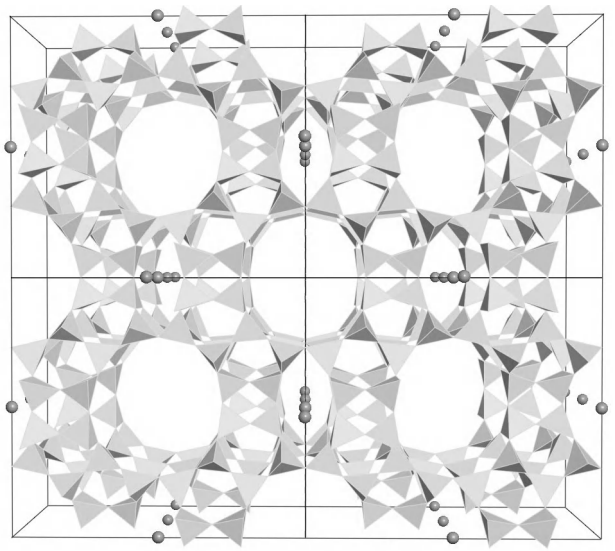


Fig. 3.2 The crystal structure of the zeolite mordenite ($NaAlSi_5O_{12}$, structure C-centred orthorhombic), showing some of the atoms as connected tetrahedra. The projection highlights the rotational and mirror symmetry in the arrangement of atoms, together with the periodic repeat of an underlying arrangement of atoms.

rotation axes and mirror planes, and the lattices and unit cells can clearly be identified. Similarly, both rotation axes and mirror planes can be seen in the crystal structure shown in Fig. 3.2.

3.2 Crystal structure: lattices, unit cell, and atomic coordinates

3.2.1 Definition of the crystal lattice

A **lattice** is an infinite periodic array of points. In principle the lattice is completely described by three **basis vectors**, which are conventionally written as **a**, **b** and **c**. The idea of a lattice is illustrated for a two-dimensional example in Fig. 3.3.

The vector between any two points on the lattice can be described as a linear combination of the three basis vectors:

$$\mathbf{t}_{[UVW]} = U\mathbf{a} + V\mathbf{b} + W\mathbf{c} \tag{3.1}$$

where U, V, W are integers. This can be compared with eqn (2.1), where we used fractional coordinates x, y, z to represent the vector to point in the unit cell. It is common to represent the vector $\mathbf{t}_{[UVW]}$, which is called a **lattice vector**, by the shorthand $[UVW]$ (the choice of rectangular brackets is important – different types of brackets have different meanings in standard crystallographic convention). The particular lattice point that is chosen to be the origin is arbitrary. In cases where there is a set of lattice vectors that are related by symmetry, such as the [100], [010], and [001] lattice vectors in a cubic lattice (introduced in Chapter 2, with the three lattice vectors being orthogonal and of equal length), the complete set is denoted by the shorthand $\langle UVW \rangle$. The set of

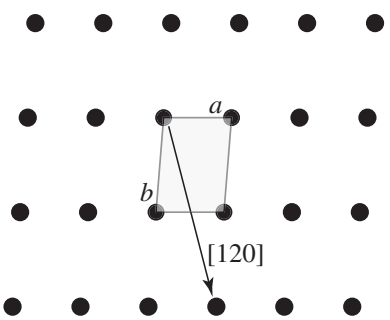


Fig. 3.3 Two-dimensional crystal lattice, showing the outline of the unit cell and an example of a lattice vector.

all lattice vectors, that is, the infinite set of all values of $[UVW]$, including the values related by symmetry, defines the lattice. This point will be formalized mathematically in Section 3.2.7.

3.2.2 The unit cell

The lattice is the starting point for a crystal, but associated with each lattice point is a region of space which contains the basis set of atoms. There are many ways of defining this region of space, but by convention we define the quantity called the **unit cell**, which is a volume of parallelepiped shape bounded by lattice points on each of its eight vertices. By definition, the unit cells pack together following the crystal lattice, and fill space completely.

The three basis vectors of the unit cell are equivalent to the lattice vectors **a**, **b**, **c**. We define a, b and c to be the lengths of the basis vectors **a**, **b**, and **c** respectively, and the angles α, β, and γ to be the angles between the pairs of basis vectors **b** : **c**, **c** : **a**, and **a** : **b** respectively. The set of six numbers are called the **lattice parameters**. Thus the edges of the unit cell have lengths a, b, c and angles α, β, γ.

3.2.3 Lattices, lattice parameters, and symmetry: the seven crystal systems

In general there need be no relationship between the lengths and directions of the three basis vectors, i.e. between the three unit cell edge lengths and the three unit cell angles. However, relationships can be generated by the symmetry of the crystal. Two examples are given in Fig. 3.4. One example is of a crystal containing a 4-fold rotation axis parallel to the **c**-axis. Since the action of the rotation will cause the **a**-axis to be identical to the **b**-axis, it is required that $a = b$ and that all three unit cell angles be equal to 90°. The second example is of a crystal containing a mirror plane. Translational symmetry can only be maintained if one of the lattice vectors is normal to the mirror plane and the other two lattice vectors lie within the mirror plane. Thus symmetry requires two of the unit cell angles to be equal to 90°. We will explore the formalism of the relevant symmetry operations for crystalline materials shortly, but for the moment we will define the different types of lattice that are possible.

For all crystal structures, any relationships between the lattice parameters are determined by the crystal symmetry. The complete set of lattices that is established by the constraints of symmetry is given in Table 3.1. This table also gives the constraints on the values of the lattice parameters for each type of lattice arising from the symmetry, together with some examples from Chapter 2.

The pair of lattices called **trigonal** and **rhombohedral** require some comments. These both have the same defining symmetry, namely the presence of a 3-fold rotation axis, yet are given different constraints on the lattice parameters. Indeed, the trigonal lattice parameters have the same constraints as the **hexagonal** lattice, namely $a = b \neq c$, $\alpha = \beta = 90°$, $\gamma = 120°$ (Table 3.1), but is not hexagonal because it has only a 3-fold rotation axis rather than the 6-fold rotation axis that defines the hexagonal class. The relationship between the trigonal and rhombohedral lattices lies in the fact that a rhombohedral lattice ($a = b = c$, $\alpha = \beta = \gamma \neq 90°$) can be recast in the form of a trigonal

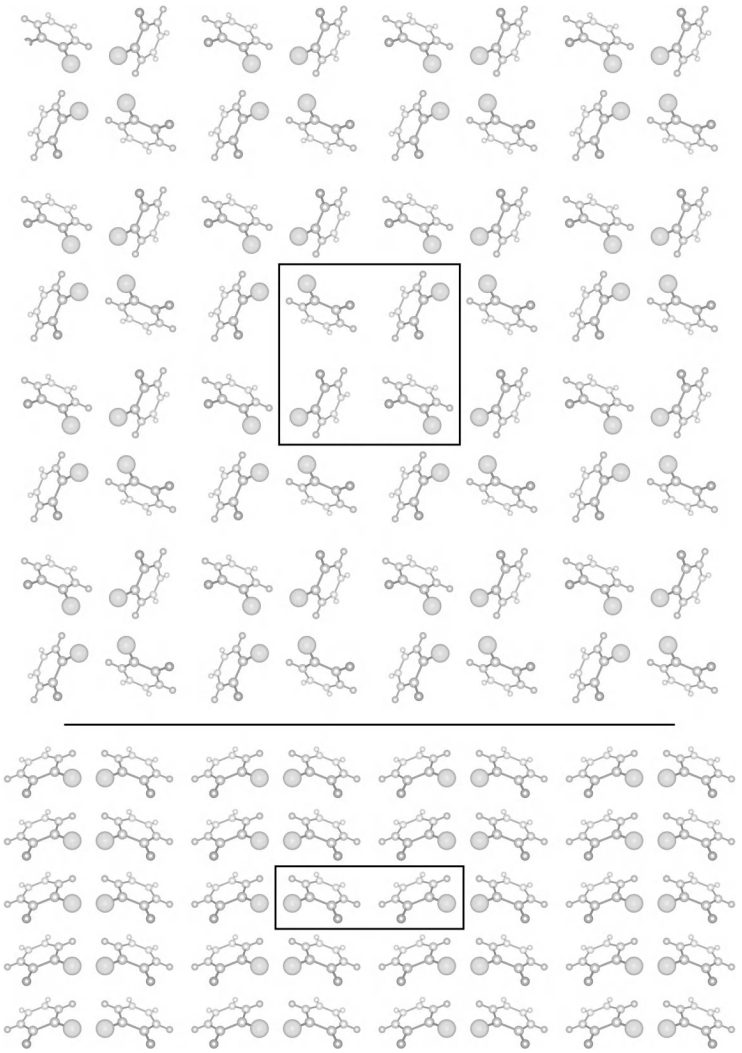


Fig. 3.4 Cartoon representations of the constraints that symmetric arrangements of atoms impose on the unit cell. *Top*, an arrangement with a 4-fold rotation axis, showing that the two lengths of the unit cell are required to be equal by symmetry. *Bottom*, an arrangement with a mirror plane normal to the plane of the diagram, showing that the unit cell vectors in the plane of the diagram are required to be orthogonal by symmetry.

lattice with a factor of three increase in the volume of the unit cell (although the inverse is not necessarily true). We discuss the relationship between the trigonal, hexagonal, and rhombohedral lattices in Appendix D: it involves some of the formal details we discuss in this chapter and the next.

We should also remark that the constraints on the values of the lattice parameters given in Table 3.1 follow standard convention. Other permutations of the relationships between the lattice parameters are formally equivalent to those given above – for example, we could have a hexagonal lattice with $a = b \neq c$ and $\beta = 120°$, but by convention the special axis is chosen to be the **c** axis in hexagonal and tetragonal, and **b** in monoclinic.

We met examples of most lattices in Chapter 2, as indicated in Table 3.1. The one case we did not meet is triclinic, which is the most complicated case because there are no relationships between the lattice parameters.

Table 3.1 The seven lattice types as defined by symmetry constraints. Note that the constraints on the values of the lattice parameters follow convention, and different combinations are allowed.

Lattice	Defining symmetry	Constraints on lattice parameters	Examples
Triclinic	None	None	—
Monoclinic	One mirror plane or one 2-fold rotation axis	$\alpha = \gamma = 90°$	naphthalene, selenium, benzoic acid
Orthorhombic	Each axis has either a mirror plane or 2-fold rotation axis or both	$\alpha = \beta = \gamma = 90°$	sulphur, iodine
Hexagonal	One 6-fold rotation or $\bar{6}$-axis	$a = b, \alpha = \beta = 90°$, $\gamma = 120°$	hcp metals, β-quartz, graphite, selenium, CdI_2
Tetragonal	One 4-fold rotation or $\bar{4}$-axis	$a = b, \alpha = \beta = \gamma = 90°$	rutile (TiO_2), ferroelectric $PbTiO_3$, white tin
Trigonal	One 3-fold rotation or $\bar{3}$-axis	$a = b, \alpha = \beta = 90°, \gamma = 120°$	α-quartz
Rhombohedral		$a = b = c, \alpha = \beta = \gamma \neq 90°$	sulphur, calcite, Al_2O_3
Cubic	4 intersecting 3-fold rotation or $\bar{3}$-axes	$a = b = c, \alpha = \beta = \gamma = 90°$	fcc, bcc metals, alkali halides

It is worth emphasizing again the fact that the different lattice types are defined by the symmetry rather than by the relationships between the lattice parameters, although it is often easier to think of the lattices in terms of the lattice parameters. This point is particularly important when differences between lattice parameters are smaller than the intrinsic resolution of a measurement.

3.2.4 Volume of the unit cell

The volume of the unit cell is equal to

$$V_{\text{cell}} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = abc \sin\alpha \sin\beta \sin\gamma \tag{3.2}$$

In the case where the lattice vectors are orthogonal, the volume of the unit cell is simply equal to $V_{\text{cell}} = abc$. For monoclinic lattices, $V_{\text{cell}} = abc \sin\beta$, and for hexagonal lattices, $V_{\text{cell}} = abc \sin\gamma = a^2c \sin 120° = \sqrt{3}a^2c/2$.

3.2.5 Conventional and primitive lattices: The 14 Bravais lattices

The main feature of a lattice is that the environment of each lattice point should be identical. The basic lattices of the seven crystal classes are not the only possible crystal lattices. One of the ideas we met in Chapter 2 was that it is possible to define conventional lattices as distinct from primitive lattices. The unit cell of a conventional lattice is not the smallest unit cell possible, and it

encompasses more than one lattice point. There are two main reasons for using conventional cells. The *first* is that they make some aspects of the formalism easier (even when accounting for the penalty of having more than one lattice point in the unit cell). The *second* reason is that the defining symmetry for all lattice types of the same class is equivalent, and it is therefore easier to have the same unit cells. For example, the defining symmetry for all cubic lattices, whether primitive, fcc or bcc, is the presence of four 3-fold rotation axes. It is rather easier to treat all cubic lattices in the same way as regards the definition of the unit cell. The one exception is that of rhombohedral or trigonal lattices, as discussed in Appendix D.

The complete set of lattices, including both primitive and conventional, is obtained by taking the seven lattices and systematically developing new lattices by placing additional lattice points at the centre of each unit cell (defining a **body-centred** lattice), at the centre of each face of each unit cell (defining a **face-centred** lattice), or at the centres of only one face of each unit cell (defining A, B or C **centred** lattices, depending on the type of face involved). Not all permutations of centred lattices with primitive lattices are independent or sensible. For example, the body-centred and face-centred tetragonal lattices are formally equivalent (see Problem 3.6), and to have an A-face centred cubic lattice would violate the essential symmetry of the cubic lattice. It also makes no sense to talk about a centred triclinic lattice.

Taking account of all equivalent permutations, there is a total of 14 possible lattices, including the seven primitive lattices described above. These are called **Bravais lattices**. The 14 Bravais lattices are shown in Fig. 3.5. The primitive lattices are designated by P. The conventional lattices are labelled as I for body-centred and F for face-centred lattices, together with the A, B, and C lattices noted above.

As we have already indicated, it is possible to describe all conventional lattices by a set of primitive lattice vectors that encompass all lattice points, but in so doing the basic symmetry inherent in the choice of conventional lattice vectors is lost in the description.

3.2.6 Atomic coordinates

The crystal structure is defined as the replication of the atomic contents of the unit cell on each lattice point. The pattern of atoms within the unit cell may often be called the **basis** or the **motif**.

The positions of the atoms in the unit cell are defined with respect to the origin of the unit cell. In an isolated system these positions will be described by vectors in a system with units of length. In a crystal, it is easier to define a position in terms of the components of the three lattice basis vectors, as already noted in eqn 2.1:

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} \tag{3.3}$$

The coordinates x, y, z are called **fractional coordinates**, and usually have values between ± 1. The use of fractional coordinates was introduced in Chapter 2 with respect to the coordinates of the atoms in the hcp structure, and fractional coordinates are always used to describe crystal structures in listings and reports of new structures.

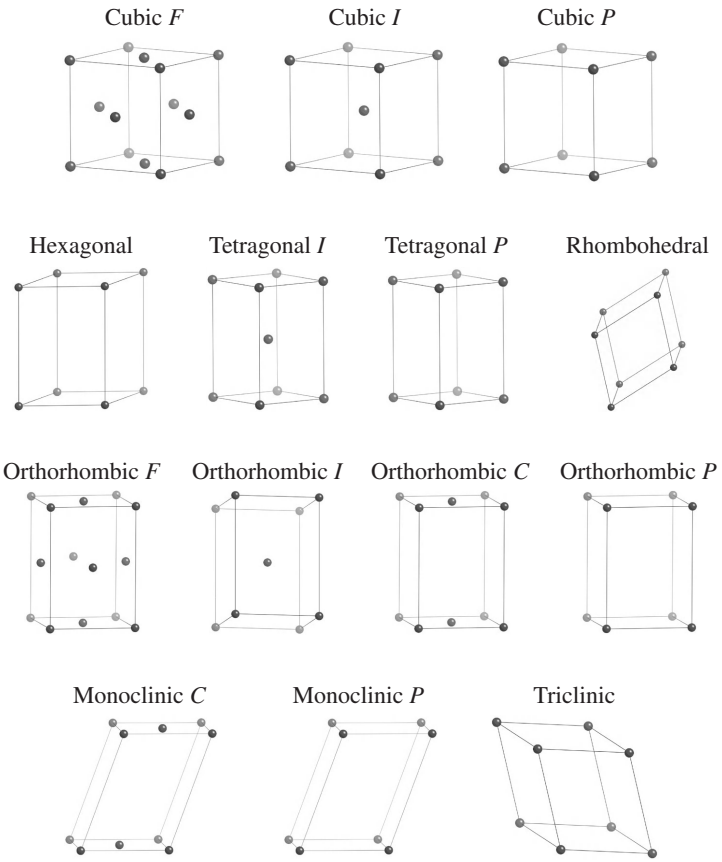


Fig. 3.5 The conventional unit cells of the 14 Bravais lattices, with the lattice points represented by the spheres. The primitive trigonal lattice is the same as the hexagonal lattice.

We will see below that there may be relationships between the positions of atoms in the unit cell because of specific symmetry. These relationships are easily represented by fractional coordinates, as will be shown later in this chapter.

3.2.7 Crystal structure as the convolution of the lattice and the atomic basis

Basic idea of convolution

The complete crystal structure is built by placing a copy of the unit cell on each lattice point, preserving the spatial arrangement of atoms. This operation is illustrated in Fig. 3.6.

The process of forming the crystal structure is an example of the mathematical operation called **convolution**. The formal process of convolution is most easily seen in one dimension. If we have two continuous functions, $g(x)$ and $h(x)$, the operation of convolution involves placing a copy of one of the functions, say $h(x)$, with an origin at each value of x and weighted by the value of $g(x)$. The resultant function is obtained by adding up all the copies of $h(x)$. The process is illustrated for two functions in Fig. 3.7. Formally, the process of

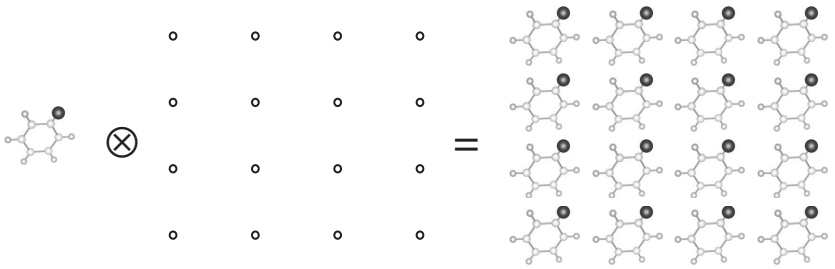


Fig. 3.6 The operation of forming a crystal structure by placing a pattern of atoms, called the **basis** or **motif**, on each lattice point.

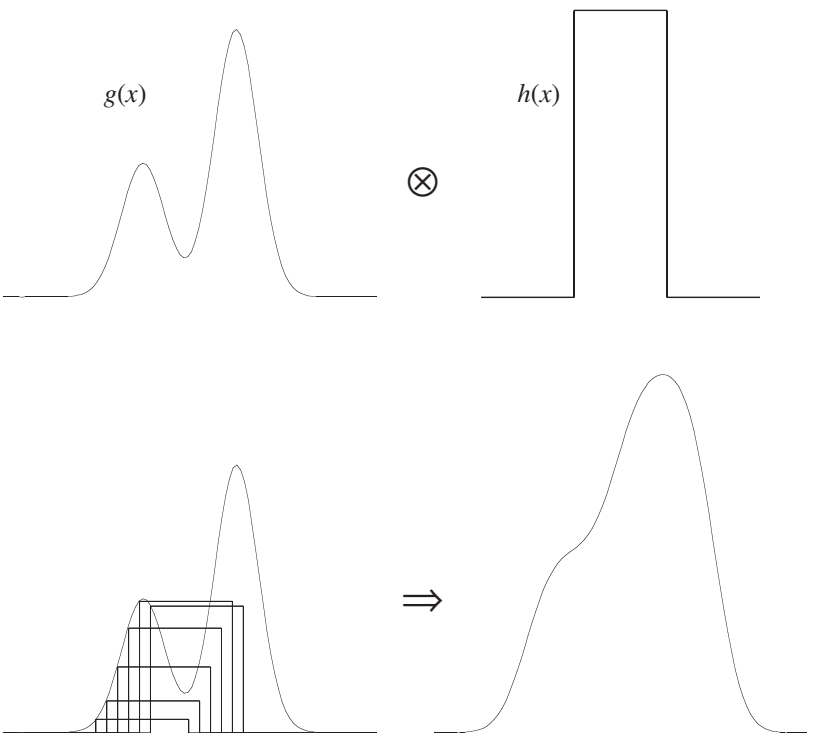


Fig. 3.7 The operation of convolution obtained by adding all copies of $h(x)$ weighted by $g(x)$.

convolution is written as

$$f(x) = g(x) \otimes h(x) = \int g(x')h(x - x')\,\mathrm{d}x' \tag{3.4}$$

Mathematical representation of a point: the Dirac delta function

The lattice is an array of points whose positions are described by the lattice vector. It is useful to be able to represent a point mathematically, and this is accomplished by a special function known as the **Dirac delta function**. Formally, this function is defined by two properties of the function:

$$\delta(\mathbf{r} - \mathbf{r}') = \begin{cases} \infty & \text{if } \mathbf{r} = \mathbf{r}' \\ 0 & \text{if } \mathbf{r} \neq \mathbf{r}' \end{cases} \tag{3.5}$$

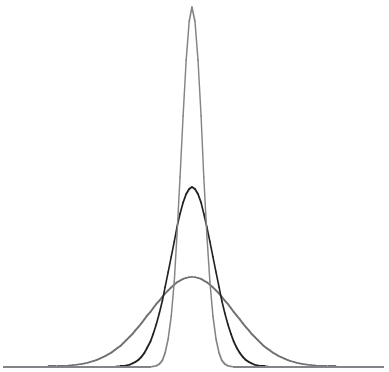


Fig. 3.8 The development of the Dirac delta function from a continuous single-peaked function. The width of the function is reduced whilst retaining a constant area, eventually reaching a function with zero width but finite area. This limiting case is the Dirac delta function.

$$\int \delta(\mathbf{r} - \mathbf{r}')\,\mathrm{d}\mathbf{r} = 1 \tag{3.6}$$

One can think of the Dirac delta function as the limiting case of a continuous single-peaked function (such as a Gaussian) with an unit integral whose width is shrunk to zero. In one dimension, this is described mathematically by the function

$$\delta(x) = \lim_{\alpha \to 0} \frac{\sqrt{\pi}}{\alpha} e^{-x^2/\alpha^2} \tag{3.7}$$

This is illustrated in Fig. 3.8.

These normalization of eqn 3.6 can be given a physical interpretation. The function $\delta(\mathbf{r} - \mathbf{r}')$ can be said to represent the instantaneous density associated with a point particle in space, namely a density that is zero everywhere except at the point where the particle is located. The integral of the density over all space, namely the normalization eqn 3.6, must be equal to 1 since the particle does exist within the volume of the integral.

Mathematical representation of the lattice

Using the Dirac delta function we can describe the lattice mathematically as

$$\mathcal{L}(\mathbf{r}) = \sum_{U,V,W} \delta(\mathbf{r} - \mathbf{t}_{[UVW]}) \tag{3.8}$$

where the set of lattice vectors, $\mathbf{t}_{[UVW]}$, was defined in eqn 3.1. This function is only non-zero at the lattice points, defined by the integers $[UVW]$, and if we integrate the space around one lattice point we obtain the result that the volume encompasses exactly one lattice point.

When one of the functions in the convolution involves a Dirac delta function, the result is a copy of the other function at a new origin. Suppose in our one-dimensional example that $g(x) = \delta(x - x_0)$. Then we have

$$f(x) = g(x) \otimes h(x) = \int \delta(x' - x_0) h(x - x')\,\mathrm{d}x' = h(x - x_0) \tag{3.9}$$

The point is that the convolution has had the effect of shifting the origin of $h(x)$ from zero to x_0, without changing the shape of $h(x)$. This result is easily generalized to three dimensions.

Convolution of lattice and atoms

The point about this analysis is that if we have a basis or motif containing a few atoms, the crystal structure is obtained by placing a copy of the basis or motif on each lattice site, Fig. 3.6. Since the lattice function is a sum of Dirac delta functions, the operation of building a crystal structure from the lattice and basis is simply a convolution process.

We also note that the basis itself can be described by a sum over a set of Dirac delta functions describing the positions of the atoms. If the second function in the convolution is also a Dirac delta function, $h(x) = \delta(x - x_1)$, we have

$$\begin{aligned} f(x) = g(x) \otimes h(x) &= \int \delta(x' - x_0)\delta(x' - x + x_1)\,\mathrm{d}x' \\ &= \delta(x - x_0 - x_1) \end{aligned} \tag{3.10}$$

The process of the convolution has been to move the atom from position x_1 in the frame of the basis to position $x_0 + x_1$ in the frame of the lattice.

Why are we so concerned to describe the crystal structure in such a formal way? The answer is that when we come to discuss the issue of diffraction (Chapter 6), which is the primary experimental method for determining the atomic structure of crystals, we will find that the analysis can be made a great deal easier by starting from the convolution description. In fact, we will find in Chapter 6 that one of the important mathematical properties of the convolution process will actually enable us to bypass some very difficult mathematical calculations.

3.3 Crystal symmetry 1. Point-symmetry operations

3.3.1 Point symmetry

There are many ways that symmetry operates on a structure. The lattice itself has a symmetry: any structure associated with one lattice point will be replicated at all other lattice points. It is this lattice symmetry that allows us to distinguish crystals from glasses. This is an example of **translational symmetry**. However, there are other types of symmetry within crystals that operate on the environment around a single point, which is called **point symmetry**. The combination of point symmetry and translational symmetry gives the overall symmetry of the crystal. We turn now to explore the details of symmetry in a formal sense, starting with point symmetry (here and Section 3.4), and then characterizing the different types of translation symmetry (Section 3.5.1), and finally considering the combinations of both types of symmetry (Section 3.5.2).

We have defined **point symmetry** as giving the symmetry seen from a particular point in the crystal. It is important to appreciate that different points in a crystal will have different point symmetry – indeed, most points have no point symmetry at all, and the existence of point symmetry will be restricted to special points, lines or planes in the unit cell.

There are four main types of symmetry operations that are associated with a single point in space, which we will represent through the effect on atomic coordinates. We will use fractional coordinates x, y, z to represent the position of the initial point, and then produce the new fractional coordinates of the new point generated by symmetry. We will denote a sign change by $\overline{x}$ rather than $-x$.

3.3.2 The four point symmetry operations

Rotation axes

Rotational symmetry is present if an object comes into coincidence with itself after rotation by an angle of $360°/n$, where n is an integer. The value of n is used to label the particular symmetry, for example, when $n = 2$ we have a 2-fold rotation axis. An example of a 4-fold rotation axis is shown in Fig. 3.9.

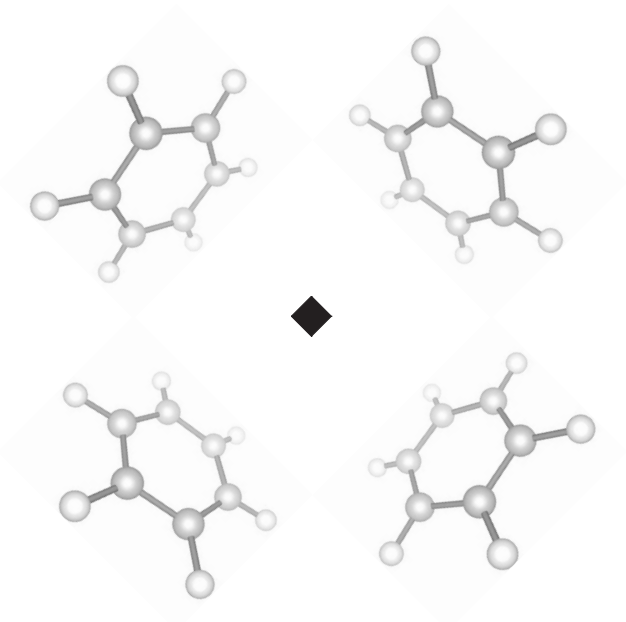


Fig. 3.9 Example of the operation of a 4-fold rotation symmetry, seen in the way the molecule is copied by 90° rotations.

In principle, rotational symmetry can exist with any values of n. However, the only values of n that are compatible with the existence of a lattice are $n = 1$, 2, 3, 4 and 6. Objects with other types of rotational symmetry cannot be packed together to completely fill space. The case of $n = 5$ is interesting because of the discovery of this type of symmetry in diffraction patterns of some metallic alloys – we will discuss this clear 'breaking of the rules' briefly at the end of the chapter.

The effects of rotational symmetry on fractional atomic coordinates are as follows. A 2-fold rotation axis parallel to [001] passing through the origin will lead to an atom at x, y, z being replicated at $\overline{x}, \overline{y}, z$. A 4-fold axis in the same place will lead to the atom at x, y, z also being replicated at $y, \overline{x}, z$; $\overline{x}, \overline{y}, z$; and $\overline{y}, x, z$. The cases of 3-fold and 6-fold axes are treated in Problem 3.8.

If the only symmetry in a crystal is a 2-fold axis, the crystal lattice will be monoclinic, because the action of a 2-fold axis is to replicate atoms within the same plane normal to the 2-fold axis. This can only be accomplished if the 2-fold axis is parallel to one of the lattice vectors and orthogonal to the other two. By convention, it is taken that the 2-fold axis is parallel to the **b** lattice vector. If the crystal has a single 4-fold axis, the lattice is tetragonal. The presence of a single 3-fold axis gives a trigonal or rhombohedral lattice, and the presence of a single 6-fold axis gives a hexagonal lattice. The cubic lattice is defined by the presence of four intersecting 3-fold axes parallel to the ⟨111⟩ directions (and not by the presence of three intersecting 4-fold axes parallel to the ⟨100⟩ directions as is commonly supposed). These conditions are summarized in Table 3.1.

There are two conventional representations of rotation axes. Standard crystallographic notation is to represent them by the symbols 2, 3, 4, or 6. A common alternative system, prefered by chemists, is the **Schoenflies notation**, which is described in Appendix C. For rotational symmetry, the Schoenflies

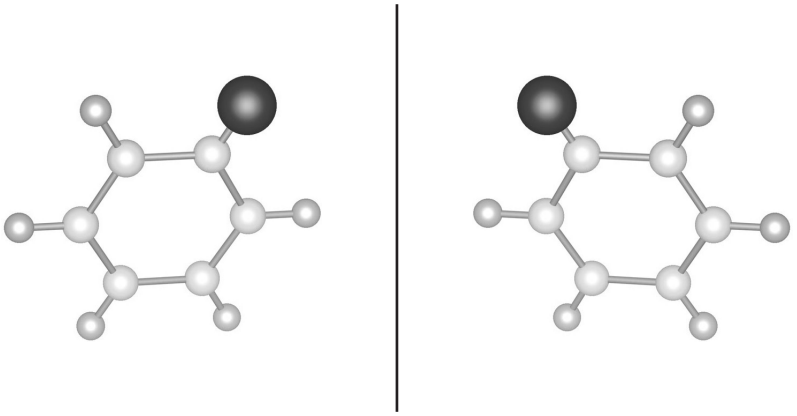


Fig. 3.10 Illustration of mirror symmetry, shown by the reflected image of the molecule.

notation uses the symbols C_n, where $n = 2, 3, 4, 6$. However, the translation between the two types of notation is not always as straightforward! Both notations are in common use.

Mirror planes

An object that can be reflected through a plane and come into coincidence with itself has **mirror symmetry**. An object can have a single mirror plane, or three orthogonal mirror planes (when combined with rotational symmetry axes there can be additional mirror planes at angles of 45° or 60°). Mirror symmetry is illustrated in Fig. 3.10.

A mirror plane normal to the [001] direction and passing through the origin of a crystal will cause an atom at x, y, z to be replicated at $x, y, \overline{z}$. A single mirror plane is another symmetry element that defines monoclinic symmetry, because the only lattice vector that will not lie in the mirror plane must be normal to it. By convention, in a monoclinic crystal the lattice vector normal to the mirror plane is the **b** axis. The presence of two or three (necessarily orthogonal) mirror planes will give orthorhombic symmetry. Mirror planes are represented by the symbol m (the corresponding nomenclature in the Schoenflies system depends on the other types of symmetry present, as noted in Appendix C.)

Centre of symmetry

An object has a **centre of symmetry** if every point at position x, y, z relative to the centre is also found at $\overline{x}, \overline{y}, \overline{z}$. A crystal with a centre of symmetry is called **centrosymmetric**. Another term for this symmetry is **inversion symmetry**. This symmetry operation is illustrated in Fig. 3.11. It is given the label $\overline{1}$, for reasons that will be discussed below.

The centre of symmetry is important for many physical properties, because if there is a centre of symmetry there can be no dipole moment associated with the positions of charged ions in the unit cell.

Rotoinversion axes

The symmetry operation of rotoinversion is a two-stage process. **First** is the operation of a rotation symmetry, and this is followed by the **second** operation of

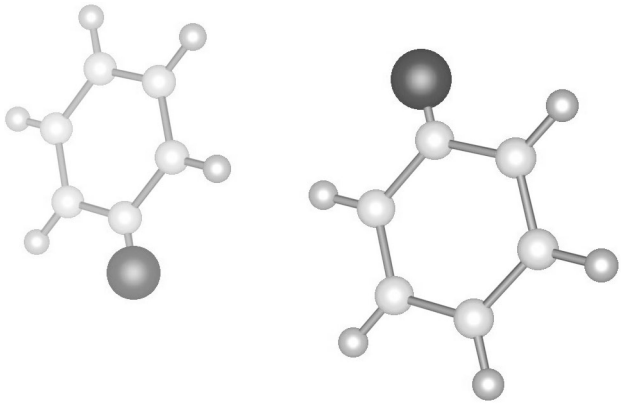


Fig. 3.11 Illustration of the operation of a centre of symmetry on a pair of molecules, whose distance in the line of sight is emphasized by the perspective and the lighter shading used to represent further distances.

a reflection through the centre. The rotation axis is then called the **rotoinversion axis**, and given the symbol $\overline{n}$ in the international crystallographic notation. It should be noted that the Schoenflies notation involves a different line of thinking for this type of symmetry, as discussed in Appendix C.

The simplest example, $\overline{1}$, is the operation of the centre of symmetry. The operation of a $\overline{2}$ rotoinversion axis along the **c** axis, is to first transform an atom at x, y, z to $\overline{x}, \overline{y}, z$ by the operation of a 2-fold rotation, and then to operate with a cente of symmetry to transform to the final position $x, y, \overline{z}$. This is actually equivalent to a mirror plane whose normal is parallel to [001]. Similarly, as a result of a $\overline{4}$ rotoinversion axis along [001], an atom at x, y, z is first transformed to $y, \overline{x}, z$ by the operation of a 4-fold rotation, and then to $\overline{y}, x, \overline{z}$ by the operation of the centre of symmetry. This point is then replicated by the same two processes to give a point at $\overline{x}, \overline{y}, z$, which in turn is then replicated at $y, \overline{x}, \overline{z}$. An example of the $\overline{4}$ rotoinversion symmetry is shown in Fig. 3.12. The examples of $\overline{3}$ and $\overline{6}$ rotoinversion axes are left to Problem 3.8. It will be demonstrated that the operation of a $\overline{6}$-axis is equivalent to that of a 3-fold rotation axis normal to a mirror plane.

3.3.3 Combination of symmetry operations

Frequently different symmetry operations may be combined. For example, a 2-fold rotation axis can be normal to a mirror plane, and the combined operation is written as $2/m$. All combinations will also generate new symmetry operations. This can easily be demonstrated for the $2/m$ combination. If we have a point at x, y, z, the 2-fold axis along the z-axis will produce a new point at $\overline{x}, \overline{y}, z$. The action of the mirror plane on these two points will be to generate two new points:

$$\begin{aligned} x, y, z &\rightarrow x, y, \overline{z} \\ \overline{x}, \overline{y}, z &\rightarrow \overline{x}, \overline{y}, \overline{z} \end{aligned}$$

The last point is generated from x, y, z by the operation of a centre of symmetry, which means that the $2/m$ also includes this symmetry operation. It is also generated from the point $x, y, \overline{z}$ by the same 2-fold axis that gave $x, y, z \rightarrow \overline{x}, \overline{y}, z$, which highlights the self-consistency.

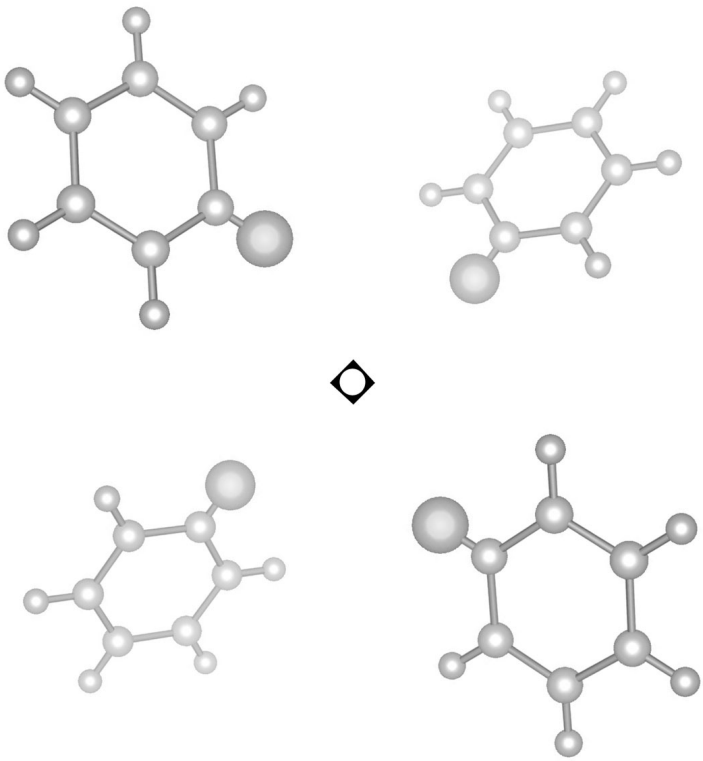


Fig. 3.12 Illustration of $\overline{4}$ rotoinversion symmetry, showing four molecules whose distance in the line of sight is emphasized by the perspective.

Some combinations of symmetry operations will actually produce the same result as another symmetry operation. We have already met an example when we noted above that a $\overline{6}$-axis is equivalent to a mirror plane normal to a 3-fold rotation axis (what we might call $3/m$). The overall result is that there are slightly fewer distinct combinations of symmetry operations than you might have expected. We now discuss the complete set of combinations appropriate for crystalline materials.

3.3.4 The 32 crystallographic point groups

Three-dimensional combinations of symmetry operations

As well as combining different symmetry operations with a single axial direction, we can also combine symmetry operations in three dimensions. These combinations of symmetry operations are called **point groups**. From the infinite number of three-dimensional point groups, only those where the rotational symmetry has $n = 1, 2, 3, 4, 6$ will occur in crystals. This restriction defines the set of **crystallographic point groups**, of which there are 32.

We have anticipated the fact that there are fewer *independent* combinations of point-symmetry operations than might have been expected, because of the fact that combinations of symmetry operations can automatically lead to the presence of additional symmetry elements. We have previously noted this for the example of $2/m$. Another example is the combination of two orthogonal 2-fold rotation axes. We start from a point at x, y, z. The first 2-fold axis, say

one that is along [100], generates the second point $x, \overline{y}, \overline{z}$. The second 2-fold rotation axis, say along [010], will operate on both points, to generate the two new points:

$$\begin{aligned} x, y, z &\rightarrow \overline{x}, y, \overline{x} \\ x, \overline{y}, \overline{z} &\rightarrow \overline{x}, \overline{y}, z \end{aligned}$$

The last point, namely that generated by the operation of *both* rotation axes, could also be generated from the original set of coordinates by operation of a third rotation axis along [001], which is the direction orthogonal to the other two rotation axes. Thus we cannot have only two 2-fold axes, since two orthogonal 2-fold rotation axes automatically generate a third 2-fold rotation axis. This combination is called 222.

Another example is when we start with a 2-fold axis along [100] and add a mirror plane whose normal is orthogonal to the 2-fold axis, say along [010] (this is in contrast to the $2/m$ condition discussed above). The action of the mirror plane on the pair of coordinates related by the 2-fold axis gives the new set of coordinates

$$\begin{aligned} x, y, z &\rightarrow x, \overline{y}, z \\ x, \overline{y}, \overline{z} &\rightarrow x, y, \overline{z} \end{aligned}$$

The operation of the combination of the 2-fold rotation axis and the orthogonal mirror plane is equivalent to that of a second mirror plane normal to [001]. Thus we have to have a combination of two mirror planes and a 2-fold axis that is orthogonal to the normals of both mirror planes. This combination is called $mm2$ (where, unlike our example, the 2-fold axis is conventionally placed along the z axis).

Labelling of the point groups

The labels representing the point groups are given three components to denote the symmetry elements along the three directions, unless there is no particular symmetry along that direction. The component is labelled with either its full symmetry or by a convenient abbreviated symbol, when the abbreviation implies additional symmetry. For example, the orthorhombic point group with a 2-fold axis and mirror plane for each axis would be labelled as $\frac{2}{m}\frac{2}{m}\frac{2}{m}$, but this is abbreviated to mmm because the three 2-fold axes are automatically generated by the three perpendicular mirror planes (Problem 3.4). The complete set of 32 crystallographic point groups is described in Table 3.2. Note that the labelling is a little complicated for the five cubic point groups.

3.4 Application of the formalism of point groups

3.4.1 Symmetry of the crystal

Point groups represent the symmetry of a point, and therefore do not directly give information about the periodicity of a crystal structure. However, they do give information about the symmetry of the environments of the atoms in the

Table 3.2 The symmetry of the 32 crystallographic point groups, labelled along the principal symmetry directions in each crystal class.

Triclinic	Point group	Schoenflies symbol	$\bar{1}$			
	1	C_1	No			
	$\bar{1}$	C_i (S_2)	Yes			
Monoclinic	Point group	Schoenflies symbol	$\bar{1}$	[010]		
	2	C_2	No	2		
	m	C_s (C_{1h})	No	m		
	$2/m$	C_{2h}	Yes	$2/m$		
Orthorhombic	Point group	Schoenflies symbol	$\bar{1}$	[100]	[010]	[001]
	222	D_2 (V)	No	2	2	2
	$mm2$	C_{2v} (C_{1h})	No	m	m	2
	mmm	D_{2h} (V_h)	Yes	$2/m$	$2/m$	$2/m$
Tetragonal	Point group	Schoenflies symbol	$\bar{1}$	$\langle 100\rangle$	$\langle 110\rangle$	[001]
	4	C_4 (V)	No	1	1	4
	$\bar{4}$	S_4	No	1	1	$\bar{4}$
	$4/m$	C_{4h}	Yes	1	1	$4/m$
	422	D_4	No	2	2	4
	$4mm$	C_{4v}	No	m	m	4
	$\bar{4}m2$	D_{2d} (V_d)	No	2	m	$\bar{4}$
	$4/mmm$	D_{4h}	Yes	$2/m$	$2/m$	$4/m$
Trigonal/ Rhombohedral	Point group	Schoenflies symbol	$\bar{1}$	$\langle 100\rangle$	$\langle 110\rangle$	[001]
	3	C_3	No	1	1	3
	$\bar{3}$	C_{3i} (S_6)	Yes	1	1	$\bar{3}$
	32	D_3	No	2	2	3
	$3m$	C_{3v}	No	1	m	3
	$\bar{3}m$	D_{3d}	Yes	$2/m$	1	$\bar{3}$
Hexagonal	Point group	Schoenflies symbol	$\bar{1}$	$\langle 100\rangle$	$\langle 110\rangle$	[001]
	6	C_6	No	1	1	6
	$\bar{6}$	C_{3h}	No	1	1	$\bar{6} = 3/m$
	$6/m$	C_{6h}	Yes	1	1	$6/m$
	622	D_6	No	2	2	6
	$6mm$	C_{6v}	No	m	m	6
	$\bar{6}m2$	D_{3h}	No	m	2	$\bar{6}$
	$6/mmm$	D_{6h}	Yes	$2/m$	$2/m$	$\bar{6}$
Cubic	Point group	Schoenflies symbol	$\bar{1}$	$\langle 100\rangle$	$\langle 110\rangle$	$\langle 111\rangle$
	23	T	No	2	1	3
	$m3$	T_h	Yes	$2/m$	1	$\bar{3}$
	432	O	No	4	2	3
	$\bar{4}3m$	T_d	No	$\bar{4}$	m	3
	$m\bar{3}m$	O_h	Yes	$4/m$	$2/m$	$\bar{3}$

crystal. This can be particularly useful in conjunction with experiments that probe the behaviour of individual atoms. For example, the point symmetry will describe the symmetry of an electric field experienced by an atom, or the shape of the electronic wave functions that may be distorted by the local electric fields. These will influence results from NMR or optical absorption resonance experiments.

The point group symmetry will also represent the macroscopic symmetry of a crystal. For example, it can describe the symmetry of a cut gemstone, or the relationship between anisotropic physical properties and the crystal axes. We will find in Chapter 6 that the point group symmetry can be used to represent the symmetry of diffraction patterns.

3.4.2 Symmetry breaking transformations

One of the important applications of the formalism of point group symmetry is to describe and quantify changes in the structure that involve a reduction in the number of symmetry operations. These changes may arise as a result of a phase transition, or dynamically through lattice vibrations, both of which are discussed in the later chapters of this book.

The use of point group symmetry to describe distortions is illustrated by the example of a crystal with orthorhombic point group *mmm* containing four atoms in the unit cell. We consider displacements along one direction, say the x-direction, as shown in Fig. 3.13. There are four independent combinations of positive and negative displacements of the atoms, each of which will preserve four symmetry elements and break four others, and these are what are shown in Fig. 3.13.

Two of these combinations involve the loss of the centre of symmetry; they will each also lose some of the other symmetry operations. The point group *mmm* has eight symmetry operations, namely the operation that changes nothing, called the **identity** operation, the three 2-fold axes, the three mirror planes, and the centre of symmetry. There are two point groups that can be

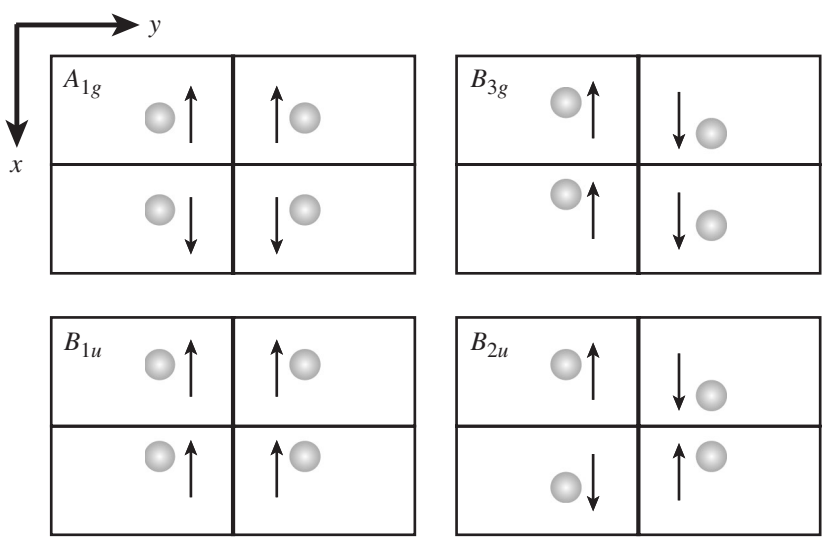


Fig. 3.13 Four sets of displacements of atoms in point group *mmm*, with symmetry labelled by the irreducible representations described in the text. The atomic motions in the A_{1g} irreducible representation do not change the symmetry, so for the corresponding picture the atoms are shown in their mean positions.

generated from mmm by the loss of the centre of symmetry, namely 222 and $mm2$, both of which have four symmetry operations. Thus the loss of the centre of symmetry will be accompanied by the loss of three other symmetry operations. There is clearly not one unique combination, and it is therefore useful to be able to quantify and characterize all possible combinations.

The characterization of the combinations of symmetry operations that can be preserved and lost in a single deformation of a structure takes us into the realm of **group theory**. We can only touch on some of the important results here, and in particular we will illustrate how the patterns of deformations that result in a lowering of symmetry can be characterized.

The main tool is the **character table**, which sorts and labels the combinations of symmetry operations that can be broken by a set of single deformations. We consider the following table of symmetry operations that lower the symmetry:

mmm	E	2_x	2_y	2_z	$\bar{1}$	m_x	m_y	m_z	
A_{1g}	1	1	1	1	1	1	1	1	mmm
B_{1g}	1	1	−1	−1	1	1	−1	−1	$2/m(x)$
B_{2g}	1	−1	1	−1	1	−1	1	−1	$2/m(y)$
B_{3g}	1	−1	−1	1	1	−1	−1	1	$2/m(z)$
A_{1u}	1	1	1	1	−1	−1	−1	−1	222
B_{1u}	1	1	−1	−1	−1	−1	1	1	$2mm(x)$
B_{2u}	1	−1	1	−1	−1	1	−1	1	$m2m(y)$
B_{3u}	1	−1	−1	1	−1	1	1	−1	$mm2(z)$

This table has one column for each symmetry operation: E stands for the identity operation; 2_x, 2_y, and 2_z represent the three 2-fold axes; $\bar{1}$ is the centre of symmetry; m_x, m_y, and m_z represent the three mirror planes. The table lists all combinations in which four of the symmetry operations are lost, together with a deformation in which the symmetry is preserved. In this table, the symbol 1 represents the symmetry operations that are preserved, and -1 represents the symmetry operations that are broken. The first column gives the label of the combination of symmetry operations, and the last column (not usually given in character tables) gives the point group symmetry that results from the loss of the particular combination of symmetry operations. The key to the labels is that the subscripts g and u (from the German **gerade** and **ungerade** for even and odd respectively) indicate that the centre of symmetry is either preserved or lost respectively, A represents the preservation of all the rotational symmetry operations, and B represents the loss of some of the rotational symmetry operations. The numbers in the subscripts simply list similar combinations. Each row in the table is called an **irreducible representation**.

The action of the irreducible representation B_{1g} is to lose two rotation axes and the mirror planes normal to these axes. This deformation preserves the centre of symmetry, one rotation axis, and the mirror plane normal to this axis. Thus the symmetry of the deformed state is $2/m$. The actions of deformations of symmetry B_{2g} and B_{3g} lead to the same point groups, but with the unique axis lying along different directions of the initial structure. The action of the irreducible representation A_{1u} is to lose the centre of symmetry and the three mirror planes, leading to the lowering of the symmetry to point group 222. The action of irreducible representation B_{1u} is to lose two of the rotation axes and the mirror plane normal to the third axis, leading to a lowering of the

symmetry to point group $mm2$. The irreducible representations B_{2u} and B_{3u} lead to the same change in symmetry, but with the unique axes along different directions.

The example of point group mmm is one of the easier groups to study because there are no subtleties. In the case where there are 3-fold, 4-fold and 6-fold rotation axes different distortions can be symmetrically equivalent; for example, the distortion of a cubic unit cell that elongates the unit cell along the x-direction is symmetrically equivalent to the distortions that elongate the unit cell along the y and z directions. Thus this type of distortion is triply degenerate, and is given the symbol T. Similarly, the two distortions that elongate a tetragonal unit cell along the x and y directions are doubly degenerate, and given the symbol E.

The use of character tables has a number of applications in the study of crystalline materials. In the context of this book, two of the important examples are in the study of displacive phase transitions (Chapter 12) and lattice vibrations (Chapter 8). In the study of phase transitions, there is a change in symmetry due to a specific distortion of the crystal structure. If the unit cell does not change size (by which we mean the lower symmetry unit cell does not span two or more unit cells of the higher symmetry structure), the distortion can be described as an irreducible representation of the point group of the higher symmetry phase. A similar description is possible if there is a change in size of the unit cell, but in this case the full description of the distortion needs also to take account of the change in translational symmetry as discussed below.

The displacements of atoms in a lattice vibration cause an instantaneous distortion of the unit cell. If the vibration has infinite wavelength (that is, if the vibration causes the same atomic displacements in each unit cell), it is possible to characterize the distortions associated with each vibration as a separate irreducible representation of the point group of the crystal. Methods exist to determine how the vibrations of a crystal decompose into the different available irreducible representations. This is useful because experimental probes such as Raman and infrared spectroscopy are sensitive only to vibrations with effectively infinite wavelength, as discussed in Chapter 10.

For vibrations that do not have infinite wavelength, similar descriptions can also apply. However, because there will be changes in the periodicity of the structure the relevant point group will be lower than that of the actual crystal. For example, a vibration travelling along the [100] direction in a cubic crystal with a finite wavelength will automatically cause a change in the translational symmetry along [100] compared to the translational symmetry along the other two crystal directions. Therefore the relevant point group from which to take the irreducible representations of the vibrations will be one of the tetragonal groups.

3.5 Crystal symmetry 2. Translational symmetry and space groups

3.5.1 Translational symmetry

We introduced the idea of **translational symmetry** when considering the symmetry invoked by the periodicity of the lattice through space. The centre

of symmetry and the rotoinversion symmetry are clearly linked to a specific point in space, but a rotation axis is a line in space and a mirror plane is a plane in space. It is therefore possible to combine translations parallel to the rotation axis or the mirror plane (usually by fraction of a lattice repeat) with these symmetry operations. These combinations extend the set of translational symmetry operations. To form these combinations may seem to be taking the formalism to new extremes, but it turns out that nature makes considerable use of these combinations in forming crystal structures.

The starting point is to merge the point groups with the 14 Bravais lattices in the same way that we constructed the point groups by combining symmetry operations. There are 61 combinations of this type, which are called **space groups**. The distribution of these combinations across the different crystal classes is

Lattice type	Lattices	Point groups	Space groups
Triclinic	1	2	2
Monoclinic	2	3	6
Orthorhombic	4	3	12
Trigonal/Rhombohedral	1	5	5
Tetragonal	2	7	14
Hexagonal	1	7	7
Cubic	3	5	15
Total	14	32	61

These space groups are simply labelled by combining the symbol for the conventional lattice, whether P, I, F, R, A, B or C, with the point group symbol. Thus, for example, the face-centred cubic structure with full symmetry is labelled $Fm\bar{3}m$. So far this is relatively straightforward, but the picture is rather complicated by the introduction of new symmetry operations that combine the rotation and reflection operations with a translation in a single operation. These new operations are called **screw axes** and **glide planes**, and when these are taken into account, the total number of space groups is increased by almost a factor of 4.

Screw axes

A **screw axis** is present when a rotation of $360°/n$, combined with a translation by a fraction $1/m$ of a lattice vector, brings the crystal back into coincidence with itself. This is labelled as n_m.

An example of a 2_1 screw axis is shown in Fig. 3.14. The initial coordinates are x, y, z. If the 2_1 axis is along [001], the action of the screw symmetry will be to generate a new set of coordinates at $\overline{x}, \overline{y}, \frac{1}{2} + z$. Similarly, a 4_1 axis along [001] will generate the set of coordinates x, y, z; $y, \overline{x}, \frac{1}{4} + z$; $\overline{x}, \overline{y}, \frac{1}{2} + z$; and $\overline{y}, x, \frac{3}{4} + z$.

Glide planes

A **glide plane** is present when a reflection in a given plane, combined with a translation along a direction parallel to the plane, brings the crystal back into coincidence with itself.

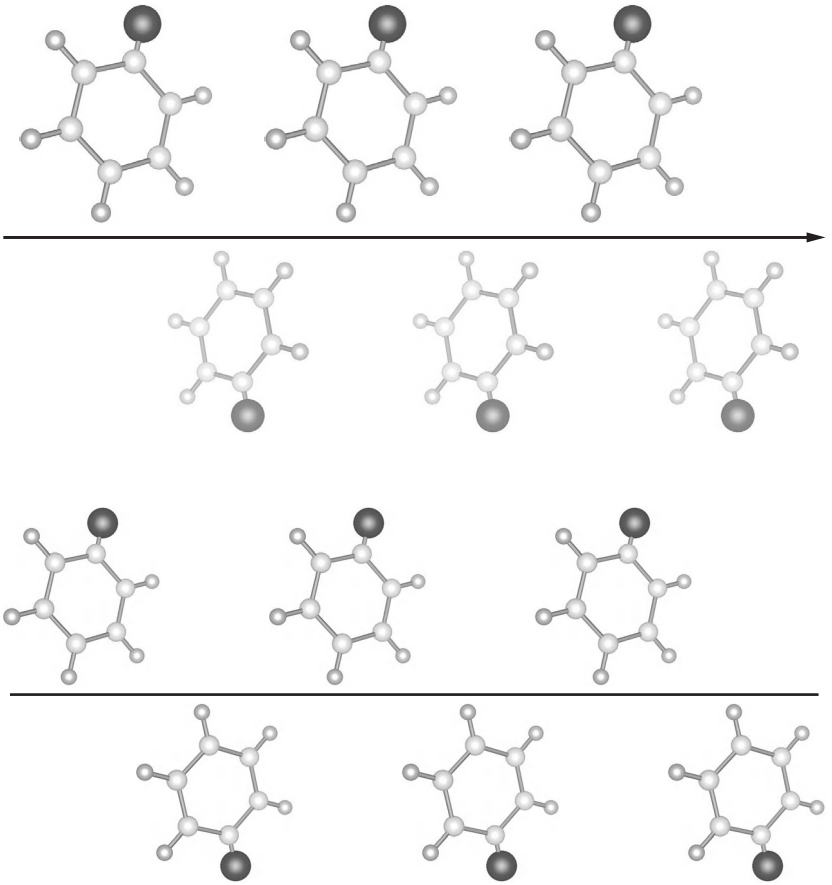


Fig. 3.14 Operation of a 2_1 screw axis on a single molecule; distance in the line of sight is emphasized by the perspective and enhanced lightening of the shading.

Fig. 3.15 Operation of a glide plane on a single molecule; distance in the line of sight is emphasized by the perspective and enhanced lightening of the shading.

An example of a glide plane is shown in Fig. 3.15. If the plane of reflection contains the **a** and **b** lattice vectors, and the translation is by half of a lattice repeat along [100], the initial coordinates at x, y, z will be replicated at $\frac{1}{2} + x, y, \overline{z}$. A glide plane is labelled according to the direction of the displacement. Thus the a, b and c glide operations involve translations of half a lattice repeat along the **a**, **b** and **c** directions respectively. Two other glide operations are the n glide which involves translation along $(\mathbf{a} + \mathbf{b})/2$ (or the other pairs of lattice vectors), and the d glide (for **diamond**) which involves translation along the diagonal of the unit cell (we will not discuss diamond glides in this book).

3.5.2 Enumeration of the space groups

It turns out that the total number of space groups, 230, is perhaps rather smaller than one might have imagined, given that the addition of different types of screw axes and glide planes significantly increases the list of symmetry operations that can be included. As in the development of point groups, certain combinations of operations can lead to the generation of others. This is most easily illustrated by the monoclinic system. The P space groups can easily be listed as the

permutations of the rotation/screw axes and mirror/glide planes to give eight space groups.

Point group	Space groups			
$2/m$	$P2/m$	$P2_1/m$	$P2/c$	$P2_1/c$
2	$P2$	$P2_1$		
m	Pm	Pc		

It should be noted that the choice of the c-glide rather than the alternative a-glide or n-glide is a matter of convention for monoclinic, just as the **b** axis is conventionally taken as the unique axis.

When we do the same thing with the C-monoclinic lattice, it turns out that the combination of the C-centre and a 2-fold axis automatically generates a set of 2_1 screw axes (this can easily be checked by the reader), so the number of C-monoclinic space groups is 5 rather than 8. With other crystal systems there are many other examples where the application of some symmetry operations automatically generate other symmetry operations, just as with point groups. The number of space groups is 230, distributed across the 14 Bravais lattices as:

Lattice type	*P*	*I*	*C*	*F*	Total
Triclinic	2	–	–	–	2
Monoclinic	8	–	5	–	13
Orthorhombic	30	9	15	5	59
Trigonal	18	–	–	–	18
Rhombohedral	7	–	–	–	7
Tetragonal	49	19	–	–	68
Hexagonal	27	–	–	–	27
Cubic	15	11	–	10	36
Total	15	39	20	15	230

3.6 Breaking the rules: aperiodic structures, incommensurate materials, and quasicrystals

3.6.1 Incommensurate or modulated structures

We will find in Chapter 6 that the experimental technique of diffraction gives information about the crystal structure. For normal ordered crystals, the periodicity and symmetry of the crystal lattice are reflected in the diffraction data, and as a result it is possible to use the signatures in the diffraction data to identify the size, shape, and symmetry of the unit cell. For many years it was recognized that some materials give diffraction data that suggest that the periodicity of their lattices are modulated by one or more superimposed waves with a repeat distance that does not match (is **incommensurate** with) the repeat distance of the crystal lattice. Often these incommensurate phases are associated with a normal phase transition (Chapter 12), with a temperature range between

the temperatures over which the two normal phases are stable. For example, quartz has an incommensurate phase that is stable for just 1.5 K sandwiched between the high-temperature hexagonal β-phase and the low-temperature trigonal α-phase. In some cases the range of stability of the incommensurate phase is much larger. In Chapter 12 we will show how an incommensurate phase can arise as a near-inevitability of the constraints on the normal phase transition due to symmetry, but there are several other mechanisms for the stability of incommensurate phases, some of which are associated with site-ordering processes, and some of which are associated with the distribution of electron energies.

Although the three-dimensionality of the crystal is broken by the incommensurate modulation, it is possible to describe the overall symmetry of an incommensurate structure in terms of higher-dimensional space groups. Not surprisingly, there are large number of space groups in each of the higher dimensions.

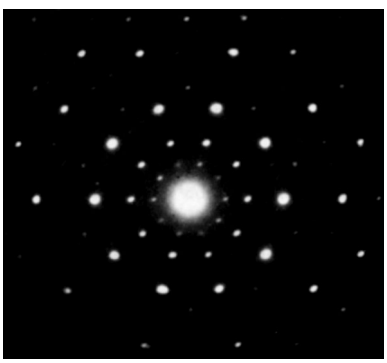


Fig. 3.16 Electron diffraction pattern of an Al–Mn alloy, showing 10-fold rotation symmetry. (Photograph courtesy of Kevin Knowles, University of Cambridge.)

3.6.2 Quasicrystals

In 1985 the worlds of condensed matter physics, materials science, chemistry, and crystallography were stunned by the publication of the results of diffraction experiments on alloys that appeared to show the existence of 5-fold and 10-fold rotational symmetry. An example is shown in Fig. 3.16. From the basic considerations discussed in this chapter, the existence of 5-fold and 10-fold rotation symmetry is inconsistent with long-range lattice symmetry, yet the form of the diffraction data looked exactly like the combination of the rotation symmetry and the long-range crystallographic order. There have now been many experiments that have confirmed these initial results, and the challenge now is to understand the origin of these intriguing results. There are a number of theoretical ideas that have been proposed. The most compelling idea is that the alloy can form units that can be described by cells with angles that are multiples of 36°. Two such objects can be tiled together to fill space completely without the generation of a regular lattice – these are known as **Penrose tiles**. The challenge then is how to decorate these objects with atoms. One mathematical approach is to follow the practice in the study of incommensurate crystals and develop a formalism based on higher-dimensional space groups, with the projection onto the three-dimensions of the experimental data giving the impression of an aperiodic structure.

Summary of chapter

- The **crystal lattice** is an infinite periodic of points. The lattice can be fully described by three basis vectors.
- The **unit cell** is the region of space bounded by eight lattice points that forms a parallelepiped. If the unit cell has the same volume as the space associated with a single lattice point, this is said to be a **primitive** unit

cell. It is often convenient to define a unit cell that has a volume that is a multiple of the volume of the primitive unit cell, when the shape of the unit cell most clearly reflects the symmetry of the lattice.

- The **lattice parameters** are the three distinct lengths of the edges of the unit cell, and the three distinct angles subtended by the edges of the unit cell.
- The constraints on the possible symmetry of a lattice lead to the definitions of seven types of lattice. These can be combined with the possibilities of forming non-primitive unit cells to give the definitions of 14 **Bravais lattices**.
- The positions of atoms within the unit cell are conveniently described using **fractional coordinates**, in which each position is represented as a combination of vectors along each of the three lattice vectors of length that is a fraction of the lattice repeat distance.
- A point is conveniently represented by the **Dirac delta function**. This enables a point to be treated using the same mathematics as used for continuous functions.
- The crystal structure can be described mathematically as a set of atom positions, the **basis** or **motif**, which are convoluted with the crystal lattice.
- **Point groups** are defined as combinations of rotation axes, mirror planes, rotoinversion axes, and a centre of symmetry. There are 32 independent **crystallographic point groups**. Point groups can be used to describe a wide range of crystal properties and behaviour.
- Distortions of a structure can be described in terms of symmetry operations that are preserved or destroyed. The different combinations of preserved and broken symmetry operations are called **irreducible representations**.
- Rotational and reflection symmetry operations can be combined with translations along a fraction of the lattice repeat to give the translational symmetry elements called **screw axes** and **glide planes** respectively.
- The point and translational symmetry operations can be combined with the Bravais lattices to give 230 **space groups**.
- Although the formalism developed to describe crystal structures is comprehensive, it is found that there are some materials that appear to break the rules. Two examples are **quasicrystals**, which appear to have 5-fold and 10-fold rotational symmetry, and **incommensurate** crystals in which the periodicity in one or more directions is modulated by a wave whose wavelength has no relation to the repeat lengths of the lattice.

Further reading

An introduction to the formalism of the crystalline state is given by both Kittel (1996) and Ashcroft and Mermin (1976). More detailed treatments can be found in Alcock (1990), Giacovazzo **et al.** (1992), Hammond (2001), Ladd (1999) and Windle (1977). Janot (1997) and Senechal (1995) are primary source references for the physics of quasicrystals.

Exercises

(3.1) Show that the presence of a plane of reflection at $x = 0$ in a crystal of orthorhombic symmetry will be accompanied by a mirror plane at $x = 1/2$. Similarly show that if there is a plane of reflection at $x = 1/4$ there will also be a plane of reflection at $x = 3/4$. (*Hint*: You will need to recognise the role of translational symmetry.)

(3.2) Show that if there is a centre of symmetry at 0, 0, 0 in an orthorhombic crystal there will also be centres of symmetry at seven other positions in the unit cell. (*Hint*: You will need to recognise the role of planes of reflection.)

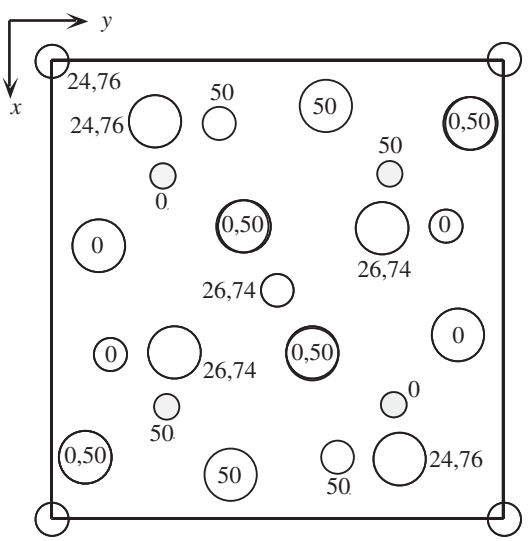
(3.3) Show that if there is a rotation axis parallel to one of the axes in an orthorhombic crystal passing through the origin there will be three other parallel rotation axes in the unit cell. Show that this is also true if the rotation axis passes not through the origin but through 1/4, 0, 0.

(3.4) Show that the point group *mmm* contains three orthogonal 2-fold axes.

(3.5) Show that the point group 4*mm* contains four mirror planes.

(3.6) Show that *I*-centred and *F*-centred tetragonal lattices are different descriptions of the same lattice.

(3.7) The figure below shows a plan of the crystal structure of the andalusite polymorph of Al_2SiO_5.

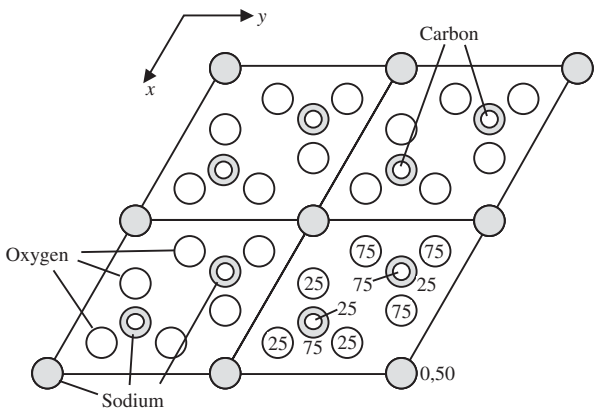


Identify all the symmetry operations, deduce the space group, and list the general positions in terms of fractional coordinates. Compare the structure with that of the sillimanite polymorph given in Problem 2.8.

(3.8) List the set of fractional atomic coordinates related by the following symmetry operations, assuming each passes through the origin and is oriented parallel to z: (a) 3-fold rotation axis; (b) 6-fold rotation axis; (c) $\bar{3}$ axis; (d) $\bar{6}$ axis. (*Note*: in each case use the axes of the lattice, for which the angle between x and y is 120°, rather than orthogonal axes.)

(3.9) List the set of fractional atomic coordinates related by the following screw symmetry operations, assuming each passes through the origin and is oriented parallel to z: (a) 3_1; (b) 6_1; (c) 6_2 axes.

(3.10) The figure below shows the hexagonal crystal structure of Na_2CO_3.



Identify the point symmetry of the sites occupied by the different atoms. (*Hint*: identify all point symmetry operations of the crystal systematically, and deduce the point symmetry of the atoms as arising from the symmetry elements that intersect each atom.) We will meet Na_2CO_3 again in Chapter 12, where you will be able to appreciate this author's interest in this material!

(3.11) Show that two orthogonal 2-fold rotation axes parallel to [100] and [010] that intersect at the origin of the unit cell will generate a third 2-fold rotation axis parallel to [001] which also passes through the origin. Compare this with the case where the 2-fold rotation parallel to [010] passes through the point 0, 0, $\frac{1}{4}$. List the general coordinates for both cases.

(3.12) List the general coordinates of the following orthorhombic space groups, ensuring that if a centre of symmetry is to present it lies on the origin of the unit cell: (a) *Pmma*,

(b) $Pccb$, (c) $P2_1mn$, (d) $Pbmb$. In each case, state the symmetry elements not explicitly stated in the space group label.

(3.13) The two space groups $Pmna$ and $Pnma$ sound similar, but are listed as separate space groups. What are the differences between these two space groups? (*Note*: it is not adequate to simply state that the mirror plane affects a different direction, since how we specify the directions in the orthorhombic group is arbitrary. The important point is the actual symmetry.)

4 The reciprocal lattice

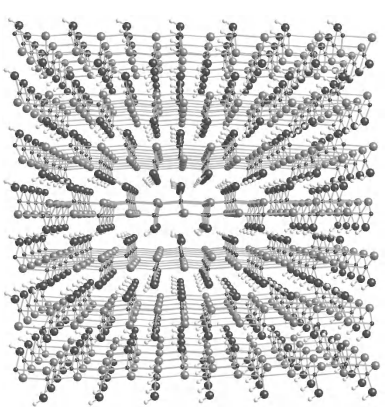


Fig. 4.1 View of the crystal structure of $Cu(OH)_2$. Even in a structure this complicated, several planes of atoms with different orientations can clearly be seen.

4.1 The concept of the reciprocal lattice

The concept of the **reciprocal lattice** arises from the need to have a formal mathematical representation of the planes of atoms within a crystal. The existence of planes of atoms plays an important role in many aspects of crystalline materials – images of planes of atoms in crystals are given in Fig. 4.1. The shapes of crystals are determined by the relative energies of surfaces defined by the crystal planes. For example, the faces of the calcite crystals shown in Chapter 1 are determined by one particular plane of atoms, as shown in Fig. 4.2. Vibrations of the crystal involve motions of planes of atoms. Diffraction of beams of radiation involve constructive reflections from planes of atoms. With such a diverse range of applications involving planes of atoms, a consistent and general representation is clearly of value, and the formalism of the reciprocal lattice fulfils this role.

Understanding the reciprocal lattice is perhaps the largest conceptual leap in the study of crystalline materials. However, there is a substantial pay-off in working to understand the reciprocal lattice, because it is one of the most powerful tools we have. Three examples will illustrate this point. *First*, the reciprocal lattice facilitates the interpretation of diffraction data, as will be discussed in Chapter 6. *Second*, the reciprocal lattice provides a method for relatively easy formal calculations involving surfaces or lattice planes. Examples of such calculations are common in calculations of physical properties, as discussed in Chapter 7. *Third*, the reciprocal lattice facilitates calculations involving waves within crystals, whether of electrons or lattice vibrations, as we will meet in Chapters 5, 8, and 12. In fact, although we may proceed by developing some formal definitions of the reciprocal lattice, the same definitions will arise naturally from many general theories, as we will show later in this book.

The starting point is to recognize that the orientation of a plane is usefully described by a vector normal to the plane. If we can identify two independent lattice vectors lying within a plane, the normal will be the vector cross-product of these vectors, as shown in Fig. 4.3. Formally, we can define two lattice vectors lying in the plane, $\mathbf{t}_1 = U_1\mathbf{a} + V_1\mathbf{b} + W_1\mathbf{c}$ and $\mathbf{t}_2 = U_2\mathbf{a} + V_2\mathbf{b} + W_2\mathbf{c}$, where U_1, V_1, W_1 and U_2, V_2, W_2 are sets of integers (Section 3.2.1). The vector normal to $\mathbf{t}_1$ and $\mathbf{t}_2$ is

$$\mathbf{t}_1 \times \mathbf{t}_2 = (U_1V_2 - U_2V_1)\mathbf{a} \times \mathbf{b} + (V_1W_2 - V_2W_1)\mathbf{b} \times \mathbf{c} + (W_1U_2 - W_2U_1)\mathbf{c} \times \mathbf{a} \quad (4.1)$$

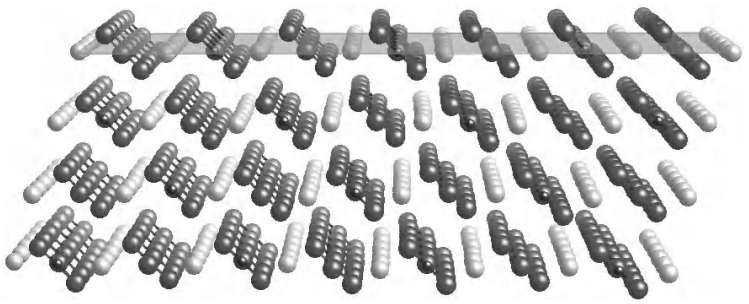


Fig. 4.2 Planes of atoms in calcite associated with the common crystal faces. The common face, which is the face seen in the crystals shown in Fig. 1.1, are marked by the plane at the top. This surface is charge-neutral since the charge of the Ca^{2+} cation is balanced by the negative charge of the CO_3^{2-} molecular anion.

It is clear from this general equation that the set of vector cross-products, $\mathbf{a} \times \mathbf{b}$, $\mathbf{b} \times \mathbf{c}$, and $\mathbf{c} \times \mathbf{a}$, will form a useful basis for the general description of plane normals. Moreover, since the prefactors $U_1V_2 - U_2V_1$, etc. are combinations of integers, they are themselves integers which can take on any values. We assign labels h, k, l to the three integers and write

$$\mathbf{t}_1 \times \mathbf{t}_2 = h\mathbf{a} \times \mathbf{b} + k\mathbf{b} \times \mathbf{c} + l\mathbf{c} \times \mathbf{a} \tag{4.2}$$

The important point to note here is that we have three basis vectors which can be combined in any integral multiples. By taking the complete set of integers h, k, l, we define a new lattice. This new lattice is called the **reciprocal lattice**.

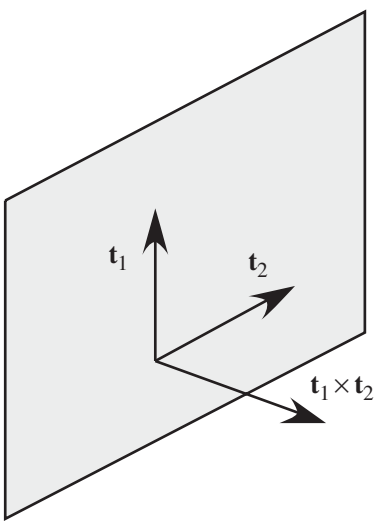


Fig. 4.3 Representation of a plane normal vector being defined as the cross-product of two vectors lying in the plane.

4.2 Definitions

Nomenclature

It is useful to normalize the vector normals. One standard normalization factor is

$$2\pi/(\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}) = 2\pi/V_{\text{cell}} \tag{4.3}$$

where $V_{\text{cell}} = \mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$ is the volume of the unit cell, as given by eqn (3.2). Sometimes the factor of 2π is not included in the normalization – it actually doesn't matter, provided that it is always understood whether the factor is included or not. With the factor of 2π included, the normalization gives the three vectors

$$\mathbf{a}^* = \frac{2\pi}{V_{\text{cell}}}\mathbf{b} \times \mathbf{c} \tag{4.4}$$

$$\mathbf{b}^* = \frac{2\pi}{V_{\text{cell}}}\mathbf{c} \times \mathbf{a} \tag{4.5}$$

$$\mathbf{c}^* = \frac{2\pi}{V_{\text{cell}}}\mathbf{a} \times \mathbf{b} \tag{4.6}$$

We then have the following products:

$$\begin{aligned} &\mathbf{a}^* \cdot \mathbf{a} = 2\pi \quad &&\mathbf{a}^* \cdot \mathbf{b} = 0 \quad &&\mathbf{a}^* \cdot \mathbf{c} = 0 \\ &\mathbf{b}^* \cdot \mathbf{a} = 0 \quad &&\mathbf{b}^* \cdot \mathbf{b} = 2\pi \quad &&\mathbf{b}^* \cdot \mathbf{c} = 0 \\ &\mathbf{c}^* \cdot \mathbf{a} = 0 \quad &&\mathbf{c}^* \cdot \mathbf{b} = 0 \quad &&\mathbf{c}^* \cdot \mathbf{c} = 2\pi \end{aligned} \tag{4.7}$$

The factor of $1/V_{\text{cell}}$ clearly ensures that the products $\mathbf{a}^* \cdot \mathbf{a}$, etc. are reduced to a common result. When the factor 2π is not included in the normalization, we have $\mathbf{a}^* \cdot \mathbf{a} = 1$. This may seem to be somewhat neater, but it can be useful to explicitly include the factor of 2π when considering waves, as in diffraction (Chapter 6) and lattice dynamics (Chapter 8).

We can now formally define the **reciprocal lattice**. It is the lattice of points formed from the infinite set of vectors

$$\mathbf{d}^*_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{4.8}$$

where h, k, l are the integers we met above, and which can take any values. The vectors $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$ are the basis vectors of the new lattice, and $\mathbf{d}^*_{hkl}$ is a vector from the origin to any point in this new lattice in a way that is analogous to the lattice vector we met earlier. This vector $\mathbf{d}^*_{hkl}$ is called a **reciprocal lattice vector**, and the set of integers h, k, l are called the **Miller indices**. Usually the plane normal to $\mathbf{d}^*_{hkl}$ is represented by the notation (hkl) – note that the notation includes the brackets.

Just as the crystal lattice gives the unit cell, the reciprocal lattice gives a corresponding **reciprocal unit cell**, defined by the three vectors $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$. The lengths of these vectors are analogous to the dimensions of the unit cell of the real crystal lattice. We also have the corresponding angles α^*, β^*, and γ^*. The product $\mathbf{a}^* \cdot (\mathbf{b}^* \times \mathbf{c}^*)$ is the volume of the reciprocal unit cell.

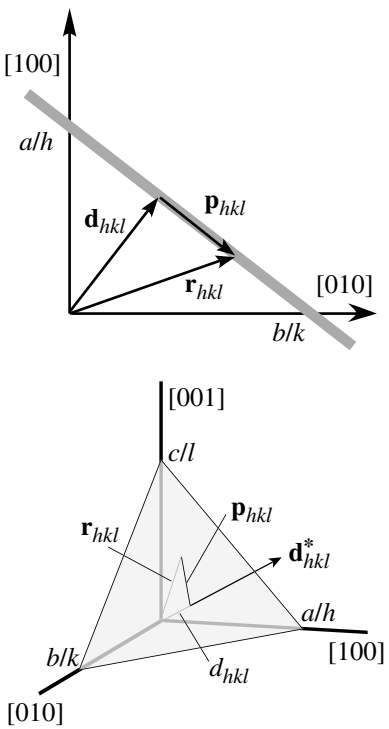


Fig. 4.4 A crystal plane (hkl) viewed edge-on. $\mathbf{r}_{hkl}$ is a vector to a point in the plane. This vector is decomposed into the sum of $\mathbf{d}_{hkl}$, a vector from the origin to the plane and normal to the plane, and $\mathbf{p}_{hkl}$, a vector lying in the plane. As demonstrated in the text, the (hkl) plane intersects the $\mathbf{a}$, $\mathbf{b}$ axes at the points a/h and b/k respectively.

4.2.1 Geometry of the reciprocal lattice and its link to the crystal lattice

We can now explore the geometry of the reciprocal lattice vectors. Consider the set of equispaced lattice planes shown in Fig. 4.4, normal to the vector $\mathbf{d}^*_{hkl}$, with separation $d_{hkl} = 2\pi/|\mathbf{d}^*_{hkl}|$. Suppose $\mathbf{r}_{hkl}$ is a vector from the origin to any point in the first plane. We can decompose $\mathbf{r}_{hkl}$ into a component parallel to $\mathbf{d}^*_{hkl}$, which we denote as $\mathbf{d}_{hkl}$, and a component lying within the plane, which we denote as $\mathbf{p}_{hkl}$. It thus follows that

$$\mathbf{r}_{hkl} \cdot \mathbf{d}^*_{hkl} = \mathbf{d}_{hkl} \cdot \mathbf{d}^*_{hkl} = d_{hkl} \times |\mathbf{d}^*_{hkl}| = 2\pi \tag{4.9}$$

We now take the special case $\mathbf{r}_{hkl} = \alpha\mathbf{a}$. We can therefore write

$$\alpha\mathbf{a} \cdot \mathbf{d}^*_{hkl} = 2\pi \tag{4.10}$$

Substituting for $\mathbf{d}^*_{hkl}$, we have

$$\alpha\mathbf{a} \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) = 2\pi\alpha h = 2\pi \tag{4.11}$$

This immediately gives $\alpha = 1/h$, leading to the conclusion that the first lattice plane normal to $\mathbf{d}^*_{hkl}$ intersects the $\mathbf{a}$ axis at a distance a/h from the origin. Similarly, $\mathbf{d}^*_{hkl}$ intersects the $\mathbf{b}$ and $\mathbf{c}$ axes at distances b/k and c/l from the origin respectively. This point is illustrated in Fig. 4.4.

We now consider the special case where $h = 1$ and $k = l = 0$, which is represented as (100). This set of planes is shown in Fig. 4.5. Since the first (hkl) plane intersects the $\mathbf{b}$ and $\mathbf{c}$ axes at distances b/k and c/l respectively, the zero values of k and l imply that the planes do not intersect these axes and

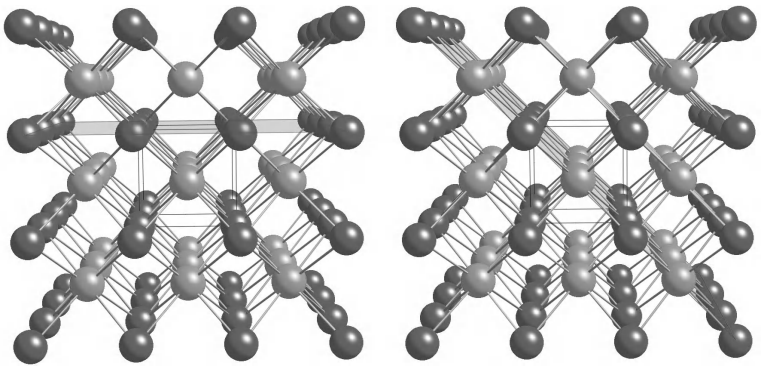


Fig. 4.5 (100) and (110) planes in the CsCl structure, shown by the shading.

are therefore parallel to them. Thus the (100) plane that includes the origin also contains the vectors **b** and **c**, and $\mathbf{d}^*_{100} = \mathbf{a}^*$ is normal to both **b** and **c**. This is an important point, because in a crystal in which the three axes **a**, **b**, and **c** are not orthogonal, the reciprocal lattice vectors $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$ will not be parallel to the lattice vectors **a**, **b**, and **c** respectively. This point is illustrated for the example of a monoclinic lattice in Fig. 4.6. The important point is that $\mathbf{a}^*$ is perpendicular to **b** and **c**, and similarly for the other two reciprocal lattice vectors. Only in cases where the crystal axes are orthogonal will the reciprocal lattice vectors lie along the same directions as the corresponding lattice vectors.

It should be appreciated that only in crystals with orthogonal axes are the reciprocal lattice vectors parallel to the corresponding lattice vectors, and only in those crystals where the volume of the unit cell is equal to abc does $a^* = 2\pi/a$, $b^* = 2\pi/b$, and $c^* = 2\pi/c$. In the monoclinic system the volume of the unit cell is equal to $abc \sin\beta$, so that $a^* = 2\pi/a \sin\beta$, $b^* = 2\pi/b$, and $c^* = 2\pi/c \sin\beta$. In the monoclinic case, we also have $\beta^* = 180° - \beta$. In the hexagonal or trigonal case ($a = b$, $\alpha = \beta = 90°$, $\gamma = 120°$), the volume of the unit cell is equal to $abc \sin\gamma = \sqrt{3}a^2b/2$, giving $a^* = b^* = 2\pi/a \sin\gamma = 4\pi/\sqrt{3}a$, $c^* = 2\pi/c$, and $\gamma^* = 60°$.

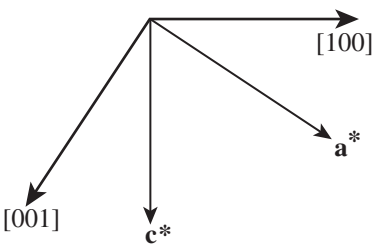


Fig. 4.6 The directions of the real and reciprocal lattice vectors in a monoclinic system. The lattice vector **b** and reciprocal lattice vector $\mathbf{b}^*$ are both normal to the plane of the diagram.

4.2.2 Relationship between real and reciprocal lattice parameters

In the general case, the real and reciprocal lattice parameters are related by the following equations:

$$a^* = \frac{2\pi}{a \sin\beta \sin\gamma^*} = \frac{2\pi}{a \sin\beta^* \sin\gamma}$$

$$b^* = \frac{2\pi}{b \sin\gamma \sin\alpha^*} = \frac{2\pi}{b \sin\gamma^* \sin\alpha}$$

$$c^* = \frac{2\pi}{c \sin\alpha \sin\beta^*} = \frac{2\pi}{c \sin\alpha^* \sin\beta}$$

$$\cos\alpha^* = \frac{\cos\beta \cos\gamma - \cos\alpha}{\sin\beta \sin\gamma} \qquad \cos\alpha = \frac{\cos\beta^* \cos\gamma^* - \cos\alpha^*}{\sin\beta^* \sin\gamma^*}$$

$$\cos\beta^* = \frac{\cos\gamma\cos\alpha - \cos\beta}{\sin\gamma\sin\alpha} \qquad \cos\beta = \frac{\cos\gamma^*\cos\alpha^* - \cos\beta^*}{\sin\gamma^*\sin\alpha^*}$$

$$\cos\gamma^* = \frac{\cos\alpha\cos\beta - \cos\gamma}{\sin\alpha\sin\beta} \qquad \cos\gamma = \frac{\cos\alpha^*\cos\beta^* - \cos\gamma^*}{\sin\alpha^*\sin\beta^*}$$

These equations are greatly simplified if we have orthogonal axes.

4.2.3 Interplanar spacing and the reciprocal lattice parameters

The interplanar spacing d_{hkl} can be obtained from the inverse of $|d^*_{hkl}|^2$, which is defined in eqn 4.8. For the general triclinic case, $|d^*_{hkl}|^2$ is given as

$$\frac{1}{4\pi^2}|d^*_{hkl}|^2 = \frac{1}{d^2_{hkl}} = h^2a^{*2} + k^2b^{*2} + l^2c^{*2} + 2klb^*c^*\cos\alpha^* + 2hla^*c^*\cos\beta^* + 2hka^*b^*\cos\gamma^* \qquad (4.12)$$

The cosines of the angles of the reciprocal unit cell can be calculated from the angles of the real-space unit cell using the relations given above, and once the angles have been calculated it will then be possible to calculate the values of a^*, b^*, and c^*.

4.2.4 Reciprocal lattice vectors and atomic structure

With the geometric link between the crystal and reciprocal lattice established, we can now embelish this by drawing in the atoms in the unit cells. The formalism is particularly useful when we consider crystal surfaces. Figure 4.7 shows a small crystallite of a two-dimensional monatomic close-packed crystal. The faces are clearly cleaved along well-defined lattice planes, and these can be indexed according to the geometric relationships we have established.

4.3 Non-primitive lattices

4.3.1 Some general principles and practical methods

We highlighted earlier the fact that there is some degree of choice in how we define the basis vectors of a lattice. In face-centred and body-centred lattices it was shown that it is convenient to define the lattice vectors in a way that highlights the symmetry, even though we have the 'penalty' of having a unit cell that encompasses more than one lattice point. The choice of description is nothing more than just a description: the existence of the lattice is quite independent of the choice of description. What is important is that the particular description of the lattice will have an impact on the way we describe the reciprocal lattice. The reciprocal lattice of any crystal will exist with an arrangement of points in reciprocal space that is independent of how we label these points. However, when we have chosen to describe the crystal lattice using

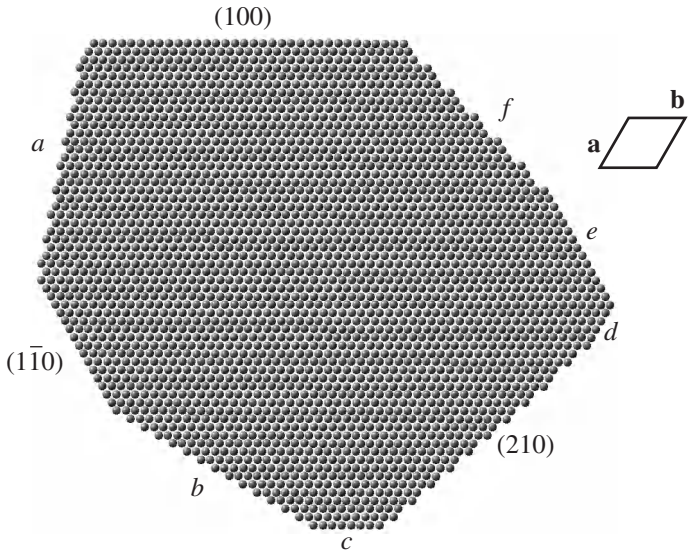


Fig. 4.7 A small crystallite of a two-dimensional close-packed plane of atoms, showing some indexed faces. The indexing of other faces is left for the problems at the end of the chapter. The unit cell is *not* drawn to scale.

a non-primitive description of the unit cell, we must be careful not to 'invent' matching reciprocal lattice points that do not really exist. A simple, if somewhat trivial, example will illustrate the dilemma. Suppose we have a crystal lattice with orthogonal lattice vectors **a**, **b**, and **c**. We are at liberty to define a non-primitive setting for the unit cell described by the lattice vectors $\mathbf{a}'$, $\mathbf{b}'$, and $\mathbf{c}'$, such that $\mathbf{a}' = 2\mathbf{a}$, $\mathbf{b}' = \mathbf{b}$, and $\mathbf{c}' = \mathbf{c}$. Now if we follow the rules outlined above, we would obtain a reciprocal lattice with $a'^* = 2\pi/2a$. However, in this setting, the reciprocal lattice point (100) is only half-way to the first reciprocal lattice point of the primitive setting, which is at distance $2\pi/a$. Can we invent new reciprocal lattice points simply by changing our definition of the unit cell in the original crystal lattice? The answer is '*no*', because the reciprocal lattice is as fundamental a construction as the real crystal lattice. We therefore need to establish a method to enable us to differentiate false reciprocal lattice points generated by the choice of the crystal unit cell from the true reciprocal lattice vectors.

We start by defining a primitive unit cell with lattice vectors **a**, **b**, and **c**. Supposed we now describe a new set of lattice vectors as linear combinations of the primitive lattice vectors, such that

$$\begin{aligned}\mathbf{a}' &= M_{11}\mathbf{a} + M_{12}\mathbf{b} + M_{13}\mathbf{c}\\ \mathbf{b}' &= M_{21}\mathbf{a} + M_{22}\mathbf{b} + M_{23}\mathbf{c}\\ \mathbf{c}' &= M_{31}\mathbf{a} + M_{32}\mathbf{b} + M_{33}\mathbf{c}\end{aligned}$$

We can write this in matrix form:

$$\begin{pmatrix}\mathbf{a}'\\ \mathbf{b}'\\ \mathbf{c}'\end{pmatrix} = \mathbf{M}\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\\ \mathbf{c}\end{pmatrix} \tag{4.13}$$

where matrix **M** has element M_{ij}. We now need to investigate how this relationship holds for other lattice quantities, such as the reciprocal lattice

vectors, (h, k, l) indices, and fractional coordinates (x, y, z). We define the matrix of the components of the lattice vectors as

$$\mathbf{L} = \begin{pmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{pmatrix}; \qquad \mathbf{L}' = \begin{pmatrix} a'_x & a'_y & a'_z \\ b'_x & b'_y & b'_z \\ c'_x & c'_y & c'_z \end{pmatrix} \tag{4.14}$$

where components are defined such that $\mathbf{a} = (a_x, a_y, a_z)$, etc. It follows that

$$\mathbf{L}^{-1} = \frac{1}{2\pi} \begin{pmatrix} a^*_x & b^*_x & c^*_x \\ a^*_y & b^*_y & c^*_y \\ a^*_z & b^*_z & c^*_z \end{pmatrix}; \qquad \mathbf{L}'^{-1} = \frac{1}{2\pi} \begin{pmatrix} a'^*_x & b'^*_x & c'^*_x \\ a'^*_y & b'^*_y & c'^*_y \\ a'^*_z & b'^*_z & c'^*_z \end{pmatrix} \tag{4.15}$$

Using these matrix representations, we can write

$$\mathbf{L}'\,\mathbf{L}'^{-1} = (\mathbf{M}\,\mathbf{L})\,\mathbf{L}'^{-1} = \mathbf{1} \tag{4.16}$$

where we have substituted $\mathbf{L}' = \mathbf{M}\,\mathbf{L}$, and written $\mathbf{1}$ as the unit matrix. A simple manipulation yields

$$\mathbf{L}'^{-1} = \mathbf{L}^{-1}\,\mathbf{M}^{-1} \tag{4.17}$$

For consistency, we need the matrix

$$\mathbf{L}^* = \frac{1}{2\pi} \begin{pmatrix} a^*_x & a^*_y & a^*_z \\ b^*_x & b^*_y & b^*_z \\ c^*_x & c^*_y & c^*_z \end{pmatrix} = (\mathbf{L}^{-1})^{\mathrm{T}} \tag{4.18}$$

Similarly we can define $\mathbf{L}'^* = (\mathbf{L}'^{-1})^{\mathrm{T}}$. We can transpose eqn 4.17 for $\mathbf{L}'^{-1}$ to give

$$\mathbf{L}'^* = (\mathbf{M}^{-1})^{\mathrm{T}}\,\mathbf{L}^* \tag{4.19}$$

This can be rewritten as

$$\begin{pmatrix} \mathbf{a}'^* \\ \mathbf{b}'^* \\ \mathbf{c}'^* \end{pmatrix} = (\mathbf{M}^{-1})^{\mathrm{T}} \begin{pmatrix} \mathbf{a}^* \\ \mathbf{b}^* \\ \mathbf{c}^* \end{pmatrix} \tag{4.20}$$

which gives the equation for the transformation of reciprocal lattice vectors between two coordinate systems, analogous to eqn 4.13 for the transformation of lattice vectors.

Now we consider the effect of the transformation of axes on the Miller indices. A point in reciprocal space is independent of how we choose to define the axes, so we have

$$(h\,k\,l) \begin{pmatrix} \mathbf{a}^* \\ \mathbf{b}^* \\ \mathbf{c}^* \end{pmatrix} = (h'\,k'\,l') \begin{pmatrix} \mathbf{a}'^* \\ \mathbf{b}'^* \\ \mathbf{c}'^* \end{pmatrix} \tag{4.21}$$

$$= (h'\,k'\,l')\,(\mathbf{M}^{-1})^{\mathrm{T}} \begin{pmatrix} \mathbf{a}^* \\ \mathbf{b}^* \\ \mathbf{c}^* \end{pmatrix} \tag{4.22}$$

By dividing out the common right-hand matrix of both sides, and then transposing, we obtain

$$\begin{pmatrix} h \\ k \\ l \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} h' \\ k' \\ l' \end{pmatrix} \tag{4.23}$$

$$\Rightarrow \quad \begin{pmatrix} h' \\ k' \\ l' \end{pmatrix} = \mathbf{M} \begin{pmatrix} h \\ k \\ l \end{pmatrix} \tag{4.24}$$

It is left as an exercise (Problem 4.4) to show that we have a corresponding relation for the fractional coordinates:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = (\mathbf{M}^{-1})^{\mathrm{T}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{4.25}$$

4.3.2 Primitive and non-primitive lattices: application to bcc and fcc lattices

We now apply the general approach to two conventional lattices, namely bcc and fcc. For both lattices, the link between the primitive and conventional unit cells is shown in Fig. 4.8. We will show how the reciprocal lattices associated with the primitive and conventional unit cells are related.

First we take the case of the bcc lattice. The matrix linking the primitive and conventional lattice vectors is

$$\mathbf{M} = \begin{pmatrix} -1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{pmatrix} \tag{4.26}$$

Hence the Miller indices of the primitive setting, $(h_\mathrm{p}, k_\mathrm{p}, l_\mathrm{p})$, are related to those of the conventional setting, $(h_\mathrm{c}, k_\mathrm{c}, l_\mathrm{c})$, by

$$\begin{pmatrix} h_\mathrm{p} \\ k_\mathrm{p} \\ l_\mathrm{p} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -h_\mathrm{c} + k_\mathrm{c} + l_\mathrm{c} \\ h_\mathrm{c} - k_\mathrm{c} + l_\mathrm{c} \\ h_\mathrm{c} + k_\mathrm{c} - l_\mathrm{c} \end{pmatrix} \tag{4.27}$$

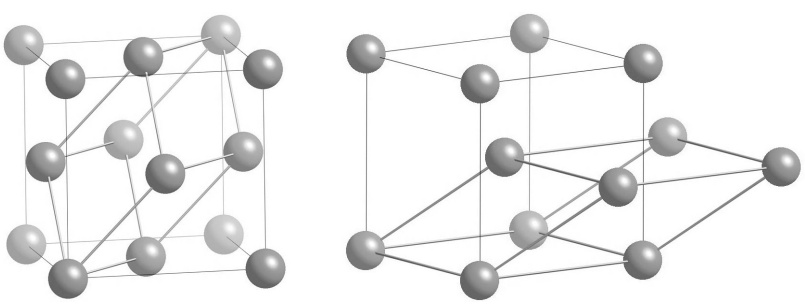


Fig. 4.8 Primitive and conventional fcc (left) and bcc (right) unit cells.

The values of (h_p, k_p, l_p) must be integers, since this is the primitive lattice. Thus the sum $h_c + k_c + l_c$ must be even. This means that the (100) and (111) points in the bcc setting are not true reciprocal lattice points, although the (110) point is.

Now we take the case of the fcc lattice. The matrix linking the primitive and conventional lattice vectors is

$$\mathbf{M} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} \tag{4.28}$$

Thus we have

$$\begin{pmatrix} h_p \\ k_p \\ l_p \end{pmatrix} = \frac{1}{2} \begin{pmatrix} h_c + k_c \\ k_c + l_c \\ h_c + l_c \end{pmatrix} \tag{4.29}$$

In order for the values of (h_p, k_p, l_p) to be integers, the values of h_c, k_c, and l_c must all be odd or all even. Thus the (100) and (110) points in the fcc setting are not true reciprocal lattice points, although now the (111) point is.

As an interesting aside, this analysis has shown that the reciprocal lattice of the bcc lattice is a fcc lattice, and vice versa. This can be a useful *aide memoire*.

4.4 The reciprocal lattice as the Fourier transform of the crystal lattice

In Chapter 6 we will meet the idea that the scattering of a beam of radiation by matter is described by the Fourier transform of the mathematical description of the matter (Fourier transforms were introduced in Chapter 1, and are discussed in Appendix B). To consider the Fourier transform of the lattice, we start by considering a one-dimensional lattice with lattice parameter a. The lattice can be described mathematically as a series of delta functions centred on each lattice point (see Chapter 3, eqn 3.8):

$$\mathcal{L}(x) = \sum_n \delta(x - na) \tag{4.30}$$

where n is an integer.

The Fourier transform of the one-dimensional lattice is given by

$$\begin{aligned} \mathcal{R}(q) &= \int \exp(iqx)\mathcal{L}(x)\,\mathrm{d}x \\ &= \sum_n \int \exp(iqx)\delta(x - na)\,\mathrm{d}x \\ &= \sum_n \exp(iqna) = \sum_n \cos(qna) \end{aligned} \tag{4.31}$$

where we have only retained the cosine term because the lattice is centrosymmetric (since $\cos -x = \cos x$ and $\sin -x = -\sin x$). The Fourier transform of the δ-function is demonstrated in Appendix B, eqn B.3.1.

For an arbitrary value of q, some of the cosine terms are negative and some are positive, and since the number of terms tends to infinity, the positive terms and the negative terms in the summation cancel to give an overall zero value for the Fourier transform. This situation is only avoided if the arguments of all the cosines are equal modulo 2π, which is achieved if

$$q = \text{integer} \times 2\pi/a \tag{4.32}$$

This gives a one-dimensional reciprocal lattice, with points separated by $a^* = 2\pi/a$. In the limit $n \to \infty$, the Fourier transform of the one-dimensional lattice will be equal to

$$\mathcal{R}(q) = \sum_h \delta(q - ha^*) \tag{4.33}$$

This argument is easily generalized to three dimensions, where the three-dimensional lattice is defined by the set of points with lattice vectors

$$\mathcal{L}(\mathbf{r}) = \sum_{U,V,W} \delta(\mathbf{r} - (U\mathbf{a} + V\mathbf{b} + W\mathbf{c})) \tag{4.34}$$

where U, V, and W are the integers defining lattice vectors (Chapter 3). The Fourier transform is

$$\mathcal{R}(\mathbf{k}) = \sum_{U,V,W} \cos(\mathbf{k} \cdot (U\mathbf{a} + V\mathbf{b} + W\mathbf{c})) \tag{4.35}$$

Again, the summation will equal zero for any general vector $\mathbf{k}$, but will be non-zero if the arguments of the cosine terms are equal modulo 2π. This is true when $\mathbf{k}$ is a reciprocal lattice vector, that is, for

$$\begin{aligned} \mathbf{k} \cdot \mathbf{r} &= (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \cdot (U\mathbf{a} + V\mathbf{b} + W\mathbf{c}) \\ &= 2\pi \times (hU + kV + lW) \end{aligned} \tag{4.36}$$

Thus the Fourier transform of the three-dimensional lattice is equal to

$$\mathcal{R}(\mathbf{k}) = \sum_{h,k,l} \delta(\mathbf{k} - (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*)) \tag{4.37}$$

The important point to draw from this analysis is that the Fourier transform of the crystal lattice is a function in reciprocal space, with values that are only non-zero when the wave vector is equal to a reciprocal lattice vector. Thus we have obtained the reciprocal lattice simply by taking the Fourier transform of the crystalline lattice.

4.5 Reciprocal space and the Brillouin zone

We first considered the reciprocal lattice as a construction that defined the set of vector plane normals. Then we showed that the reciprocal lattice vectors have a definite relation to the lattice vectors. Finally, we showed that the reciprocal lattice is the Fourier transform of the crystal lattice. The crystal lattice is defined

in the space we are used to thinking about, and one in which we can define any general vector without it having to have any relationship to the crystal lattice. The same is true of the space containing the reciprocal lattice. Since the dimensions of the vectors of the reciprocal lattice are the inverse of length, any vector that has the same dimensions can be described in the same space as the reciprocal lattice. We call this space the **reciprocal space**.

There is one important set of vectors that are naturally described as vectors in reciprocal space, and these are the wave vectors that characterize the wavelength and direction of a travelling wave. Specifically, a travelling wave can be defined by a variable u at any position $\mathbf{r}$ that varies as

$$u(\mathbf{r}) = u_0 \exp(i\mathbf{k} \cdot \mathbf{r}) \tag{4.38}$$

where $\mathbf{k}$ is the wave vector, and is related to the wavelength λ by $|\mathbf{k}| = 2\pi/\lambda$. u_0 is the amplitude. In a wave of this form, all points in a plane normal to $\mathbf{k}$ that contains the position $\mathbf{r}$ will have the same value of $\mathbf{k} \cdot \mathbf{r}$ and hence the same value of u. Again we are talking about planes in real space defined by a plane normal that is a vector in reciprocal space.

This is not just a matter of mathematical formalism, as we will show shortly. But first we should mention two types of waves that are important in crystals, namely electron waves and atomic vibrations. In both cases, the important quantity that defines the wave is its wave vector, which corresponds to a vector in reciprocal space. There is an important result that follows from the wave equation. Suppose we have a position vector $\mathbf{R}$ that corresponds to a point on the crystal lattice: $\mathbf{R} = U\mathbf{a} + V\mathbf{b} + W\mathbf{c}$. Suppose also that we have two wave vectors that are related by a reciprocal lattice vector:

$$\mathbf{k}' = \mathbf{k} + \mathbf{G}; \qquad \mathbf{G} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{4.39}$$

Since

$$\mathbf{G} \cdot \mathbf{R} = 2\pi(hU + kV + lW) \tag{4.40}$$

we have

$$\mathbf{k}' \cdot \mathbf{R} = \mathbf{k} \cdot \mathbf{R} + \mathbf{G} \cdot \mathbf{R} = \mathbf{k} \cdot \mathbf{R} + 2\pi(hU + kV + lW) \tag{4.41}$$

The last term is equal to 2π times an integer, and thus

$$\exp(i\mathbf{k}' \cdot \mathbf{R}) = \exp(i\mathbf{k} \cdot \mathbf{R}) \tag{4.42}$$

This result shows that two waves whose wave vectors differ by a reciprocal lattice vector have an identical effect on a crystal lattice point $\mathbf{R}$. This means that we often only need to consider the set of wave vectors that lie within a volume equal to that of the reciprocal unit cell. How we define this volume is arbitrary, but there is one construction that is particularly useful, which is called the **Brillouin zone construction**. This recognizes the fact that since the origin of reciprocal space and the nearest reciprocal lattice points are equivalent, there is likely to be a special status associated with the points half-way between them. This is reinforced by the fact that $\mathbf{k}$ and $-\mathbf{k}$ give waves that are simply the complex conjugates of each other. The Brillouin zone construction is illustrated in two dimensions in Fig. 4.9. The boundaries are defined by the planes that

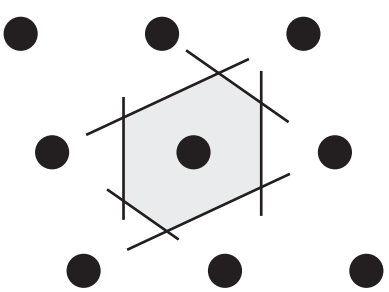


Fig. 4.9 Construction of the Brillouin zone.

bisect the vectors from the origin of the reciprocal lattice to its closest reciprocal lattice points. This gives a shape that can be packed to fill space, and which is called the **Brillouin zone**. Since there is one shape per reciprocal lattice point, the volume of the Brillouin zone is equal to that of the reciprocal unit cell.

The points that are on the boundaries of the Brillouin zone and exactly halfway between two reciprocal lattice points have a special significance. We denote this wave vector as $\mathbf{k}_{\mathrm{ZB}}$, where the subscript ZB denotes the **zone boundary**. The reciprocal lattice has a centre of symmetry, so the wave vector $-\mathbf{k}_{\mathrm{ZB}}$ is equivalent to $\mathbf{k}_{\mathrm{ZB}}$. Moreover, the points $-\mathbf{k}_{\mathrm{ZB}}$ and $\mathbf{k}_{\mathrm{ZB}}$ are separated by a reciprocal lattice vector. Now consider a wave vector close to and parallel to $\mathbf{k}_{\mathrm{ZB}}$, which we write as $(1-\xi)\mathbf{k}_{\mathrm{ZB}}$. Since the reciprocal lattice has a centre of symmetry, this is equivalent to $-(1-\xi)\mathbf{k}_{\mathrm{ZB}}$. An equivalent point is obtained by adding $2\mathbf{k}_{\mathrm{ZB}}$, which is equivalent to a reciprocal lattice vector, giving the point $(1+\xi)\mathbf{k}_{\mathrm{ZB}}$. Thus we have the result that $(1+\xi)\mathbf{k}_{\mathrm{ZB}}$ is equivalent to $(1-\xi)\mathbf{k}_{\mathrm{ZB}}$. The zone boundary point $\mathbf{k}_{\mathrm{ZB}}$ is thus a special point of reflection. If we imagine any function $f(\mathbf{k})$, we have

$$\left.\frac{\partial f(\mathbf{k})}{\partial \mathbf{k}}\right|_{\mathbf{k}=\mathbf{k}_{\mathrm{ZB}}} = 0 \tag{4.43}$$

We will meet this condition when we consider lattice dynamics, where we will find that (except in some cases) the derivative of the frequency with respect to wave vector is zero at these special wave vectors. In these cases, the waves are standing waves with zero group velocity. In the case of electrons in crystals, the derivative of the electron energy with respect to wave vector is also zero at these special wave vectors.

Summary of chapter

- The **reciprocal lattice** is the Fourier transform of the real-space lattice. The basis vectors of the reciprocal lattice vectors are each normal to a pair of basis vectors of the real-space lattice.
- Each **reciprocal lattice vector** is normal to a particular set of planes of lattice points in the real-space lattice. The length of the reciprocal lattice vector is the inverse of the spacing of the corresponding planes in the real-space lattice.
- In cases where it is convenient to use a non-primitive description of the crystal lattice, there is a matrix that can be used to transform the lattice vectors of the primitive and non-primitive lattices. This matrix can be converted to transform the reciprocal lattice vectors between the two descriptions, and hence to deduce the form of the reciprocal lattice corresponding to the non-primitive description of the real-space lattice.
- It is convenient to describe the space surrounding each reciprocal lattice point using the formalism of the **Brillouin zone**.
- Two waves whose wave vectors differ by a reciprocal lattice vector have identical effects on a real-space lattice point. As a result, the set of wave vectors contained within a single Brillouin zone are sufficient to describe the waves associated with lattice vibrations and electronic wave functions.

Further reading

Because of its central role in solid state physics and chemistry, the reciprocal lattice is described in many texts. Both Kittel (1996) and Ashcroft and Mermin (1976) give a good introduction. Other useful references are Elliott (1998), Giacovazzo *et al.* (1992), Hammond (2001), and McKie and McKie (1986).

Exercises

(4.1) Show that two lattice vectors, $[U_1 V_1 W_1]$ and $[U_2 V_2 W_2]$, lie in a plane parallel to (hkl), where $h = V_1 W_2 - V_2 W_1$, $k = W_1 U_2 - W_2 U_1$ and $l = U_1 V_2 - U_2 W_1$. Similarly, show that the intersection of two lattice planes, $(h_1 k_1 l_1)$ and $(h_2 k_2 l_2)$, is a vector parallel to $[U V W]$, where $U = k_1 l_2 - k_2 l_1$, $V = l_1 h_2 - l_2 h_1$ and $W = h_1 k_2 - h_2 k_1$.

(4.2) Index the unindexed surfaces of the crystallite shown in Fig. 4.7.

(4.3) Show that the volume of the reciprocal unit cell is equal to $(2\pi)^3 / V_{\text{cell}}$, where V_{cell} is the volume of the unit cell. (*Note*: there are several ways of approaching this, one being the brute-force manipulation of equations, and the other is to write the components of the lattice vectors as a 3×3 matrix, to note that the determinant of the matrix is the volume enclosed by the three vectors, and to use the mathematical result for square matrices **A** and **B** that $\det[\mathbf{A} \times \mathbf{B}] = \det[\mathbf{A}] \times \det[\mathbf{B}]$.)

(4.4) Derive eqn 4.25.

(4.5) For two descriptions of the unit cell, described by eqn 4.13, show that $\det[\mathbf{M}]$ is equal to the factor difference in the volumes of the unit cells in the two descriptions. (*Note*: use the result for the determinant of the product of two matrices given in Problem 4.3.)

(4.6) Show that the points (100) and (110) in the reciprocal space of a face-centred cubic lattice are equivalent. Similarly, show that the points (100) and (111) in the reciprocal space of a body-centred cubic lattice are equivalent.

(4.7) Show that the reciprocal lattice points of a C-centred lattice are restricted to the condition $h + k = 2n$, where n is an integer. Generalize this result for A and B-centred lattices.

(4.8) Show that when using the trigonal setting of the rhombohedral lattice (Appendix D), the reciprocal lattice points are limited to the condition $h - k + l = 3n$, where n is an integer.

Atomic bonding in crystals

5

5.1 Bonding and the variety of crystal structures

The diversity of crystal structures is due to the factors associated with bonding. We have met four different types of crystal structures for materials containing equal numbers of two types of atoms (NaCl, CsCl, cubic and hexagonal ZnS). There is a plethora of structures for materials of the general form AB_2. We have met the cubic structure of CaF_2, and we have noted that SiO_2 can be found in many different structures. The high-pressure phase of SiO_2 has the rutile structure of TiO_2, and this compound exists in other different crystal structures.

Part of the rationalization for the different structure types comes from the relative sizes of the atoms. However, this only tells part of the story. It is known that under different conditions of temperature or pressure the same substance can exist in different structures, in which case relative sizes of atoms are not the overriding factor. Bonding is more complex an issue than the simple packing of spheres into the most efficient arrangements, although some aspects of atomic bonding may reduce to this. In this chapter we explore *some* of the issues associated with how different aspects of binding lead to the stabilities of crystal structures. We have to start from the study of equilibrium thermodynamics, since this provides a 'broad brush' understanding of crystal stability. The details of bonding, which we will then go on to discuss, link the macroscopic general principles to the specific interatomic interactions. The thermodynamic arguments will show that even the details of bonding only provide part of the overall picture, and we will have to explore the link between interatomic interactions and atomic vibrations to complete the picture.

5.2 Thermodynamic preamble: the context of the binding energy

Formally, the stability of a crystal structure with respect to other possible structures (whether drastically or subtlely different) is determined by the **Gibbs free energy**, G. This is written as

$$G = U + PV - TS \tag{5.1}$$

where U is the internal energy, P the pressure, V the volume, T the temperature, and S the entropy. There are two other important thermodynamic

quantities: the **enthalpy** H is given by

$$H = U + PV \tag{5.2}$$

and the **Helmholtz free energy**, F, given by

$$F = U - TS \tag{5.3}$$

The Helmholtz free energy also has the alternative symbol A. The four thermodynamic potentials are also written in differential form:

$$\begin{aligned} \mathrm{d}U &= T\,\mathrm{d}S - P\,\mathrm{d}V \\ \mathrm{d}H &= T\,\mathrm{d}S + V\,\mathrm{d}P \\ \mathrm{d}F &= -S\,\mathrm{d}T - P\,\mathrm{d}V \\ \mathrm{d}G &= -S\,\mathrm{d}T + V\,\mathrm{d}P \end{aligned}$$

The various quantities in the thermodynamic potentials are related to their differentials by standard thermodynamic relationships, called the **Maxwell relations**:

$$S = -\left.\frac{\mathrm{d}F}{\mathrm{d}T}\right|_V = -\left.\frac{\mathrm{d}G}{\mathrm{d}T}\right|_P \tag{5.4}$$

$$V = \left.\frac{\mathrm{d}H}{\mathrm{d}P}\right|_S = \left.\frac{\mathrm{d}G}{\mathrm{d}P}\right|_T \tag{5.5}$$

$$T = \left.\frac{\mathrm{d}U}{\mathrm{d}S}\right|_V = \left.\frac{\mathrm{d}H}{\mathrm{d}S}\right|_P \tag{5.6}$$

$$P = -\left.\frac{\mathrm{d}U}{\mathrm{d}V}\right|_S = -\left.\frac{\mathrm{d}F}{\mathrm{d}V}\right|_T \tag{5.7}$$

Some of these relations will prove to be of considerable value in later discussions, which is why we briefly review them here. Note that the fact that S and V are positive quantities means that the derivatives of the free energy with respect to temperature and pressure must be negative and positive respectively. Moreover, since entropy increases on heating, there will be a downwards curvature of the free energy with respect to temperature, and the fact that the volume decreases with pressure means that $\mathrm{d}G/\mathrm{d}P$ will decrease on increasing pressure as will be seen in Fig. 5.2.

The various components of the free energy reflect different aspects of the properties and behaviour of the crystal. The entropy arises from two main sources. One, which we will discuss only fleetingly in places, is due to disorder of atom positions. The second source is the vibrations of the atoms, which we will discuss in some detail in Chapters 8 and 9. The internal energy also has two main sources, namely the potential energy of interactions between atoms, and the energy of atomic vibrations. In this chapter we will focus specifically on the energies associated with interatomic interactions, taking both an empirical perspective, and a more formal view from quantum mechanics.

The concept of the free energy has considerable power. The important point about the free energy is that the equilibrium state of a material is that for

which the free energy is a minimum. Moreover, the free energy is a continuous function of external variables such as T or P, which means that it will not undergo any discontinuous change. Therefore, at a temperature and pressure at which a structure undergoes a phase transition, the free energies of both phases will be equal. This is illustrated by considering a pressure–temperature phase diagram of the form given in Fig. 5.1. This involves two phases, which we label α and β. If it is possible to understand the nature of the free energy, whether in general terms or with a specific theoretical model, it is possible to understand why specific crystal structures are stable over any particular range of temperature or pressure.

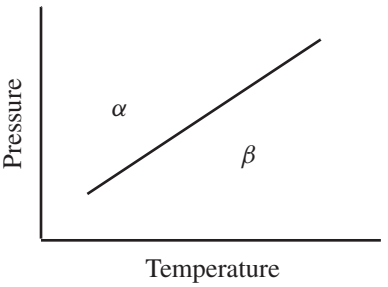


Fig. 5.1 Phase diagram with phase boundary in P and T with two phases, α and β.

In Fig. 5.2 we show general representations of the curves of free energy versus temperature for both phases at a fixed pressure, and the curves of free energy versus pressure at a fixed temperature. We consider first the phase transition at a fixed pressure, which we arbitrarily set as zero. The stable phase at low temperature, denoted as α in Fig. 5.1, will have the lower internal energy, and the phase that is stable at high temperature, which is denoted as β, will have the higher entropy. Thus at some temperature T_c the free energies of the two phases will be identical:

$$\begin{aligned} G_\alpha|_{T=T_c,P=0} = G_\beta|_{T=T_c,P=0} \quad &\Rightarrow \quad U_\alpha - T_c S_\alpha = U_\beta - T_c S_\beta \\ &\Rightarrow \quad T_c = \frac{U_\alpha - U_\beta}{S_\alpha - S_\beta} = \frac{\Delta U}{\Delta S} \end{aligned} \tag{5.8}$$

The values of ΔU and ΔS used in this equation should be those at the fixed pressure and $T = T_c$. However, if the dependence of U and S on pressure and temperature is not significant – and we note that both will change in the same sense so that the ratio $\Delta U/\Delta S$ will vary even less – measurements or calculations of U and S at some general temperature and pressure can be used to give a reasonable estimate of T_c.

We now consider the case of a high-pressure phase transition at low temperature (which we set arbitrarily to zero). The low-pressure phase, α, will have the lower internal energy at low pressures. However, it will also have the higher volume, and at a high pressure the high-pressure phase, β, will have the lower enthalpy even though it has the higher internal energy. At some pressure,

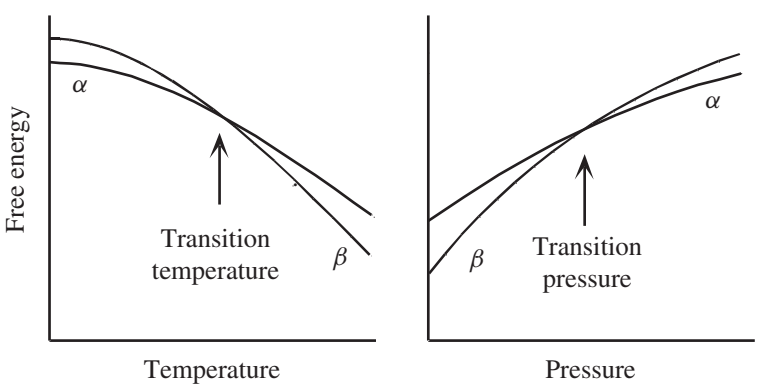


Fig. 5.2 Plots of crossing of free energies with variation of T or P for two phases of a material. Note the slopes that give S and V respectively.

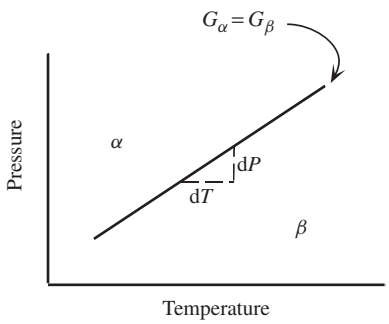


Fig. 5.3 Phase diagram in P and T, which represents the line where the free energies of both phases are the same. The Clausius–Clapeyron equation is found by considering small incremental changes in T and P along the phase boundary.

which we denote as P_c, the free energies of the two phases will be equal:

$$G_\alpha|_{P=P_c,T=0} = G_\beta\big|_{P=P_c,T=0} \Rightarrow U_\alpha + P_c V_\alpha = U_\beta + P_c V_\beta$$
$$\Rightarrow P_c = -\frac{U_\alpha - U_\beta}{V_\alpha - V_\beta} = -\frac{\Delta U}{\Delta V} \tag{5.9}$$

where the values of ΔU and ΔV are strictly those at the transition pressure. However, simply by knowing that the difference between the volumes of the two phases has an opposite sign to the difference between the internal energies is enough to understand that the phase transition will happen at a positive pressure. A rough measure or calculation of ΔU and ΔV will be enough to give an estimate of the transition pressure P_c.

Now we consider a phase diagram across both pressure and temperature, as illustrated in Fig. 5.3. For all points along the phase diagram $G_\alpha = G_\beta$. We take one point along the phase boundary, and envisage moving along the phase boundary by infinitesimal increments of temperature and pressure dT and dP respectively. This will give changes in the free energies of the two phases:

$$dG_\alpha = -S_\alpha\, dT + V_\alpha\, dP$$
$$dG_\beta = -S_\beta\, dT + V_\beta\, dP$$
$$\Rightarrow \frac{dP}{dT} = \frac{S_\alpha - S_\beta}{V_\alpha - V_\beta} = \frac{\Delta S}{\Delta V} \tag{5.10}$$

where we have used the result that $dG_\alpha = dG_\beta$, and assumed that the entropy and volume of each phase do not change with the infinitesimal changes in T and P. This relation is a form of the well-known **Clausius-Clapeyron** equation, relating the ratio of the changes in entropy and volume to the slope of the phase diagram (often this relation contains the substitution $L/T = \Delta S$, where L is the latent heat of the phase transition). It might be expected that a phase with higher volume is likely to be less stiff than the phase with lower volume. Because of this, the phase with higher volume is likely to be somewhat more flexible, with lower-frequency vibrations, and hence with higher entropy (this statement will be properly quantified in Chapters 8 and 9). Thus we expect the ratio $\Delta S/\Delta V$ to be positive (although some exceptions are known).

We have gone through the thermodynamic arguments to highlight how much we can understand about structural stability from qualitative reasoning. Our task now is to flesh out this qualitative understanding by considering the forces between atoms. In this chapter we will focus on the bonding contribution to the internal energy. The vibrational contributions to the internal energy and the entropy will be discussed in later chapters, but these will themselves also depend on the interactions between atoms in the crystals.

5.3 Lattice energy

Many models of the interactions between atoms in solid and liquid phases have been developed in terms of interactions between pairs of atoms. In this case, the energy of a pair of atoms will depend only on the distance between them. Formally, we denote the interaction between a pair of atoms, labelled i and j,

as $\phi_{ij}(r_{ij})$, where r_{ij} is the separation between the two atoms, and we note that ϕ_{ij} depends on the types of atoms as well as on their separation. The potential energy of the crystal with all atoms at rest, called the **lattice energy**, is therefore given as

$$U_{\text{Lattice}} = \sum_{\langle ij \rangle} \phi_{ij}(r_{ij}) \tag{5.11}$$

The angle brackets in the summation denote the fact that in summing over i and j we should not count any interaction twice. It is not difficult to extend the definition of the lattice energy to account for interactions that depend on groups of atoms, some of which we will meet below.

The concept of the lattice energy has widespread application in the study of crystals. Because the lattice energy considers atoms to be at rest, it clearly neglects all effects of dynamics, and hence corresponds to zero temperature. Moreover, the quantum mechanics of vibrations (Chapter 9) gives the result that the ground state can never have atoms completely at rest – this is the so-called **zero point motion** – and the lattice energy is therefore technically the classical energy at zero temperature. However, for many properties, the effects of dynamics are less significant than the basic uncertainties in the model representations of forces between atoms, so the neglect of dynamics can be justified as a reasonable compromise.

The lattice energy gives three main areas of application. The first is the pure energy. In some cases it is possible to measure the energy of a crystal, for example by measuring the energy of sublimation or energy of solution of a material. Moreover, for a material that exists in different phases, the difference in energy between the two phases can be determined from the phase diagram, and this difference can be calculated from the lattice energies of the two phases.

The second area of application of the lattice energy is that the minimum of the lattice energy can be assumed to be a close approximation to the minimum of the free energy, and hence it is possible to calculate a structure that is close to the equilibrium structure of a material by minimization of the lattice energy. For example, we consider the special case when the potential energy functions depend on the separations of pairs of atoms, r_{ij}. In this case, the dependence of the energy on all structural parameters p_l (essentially the lattice parameters and atomic coordinates) can be expressed through the dependence of the atomic separations on the structural parameters. Thus we can write the set of equilibrium conditions, namely that for which the lattice energy is a minimum with respect to each structural parameter, as

$$\frac{\partial E}{\partial p_l} = \sum_{i,j} \frac{\partial \phi_{ij}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial p_l} = 0 \tag{5.12}$$

A minimization algorithm is required to solve this equation, and some comments on this are given in Appendix F. In practice, the lattice parameters and atomic coordinates are adjusted until eqn (5.12) is satisfied, giving the crystal structure with the lowest lattice energy. It follows that in this structure there are no residual forces on the atoms and no stresses on the structure.

The third application of the lattice energy is in calculations of the numerical coefficients for physical properties that arise as derivatives of the energy. We

have already met a number of these, such as the dielectric and piezoelectric effects. The more formal treatment is given in Chapter 7, and the details of the calculations are given in Appendix K.

5.4 Models of bonding

5.4.1 Coulomb energy

The principal binding energy in solids arises from the **Coulomb energy**. In ionic solids, this binding energy simply comes from the interactions between charged ions. However, the short-range repulsion interactions between neighbouring ions – the interactions that keep neighbouring ions apart and which balance the Coulomb attraction at short range – and the binding mechanisms in covalent materials will require us to consider explicitly the Coulomb interactions involving the electrons, namely the interactions between electrons and atomic nuclei, and between different electrons. The interactions involving the electrons will require us to take account of quantum mechanics, and it is convenient here to separate those effects that are solely due to the Coulomb energy from those that also depend on quantum mechanics.

The Coulomb energy between two ions i and j with charges Q_i and Q_j respectively is

$$U(r_{ij}) = \frac{Q_i Q_j}{4\pi\epsilon_0 r_{ij}} \tag{5.13}$$

Whilst this equation is well-understood, its implementation in calculating the energy of a crystal as far from trivial. The point is illustrated by calculation of the Coulomb energy for a crystal with the NaCl structure, where ions have charges $\pm Q$. Starting from the ion at the origin, we can sum the Coulomb energy neighbour by neighbour. This procedure is shown graphically in Fig. 5.4. The first neighbours are the ions of opposite charge at $\frac{1}{2}, 0, 0$ and related positions. This gives a negative contribution to the energy. The second neighbours are the same type of ions at the sets of positions related to $\frac{1}{2}, \frac{1}{2}, 0$, and these give a positive contribution to the overall energy. Although the $1/r$ factor means that each interaction is lower in value than for the nearest-neighbour interactions, there are 12 second neighbours compared to six nearest neighbours, so the net contribution is enough to change the sign of the running total. The extension of the sum to further neighbours is shown in Fig. 5.4. What is clear from the way the sum is proceeding is that it is not converging. In fact, the non-convergence of the sum when taken term by term can only be overcome by arranging the counting is special ways.

The Coulomb sum is most easy to rearrange for crystals of high symmetry. For simple cubic crystals, the Coulomb sum for one mole of formula units (e.g. one mole of cation/anion pairs) can be represented by

$$U_{\text{Madelung}} = \alpha \frac{N_A Q_+ Q_-}{4\pi\epsilon_0 R} \tag{5.14}$$

where R is the nearest-neighbour cation–anion separation, and α is a number known as the **Madelung constant**. Values of the Madelung constant for several simple cubic structures are listed below, with the relationship between the

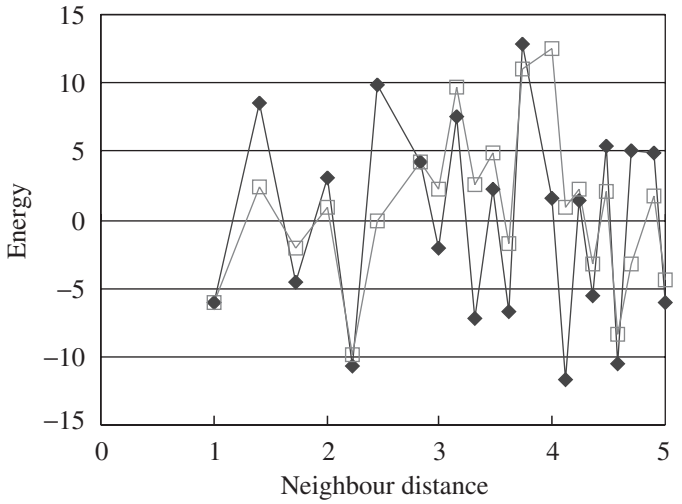


Fig. 5.4 Graphical representation of the sum over Coulomb interactions in the NaCl lattice, starting from the nearest-neighbour interactions, and building the sum neighbour by neighbour. The plot shows the individual terms (filled diamonds) and the running total (open squares) for the summation of 1/distance as the sum extends to further neighbours. The limiting value of the total as the neighbour distance tends to infinity is 1.75.

nearest-neighbour distance R and the unit cell parameter a:

Structure type	α	R
NaCl	1.74756	$a/2$
CsCl	1.76267	$\sqrt{3}a/2$
Cubic ZnS	1.63806	$\sqrt{3}a/4$
CaF_2	2.51939	$\sqrt{3}a/4$

In fact, because we have not included constants such as ionic charge in the term-by-term sum shown in Fig. 5.4, the sum should be converging towards the value of the Madelung constant. This is much smaller than the oscillations in the term-by-term summation, and highlights the problems of convergence.

For more complicated crystals, there is no simple representation of the Coulomb contribution to the lattice energy. The most common approach to compute the Coulomb lattice sum is that due to Ewald. The starting point is to represent $1/r$ by a definite integral:

$$\frac{1}{r} = \frac{2}{\sqrt{\pi}} \int_0^\infty \exp(-r^2\rho^2)\,\mathrm{d}\rho \tag{5.15}$$

This sum is then separated into two indefinite integrals:

$$\frac{1}{r} = \frac{2}{\sqrt{\pi}} \int_0^g \exp(-r^2\rho^2)\,\mathrm{d}\rho + \frac{2}{\sqrt{\pi}} \int_g^\infty \exp(-r^2\rho^2)\,\mathrm{d}\rho \tag{5.16}$$

These two terms are known as the **error function** and **complementary error function** respectively. Values of these functions have been tabulated. The second function has the property that it decreases rapidly with increasing r (see Problem 5.2). Thus the second term, when used in the lattice sum, convergences well when summing neighbour by neighbour. The first term does not converge rapidly as it stands, but it can be transformed to reciprocal space, and when the corresponding lattice sum is performed in reciprocal space it also converges

reasonably quickly. The speed of the convergence of both terms in reciprocal and real space respectively is controlled by the value of g, which is chosen to ensure that both sums have more-or-less equal numbers of terms.

5.4.2 Repulsive interactions

In ionic crystals the Coulomb summation favours as small a crystal volume as possible. The reduction in volume is then limited by the close contact of atoms when the electron orbitals of the two ions begin to overlap. The electrons in the overlap region interact by the mutual Coulomb repulsion, which raises their energy. Moreover, the Pauli exclusion principle also splits the energy levels of the electrons in the overlap region, with an overall increase of kinetic and potential energies of the electrons. The net result is that the interaction energy increases as the atoms get closer together.

The repulsive energy is frequently parametrized using a form proposed by Born and Mayer:

$$\phi_{\text{Repulsive}}(r_{ij}) = B_{ij} \exp(-r_{ij}/\rho_{ij}) \tag{5.17}$$

where B_{ij} and ρ_{ij} are constants. This form has some justification from quantum mechanics. For simple ionic crystals a combination of the Born–Mayer repulsive energy and the Coulomb energy can be adequate to explain many crystal properties. Another representation of the repulsive interaction, which is more of historical interest but better lends itself to simple calculations, is

$$\phi_{\text{Repulsive}}(r_{ij}) = A_{ij} r_{ij}^{-n} \tag{5.18}$$

where n is an integer with value typically between 6 and 12.

5.4.3 Combination of Coulomb and Born–Mayer interactions: example of alkali halides

The Madelung representation of the Coulomb energy can be easily combined with the Born–Mayer interaction restricted to nearest neighbours (a reasonable restriction). If the cations and anions both have m nearest neighbours ($m = 4$ for the cubic ZnS structure, $m = 6$ for the NaCl structure, and $m = 8$ for the CsCl structure), the lattice energy per mole of ion pairs can be written as

$$\begin{aligned} U_{\text{Lattice}} &= U_{\text{Madelung}} + U_{\text{Repulsive}} \\ &= \alpha \frac{N_{\text{A}} Q_+ Q_-}{4\pi\epsilon_0 R} + m N_{\text{A}} B \exp(-R/\rho) \end{aligned} \tag{5.19}$$

The equilibrium value of R is obtained from the minimization of U_{Lattice}:

$$\frac{\partial U_{\text{Lattice}}}{\partial R} = -\frac{U_{\text{Madelung}}}{R} - \frac{U_{\text{Repulsive}}}{\rho} = 0 \tag{5.20}$$

If we know the values of the charges, and the two parameters B and ρ in the Born–Mayer function, we can obtain the equilibrium value of R as the solution to this equation (noting that the equation is not linear, since both U_{Madelung} and $U_{\text{Repulsive}}$ are functions of R). Whilst this example appears to be rather trivial,

an analogous calculation can be performed on any complex crystal to determine the complete set of lattice parameters, or to determine the equilibrium atomic positions. Of course, this problem cannot be tackled by hand, but with modern computers the equilibrium structure of a complex material can be calculated in a matter of seconds.

On the other hand, if R is known from experiment, and one has additional data (such as the lattice energy and the bulk modulus) it is possible to determine the free parameters in the energy expressions, such as the values of ionic charge and the parameters B and ρ in eqn 5.19. Expressions for the lattice energy and equilibrium value of R are given above. The bulk modulus is defined as

$$B = -V\frac{\partial P}{\partial V} = V\frac{\partial^2 U}{\partial V^2} \tag{5.21}$$

For a cubic crystal where the volume occupied by one ion pair is equal to γR^3 (values of γ are $16/\sqrt{27}$, 2, and $8/\sqrt{27}$ for the ZnS, NaCl, and CsCl structures respectively) the bulk modulus can be rewritten as (Problem 5.3)

$$\begin{aligned} B &= \frac{1}{9N_{\mathrm{A}}\gamma R}\frac{\partial^2 U_{\mathrm{Lattice}}}{\partial R^2} \\ &= \frac{1}{9N_{\mathrm{A}}\gamma R}\left(\frac{U_{\mathrm{Madelung}}}{R^2} + \frac{U_{\mathrm{Repulsive}}}{\rho^2}\right) \end{aligned} \tag{5.22}$$

where the factor of N_{A} is used because the earlier expressions have been set to be equal to energy per mode of ion pairs.

We now have three equations with the variables U_{Madelung}, $U_{\mathrm{Repulsive}}$, and ρ, which can be solved separately. The value of U_{Madelung} will give the values of Q_+ and Q_-, and by combining the values of ρ and $U_{\mathrm{Repulsive}}$ it will be possible to determine the value of B in the Born–Mayer interaction. The analysis for alkali halides is left for Problems 5.5 and 5.6 – it can be demonstrated that consistency with the data is achieved with ionic charges that are close to their formal values, and with values of the ratio R/ρ of order 10.

This example has shown that it is possible to obtain exact values for the few parameters in the model interatomic potential energy functions. A similar approach can be used for more complex crystalline materials, using as experimental data the derivatives of the lattice energy with respect to all structural parameters, and any number of physical properties such as elastic constants or dielectric constants. In general there will not be an exact solution of the full set of equations, and the best set of parameters in the model interatomic potential are those that give the lowest values of the first derivatives of the lattice energy and the closest agreement between the calculated and observed physical properties. The general point can therefore be appreciated from the specific example.

5.4.4 Dispersive interactions: binding in molecular crystals

In the models discussed above, the binding energy is primarily due to the Coulomb interaction. In some materials, the direct Coulomb energy is much less significant. These include the crystals of the rare gas atoms (Ne, Ar, Kr, Xe), crystals composed of symmetric molecules (such as N_2, CH_4, C_{60}, and many

organic molecules), and some layer compounds (such as graphite). In these cases, the primary binding energy is the **dispersive interaction**, which is the most important term of a series of interactions between neutral atoms.

The origin of this interaction comes from the fact that the instantaneous distribution of electrons in an atom is not uniform: a uniform distribution is only a representation of the average. Without going into details, we simply state that the interactions between two neutral atoms arise from the energies associated with correlated fluctuations of the electron densities of the two atoms. These fluctations give rise to instantaneous diopole (and higher multipole) moments of the electron density, and the electrostatic interactions between these correlated moments give rise to a negative energy. The most important of these is the interaction between coupled fluctuations of the dipole moments, which give rise to an energy that varies as $-r^{-6}$. The interaction is always attractive.

In modelling the energies of molecular crystals, the dispersive interaction is combined with a repulsive term. One example is the **Lennard–Jones** potential:

$$\phi_{ij}(r_{ij}) = -4\varepsilon\left[\left(\frac{\sigma}{r}\right)^6 - \left(\frac{\sigma}{r}\right)^{12}\right] \tag{5.23}$$

The meaning of the parameters ε and σ is explored in Problem 5.1. Alternatively, and more usually, the dispersive interaction is combined with the Born–Mayer repulsive energy to give the function known as the **Buckingham interaction**:

$$\phi_{ij}(r_{ij}) = -\frac{C_{ij}}{r_{ij}^6} + B_{ij}\exp(-r_{ij}/\rho_{ij}) \tag{5.24}$$

The Buckingham potential has a defined minimum, and because of this it is actually a useful parametrization of any potential energy curve. The combination of a short-range repulsion and a longer-range attractive energy is seen as a net potential energy that has a minimum at some distance, a rapid rise in energy for shorter distances, and a more gradual rise for longer distances, similar to the potential energy curve shown in Chapter 1. (The Buckingham potential also has a maximum at a low value of r, and its value tends to $-\infty$ as $r \to 0$. This does not often cause problems, but can do so in some applications.)

5.4.5 Shell models

The dispersive interaction arises because the electronic charge distribution of an atom is polarizable. The dispersive interaction arises from fast correlated fluctuations of the electron charge distribution. However, the polarizability is also important in that electric fields in the crystal, whether internal to the crystal or externally applied, can distort the electronic charge distribution. This effect is particularly important for the calculation of dielectric constants, but will also have an impact on lattice energies with respect to polarizable ions that are not on centres of symmetry. The most successful representation of the ionic polarizability is the simple **shell model**. The ion is represented by two charged components, one the nucleus and the tightly bound inner electrons, called the **core**, and the other as a massless **shell** representing the outer valence electrons. The core and shell interact via a harmonic energy:

$$\phi_{\text{shell}} = \frac{1}{2}kd^2 \tag{5.25}$$

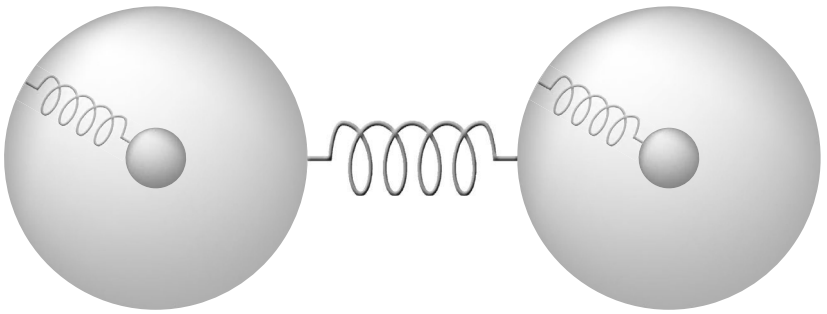


Fig. 5.5 Simple representation of the shell model, where the springs represent short-range interactions.

where d is the separation of the centres of the core and shell. The parameter k is related to the polarizability.

The forces between atoms are represented by the forces between cores and shells. The charges on the core and shell enter the Coulomb energy separately. Usually it is assumed that short-range interactions only involve the shells and not the cores, particularly because the repulsive interactions concern the outer electrons. The full shell model is represented in Fig. 5.5.

There are more sophisticated versions of the shell model, allowing, for example, deformations of the shell or variations in its radius.

5.4.6 Hydrogen bonds

In some ways hydrogen atoms are similar to other atoms, but there are cases where hydrogen atoms play a unique role in crystal binding. This is because hydrogen atoms, particularly when electron-deficient, are smaller than other ions, and therefore can take part in bonds that are shorter and stronger than other bonds. The special bonds involving hydrogen are called **hydrogen bonds**, but there are several ways in which hydrogen bonds are manifest.

The simplest type of hydrogen bond is where the hydrogen cation lies between two anions. Because of the small size of the hydrogen cation, the two anions can get quite close, and for the same reason there is only space for two neighbouring anions. In this case the hydrogen acts as the glue between two anions.

In other types of hydrogen bonds, the hydrogen atom is closer to one anion than another. Often the closest anion will be oxygen, and one can identify the pair of oxygen and hydrogen as the hydroxyl, OH^-, molecular ion, but molecular groups such as N–H will also take part in hydrogen bonds. The characteristic feature of hydrogen bonds is some degree of directionality of the bonds. For example, the OH^- will point towards the third bonded atom. This point is illustrated in the case of common ice, whose structure is shown in Fig. 5.6. In this case, each water molecule is hydrogen-bonded to four others in a tetrahedral arrangement (similar to the arrangement of tetrahedra in some of the phases of silica). There is orientational disorder of the water molecules, with only one hydrogen atom lying between neighbouring oxygen atoms.

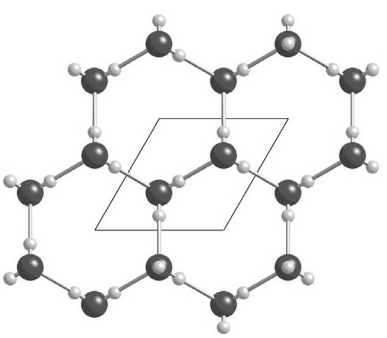


Fig. 5.6 Crystal structure of common ice showing disordered positions of hydrogen atoms (small spheres) in the hexagonal network of oxygen atoms (large spheres).

Another example of hydrogen bonding with disorder is the ferroelectric KH_2PO_4 (we met other types of ferroelectric materials in Chapter 2). In this case there is one hydrogen atom between each pair of oxygen atoms on neighbouring PO_4 tetrahedra, as shown in Fig. 5.7. The hydrogen atom jumps between

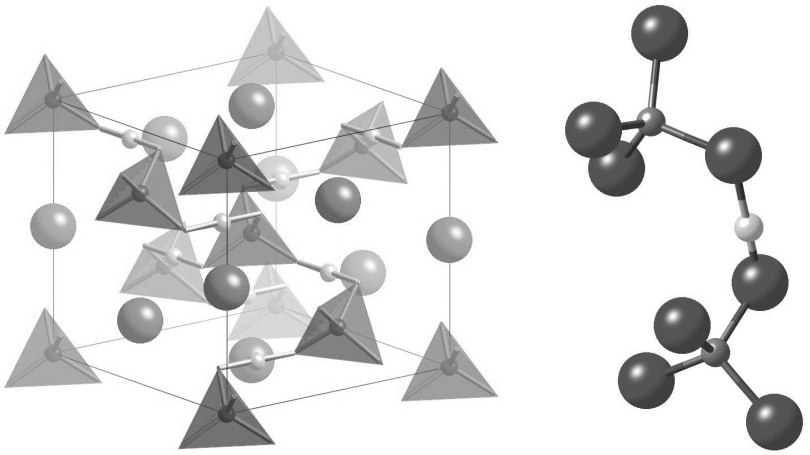


Fig. 5.7 Crystal structure of KH_2PO_4 with disordered hydrogen bond. In the left-hand plot, the PO_4 tetrahedra are shown as shaded units, the large spheres represent the K cations, and the small spheres represent the hydrogen atoms. In the right-hand plot, which focuses on a single hydrogen bond, the large atoms are the oxygen atoms, the P atoms are those bonded to four oxygen atoms, and the hydrogen atom lies between two oxygen atoms closer to one atom than the other.

two potential energy minima along the O. . .O bond. At low temperatures, the positions of the hydrogen atoms order, and this ordering process gives rise to a ferroelectric phase transition. Phase transitions of this type are common, particularly in ferroelectric materials.

5.4.7 Empirical representations of covalent and metallic bonding

Covalent and metallic bonding have been conspicuously absent from the discussion so far. This is because the bonding is not easily rationalized in terms of simple pair functions. It is possible to represent interactions between atoms in metallic and covalent systems in terms of simple empirical functions, but these are not as easily justified as the Coulomb, Born–Mayer, and dispersive functions discussed above in the context of ionic crystals.

Metallic bonding

Traditionally much of solid state physics has been concerned with describing the distribution of electron energies in metals, and the response of this distribution to applied fields. For metals, unlike insulators, the distribution of electron energies is the primary factor in determining properties such as electrical and thermal conductivity, as well as magnetism. Clearly the distribution of electrons will also account for the cohesion of metals.

Metallic bonding arises from the delocalization of some of the electrons in the crystal. This gives rise to the simple picture of a uniform gas of electrons and a periodic array of positively charged atom cores (nuclei and bound electrons). The Coulomb energy can be computed for the interactions between the atom cores and between the electrons and atom cores, using methods similar to those used in evaluating the Coulomb energy of ionic crystals. As for ionic solids, the result will be an energy that varies as $-1/R$, where R is the separation of nearest-neighbour atom cores.

The Coulomb interaction between the atom cores and electrons needs to be balanced by a force that prevents the atom cores from becoming too close. In a

uniform electron gas, the energy of each electron is related to its wave vector by

$$E = \frac{\hbar^2 k^2}{2m} \tag{5.26}$$

The Pauli exclusion principle prevents two electrons from having the same quantum state. This means that only two electrons can have the same wave vector **k**, each with opposite spin. If there are N unit cells in the crystal, there will be N wave vectors in the reciprocal unit cell or Brillouin zone. Because the energy of each electron increases as k^2, the electrons occupy the set of smallest wave vectors possible, and occupy all the wave vectors within a sphere about the origin of reciprocal space until all electrons are accounted for. The surface of the sphere is called the **Fermi surface**, and the surface will have radius k_F, where the subscript F represents the Fermi surface. The energy of the electrons with wave vectors on the Fermi surface will be equal to $\epsilon_F = \hbar^2 k_F^2/2m$. The physics of the Fermi surface unlocks an understanding of the physics of metals, and is covered in some detail in most standard textbooks on solid state physics.

In a crystal of volume V (assumed to be cube-shaped), the distribution of electron wave vectors will form a grid in reciprocal space. This is illustrated in Fig. 5.8. Each grid point has an associated volume of $(2\pi)^3/V$. If the crystal contains n electrons, they will occupy a total volume of $n(2\pi)^3/2V$ in reciprocal space. Equating this with the volume of the Fermi sphere, $(4\pi/3)k_F^3$, we have

$$k_F^3 = \frac{3\pi^2 n}{V} \tag{5.27}$$

The energy of the n electrons is given by the integral over the volume of the Fermi sphere. This yields (Problem 5.9) the mean energy per electron of

$$\frac{E}{n} = \frac{3\hbar^2 k_F^2}{10m} = \frac{3}{5}\epsilon_F \tag{5.28}$$

Since $k_F \propto V^{-1/3}$, and $V^{1/3} \propto R$, we have $E/n \propto R^{-2}$. This energy is positive, and balances the negative Coulomb energy. Thus the binding energy of the metal is given as

$$U = U_{\text{Coulomb}} + U_{\text{electrons}} = -\frac{a}{R} + \frac{b}{R^2} \tag{5.29}$$

where a and b are constants. Clearly this function can be minimized to give an equilibrium value of the atomic separation R.

This description of metallic binding gives a rough idea of the basic principles, but is a gross oversimplication. It neglects interactions between the electrons, and it neglects any inhomogeneity of the electron gas. Atom cores are not points, and therefore the electron gas will only exist in the space between the atoms and not everywhere in the metal. However, the basic idea summarized by eqn 5.29 can accommodate a more accurate picture.

Of course, what this approach has not given is a picture of interactions between specific pairs of atoms. If two atom cores are displaced with respect to each other, there will be a local change in the electron distribution which will give a change in energy that is related to the size of the displacements. It is possible to write this change of energy in terms of interactions between atoms with some degree of justification.

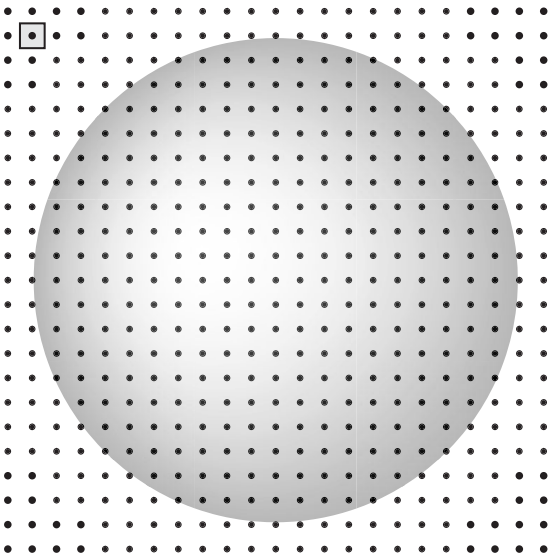


Fig. 5.8 Two-dimensional representation of the distribution of allowed electron wave vectors in reciprocal space. The square shows the volume $(2\pi)^3/V$ associated with each grid point. The sphere represents the Fermi surface, and encompasses all grid points that are occupied by electrons in the ground state. For clarity the number of grid points shown within the Fermi surface is many orders of magnitude smaller than in real materials.

Covalent bonding

Covalent bonds, like metallic binding, involve delocalized distributions of electrons. Unlike metals, the delocalization occurs across networks of bonds rather than throughout space. Another important difference between covalent and metallic systems lies in the energy distribution of the electrons. In a metal, there are energy states lying just above the Fermi energy, and electrons can easily be excited into theses states (and in this lies much of the physics of metals). However, in covalent systems the energy band lying above the highest-energy filled state is at a significantly higher energy, and it is much harder for electrons to be excited into this band. This factor is the origin of many of these difference between metals and covalent systems.

The simplest examples of covalent structures are the elemental crystals of silicon and carbon (Section 2.2.4). The structures of silicon and carbon in its diamond phase give three-dimensional covalent networks. In the graphite form of carbon, this network is two-dimensional. Molecular crystals, such as I_2, N_2, and C_{60} (Sections 2.3.3 and 2.7.3), also have covalent bonds within the molecules, but without forming a network of covalent bonds. Some crystals containing different types of atoms can also be thought of in terms of networks of covalent bonds (Section 2.3.3). The primary examples are the phases of silica, SiO_2, which form infinite networks of corner-linked SiO_4 tetrahedra (Sections 2.3.3 and 2.7.3). However, we very quickly hit an ambiguity, because the distinction between covalent and ionic crystals is not clear-cut. In silica, the silicon and oxygen ions appear to be charged (calculations typically suggest that there is just over one electron unit of charge on the oxygen atoms, and a charge of just above $+2e$ on the silicon) but with strong bonds between the Si and O atoms that involve electron delocalization. In fact, this is the most common situation, even in simple crystals such as MgO with the NaCl structure.

The electron distribution in a covalent crystal, whether close to that of an ionic crystal or fully shared by all atoms, is determined by the details of the atoms

and their coordination in the structure. Moreover, the electron distribution can be changed by changing pressure, or as a result of a phase transition through changing temperature. It is possible for a material to have metallic bonding in one phase and be an insulator in another.

It is often convenient to represent covalent bonds by an empirical function, such, as the **Morse potential**, which has the form

$$\phi(r) = \epsilon[\exp(2\alpha(r - r_0)) - 2\exp(\alpha(r - r_0))] \tag{5.30}$$

where r_0 is the equilibrium separation of the atoms, ϵ is the energy of the bond at $r = r_0$, and α gives the curvature of the potential energy at $r = r_0$ and determines the frequency of the bond-stretching vibration. In cases where the covalent bonding enforces a specific shape of the distribution of atoms, such as tetrahedral bonding of carbon and silicon atoms in their elemental crystal structures, or groups such as SiO_4 in silicates, the forces that determine the shape can be represented by bond-bending terms such as

$$\phi(\theta) = \frac{1}{2}K(\theta - \theta_0)^2 \tag{5.31}$$

where θ is the bond angle, and θ_0 is the equilibrium angle with values such as 90° or 109.46° for octahedral and tetrahedral coordination respectively.

5.5 Quantum mechanical view of chemical bonding

5.5.1 The need to take a proper quantum-mechanical view

Chemical bonding is essentially quantum mechanical in origin. We have acknowledged this throughout our discussion of models of bonding above. The repulsive interaction arises from the overlap of electronic wave functions, and in ionic systems quantum mechanical calculations of systems with closed shells give rise to interactions that vary with separation which are closely represented by the Born–Mayer potential function. In ionic systems, overlap of the electron distributions of close atoms does not significantly change the electron distributions, but it does change the energies because of the interaction between the electrons in the overlap region. Metallic and covalent bonding are clearly quantum mechanical, and it is very difficult to do much beyond using empirical, and only partially justified, potential energy functions without needing to tackle the quantum mechanics head on. Even when we do so, the problem is so difficult that it is often necessary to resort to the use of empirical terms, but by doing so via the quantum mechanical treatment it means that the empirical component has a proper degree of physical justification.

In this section we will introduce some of the basic ideas in the application of quantum mechanics to the study of crystalline materials. The main problem is that the number of electrons in a crystal will be of the order of Avogadro's number, but the basic Schrödinger equation can only be solved exactly for a single electron. As a result, it is tempting to try to represent the many-electron wave function in terms of single-electron wave functions. Initially it is also tempting to neglect the Coulomb interactions between electrons, which turns out not to

be as bad an approximation as one might have suspected – in fact, this approach is inherent in much of the traditional development of the quantum mechanics of solids, and the problems can often be avoided until discussions of magnetic ordering. Indeed, it is when the direct Coulomb interaction between electrons is included that the theory becomes much harder to solve, and there is a need to resort to approximate methods.

In fact the outlook for solving the quantum mechanics of crystalline materials is rather more optimistic than the previous paragraph may indicate. The power of modern computers, combined with the development of efficient algorithms, has suddenly transformed the problem from one of great difficulty into one that can be solved using standard desktop computing tools. Our aim in this section cannot extend much beyond that of giving a very simple introduction, but most of the basic ideas now in common use will be touched upon here.

5.5.2 Born–Oppenheimer approximation

In everything that will be discussed here, it is implicit that the atomic nuclei are treated classically, and the wave functions only contain the electrons as dynamic variables. Thus the Schrödinger equation is solved only for the electrons, and the atomic nuclei are assumed to be fixed in position. This is the heart of the Born–Oppenheimer approximation. The justification is that there is a huge difference in masses between the nuclei and electrons. As a result, if there is any movement of the nuclei with respect to each other, the distribution of the electrons can adjust so quickly as to be almost instantaneous. The one case where the Born–Oppenheimer approximation may be unjustified is in the study of hydrogen bonds, particularly those where there are two positions for the hydrogen atom. In these cases it is believed that the hydrogen atoms can tunnel between energy minima, and a complete description of the material may need to take account of this quantum tunnelling. Other light atoms, such as lithium, can also undergo tunnelling dynamics, although usually only at low temperatures and in structures where there is more than one potential energy minimum available for the atom. Often this is the case when the lithium atom goes into a structure as an impurity. But in general calculations of materials, there is no need to treat the motions of nuclei within the framework of quantum mechanics.

5.5.3 Bloch's theorem for electrons in a periodic structure

The periodicity of the crystal structure should also give rise to a periodicity of the electron wave functions in a crystal, since the potential energy in the Schrödinger equation will be periodic. Specifically the wave function for a single electron can be written as

$$\psi(\mathbf{r}) = u(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \tag{5.32}$$

where $u(\mathbf{r})$ is a periodic function such that $u(\mathbf{r}) = u(\mathbf{r} + \mathbf{R})$ if $\mathbf{R}$ is a lattice vector. This gives the periodicity condition

$$\psi(\mathbf{r} + \mathbf{R}) = \exp(i\mathbf{k} \cdot \mathbf{r})\psi(\mathbf{r}) \tag{5.33}$$

This condition, known as **Bloch's theorem** will be implicit in the following treatment of the quantum mechanics of electrons in crystalline materials, even if the point is not made explicit. The important point is that the wave functions of the crystal can be expressed in terms of their wave vector **k**, exactly as was seen for the free electron gas in the simple model of metallic binding.

5.5.4 Simple view of bonding in molecules

The simplest basic approach in understanding bonding from a quantum mechanical approach can be illustrated by some ideas in chemical molecular bonding. The starting point is the electronic wave functions of the individual atoms. When the atoms come close to each other, the wave functions overlap, and there are interactions between the two atoms. The process is illustrated by the bonding of the two hydrogen atoms. Suppose that the electronic wave functions of the two atoms are given by ϕ_1 and ϕ_2. The starting point is to write the Schrödinger equation for one electron as

$$\hat{\mathcal{H}}\psi = \left(-\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\epsilon_0|\mathbf{r}-\mathbf{R}_1|} - \frac{e^2}{4\pi\epsilon_0|\mathbf{r}-\mathbf{R}_2|}\right)\psi = E\psi \qquad (5.34)$$

where $\mathbf{r}$ is the position of the electron, and $\mathbf{R}_1$ and $\mathbf{R}_2$ are the positions of the two nuclei. The central approximation in this equation is that the potential energy of the electrons arises only by the interactions of the electrons with both nuclei, but there is complete neglect of interactions between the two electrons. We flagged this approximation above, and will discuss it again below. However, it turns out that although it seems to be a very drastic approximation, considerable progress can be made in spite of it.

The solution to the two-atom Hamiltonian for a single electron can be written as a linear combination of the wave functions of the two atoms (commonly called the **linear combination of atomic orbitals, LCAO**):

$$\psi_\mathrm{s} = \frac{1}{\sqrt{2}}(\phi_1 + \phi_2) \qquad (5.35)$$

$$\psi_\mathrm{a} = \frac{1}{\sqrt{2}}(\phi_1 - \phi_2) \qquad (5.36)$$

where ψ_s is the **symmetric** combination, and ψ_a is the **antisymmetric** combination. These two states are represented in Fig. 5.9. It turns out that the two solutions for the energy, namely E_s and E_a, have the order

$$E_\mathrm{s} < E_\mathrm{a} \qquad (5.37)$$

The symmetric state has the lowest energy, which is easily rationalized by consideration of Fig. 5.9. The antisymmetric state localizes each electron on one or other of the atoms, with a value of $\psi_\mathrm{a} = 0$ at the point half-way between the atoms. There is therefore a minimum of electron density along the bond between the two atoms. One the other hand, the symmetric wave function gives an electron density between the two atoms, and the electrons in the bonding region have lower energy because of their interaction between the two nuclei.

We can summarize this discussion by noting that as the two atoms are brought together, the electronic wave functions of the two atoms overlap and combine

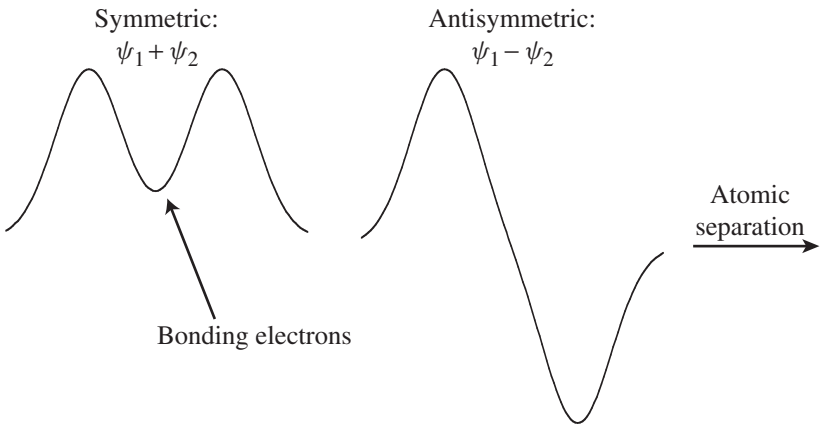


Fig. 5.9 Representation of the symmetric and antisymmetric wave functions of a single electron in the hydrogen molecule.

in either the symmetric or antisymmetric state. The energies of the two states then differ from the energies of the isolated atoms, with the symmetric state having the lower energy.

5.5.5 Tight-binding methods

We now extend this discussion to include binding of the atoms in a solid. This approach is known as the **tight binding method**. We illustrate the basic idea by considering a one-dimensional chain of identical atoms. Again, we start from the atomic orbitals, labelling the wave function of the electron on atom n as ψ_n. The Hamiltonian for a single electron is written as

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2m}\nabla^2 - \frac{ze^2}{4\pi\epsilon_0}\sum_j \frac{1}{|\mathbf{r} - \mathbf{R}_j|} \tag{5.38}$$

where ze is the charge on each nucleus. There is again a complete neglect of interactions between electrons.

The wave function is now written as a linear combination of the individual atomic orbitals:

$$\psi = \sum_n c_n \phi_n \tag{5.39}$$

where the coefficients c_n are normalized such that

$$\sum_n |c_n|^2 = 1 \tag{5.40}$$

We assume an arrangement of N atoms in a chain whose ends are connected to form a loop, such that atom $n = N$ is equivalent to atom $n = 0$. This topological arrangement means that we do not need to take account of the ends of the chain. This topological arrangement is known as a **periodic boundary condition**, and we will use this device in our study of lattice dynamics in Chapter 8. With the periodic boundary conditions, the values of c_n can be written as

$$c_n = \exp(ikna)/N^{1/2} \tag{5.41}$$

where k is any wave vector, and the factor $N^{-1/2}$ ensures normalization. Any linear combination of atomic orbitals with these coefficients is consistent with Bloch's theorem for electronic wave functions in a periodic potential, which states that, for a three-dimensional wave function

$$\psi_k(\mathbf{r}) = u_k(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \tag{5.42}$$

where $u_k(\mathbf{r})$ is a function with the periodicity of the crystal.

We now calculate the energy, which follows from the identity

$$\int \psi^* \hat{\mathcal{H}} \psi \, \mathrm{d}\mathbf{r} = E \int \psi^* \psi \, \mathrm{d}\mathbf{r} = E \tag{5.43}$$

Substituting for the wave functions, we have

$$E = \frac{1}{N} \int \sum_n \exp(-ikna) \phi_n^* \hat{\mathcal{H}} \sum_{n'} \exp(ikn'a) \phi_{n'} \, \mathrm{d}r \tag{5.44}$$

We now separate different terms involving integrals of different pairs of atoms. The first term is for $n = n'$:

$$\alpha = \int \phi_n^* \hat{\mathcal{H}} \phi_n \, \mathrm{d}r \tag{5.45}$$

Since the atomic orbitals have the same form for each atom, α is independent of n. We also define the terms for neighbours, $m = n - n'$:

$$\gamma_m = \int \phi_n^* \hat{\mathcal{H}} \phi_{n \pm m} \, \mathrm{d}r \tag{5.46}$$

With these definitions, the energy can be written as

$$E = \alpha + \sum_m (\exp(ikma) + \exp(-ikma)) \gamma_m \tag{5.47}$$

$$= \alpha + 2 \sum_m \gamma_m \cos(kma) \tag{5.48}$$

The energy of the wave function clearly depends on the wave vector. A representation is given in Fig. 5.10. The dependence of energy on wave vector is called the **band structure**. The spread of electron energies in the band structure shown in Fig. 5.10 will decrease if the atoms are pulled further apart, as the potential energy term in the Hamiltonian associated with electrons seeing distant nuclei becomes smaller. The ideas here are easily generalized for three-dimensional crystals.

The important generalization of these ideas is to go beyond the implicit assumption that the electrons are s electrons to include also the p and d electrons for many important applications. The γ coefficients of eqn 5.46, which are called **hopping integrals** or **transfer integrals**, can be defined for s, p, and d electrons, mixing across different types of electrons. Of course, this gives a large number of coefficients, whose values may have to be determined empirically. The problem gets more complex if we have more than one type of atom. As a result of the generalization of the model, the band structure graph of Fig. 5.10 will become more complex. However, the basic idea can be appreciated in general terms.

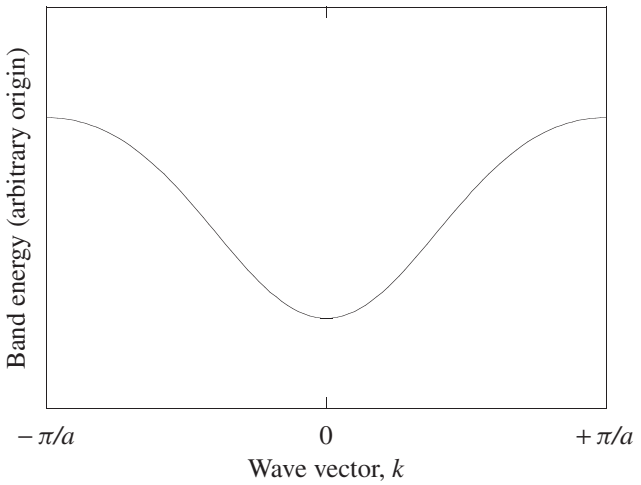


Fig. 5.10 Representation of the band structure of a one-dimensional chain of atoms as given by eqn 5.48.

5.5.6 Electron–electron interactions: Hartree–Fock and beyond

The tight-binding approach is an example of an **independent electron model**, since interactions between electrons are neglected. To include interactions between electrons adds a huge level of complexity to any calculations. Moreover, the success of the tight-binding approach suggests that we do not need to include interactions between electrons to gain a reasonable level of understanding of the quantum mechanics of bonding in solids. A more complete theory is needed when accurate calculations are required.

The problem with understanding electron interactions can be appreciated by writing down the Hamiltonian for all the electrons in a three-dimensional solid:

$$\hat{\mathcal{H}} = \sum_i \left(-\frac{\hbar^2}{2m}\nabla_i^2 - \frac{e^2}{4\pi\epsilon_0}\sum_l \frac{Z_l}{|\mathbf{r}_i - \mathbf{R}_l|} \right) + \frac{e^2}{4\pi\epsilon_0}\sum_{\langle i,j\rangle} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (5.49)$$

where i and j label the electrons, and l labels the nuclei. $Z_l e$ is the charge of the nucleus l at position $\mathbf{R}_l$. The positions of the electrons are denoted by $\mathbf{r}_i$. Two difficulties arise from the third term. First, this term involves pairs of electrons, which makes it impossible to obtain an analytical solution for all electrons. The Schrödinger equation, in common with any other quantum mechanical or classical equation of motion, cannot be solved for many particles at once. Only if the third term was neglected, as in the simple tight-binding approach above, would it be possible to treat the electrons separately. The second difficulty is that the wave function for the system involves all electrons in a single function, and it would be much easier if it were possible to find single-particle solutions to the Schrödinger equation. However, the third term in the Hamiltonian comes back to give yet more problems in this regard. Herein lie the technical problems encountered when looking for a solution of the Schrödinger equation for many electrons.

Historically, the first approach was to form an approximation for the interaction between different electrons and then solve for single-particle wave

functions. Specifically, one could aim to write this term as

$$\frac{e^2}{4\pi\epsilon_0}\sum_{\langle i,j\rangle}\frac{1}{|\mathbf{r}_i-\mathbf{R}_j|}=\frac{e}{4\pi\epsilon_0}\int\rho(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}' \tag{5.50}$$

where $\rho(\mathbf{r}')$ is the charge density, and can be represented as

$$\rho(\mathbf{r}')=e\sum_j|\psi_j(\mathbf{r}')|^2 \tag{5.51}$$

The sum is over all occupied one-electron states. Substituting into the Hamiltonian gives a set of one-electron equations:

$$\begin{aligned}&-\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r})-\frac{e^2}{4\pi\epsilon_0}\sum_j\frac{Z_j}{|\mathbf{r}_i-\mathbf{R}_j|}\psi_i(\mathbf{r})\\&+\left(\frac{e^2}{4\pi\epsilon_0}\sum_j\int\frac{|\psi_j(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}\right)\psi_i(\mathbf{r})=\epsilon_i\psi_i(\mathbf{r})\end{aligned} \tag{5.52}$$

The equations for all the electron states $\psi_i(\mathbf{r})$ are known as the **Hartree equations**. Each equation still contains all other single-particle states, but it is possible to solve the set of equations by an iterative scheme. A trial set of functions $\psi_i(\mathbf{r})$ is proposed, and these are used to compute the electron density $\sum_j|\psi_j(\mathbf{r}')|^2$. This is held fixed in the equations, which are then solved for a new set of $\psi_i(\mathbf{r})$. The new set is used to compute the new values of $\sum_j|\psi_j(\mathbf{r}')|^2$, and the whole procedure repeated until convergence. This is an example of a **self-consistent procedure**.

Although the Hartree scheme appears reasonable, it has a number of failings. The most obvious one is that the equation for a given $\psi_i(\mathbf{r})$ contains the contribution from this wave function in the electron density sum $\sum_j|\psi_j(\mathbf{r}')|^2$. Thus the equations contain terms in which each electron interacts with its own contribution to the electron density. Another problem is that the Hartree equations are consistent with an overall wave function Ψ that is a simple product of the single-particle functions:

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)=\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)\cdots\psi_N(\mathbf{r}_N) \tag{5.53}$$

This function is incompatible with one of the central principles of quantum mechanics, in which the sign of Ψ should change when any two electrons are exchanged:

$$\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_i,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_N)=-\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_i,\ldots,\mathbf{r}_N) \tag{5.54}$$

One representation of the overall wave function that is consistent with the antisymmetry principle is the **Slater determinant**:

$$\begin{aligned}\Psi&=\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)\cdots\psi_N(\mathbf{r}_N)-\psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)\cdots\psi_N(\mathbf{r}_N)+\cdots\\&=\begin{vmatrix}\psi_1(\mathbf{r}_1)&\psi_1(\mathbf{r}_2)&\cdots&\psi_1(\mathbf{r}_N)\\\psi_2(\mathbf{r}_1)&\psi_2(\mathbf{r}_2)&\cdots&\psi_2(\mathbf{r}_N)\\\vdots&\vdots&&\vdots\\\psi_N(\mathbf{r}_1)&\psi_N(\mathbf{r}_2)&\cdots&\psi_N(\mathbf{r}_N)\end{vmatrix}\end{aligned} \tag{5.55}$$

It can be shown that the solution of the Schrödinger equation for the Slater determinant can be written as

$$\begin{aligned}
&\int \Psi^* \hat{\mathcal{H}} \Psi \, \mathrm{d}\mathbf{r}_1 \cdots \mathrm{d}\mathbf{r}_N \\
&\quad = \sum_i \int \psi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0} \sum_j \frac{Z_j}{|\mathbf{r} - \mathbf{R}_j|} \right) \psi_i(\mathbf{r}) \, \mathrm{d}\mathbf{r} \\
&\qquad + \frac{e^2}{4\pi\epsilon_0} \sum_{\langle i,j \rangle} \int \frac{1}{|\mathbf{r} - \mathbf{r}'|} |\psi_i(\mathbf{r})|^2 |\psi_j(\mathbf{r}')|^2 \, \mathrm{d}\mathbf{r} \mathrm{d}\mathbf{r}' \\
&\qquad - \frac{e^2}{4\pi\epsilon_0} \sum_{\langle i,j \rangle} \int \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}') \psi_j^*(\mathbf{r}') \psi_j(\mathbf{r}) \, \mathrm{d}\mathbf{r} \mathrm{d}\mathbf{r}'
\end{aligned} \tag{5.56}$$

The last term is the new term over and above the Hartree equations. When the energy is unpacked in terms of the single-electron functions, a new set of equations emerges:

$$\begin{aligned}
&-\frac{\hbar^2}{2m} \nabla^2 \psi_i(\mathbf{r}) - \frac{e^2}{4\pi\epsilon_0} \sum_j \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} \psi_i(\mathbf{r}) \\
&\quad + \frac{e^2}{4\pi\epsilon_0} \left(\sum_j \int |\psi_j(\mathbf{r}')|^2 \frac{1}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}' \right) \psi_i(\mathbf{r}) \\
&\quad - \frac{e^2}{4\pi\epsilon_0} \sum_j \int \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi_j^*(\mathbf{r}') \psi_i(\mathbf{r}') \psi_j(\mathbf{r}) \, \mathrm{d}\mathbf{r}' = \epsilon_i \psi_i(\mathbf{r})
\end{aligned} \tag{5.57}$$

The last term acts as an integral operator, and is called the **exchange interaction**. It looks as if it will make the equations much harder to solve, and so it does!

Although there are clear improvements in principle over the Hartree model, this model, known as the **Hartree–Fock model**, is still known to be incomplete. Calculations on the uniform electron gas in the high-density limit have given expressions for terms beyond the Hartree–Fock energy. These are called, for want of a better word, the **correlation energy**, although electron correlations are present in the basic Hartree–Fock approach. It is possible to correct Hartree–Fock calculations for these effects using a variety of different schemes.

5.5.7 Representation of electronic wave functions

Ideally the electronic wave functions should be determined directly as the eigenvectors of the Schrödinger equation. However, with the problems we have already identified with solving the Schrödinger equation for many electrons, this ideal is unrealistic for practical solutions. Instead, when representing the many-electron wave function in terms of single-particle functions, it is also useful to make an assumption about the form of these functions and adjust the form to give the best solutions of the equations. A single-particle function can be represented by the form

$$\psi = \sum_j c_j \phi_j(p_{j1}, p_{j2}, \ldots) \tag{5.58}$$

where c_j is an adjustable coefficient on the j-th contribution to the single-electron function, and p_{j1}, p_{j2}, ... may be coefficients within the functions ϕ_j whose values can also be adjusted. There are two types of functions in common use. One is the plane wave representation. Noting that the wave functions are periodic functions of the crystal, it is possible to form an expansion in terms of the reciprocal lattice vectors **G**, and write

$$\psi(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{G}} \exp(i\mathbf{G} \cdot \mathbf{r}) \tag{5.59}$$

The prefactors $c_{\mathbf{G}}$ can be treated as adjustable variables. The number of terms in the expansion can also be adjusted, although it would not be wise to truncate the series before it has properly converged.

The other representation of the wave functions is in terms of localized functions, typically centred on the atom positions. These functions could be simple radial functions, or the known eigenfunctions of the atomic Schrödinger equation. In these cases, not only can the occupation be adjusted, but also the spatial extent of the functions.

In principle all electrons should be treated within any calculation, but in practice to do so is quite demanding. Calculations can be made easier by treating the electrons that lie close to the core of the atoms separately from the outer electrons. In a further approximation, the inner core electrons can be removed from the calculation by replacing them with a potential energy function that closely approximates the effect of the inner electrons on the outer electrons. This potential energy function, called the **pseudopotential**, can be deduced with reasonable accuracy.

In order to determine the best values of the expansion coefficients in eqn 5.58, use is made of the **variational theorem**. The essential point is that the best wave function is that for which the energy is a minimum. Clearly the exact electron wave functions define the ground state of the system. If the ground-state wave function is written as Ψ_0, and the ground-state energy as E_0, the Schrödinger equation is written as

$$\hat{\mathcal{H}}\Psi_0 = E_0\Psi_0 \tag{5.60}$$

When we consider a small change in Ψ_0 to $\Psi = \Psi_0 + \delta\Psi$, such as would arise from making an approximation to the wave function, we find that the energy can be written as

$$E = E_0 + O(\delta\Psi)^2 \tag{5.61}$$

This result is derived in Appendix G. The important point from this equation is that a first-order change in the wave function will produce a second-order change in the energy. The main point of the variational theorem is that it makes it possible to propose a trial form for the wave function with any number of adjustable parameters. The values of these parameters can be tuned to give the lowest energy, and this process will result in the form of the wave function that best resembles (within the constraints of the form of the trial function) the true wave function. The variational theory implies that the energy can be obtained with higher accuracy than that of the wave function.

5.5.8 Practical calculations of binding energies from quantum mechanics

The Hartree–Fock equations, with or without corrections for the correlation energies, can be solved using an extension of the methods discussed in this section. More recently, an alternative formalism called **density functional theory** (DFT) has become rather more popular. The idea here is to recast the basic equations in terms of the electron density rather than the wave function directly, although, of course, the electron density is still obtained from the underlying wave function. Although it is possible to write the potential energy terms, i.e. the interactions between the electron distribution and the atomic nuclei, and the Coulomb interactions between the electrons, as functionals of the electron density, the exchange and correlation energies cannot be readily expressed in this way. Thus it will be necessary to introduce some approximations, which are discussed below. The advantage of the DFT approach is that the equations can be written down in terms of single-electron wave functions relatively easily.

The simplest solution to the problem posed by the exchange and correlation energies is the **local density approximation** (LDA). It is assumed that it should be possible to evaluate the exchange-correlation energy at any point using results obtained separately for a uniform electron density. It is possible to go beyond this by including terms that take account of spatial variation of the electron density through the gradient of the density.

DFT methods can be used with plane-wave and localized basis sets for the electrons, and the parameters adjusted in accordance with the variational theorem. DFT methods also work with and without the use of pseudopotentials to approximate the core electrons. With such schemes it is now possible to calculate the ground-state energy of relatively complex crystals, and relax the structure to that of the lowest energy using algorithms similar to those used in minimization of the lattice energy constructed from empirical functions.

Summary of chapter

- The **lattice energy** is the potential energy of the crystal structure with all atoms at rest. It can be calculated from simple models of the individual bond energies.
- The **Coulomb energy** is not easy to sum owing to the fact that it only converges when the sum over the crystal is taken to infinite distance. There are techniques for performing the sum of the Coulomb energy over lattices. For some of the simpler cubic structures, the sum can be represented by the **Madelung energy**, which is inversely proportional to the nearest-neighbour distance, and is proportional to a constant that is defined separately for each structure type.
- The energy associated with repulsions between neighbouring atoms at short separations can be represented by an exponential function, the **Born–Mayer** potential, or by a function of the form $1/r^n$, where $n = 12$.
- The **dispersive interaction**, which depends on interatomic separation as $-1/r^6$, arises from correlated fluctuations of the electron densities of the two atoms which give rise to short-lived electric dipole moments.

The dispersive interaction can be combined with a $1/r^{12}$ term to give the **Lennard–Jones** function, or the Born-Mayer term to give the **Buckingham** potential.

- Atomic polarizability can be represented by the **shell model**, in which the atom is treated as consisting of an inner core that carries all the mass and an outer shell. The core and shell interact through a harmonic force.
- Metallic and covalent bonds can be represented by empirical functions, but the true nature of these bonds can only be properly represented using quantum mechanical models.
- The **tight-binding model**, which accounts for the overlap of atomic wave functions, provides a simple intuitive description of the quantum mechanics of atomic bonding. The resultant wave function can be written in terms of plane waves characterized by the wave vectors contained within the Brillouin zone.
- A complete understanding of the electronic structure of materials needs to account for the exchange and correlation energies. One approach that is commonly used to describe the electronic structure is **density functional theory**.

Further reading

Many aspects of bonding are discussed in textbooks on solid state physics and chemistry over a wide range of detail. Both Kittel (1996) and Ashcroft and Mermin (1976) give a reasonably thorough introduction. Other aspects of the topic, including the role of quantum mechanics and electronic structure, are given by Alcock (1990), Burdett (1995), Borg and Dienes (1992), Dove (1993), Elliott (1998), Hazen and Finger (1982), Ladd (1994), Nichols (1995), Pettifor (1995), Rao and Gopalakrishnan (1986) and Sutton (1993). It is worth consulting a number of texts, because it is instructive to see how the topic of bonding in solids can be approached from a number of different directions.

Exercises

(5.1) Crystals of neon have a ccp structure. The potential energy of the crystal has been measured as $-1.88\,\mathrm{kJ\,mol^{-1}}$, and the lattice parameter a is 4.466 Å. Assuming that we only need to consider nearest-neighbour interactions, and that the bonding can be described by the Lennard-Jones potential (eqn 5.23), use these data to obtain numeric values for the parameters ϵ and σ.

(5.2) The complementary error function is defined as

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-u^2)\,\mathrm{d}u \qquad (5.62)$$

Show that the second term in the expansion of $1/r$ is given by

$$\frac{2}{\sqrt{\pi}} \int_g^\infty \exp(-r^2\rho^2)\,\mathrm{d}\rho = \frac{1}{r}\mathrm{erfc}(gr) \qquad (5.63)$$

Use a spreadsheet or other programming tool to show that this function falls to zero much faster than $1/r$. A good numerical approximation is

$$\mathrm{erfc}(x) = \exp(-x^2) \sum_{n=1}^{5} a_n t^n \qquad (5.64)$$

where $t = (1 + bx)^{-1}$, and $a_1 = 0.25483$, $a_2 = -0.28450$, $a_3 = 1.42141$, $a_4 = 1.45315$, $a_5 = 1.06141$ and $b = 0.32759$.

(5.3) Derive eqn 5.22. (*Hint*: use the relationship $\partial/\partial V = (\partial V/\partial R)^{-1}\partial/\partial R$.)

(5.4) Show that a ccp crystal with nearest-neighbour Lennard–Jones interactions (eqn 5.23) has bulk modulus $B = 6\varepsilon/\sigma^3$.

(5.5) Given the lattice energies and lattice parameters for the alkali halides with the NaCl structure listed in the table below, calculate the Madelung energy, assuming each ion has formal charge (i.e. the charges given by the valence, so that the cation charge is 1 unit of positive electron charge, and the anion charge is -1 unit of electron charge), and hence the Born–Mayer energy.

		F	Cl	Br	I	
Li	a:	4.028	5.140	5.502	6.000	Å
	E:	1014.3	832.6	794.5	743.9	kJ mol^{-1}
	B:	67.1	29.8	23.8	17.1	GPa
Na	a:	4.634	5.640	5.978	6.474	Å
	E:	897.5	764.4	726.7	683.2	kJ mol^{-1}
	B:	46.5	24.0	19.9	15.1	GPa
K	a:	5.348	6.294	6.596	7.066	Å
	E:	794.5	694.0	663.5	627.5	kJ mol^{-1}
	B:	30.5	17.4	14.8	11.7	GPa
Rb	a:	5.630	6.582	6.890	7.342	Å
	E:	759.3	666.8	638.8	606.6	kJ mol^{-1}
	B:	26.2	15.6	13.0	10.6	GPa

Given the condition for equilibrium, calculate the value of ρ for each case, and hence the coefficients B in the Born–Mayer interaction. For each crystal, calculate the value of the bulk modulus and compare with the experimental value. (*Hint*: a case for using a spreadsheet program.)

(5.6) Rather than treat the ions as having formal charges, it is possible to allow their values to vary to give agreement across a range of experimental data. Manipulate eqns 5.19–5.22 to obtain values for ρ, $U_{\text{repulsive}}$, and U_{Madelung} from the experimental data for lattice energy, lattice parameter, and bulk modulus for the alkali halides given in the previous question, and use these to obtain values of B in the Born–Mayer repulsive interaction and the ionic charges.

(5.7) Use a spreadsheet with a minimiser function to obtain the value of lattice parameter that gives the minimum of the lattice parameters for NaCl in both the standard and CsCl structures. Compare the Na. . .Cl distances in both phases in the light of the idea of ions having constant radii. Assuming that the differences in energy and volume are independent of pressure, use eqn 5.8 to calculate the pressure at which NaCl transforms between the two structures at zero temperature. Add the term $+PV$ to the lattice energy to give the lattice enthalpy, and minimize the lattice enthalpy for the two phases of NaCl over a range of pressures. Use these calculations to deduce the transition pressure, and compare this result with your earlier estimate.

(5.8) Use the same spreadsheet to calculate the equilibrium energy and volume of NaCl in the cubic ZnS structure, and compare the results with those obtained for the standard and CsCl structures.

(5.9) Derive eqn 5.28 by integrating the energy versus wave vector relationship, eqn 5.26, for a spherical distribution of wave vectors up to the Fermi wave vector k_F.

(5.10) Al_2SiO_5 exists in three polymorphs, with thermodynamic data (per mole of formula units):

	Volume m^3 mol^{-1}	Energy kJ mol^{-1}	Entropy J K^{-1} mol^{-1}
Andalusite	4.29×10^{-5}	0	912.3
Sillimanite	5.02×10^{-5}	4.593	922.4
Kyanite	4.79×10^{-5}	9.593	927.2

Calculate the phase diagram for the three polymorphs of Al_2SiO_5. (You met the crystal structures of andalusite and sillimanite in Problems 3.7 and 2.8 respectively.)

Diffraction

6

6.1 Basics of diffraction

6.1.1 Use of radiation beams

Having explored many of the principles of crystal structures, we now need to consider how crystal structures can be determined. Inevitably this information is obtained by the use of beams of radiation. The strict requirement is that the wavelength of the radiation will need to be smaller than typical interatomic distances if there is to be adequate resolution in the reconstructed image of the atoms in the crystal. There are several candidate beams, such as electromagnetic radiation in the X-ray spectrum, beams of electrons, or beams of neutrons. Here we meet wave–particle duality with a vengeance. The beams are treated as waves according to the standard relationships between momentum, $\mathbf{p}$, wave vector, $\mathbf{k}$, velocity, v, and, wavelength, λ:

$$\mathbf{p} = \hbar\mathbf{k}; \quad k = |\mathbf{k}| = 2\pi/\lambda; \quad |\mathbf{p}| = mv = h/\lambda \tag{6.1}$$

On the other hand, although the beams are scattered from the sample as a result of the way that the arrangement of atoms impacts on the phases of the scattered waves, the electrons and neutrons (and also the photons) are detected as individual particles.

In addition to the constraints on the wavelengths of the beams, there is also the requirement that the beam of radiation should not be too strongly absorbed or cause significant damage to the crystal. This rules out the use of beams of neutral atoms or protons. All beams of radiation will be absorbed to some extent, or will cause damage, but the effects of these processes can be minimized with X-rays, electrons, or neutrons.

We need to establish some terms. We will assume that all atoms of the same type will scatter a beam of radiation in the same way. This is called coherent scattering. This point is particularly important in neutron diffraction, where different isotopes may scatter quite differently, or where the scattering depends on the relative orientations of the magnetic moment of the neutron and the magnetic moment of the atomic nucleus. In this chapter we will *appear* to only consider scattering processes in which the wavelength of the radiation, and hence the energy, is unchanged through the scattering process (we will extend this restriction at one particular point in this chapter). Such a process is called elastic scattering (and has nothing to do with the term in classical mechanics, where it is taken to mean processes in which the total kinetic energy is conserved). The possibility for the energy of the scattered beam to change

leads to **spectroscopy**, which is discussed in Chapter 10. Finally, **diffraction** is the process of coherent elastic scattering from the **long-range order** in a structure, which, for a crystal, means the periodic structure given by the average positions of the atoms.

6.1.2 Bragg's law

The structure of a material can be deduced when beams of radiation are scattered elastically with a coherent superposition of the waves. The simplest representation of this process is encompassed in **Bragg's law of diffraction**, as illustrated in Fig. 6.1. The waves are reflected from planes of atoms separated by distance d (defined in Chapter 4), and the coherent superposition of the waves occurs when the lengths of the paths taken by the components reflected from neighbouring planes differ by an integral number of wavelengths. This is shown in the construction in Fig. 6.1. It is straightforward to show from this construction that diffraction from planes with a given index hkl (Chapter 4) follows for the case when the path difference is one wavelength as

$$\sin\theta_{hkl} = \lambda/2d_{hkl} \tag{6.2}$$

This is **Bragg's law**. The important point established by Bragg's law is the relationship between the scattering angle θ, which in diffraction is called the **Bragg angle**, and the d-spacing, and this is used throughout the analysis in this chapter.

6.1.3 Single-crystal and powder diffraction measurements

Diffraction, as implied by the Bragg law, occurs as reflections of beams of radiation from distinct crystal planes, and will only occur when the normal to the plane is aligned at a specific angle to the incident beam consistent with the wavelength of the radiation. It is also a requirement that the measuring device, whether film or detector, is aligned with the geometry of the scattered beam.

The best way to determine the crystal structure is to perform a diffraction experiment using a near-perfect single-crystal sample. However, there are sometimes advantages in performing diffraction measurements from polycrystalline

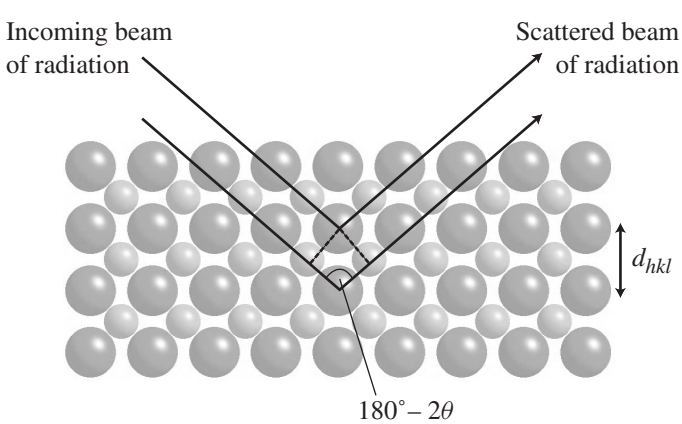


Fig. 6.1 Geometric construction of Bragg reflection.

samples. These measurements can be a lot quicker, and there are cases when single crystals are not available. In a polycrystalline sample there will be a uniform distribution of the orientations of the small crystallites. This means that the condition of the orientation of the crystal planes with respect to the incident radiation beam will be automatically fulfilled for some of the crystallites, so the experiment needs only to vary the scattering angle and not be concerned about the orientation of the sample. In some polycrystalline methods that exploit the possibility of focusing the beam in order to increase the intensity of the diffracted beam there may be a coupled motion of the sample with the detector, but this will only involve a single angle and the coupling only depends on the Bragg angle and has nothing to do with the crystal planes.

6.1.4 Diffraction and crystal structures

Bragg's law is interesting in that it gives a geometric picture of how beams are coherently scattering from lattice planes with strict conditions on wavelength and angle. However, it provides no information about how different atoms and different structures have an affect on the scattering process other than through their effect on the lattice spacing. The purpose of this chapter is to develop a quantitative theory of the diffraction of radiation. We take this in two stages. The first stage is to determine the distribution of scattered radiation assuming a known crystal structure. Once we have established how radiation is scattered by a crystal, we then turn to the inverse problem of deducing the crystal structure from measurements of scattering radiation. There is an important problem called the **phase problem**, that will become apparent in developing the formalism. This means that the inverse problem is not trivial – in fact the solution of the inverse problem has proven to be a considerable intellectual challenge that has been acknowledged through the award of several Nobel prizes in physics and chemistry.

6.2 Beams of radiation and measurement of diffraction patterns

6.2.1 Laboratory X-ray methods

X-rays are the part of the electromagnetic spectrum with wavelengths between 0.1 and 100 Å, and hence energies in the range 10^2–10^5 eV. To appreciate this energy scale it should be noted that the kinetic energy of an atom in a gas at room temperature is 0.025 eV. X-rays lie between the ultraviolet and gamma ray portions of the electromagnetic spectrum.

In the laboratory, X-rays are produced when a beam of electrons strikes a metal target. The X-rays are actually produced by two distinct mechanisms. First, the electrons are rapidly decelerated when they enter the metal, and an accelerating or decelerating charge will automatically produce electromagnetic radiation. This mechanism, called **Bremsstrahlung**, produces a broad distribution of X-ray energies commensurate with the energies of the incident electrons. The spectrum of X-rays produced from a laboratory X-ray source is represented in Fig. 6.2. The Bremsstrahlung radiation is the continuous distribution, which has a minimum wavelength that corresponds to the energy of the incident

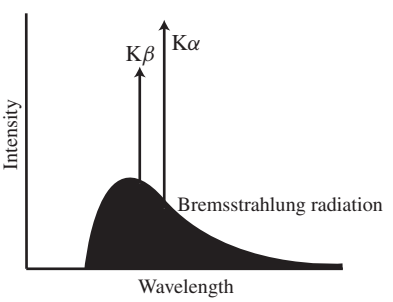


Fig. 6.2 Schematic plot of characteristic X-ray spectrum.

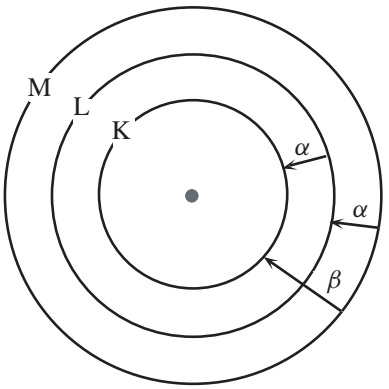


Fig. 6.3 Schematic plot of energy levels of an atom showing transitions that give characteristic X-ray energies.

electrons being lost immediately. The second mechanism of X-ray production is shown by the peaks in Fig. 6.2. On hitting the atoms in the target, some electrons knock out one of the inner electrons. X-rays are given off when an electron in one of the outer shells drops down into the empty electron level and loses its energy as an X-ray photon of characteristic energy. The set of characteristic X-ray emissions is shown schematically in Fig. 6.3. The important case for diffraction experiments is that of electrons falling from the L to the K shells, i.e. from the second to the first shell, and this transition is called Kα. There are in fact two transitions of this type, called Kα_1 and Kα_2 with very similar energies (reflecting the fact that there is a splitting of energy levels within each shell, due to the dependence of energy on angular momentum.) The intensity of the Kα_1 contribution is twice as large as the Kα_2 contribution. Some characteristic X-ray wavelengths are given in Table 6.1.

According to Bragg's law, diffraction will occur for specific combinations of diffraction angle and wavelength. In an experiment, measurement of the angle can be performed with much greater precision than a direct measurement of the X-ray wavelength (unless a subsequent diffraction process is then used in the measurement of the wavelength). Therefore experiments are performed using a fixed wavelength, treating the scattering angle as the variable quantity. The raw spectrum of X-rays is not suitable for this type of experiment because it contains a continuous distribution of wavelengths, including several strong peaks. The common practice is to use absorbing filters to remove most of the X-rays apart from the Kα peak. The filters are not perfect, but if they can remove all the other characteristic peaks (which they can), the small amount of Bremsstrahlung radiation left in the beam will be weak compared to the Kα component and will not cause any problems. Alternatively, a specially mounted crystal in the form of a flat plate, called a **monochromator**, can be used to diffract only the Kα peak. By using a slightly curved crystal as the monochromator it becomes possible to filter out the Kα_2 component, leaving only the Kα_1 wavelength.

6.2.2 Measurement of the intensity of scattered X-ray beams

The traditional method of recording and measuring X-ray diffraction patterns is to use photographic film that has been optimized for X-rays. This is ideal for obtaining a diffraction pattern from a single crystal across a wide sweep of reciprocal space. The greyness of each reflection can be measured by optical methods and converted into a relative intensity. Several diffraction photographs

Table 6.1 X-ray wavelengths for selected material.

	Wavelength Å			
Element	Kα_1	Kα_2	Mean Kα	Kβ_1
Cr	2.28962	2.29351	2.29092	2.08480
Fe	1.93597	1.93991	1.93728	1.75653
Co	1.78892	1.79278	1.79021	1.62075
Cu	1.54051	1.54433	1.54178	1.39217
Mo	0.70926	0.71354	0.71069	0.63225
Ag	0.55936	0.56378	0.56083	0.49701

may be required in order to record the range from very strong to very weak, since the strong reflections can easily blacken (i.e. saturate) the film. The distribution of measurements of individual reflections on the X-ray film from a single-crystal experiment depends on the type of camera being used. In the simplest cameras the distribution of X-ray reflections has to be interpreted by use of geometric relationships, but there is a complex type of camera called the **precession camera** that involves simultaneous movements of the crystal and film in a precession type motion in order to give an indistorted impression of a single layer of reciprocal space on the film. An example of a diffraction measurement using a precession camera is shown in Fig. 6.4.

In a simple powder diffraction experiment with film, a strip of film can be wrapped around the sample. This method, called the **Debye–Scherrer** method, gives a quick measurement of the powder diffraction pattern, but it is not particularly accurate with respect to the diffraction angle because there is no focussing or collimation. The Debye–Scherrer camera and an example of a powder diffraction photograph is shown in Fig. 6.5. It is possible to improve the accuracy of measurements of powder diffraction by film methods using focusing cameras.

The modern way of measuring X-ray intensities is to use an electronic detector. This has the advantage over films of giving the intensity directly and more accurately. In a single-crystal experiment the X-ray source is in a fixed position, the crystal will be at a fixed position and with a variable orientation, and the detector will be attached to mechanical arms that allow it to be moved across a spherical surface. In a powder diffraction experiment the detector will move in a plane around a semicircle. The disadvantage of using a single detector is that it can only record a single point in reciprocal space at one time. This problem can be overcome using area detectors, which contain a large number of pixels each capable of recording its own signal. An example of a powder diffraction measurement using a focusing arrangement is shown in Fig. 6.6.

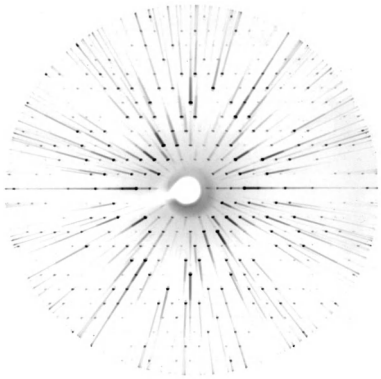


Fig. 6.4 Example of a single-crystal X-ray diffraction photograph recorded on an instrument called the **precession camera**, in which correlated motions of the crystal, photographic film and masks enable the final photograph to give an undistorted record of a layer of reciprocal space. This particular photograph was obtained from a crystal of beryl. (Photograph courtesy of Tony Abraham, University of Cambridge.)

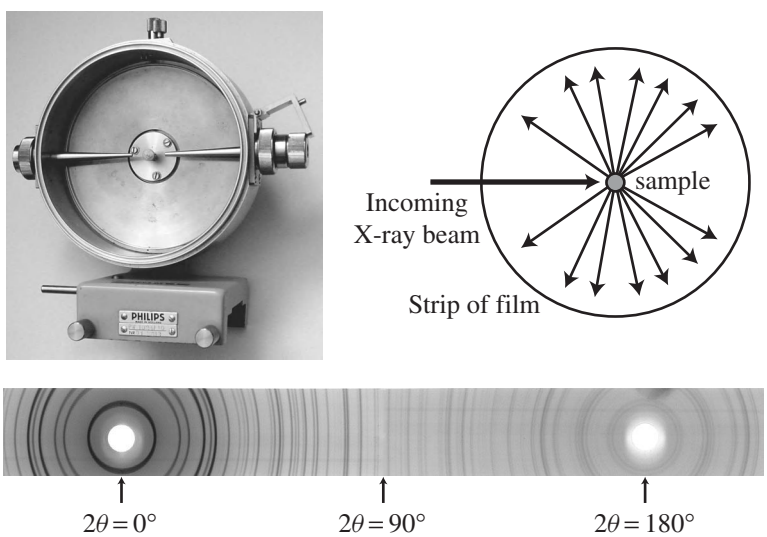


Fig. 6.5 Debye–Scherrer camera for the simple recording of an X-ray powder diffraction photograph, together with an example recorded for a powder sample of beryl, the same material used in the single-crystal diffraction image shown in Fig. 6.4. (Diffraction photograph courtesy of Tony Abraham, University of Cambridge.)

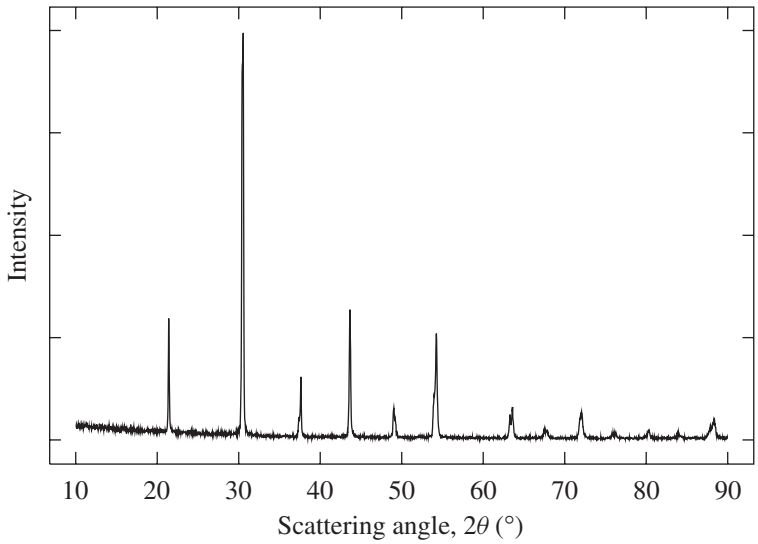


Fig. 6.6 X-ray powder diffraction measurement of a tetragonal ferroelectric perovskite, showing sharp Bragg peaks that are split because of the lowering of symmetry from cubic. (Data kindly provided by Matthew Tucker, University of Cambridge.)

6.2.3 Synchrotron X-ray sources

The fundamental limitation on the use of laboratory X-ray tubes is that the maximum intensity is limited by the need to prevent overheating of the target. X-ray production is not very efficient, and most of the energy of the incident electron beam is lost as heat. This can partly be overcome using rotating targets, so that the electron beam does not strike a single area of the target. However, there is still a limitation on the maximum intensity. Intensity can be important for studies of time-dependent processes by diffraction. A second limitation is that the usable wavelengths given by an X-ray tube are limited to those of the energy level transitions of the target material. There are cases when a much finer control on the wavelength is useful, particularly when (as will be discussed in Section 6.8.1) it is required to have a wavelength that is close to an absorbtion wavelength of the material under study.

These problems can be overcome using **synchrotron sources**. A synchrotron is a ring of magnets of some tens of metres radius which accelerate a beam of electrons in a circular orbit. An electron travelling in a circular orbit is always accelerating in a radial direction and decelerating in the direction of the tangent. Thus the circular beam is continuously losing energy as photons: this process is called **synchrotron radiation**. The production of a usable photon beam can be enhanced by the use of special devices inserted into the electron ring that give a special accelerating kick to the electron beam and hence a large enhancement to the intensity of the emitted beam of X-rays.

The synchrotron radiation is emitted as a continuous distribution of X-ray wavelengths. The specific wavelength to be used in a diffraction experiment can be selected from this distribution by Bragg reflection from a monochromator crystal. It is also possible to perform diffraction measurements at a fixed scattering angle using a detector that is able to discriminate between different energies, but the precision of this type of measurement is much lower than that possible when using a fixed wavelength and using the scattering angle as a variable.

Fig. 6.7 The European Synchrotron Radiation Facility (ESRF, seen as the large ring structure in the bottom left corner of the photograph) and the neutron scattering facility of the Institute Laue-Langevin (ILL, cluster of buildings to the right of the synchrotron ring – the large round building is the actual nuclear reactor). The two facilities are nestled together at the meeting of two rivers in Grenoble. (Photograph by courtesy of the ILL.)

Any synchrotron source will be run as a national or international science facility. One example is the European Synchrotron Radiation Facility in Grenoble. This is seen as the large ring structure in Fig. 6.7. As a large-scale facility it has a broad suite of instruments. Some diffractometers will be optimized for intensity, others for resolution. Some will be optimized for materials with large d-spacings. Other instruments will be designed for specific sample environments. For example, diffraction experiments under conditions of high pressure will need the X-ray beam to have a small wavelength in order to reduce absorption of the beam by the pressure-transmitting mechanism and to maximize the range of d-spacings that can be measured with the limited range of θ allowed by the high-pressure equipment. The advantage of having a facility operating with optimized instruments is that it is straightforward to run good experiments. The disadvantages are that demand for these instruments will inevitably exceed the available resources, and experiments that are approved by whatever selection procedures are in place will need to be scheduled subject to the constraints imposed by scheduling many other users. Thus it is never possible to reproduce the flexibility that comes with the use of laboratory equipment.

6.2.4 Neutron beams

The significant alternative to X-rays for diffraction measurements to obtain information about crystal structures is the use of beams of neutrons. Neutrons are subatomic particles with no charge and a mass that is very close to that of the proton (1.675×10^{-27} kg). The neutron also has a magnetic moment which arises from its internal structure as a small group of quarks. The neutron is not a stable particle, but its lifetime (around 890 s) is much longer than the time scale of a measurement.

Beams of neutrons can be produced by two mechanisms. The first is to generate beams of neutrons within a nuclear reactor. The design of a nuclear reactor has fuel elements containing fissile material such as uranium within a

medium of lighter materials (such as graphite or water) called the **moderator**. The neutrons are produced in fission reactions that are catalysed by other neutrons, and slowed down by collisions with atoms in the moderator. In a reactor designed to generate power, the neutrons are maintained within the core of the reactor in order to sustain a continued series of fission reactions. However, in a neutron source it is necessary to reflect the beams of neutrons down tubes that lead to the instruments outside the main reactor vessel.

As in synchrotron radiation sources, the beams of neutrons are produced with a broad distribution of neutron energies that is characteristic of the temperature of the moderator. For many applications the ideal energies of neutrons are those close to room temperature, and these energies can be obtained by controlling the temperature of the moderator. The wavelength of a neutron of energy E is obtained by extending eqn 6.1:

$$E^2 = \frac{p^2}{2m} \quad \Rightarrow \quad \lambda = \frac{(2\pi\hbar)^2}{\sqrt{2mE}} \tag{6.3}$$

The neutron energy is typically expressed in units of meV ($1\,\text{eV} = 96.484\,\text{kJ mol}^{-1}$), so that energy can be converted to wavelength through the numerical relationship $\lambda = 9.05/E^{1/2}$ Å. For example, a neutron with energy corresponding to room temperature, 25 meV, will have a wavelength of 1.8 Å. Although for many applications neutron beams with thermal energies are ideal, there are experiments for which higher or lower-energy beams are required.

Usually a single wavelength is selected from the broad distribution by Bragg reflection from a crystal monochromator. In this situation the diffraction equipment using the neutron beam is similar in principle to the diffraction equipment used with laboratory X-ray beams, but on a somewhat larger scale. A neutron powder diffractometer is shown in Fig. 6.8. This is one of the suite of instruments at the reactor-based neutron facility of the Institute Laue–Langevin

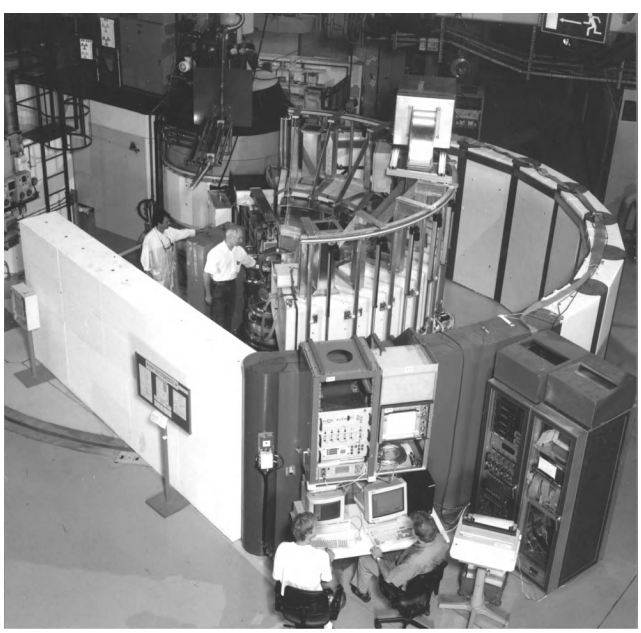


Fig. 6.8 Powder diffractometer at the Institute Laue–Langevin neutron source. The striking feature is the wide bank of detectors in the horizontal plane. (Photograph by courtesy of the Institute Laue–Langevin.)

at Grenoble. The Institute is positioned next to the European Synchrotron Radiation Facility, and can be seen in Fig. 6.7.

Reactor neutron sources have been extremely successful for many years. The first experiments were performed in the 1950s, and from the 1960s onwards a number of reactor neutron sources have become operational around the world. The availability of a broad spectrum of neutron wavelengths affords a great deal of flexibility similar to that obtained with synchrotron X-ray sources. However, two problems have emerged with regards to the development of improved reactor neutron sources. The first is the potential for political problems in gaining approval for the construction of new reactor sources and in obtaining licenses for the use of enriched fissile fuel material. The second problem is that it becomes increasingly difficult to design nuclear cores that give an increased flux of neutron beams. The problems arise from the need to remove heat from the core, and the attendent safety issues that need to be resolved. Although it is possible to design improved sources, it appears that the costs escalate to unacceptable levels. For these reasons it is likely that there will be a shift towards the use of the second method of producing beams of neutrons.

This second source of neutron beams is called a **spallation source**. In this case a beam of high-energy protons is fired into a heavy-metal target (such as tantalum or uranium, but future sources are likely to use mercury). The protons strike the nuclei and knock neutrons off. These neutrons have energies that are too high to be useful for many applications, but can easily be slowed down by passing through a moderating material (such as cold methane) to remove the kinetic energy. Although it is possible to build spallation sources that produce a continuous beam of neutrons, so that the ways the beam can be used are very similar to the ways that reactor beams are used, it is more common to use a pulsed beam of protons to produce a pulsed beam of neutrons. This pulsed beam will have a spectrum of wavelengths that is determined by the type of moderator being used. The UK ISIS spallation neutron facility is shown in Fig. 6.9.

Fig. 6.9 Inside the experimental hall at the ISIS spallation neutron source. The proton beam is generated in a synchrotron stage in a building adjacent to the wall at the top of the photograph. The proton beam extracted from the synchrotron ring travels towards the target stage, which is housed in the cylindrical assembly seen in the upper centre of the picture. Many different kinds of instruments are clustered around the target station, most of which have their detectors positioned below floor level. (Photograph by courtesy of the ISIS Facility.)

It is possible to use rather different approaches to using the neutrons that will exploit the pulse nature. With modern electronics and detectors it is possible to measure the time it takes for neutrons in a single pulse to travel from the source through the sample and into the final detector with very high accuracy. The time it takes for a neutron to travel from the source to the detector is equal to the distance divided by its velocity. Given that the distance is also known to high accuracy, it is possible to determine the velocity to a similarly high accuracy.

We use the design of a powder diffractometer as an example of the use of the time-of-flight approach. A pulse of neutrons containing a spectrum of wavelengths will be produced in one cycle of the proton beam. This pulse of neutrons will travel towards the diffractometer, spreading out in space due to the spread of velocities. As each neutron is scattered from the sample and reaches the detector its time is recorded. The detector is fixed in position, and for each pulse it will record the intensity as a function of flight time. The total distance travelled by each neutron, which is equal to the sum of the distance between the moderator and sample and the distance between the sample and detector, is represented by L, and the flight time is represented by t. The detector is held at a fixed Bragg angle of θ. The neutron velocity is equal to $v = L/t$. From the de Broglie equation, the relationship to the wavelength of the neutron beam is

$$\lambda = \frac{h}{mv} = \frac{ht}{mL} \tag{6.4}$$

From Bragg's law (eqn 6.2),

$$\begin{aligned} \lambda = 2d \sin\theta &= ht/mL \\ \Rightarrow d &= ht/2mL \sin\theta \end{aligned} \tag{6.5}$$

As a result, a measurement of the flight time can easily be converted to a measurement of the corresponding d-spacing of the planes that the neutrons are being diffracted from. The limitations in the accuracy of the measurement are determined by the accuracy of the measurements of flight time, distance, and angle. The effects of the accuracy of the latter can be reduced by using large values of the scattering angle, as shown in Problem 6.1. The main limitations on the accuracy of the distance is the thickness of the moderator, and the effects of this can be minimized by the use of long flight paths. It is shown in Problem 6.1 that this requires the use of flight paths of 10–100 m, which are quite feasible, and the loss of neutron beams can be minimized using special tubes called **wave guides** which exploit the ability of neutrons to reflect from the walls at shallow angles. A modern powder diffractometer at a spallation neutron source is shown in Fig. 6.10. An example of a powder diffraction pattern obtained using time-of-flight neutron methods is shown in Fig. 6.11.

The use of neutron beams suffers from the same limitations as synchrotron beams with regard to the availability being limited to central facilities. Unlike the case of X-ray experiments, there is no laboratory alternative for neutron diffraction.

6.2.5 Comparison of the characteristics of X-ray and neutron beams

The formalism that will be developed in this chapter will apply equally to X-ray or neutron diffraction, and many experiments involving the two types of

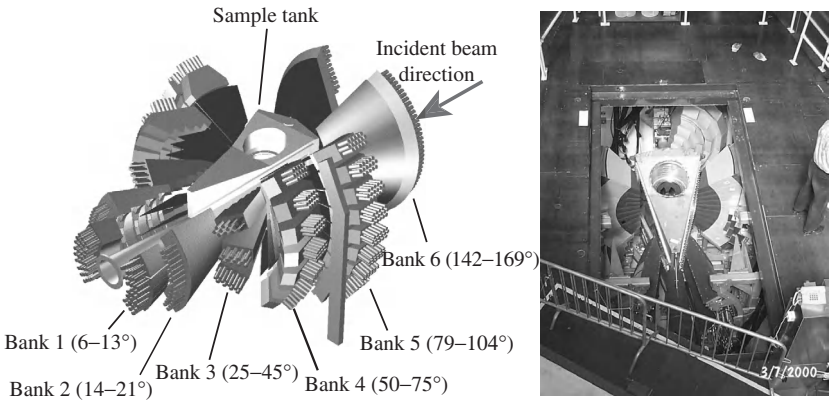


Fig. 6.10 State-of-the art neutron powder diffractometer at the ISIS spallation neutron source. (Photograph and diagram by courtesy of the ISIS Facility.)

radiation are very similar in concept. As noted above, laboratory X-ray sources have many operational advantages over synchrotron X-rays and neutrons. However, X-ray and neutron diffraction have complementary strengths and weaknesses, depending on the actual system under study, and there may often be compelling reasons to choose one type of radiation over the other, or to obtain data from both types of radiation in a single study.

The primary difference between the techniques of X-ray and neutron diffraction arise from the fact that the scattering mechanisms are different. X-rays are scattered by the electrons and will therefore provide information about the electron density in the scattering material. Neutrons are scattered by two mechanisms. First, neutrons are scattered by the atomic nuclei, and will therefore provide information about the positions of the nuclei in the scattering material. To obtain information about the shape of the electron distribution about an atomic nucleus, which is of interest in the chemistry of bonding, combined X-ray and neutron diffraction studies can be of considerable value. Second, in magnetic materials, the neutrons can be scattering by the atomic magnetic moments through the interaction with the magnetic dipole moment of the neutron. The magnetic moments of solids, except at very low temperatures, arise from the alignment of the spins of the electrons, and so there can be a component of scattering of neutrons directly from the electrons. Unlike X-ray diffraction, the magnetic scattering of neutrons does not involve all the electrons, but only the electrons in the partially filled energy levels in which the electrons are aligned to give the magnetic moment.

Since X-rays are scattered from the electrons, it will follow (and we will show formally in Section 6.4.2) that the scattering power of each atom is proportional to the number of electrons. This means that there is a consistent variation of the scattering power with atomic number, so that heavy atoms will scatter X-rays more strongly than will light atoms. Also, the scattering of X-rays will be different from neutral atoms and their corresponding charged ions (e.g. from neutral Si or from the Si^{4+} cation). On the other hand, there is no corresponding correlation in the power of neutron scattering from atomic nuclei and the atomic number. In fact there is a wide distribution in the scattering

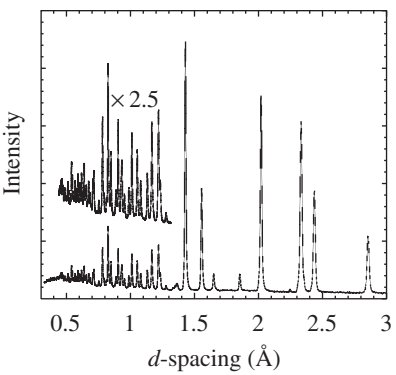


Fig. 6.11 Neutron powder diffraction data for a sample of spinel, obtained on a medium-resolution time-of-flight diffractometer at the ISIS spallation neutron source.

power, and the sign of the scattering amplitude can be either positive or negative since the neutron–nucleus scattering may or may not give a change of phase of π to the neutron wave (Section 6.4.3).

There are three important practical implications to this comparison of the scattering power of X-rays and neutrons. *First*, neutrons will be better at distinguishing between atoms that have similar numbers of electrons, such as Si^{4+} and Al^{3+} cations. On the other hand, there will also be other pairs of atoms that have very similar scattering power for neutrons (e.g. Cl and Br) but very different atomic number, so these will be better distinguished by X-ray methods. *Second*, in structures where there is an atom that dominates most of the X-ray scattering (these are typically atoms with a lot of electrons), the positions of the other atoms (those with few electrons, such as oxygen) may be difficult to deduce. In neutron scattering the scattering powers of most atoms are broadly in the same range, so there is rarely a problem with one atom dominating the scattering. However, there are some atoms (vanadium is an example) where the scattering power of neutrons is almost zero. *Third*, the hydrogen atom is virtually invisible to X-rays but will scatter neutrons with the same order of strength as any other atom. For crystallographic problems where the positions of hydrogen atoms are important, for example when there are C–H and O–H bonds, neutron scattering is the main way of providing information about the bond lengths.

We will show later (Section 6.4.2) that the X-ray scattering power of an atom falls off with increasing scattering angle to an extent that is determined by the inverse of the size of the atom, whereas with nucleus–neutron scattering the scattering power is independent of the scattering angle (because the nucleus is tiny compared to the wavelength of the neutron beam). For studies where we need scattering at high angles, for example to obtain accurate bond lengths in disordered systems, neutron scattering may be the preferred method.

The scattering of X-rays by electrons is a stronger process than the scattering of neutrons by atomic nuclei. This is coupled with the fact that X-ray beams are much more intense than neutron beams. These two factors have implications for the design of experiments. It is possible to use much smaller samples for X-ray diffraction than in neutron diffraction. For example, X-ray experiments use samples that are likely to be much smaller than 1 mm^3, whereas the samples for neutron diffraction are likely to be several mm^3 (or even a few cm^3) in size. The other side of this issue is that neutrons can penetrate into materials much further than X-rays. For example, there may be the haunting possibility that X-ray diffraction is determined by the surface rather than the bulk. This may be important when studying phase transitions (strictly a bulk issue; see Chapter 12), which can be significantly affected by surfaces. In the design of experiments, the greater penetrating power of neutrons gives rather more flexibility in the use of complex equipment for control of the sample environment. For example, in studies of diffraction from materials at high pressures, neutrons will penetrate through the pressure-generating anvils more easily than laboratory X-rays (the solution to this problem for X-ray diffraction is to use high-energy X-ray beams generated by a synchrotron source). The need to reduce the thickness of the environment equipment will lead to the reduction in the size of this equipment, which may mean that large gradients in temperature or pressure are hard to avoid. All these difficulties can be solved by careful design of experiments, but it

does suggest that sometimes the larger scale of neutron diffraction experiments can be exploited in certain types of difficult diffraction experiments.

6.2.6 Beams of electrons

The third common radiation for diffraction experiments is a beam of electrons. Electrons with sufficiently high energy can pass through thin samples of matter, and have wavelengths that are much smaller than interatomic distances. For example, a high-power electron beam of energy 1 MeV will have a wavelength of 0.0035 Å.

Electron diffraction methods are very different from X-ray or neutron diffraction, both in formalism and in techniques. They are also used for quite different applications. Because of the short wavelength, only a single setting of the crystal orientation is needed to obtain diffraction across a wide area of reciprocal space – this is discussed in Appendix H. For this reason, electron diffraction is commonly used for surveys of the scattering in reciprocal space. However, because the link between the intensity of the diffracted beam and the crystal structure is not as straightforward as for X-ray or neutron diffraction, it is much less often used for the determination of crystal structures. For this reason we will focus much more on these other types of radiation.

The main virtue of electron diffraction is that the diffracted beams can be combined within the instrument in order to generate an image with a high resolution. It is possible to produce a resolution that is of the order of the size of an atom. This allows for imaging of the long-range structure of a material. Often there is more interest in imaging short-range features, such as imperfections, microstructures, or interfaces between different material. Examples are given in Fig. 1.12 and Appendix A.

6.3 Basics of the theory of diffraction

6.3.1 The wave equation

A beam of radiation can be described by three quantities: its wavelength λ, the direction in which the beam is travelling, and its amplitude A. The first two are conveniently expressed by the wave vector $\mathbf{k}$, which is a vector pointing along the direction of the beam with magnitude $|\mathbf{k}| = 2\pi/\lambda$. In three directions, the wave can be described in complex exponential form as

$$\psi = A \exp(i\mathbf{k} \cdot \mathbf{x}) \tag{6.6}$$

where $\mathbf{x}$ is a position in space. We will find it convenient to use the complex exponential form rather than the simpler sine or cosine form. The quantity $\mathbf{k} \cdot \mathbf{x}$ has the units of an angle (given in radians) and is called the phase angle. In an experiment the quantity that is usually measured, whether by a radiation detector or on a photographic film, is the intensity of the beam, which is proportional to the square of the amplitude.

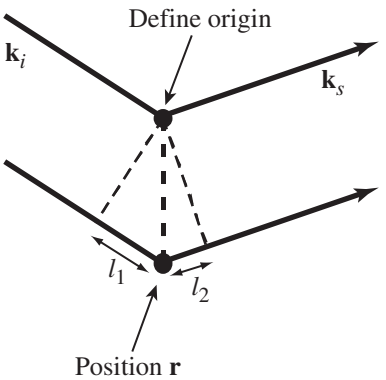


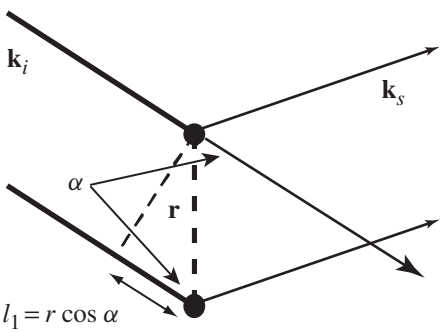
Fig. 6.12 Scattering of a beam of radiation from two particles separated by the vector $\mathbf{r}$. The wave vectors of the incoming and scattered beams are $\mathbf{k}_i$ and $\mathbf{k}_s$ respectively.

6.3.2 Scattering of radiation from two particles

Consider the situation illustrated in Fig. 6.12, which shows the scattering of a beam of radiation with initial wave vector $\mathbf{k}_i$ by two particles separated by the vector $\mathbf{r}$. We consider the wave that is scattered with wave vector $\mathbf{k}_s$. Note that the direction of the wave is changed, but not its wave vector, i.e. $|\mathbf{k}_i| = |\mathbf{k}_s|$.

The two rays come in with same phases, but in the scattering process they have different path lengths which lead to the two scattering rays having different phases.

Consider the change of phase given by the path length difference $l_1 = r\cos\alpha$.



$\cos\alpha$ can be obtained from the vector dot product of $\mathbf{k}_i$ and $\mathbf{r}$:

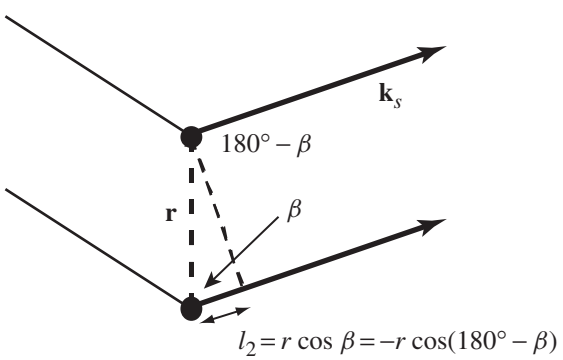
$$\mathbf{k}_i \cdot \mathbf{r} = |\mathbf{k}_i| r\cos\alpha = \frac{2\pi}{\lambda} r\cos\alpha \tag{6.7}$$

Thus we obtain

$$l_1 = r\cos\alpha = \frac{\lambda}{2\pi}\mathbf{k}_i \cdot \mathbf{r} \tag{6.8}$$

This path length gives a phase difference of $2\pi l_1/\lambda = \mathbf{k}_i \cdot \mathbf{r}$.

The second contribution to the change of phase is given by the path length difference $l_2 = r\cos\beta$.



$\cos\beta$ can be obtained from the vector dot product of $\mathbf{k}_s$ and $\mathbf{r}$:

$$\mathbf{k}_s \cdot \mathbf{r} = |\mathbf{k}_s| r \cos(180^\circ - \beta) = -\frac{2\pi}{\lambda} r \cos\beta \tag{6.9}$$

Thus we obtain

$$l_2 = r\cos\beta = -\frac{\lambda}{2\pi}\mathbf{k}_s \cdot \mathbf{r} \tag{6.10}$$

This path length gives a phase difference of $2\pi l_2/\lambda = -\mathbf{k}_s \cdot \mathbf{r}$.

The total phase difference is the sum of the two separate phase differences:

$$\frac{2\pi l_1}{\lambda} + \frac{2\pi l_2}{\lambda} = (\mathbf{k}_i \cdot \mathbf{r}) - (\mathbf{k}_s \cdot \mathbf{r}) = (\mathbf{k}_i - \mathbf{k}_s) \cdot \mathbf{r} \tag{6.11}$$

We define the change in wave vector as

$$\mathbf{Q} = \mathbf{k}_i - \mathbf{k}_s \tag{6.12}$$

Q is called the **scattering vector**, and is a useful shorthand because the overall phase change depends only on **Q**, not on either $\mathbf{k}_i$ or $\mathbf{k}_s$ separately.

If the scattered wave from the atom at the origin has equation

$$\psi_1(\mathbf{x}) = \exp(i\mathbf{k}_s \cdot \mathbf{x}) \tag{6.13}$$

the wave scattered from the second atom will be out of phase by the amount we have calculated:

$$\psi_2(\mathbf{x}) = \exp(i\mathbf{k}_s \cdot \mathbf{x}) \exp(i\mathbf{Q} \cdot \mathbf{r}) \tag{6.14}$$

The total scattering is given by the sum of the two rays:

$$\psi_1(\mathbf{x}) + \psi_2(\mathbf{x}) = \exp(i\mathbf{k}_s \cdot \mathbf{x}) \times (1 + \exp(i\mathbf{Q} \cdot \mathbf{r})) \tag{6.15}$$

Thus the amplitude of the scattered beam is modified by the phase factor:

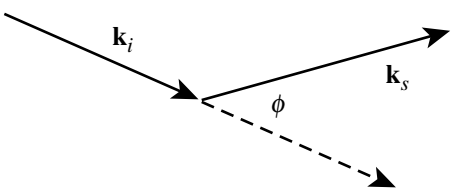
$$F(\mathbf{Q}) = 1 + \exp(i\mathbf{Q} \cdot \mathbf{r}) \tag{6.16}$$

where the first term represents the component from the first atom and the second term represents the component from the second atom.

To link **Q** to more familiar concepts in diffraction, we note that

$$|\mathbf{Q}|^2 = |\mathbf{k}_i - \mathbf{k}_s|^2 = k_i^2 + k_s^2 - 2\mathbf{k}_i \cdot \mathbf{k}_s \tag{6.17}$$

The angle between $\mathbf{k}_i$ and $\mathbf{k}_s$ is equal to 2θ, and the vectorial relationship between $\mathbf{k}_i$, $\mathbf{k}_s$, and **Q** can be represented by



Given that the magnitudes of the incoming and scattered wave vectors are $2\pi/\lambda$, we have

$$\mathbf{k}_i \cdot \mathbf{k}_s = \left(\frac{2\pi}{\lambda}\right)^2 \cos 2\phi \tag{6.18}$$

and

$$\begin{aligned} |\mathbf{Q}|^2 &= 2\left(\frac{2\pi}{\lambda}\right)^2 - 2\left(\frac{2\pi}{\lambda}\right)^2 \cos 2\theta \\ &= \frac{8\pi^2}{\lambda^2}(1 - \cos 2\theta) = \frac{16\pi^2}{\lambda^2}\sin^2\theta \end{aligned} \tag{6.19}$$

Thus

$$Q = |\mathbf{Q}| = \frac{4\pi}{\lambda}\sin\theta \tag{6.20}$$

When we apply these equations to Bragg scattering from lattice planes of spacing d, the Bragg equation $\lambda = 2d\sin\theta$ leads to $Q = 2\pi/d$. This anticipates the result obtained in Section 6.7.

6.3.3 Scattering of radiation from a collection of particles

The theory developed above can easily be generalized for a collection of particles. If we define the positions of each particle with respect to an origin as $\mathbf{r}_j$, we simply add together the phase shifts from each particle:

$$F(\mathbf{Q}) = \sum_j \exp(i\mathbf{Q}\cdot\mathbf{r}_j) \tag{6.21}$$

The intensity of the scattering beam is given as

$$|F(\mathbf{Q})|^2 = \left|\sum_j \exp(i\mathbf{Q}\cdot\mathbf{r}_j)\right|^2 = \sum_{i,j}\exp[i\mathbf{Q}\cdot(\mathbf{r}_i - \mathbf{r}_j)] \tag{6.22}$$

We note two important points. First, the measured intensity contains information about the separations of all pairs of particles, $\mathbf{r}_i - \mathbf{r}_j$. Second, the intensity $|F(\mathbf{Q})|^2$ is a real positive value, whereas $F(\mathbf{Q})$ is a complex number. By measuring $|F(\mathbf{Q})|^2$ rather than $F(\mathbf{Q})$ we have lost all information about the phase of $F(\mathbf{Q})$. This quantity contains the information about the positions of particles relative to a common origin, and this information cannot be extracted directly from measurements of $|F(\mathbf{Q})|^2$. This is the so-called **phase problem**, which we will need to come back to later in this chapter (Sections 6.7.4 and 6.9).

To generalize, we need to take account of situations where there are different types of particles. Each particle will scatter the radiation beam by a different amount. This is accounted for by weighting the components of the amplitude:

$$F(\mathbf{Q}) = \sum_j f_j \exp(i\mathbf{Q}\cdot\mathbf{r}_j) \tag{6.23}$$

The factor f_j is known by a variety of names, such as **scattering factor** or **form factor**. In neutron scattering it is called the **scattering length**, reflecting the roots

of neutron scattering in nuclear physics where the square of the scattering length gives a cross-sectional area. In X-ray scattering, where we consider scattering from atoms, the factor f_j is called the **X-ray atomic scattering factor**, and is actually a function of $Q = |\mathbf{Q}|$, for reasons we explain shortly (Section 6.4.2).

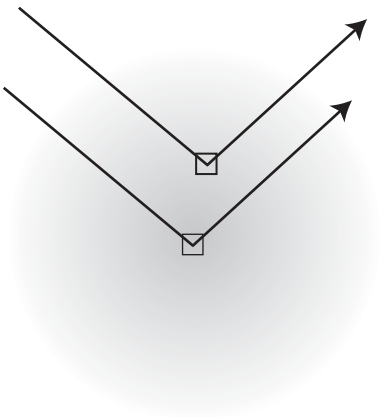


Fig. 6.13 Two rays of a radiation beam scattered from volume elements in a continuous distribution of particles, in this case shown as a representation of the distribution of electrons in an atom.

6.4 Scattering of radiation from a continuous distribution of particles

6.4.1 General principle

To develop the theory for a continuous distribution of particles, such as for a distribution of electrons around an atom, we consider scattering from small volume elements within the distribution, as illustrated in Fig. 6.13. If we define an origin, which in our example will be the centre of the atom, we consider scattering of the radiation beam from a small volume element $\mathrm{d}V$ at position $\mathbf{r}$. The density of particles in this volume element is $\rho(\mathbf{r})$. Thus the amplitude of scattering from the volume element will be

$$\mathrm{d}F(\mathbf{Q}) = \rho(\mathbf{r}) \exp(i\mathbf{Q} \cdot \mathbf{r})\, \mathrm{d}V \tag{6.24}$$

To obtain the scattering from the whole distribution, we simply integrate over the whole volume, denoting $\mathrm{d}V$ by $\mathrm{d}\mathbf{r}$:

$$F(\mathbf{Q}) = \int \rho(\mathbf{r}) \exp(i\mathbf{Q} \cdot \mathbf{r})\, \mathrm{d}\mathbf{r} \tag{6.25}$$

The important result is that the amplitude of the scattering now provides information about the spatial dependence of the particle density.

6.4.2 X-ray atomic scattering factor

X-rays are scattered by electrons. When we consider scattering of X-rays by atoms we can think of scattering from the continuous distribution of electrons around the atom. If we denote the electron density as $\rho_{\mathrm{el}}(\mathbf{r})$, the X-ray atomic scattering factor is given as

$$f(\mathbf{Q}) = \int \rho_{\mathrm{el}}(\mathbf{r}) \exp(i\mathbf{Q} \cdot \mathbf{r})\, \mathrm{d}\mathbf{r} \tag{6.26}$$

In the limit $\mathbf{Q} \to 0$, where the X-rays are scattered without deflection, we have

$$f(\mathbf{Q} = 0) = \int \rho_{\mathrm{el}}(\mathbf{r})\, \mathrm{d}\mathbf{r} = Z \tag{6.27}$$

where Z is the total number of electrons in the atom or ion. For all atoms and ions of interest $\rho_{el}(\mathbf{r})$ can be calculated using quantum mechanics, enabling the scattering factor to be calculated. As an aside, we note that in practice there will not be an analytical function for $f(Q)$, and so for practical use the numeric values of $f(Q)$ will be fitted to an appropriate functional form. Typically this function is represented as

$$f(Q) = \sum_{i=1}^{4} a_i \exp(-b_i \sin\theta/\lambda) + c \tag{6.28}$$

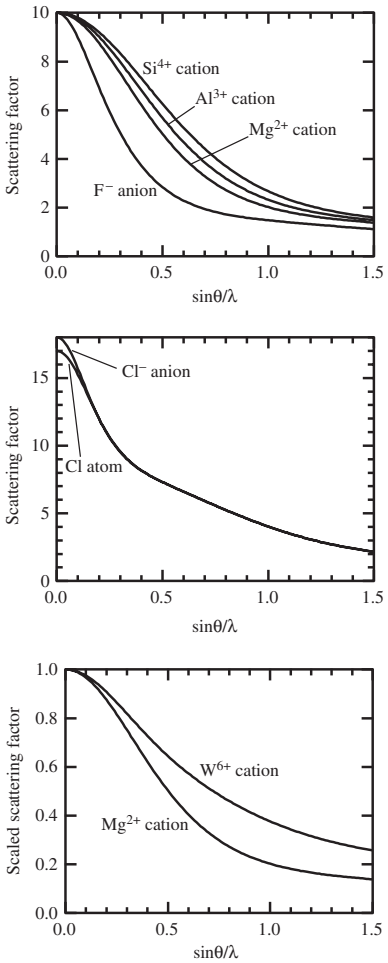


Fig. 6.14 Examples of X-ray scattering factors (see text).

and values for the coefficients have been tabulated for most atoms and ions in the International Tables for Crystallography (Volumes IV or C).

Figure 6.14 shows the X-ray atomic scattering factors for a number of neutral and charged atoms. One interesting comparison is for F^-, Mg^{2+}, Al^{3+}, and Si^{4+}. These cations all have the same number of electrons, so that the values of the scattering factors are equal at $\sin\theta/\lambda = 0$. However, the ions have different sizes, and this is reflected in the way that the scattering factors fall off with increasing $\sin\theta/\lambda$.

Also shown in Fig. 6.14 is a comparison between the scattering factors for neutral Cl and the Cl^- anion. The Cl^- anion has one more electron, so its scattering factor at $\sin\theta/\lambda = 0$ is larger by one electron unit. This extra electron is localized to the outer electron shell of the Cl atom, which means that its contribution to the atomic scattering factor is restricted to lower values of $\sin\theta/\lambda$. Thus for increasing values of $\sin\theta/\lambda$ the scattering factors for the neutral atom and anion become very similar.

Finally, in Fig. 6.14 is a comparison of the scattering factors for the Mg^{2+} and W^{6+} cations. These have similar ionic radii. This means that the outer electrons in both cases will have contributions to the atomic scattering factors of similar form, in particular with a similar dependence on $\sin\theta/\lambda$. However, the W^{6+} cation contains many more electrons deep inside the atom, which will give a large contribution to the scattering factor at larger values of $\sin\theta/\lambda$.

By these simple examples we have been able to understand the main features of atomic scattering factors, and how they are related to the electronic structure of atoms and ions.

6.4.3 Neutron scattering factors

In principle there is a similar spatial dependence for the scattering of neutrons from nuclei. However, the sizes of atomic nuclei are typically 10^{-5} times the size of the atom, and the wavelengths of neutrons in diffraction experiments are similar to the atomic size. In this case, the scattering function is so wide in Q that it can be treated as a constant for all values of Q of practical interest.

Whereas the X-ray scattering factors in the limit $Q \to 0$ scale as the number of electrons, there is no correlation between the neutron scattering length at the atomic number. The range of coherent neutron scattering lengths is plotted in Fig. 6.15.

6.5 Diffraction and Fourier analysis

6.5.1 Scattering processes as Fourier transforms

The equations we have developed are examples of the mathematical technique of Fourier transformation. Although we have not explicitly used Fourier transforms to develop the theory of diffraction, the final result has been that the diffraction amplitude is the Fourier transform of the diffracting object.

In application to X-ray diffraction, the amplitude of the scattered signal is given by the Fourier transform of the electron density. If we have measurements of the X-ray amplitudes over a wide range of scattering vectors, we could

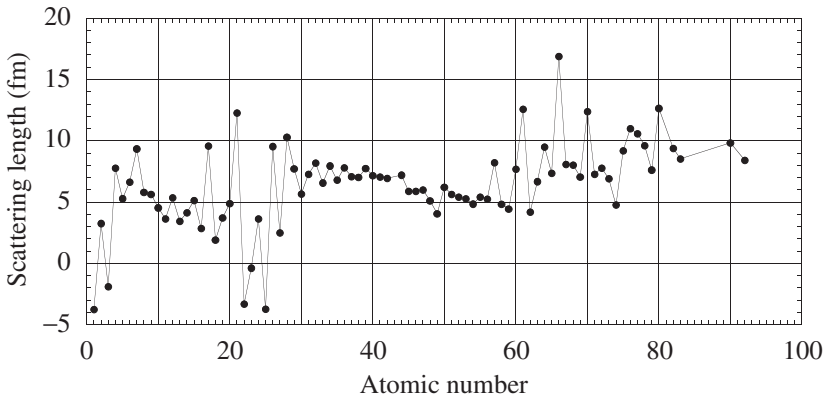


Fig. 6.15 Range of values of the neutron coherent scattering length plotted as a function of atomic number, showing little correlation between the scattering length and atomic number. (Data taken from http://www.ncnr.nist.gov/resources/n-lengths/.)

reconstruct the electron density by reverse Fourier transformation. The problem that we have already mentioned, and which we will return to later (Sections 6.7.4 and 6.9), is that we can only measure the magnitude of the amplitude of the scattered radiation and we have no information about the phase.

We noted in Section 3.2.7 that a point particle can be described by the Dirac delta function. The instantaneous density of particles in the sample of volume V is therefore given as

$$\delta(\mathbf{r}) = \frac{1}{V}\sum_j \delta(\mathbf{r} - \mathbf{r}_j) \tag{6.29}$$

From eqn 6.25, which describes the scattering of radiation from a continuous distribution of particles, we have

$$\begin{aligned} F(\mathbf{Q}) &= \frac{1}{V}\sum_j \int \delta(\mathbf{r} - \mathbf{r}_j)\exp(i\mathbf{Q}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} \\ &= \sum_j \exp(i\mathbf{Q}\cdot\mathbf{r}) \end{aligned} \tag{6.30}$$

The details of this Fourier transform are given in Appendix B, Section B.3.1. The result of the Fourier transform, not surprisingly, is exactly the same as the result for scattering from a collection of particles derived earlier, eqn 6.21.

6.5.2 Fourier transforms and convolution

The convolution theorem

The power of Fourier transforms in diffraction is in the use of convolution. In Section 3.2.7 we noted that the crystal structure can be described as the convolution of the mathematical functions describing the atomic coordinates and the lattice. In fact we will need to extend this description, and the fact that we take this approach will make subsequent analysis much easier than it might otherwise be.

Suppose we have three functions, $f(x)$, $g(x)$, and $h(x)$, that are related through the convolution operation as $f(x) = g(x) \otimes h(x)$. The respective Fourier transforms, $F(k)$, $G(k)$, and $H(k)$, are related by a simple relation:

$$F(k) = \int (g(x) \otimes h(x)) \exp(ikx)\, \mathrm{d}x = G(k) \times H(k) \tag{6.31}$$

This simply states that the convolution of two functions in one space corresponds to the product in the Fourier space. Equation 6.31 is derived in Appendix B, Section B.4.

The point we will establish in the following sections is that since the crystal structure can be described as a convolution of a number of constituent parts (lattice, basis, electrons, and thermal motion), the Fourier transform will simply be equal to the product of the Fourier transforms of the individual constituents. This will make the formal analysis of the diffraction from complex objects considerably easier! However, before we proceed to use the **convolution theorem**, as eqn 6.31 is known, we will pull together all the results established so far to consider the analysis of the structure of glasses from diffraction data.

6.6 Application: the structure of glasses revealed by neutron scattering

We noted above that the intensity of a scattered beam of radiation contains information about the *relative* positions of particles rather than their *absolute* positions. If we have a collection of particles, we cannot tell from the scattered beam where the particles are, but we can obtain information about how close the particles are to each other. In the study of disordered systems, such as liquids and glasses, this is actually all we need, for any order only exists over small length scales.

Figure 6.16 shows the intensity of scattering of neutrons from silica (SiO_2) glass. Unlike diffraction from crystals, as seen above in Figs 6.6 and 6.11, which contain sharp Bragg peaks, diffraction from glass shows very broad features. However, it turns out that it is possible to relate the diffraction pattern of silica glass to that of the cristobalite crystalline phase of silica (Section 2.3.3) by interpreting the features of the silica glass diffraction as broadened Bragg peaks of the cristobalite diffraction pattern. For example, the value of Q for the first peak in the measured diffraction pattern of silica glass is very close to the value for the strong 111 peak observed in diffraction from a powdered sample of cristobalite. Thus one might, as a cautious first approximation, interpret the diffraction pattern of the glass as $S_{\mathrm{glass}}(Q) \approx S_{\mathrm{crystal}}(Q) \otimes \mathcal{R}(Q)$, where $\mathcal{R}(Q)$ is the broadening function which reflects the length scale on which the local structure of the glass is similar to that of the crystal. The Fourier transforms of the scattering intensity are shown in Appendix J to yield the neutron-weighted $g(r)$ functions directly. The result for the Fourier transform of the data shown in Fig. 6.16 is given in Fig. 6.17. Since the diffraction patterns of the crystal and glass are related by a convolution operation, the resultant $g(r)$ functions for the two phases will simply differ by the factor given by the Fourier transform of $R(Q)$. This is consistent with the results shown in Chapter 2.

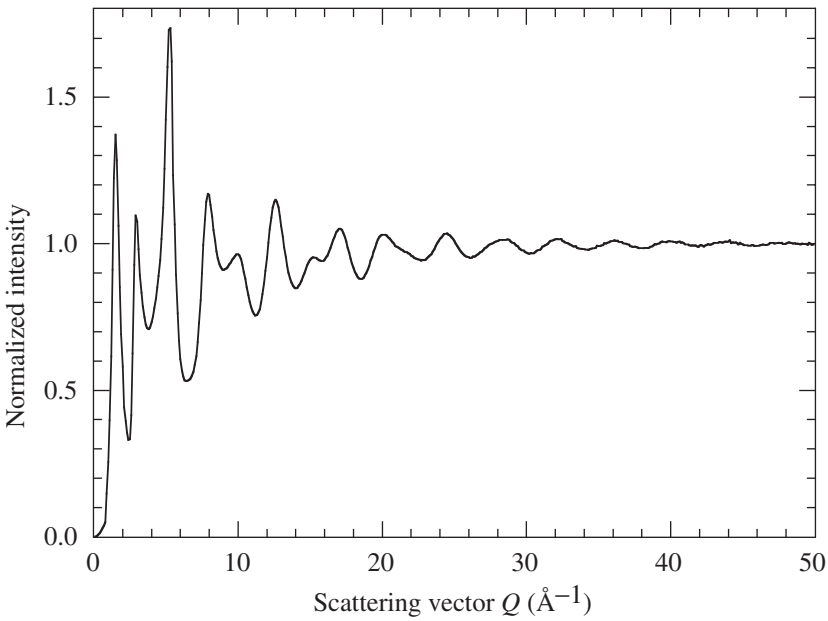


Fig. 6.16 Neutron diffraction pattern from silica glass. (Data obtained by Matthew Tucker, David Keen and the author.)

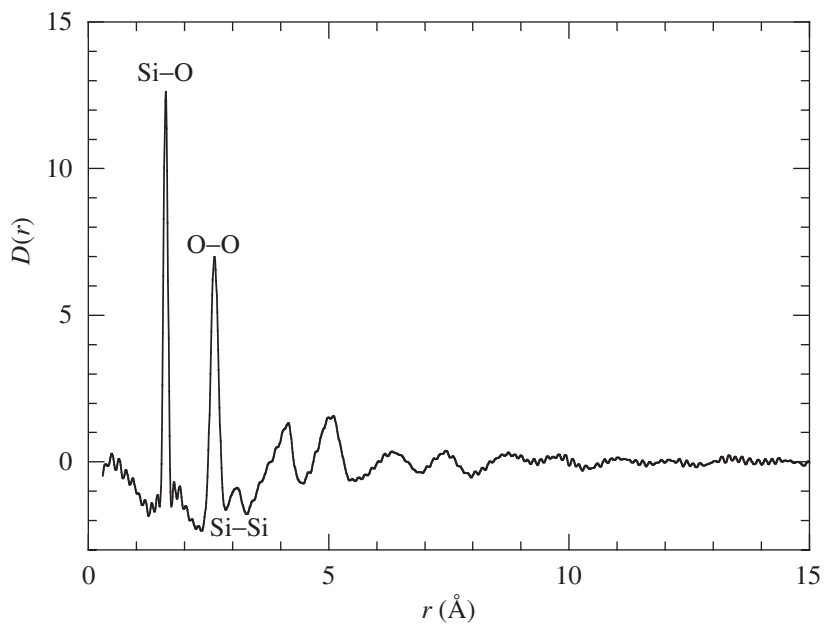


Fig. 6.17 Experimental neutron diffraction measurement of $D(r) = 4\pi r \sum_{i,j} c_i c_j \times b_i b_j g_{ij}(r)$ for silica glass (see Appendix J for a discussion of the function $D(r)$). The peaks correspond to specific interatomic distances (e.g. Si–O at 1.6 Å). (Data from Matthew Tucker, David Keen and the author.)

6.7 Diffraction from crystalline materials

6.7.1 Fourier transform of the perfect crystal

The use of the convolution theorem

We can think of the crystal structure in terms of three convolution operations:

$$\begin{aligned}&\text{Crystal structure}\\&\quad=\text{lattice}\otimes\text{unit cell}\otimes\text{electrons in atoms}\otimes\text{thermal motion}\end{aligned}\tag{6.32}$$

When we take the Fourier transform of the crystal structure, the convolutions transform to products of individual Fourier transforms (eqn 6.31):

$$\begin{aligned}\text{FT(Crystal structure)}=\text{FT(lattice)}\times\text{FT(unit cell)}\times\text{FT(electrons in atoms)}\\\times\text{FT(thermal motion)}\end{aligned}\tag{6.33}$$

Therefore, to build up the Fourier transform of the crystal we can treat each of these Fourier transforms separately.

Fourier transform of the lattice

The Fourier transform of the lattice was discussed in Section 4.4, and was shown to be equivalent to the reciprocal lattice. Reproducing the main result (eqn 4.37), this can be written as

$$\mathcal{R}(\mathbf{Q})=\sum_{h,k,l}\delta(\mathbf{Q}-(h\mathbf{a}^*+k\mathbf{b}^*+l\mathbf{c}^*))\tag{6.34}$$

Fourier transform of the individual atoms

The unit cell contains a group of atoms, where each atom is defined by a position and has a distribution of electron density. If the position of each atom is represented by a Dirac delta function, one can think of the electron density of the whole group as being described by the convolution of the delta functions by the functions describing the electron density about each atom, $\delta_j(\mathbf{r})$. The Fourier transform of the electron density about a single atom is the atomic scattering factor $f_j(\mathbf{Q})$ which we discussed earlier (Section 6.4.2).

An atom at the position $\mathbf{r}_j$ is represented by the delta function $\delta(\mathbf{r}-\mathbf{r}_j)$. This has the Fourier transform $\exp(i\mathbf{Q}\cdot\mathbf{r}_j)$. The electron density associated with the atom at position $\mathbf{r}_j$ can be represented by the convolution $\rho_j(\mathbf{r})\otimes\delta(\mathbf{r}-\mathbf{r}_j)$. The resultant Fourier transform is the product $f_j(\mathbf{Q})\exp(i\mathbf{Q}\cdot\mathbf{r}_j)$.

Taking account of thermal motion

The effect of thermal motion is to blur the position of the atom. In effect, the position of the atom needs to be convoluted with the spread of displacements that arise from the thermal motion, so in the Fourier transform of the unit cell we have to incorporate the Fourier transform of the spread of displacements as a multiplicative factor. If the distribution of atom displacements from the average position is written as $p(\mathbf{r})$, the distribution of the electrons in the atom within the crystal can be represented by the new convolution $\rho_j(\mathbf{r})\otimes\delta(\mathbf{r}-\mathbf{r}_j)\otimes p(\mathbf{r})$.

If we assume that the thermal motion is isotropic and harmonic, $p(\mathbf{r})$ will be a Gaussian function whose width is characterized by the mean-squared atomic displacement, $\langle u_j^2 \rangle$. The Fourier transform of the spread of distributions will also be a Gaussian distribution (the Fourier transform of a Gaussian function is given in Appendix B, Section B.3.4):

$$T_j(\mathbf{Q}) = \exp\bigl(-8\pi^2 \langle u_j^2 \rangle \sin^2\theta/\lambda^2\bigr) = \exp(-B_j \sin^2\theta/\lambda^2) \qquad (6.35)$$

It is common to use the shorthand

$$B_j = 8\pi^2 \langle u_j^2 \rangle \qquad (6.36)$$

Because $\langle u_j^2 \rangle$ is found to be proportional to temperature (except at low temperatures), this is called the **temperature factor**. Alternative names are the **Debye–Waller factor**, after the people who first developed the theory, and the **atomic displacement parameter**. The links to lattice dynamics, which are covered in Chapters 8 and 9, are given in Appendix N.

Fourier transform of the unit cell

The actual unit cell consists of several atoms, whose Fourier transform are simply added together. The end result for the Fourier transform of the unit cell is the **structure factor**:

$$F(\mathbf{Q}) = \sum_j f_j(\mathbf{Q}) \exp(i\mathbf{Q} \cdot \mathbf{r}_j) T_j(\mathbf{Q}) \qquad (6.37)$$

Fourier transform of the crystal

The final result for the Fourier transform of the crystal is the product of the structure factor and the reciprocal lattice. Since the delta functions in the reciprocal lattice function are only non-zero when $\mathbf{Q} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$, the Fourier transform of the crystal need only be considered at these values of $\mathbf{Q}$.

We write the atomic position of an atom in the unit cell as $\mathbf{r}_j = x_j\mathbf{a} + y_j\mathbf{b} + z_j\mathbf{c}$ (eqn 3.3), and thus we have the product

$$\begin{aligned} \mathbf{Q} \cdot \mathbf{r}_j &= 2\pi(h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \cdot (x_j\mathbf{a} + y_j\mathbf{b} + z_j\mathbf{c}) \\ &= 2\pi(hx_j + ky_j + lz_j) \end{aligned} \qquad (6.38)$$

The structure factor can now be written in its usual form:

$$\begin{aligned} F(hkl) = \sum_j f_j(Q_{hkl}) &\exp(2\pi i(hx_j + ky_j + lz_j)) \\ &\times \exp(-B_j \sin^2\theta_{hkl}/\lambda^2) \end{aligned} \qquad (6.39)$$

Here we have assumed that the shape of the atoms is spherical, depending only on the magnitude of $\mathbf{Q}_{hkl}$, i.e. $Q_{hkl} = |\mathbf{Q}_{hkl}| = 4\pi \sin\theta_{hkl}/\lambda$. In a first approximation it can also be assumed that all atoms have the same value of B, which is then known as the **overall temperature factor**. In more sophisticated models of the crystal structure each atom can be treated with its own **anisotropic temperature factor**, which is represented by six parameters per atom.

6.7.2 The effect of particle size on the diffraction pattern

By the use of the convolution approach we can also account for the effect of crystal size. The formalism developed in this chapter is strictly appropriate for infinite systems. This is inherant in the definition of the lattice. A crystal of finite size is effectively an infinite crystal multiplied by a three-dimensional equivalent of the slit function as the function that represents the shape of the crystal. Therefore, the diffraction pattern can be thought of as the diffraction from an infinite crystal convoluted with the Fourier transform of the three-dimensional slit function, which is derived in Appendix B, Section B.3.2. This will lead to broadening of the diffraction peaks.

The effect of particle size is often of no practical consequence in single-crystal X-ray and neutron diffraction measurements, because the broadening of the diffraction reflections is smaller than the resolution of the instrument. In powder diffraction, where particles are necessarily small, the broadening of the diffraction lines can be a problem with high-resolution instruments. It is not necessary for a powder to be this fine for the powder diffraction method to work.

It is in the electron microscope where the particle size broadening is exploited to best effect. The sample in electron diffraction has to be made particularly thin in the direction parallel to the electron beam in order to allow the beam to pass through the sample. The finite size in this direction leads to a great deal of broadening of the Bragg peaks in the same direction, which enables a wide diffraction pattern to be obtained with a single setting of the crystal. This is discussed in Appendix H.

6.7.3 The inverse transform: obtaining the electron density from X-ray diffraction measurements of the structure factor

With X-ray diffraction, the inverse Fourier transform of the set of structure factors gives the electron density:

$$\rho(x, y, z) = \sum_{h,k,l} F(hkl)\exp[-2\pi i(hx + ky + lz)] \tag{6.40}$$

When we obtain a measurement of diffraction intensities, we can only obtain data over a limited range of values of Q. For example, if we use Cu Kα radiation with $\lambda = 1.542\,\text{Å}$, we only have data up to a maximum value of $Q_{\max} = 4\pi/\lambda = 8.15\,\text{Å}^{-1}$. This situation is equivalent to multiplying the diffraction pattern by a three-dimensional equivalent of a slit function. Thus the resultant Fourier transform of the diffraction amplitude that is supposed to give the electron density will actually give the electron density convoluted with the Fourier transform of the slit function, which is known as the **sinc function** (Appendix B, Section B.3.2):

$$y(aQ) = \sin aQ/aQ \tag{6.41}$$

with full width of order $2\pi/Q_{\max} \sim 0.75\,\text{Å}$. This means that all detail in the calculated electron density is blurred by this amount, i.e. we can resolve the structural details only down to this length scale. If we consider the example of a

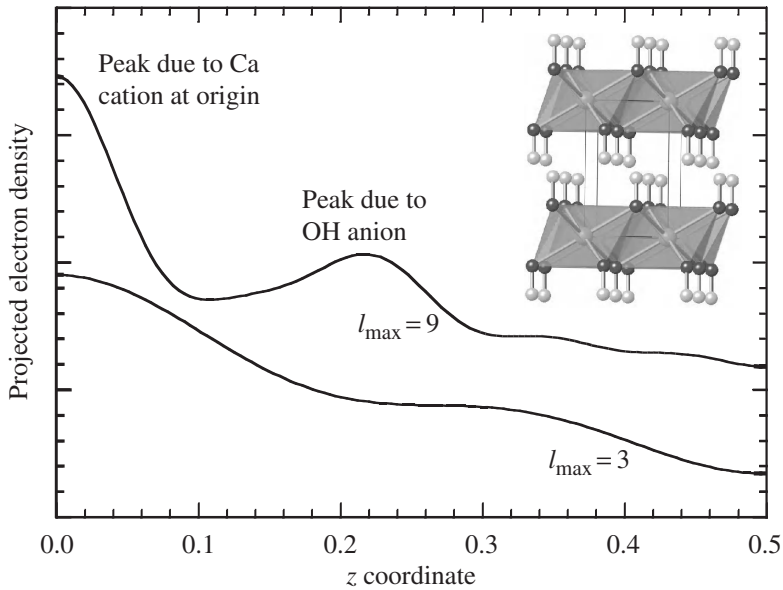


Fig. 6.18 One-dimensional construction of the projection of the electron density along [001] in $Ca(OH)_2$ by summing $\rho(z) = \sum_l F(00l) \exp(2\pi l z)$. The two plots (displaced vertical axes) show the effect of the restricting range of Q by limiting the sum to maximum values of l. The structure of $Ca(OH)_2$ has the Ca cation at the origin, and sheets of OH anions parallel to (001).

silica glass, the shortest distance in $g(r)$ is the Si–O bond length of 1.6 Å. To get reasonable resolution on this length scale it is possible to use neutron scattering with $Q_{max} = 40\,\text{Å}^{-1}$, giving a resolution of 0.15 Å (equivalent to 10% of the bond length). In fact, this argument is somewhat overpessimistic, since it is possible to obtain the midpoint of a distribution to much higher precision, which is why diffraction is capable of giving useful structural information!

The effect of limiting the range of values of Q is shown in Fig. 6.18, where the $00l$ reflections from $Ca(OH)_2$ are combined to give the projection of the electron density along [001]. The full three-dimensional crystal structure is also shown in this figure. The hydrogen atoms are unlikely to be resolved in the X-ray data, and so we expect to see one peak at the origin corresponding to the Ca atom and one peak at $z \sim 0.22$ corresponding to the O atom. The results in the figure show that the projection of the electron density becomes sharper (to within the size of the atoms) as the range of reflections is increased.

6.7.4 The phase problem

In order to obtain the electron density from X-ray diffraction, we need measurements of $F(hkl)$. A measurement of the intensity of the reflections will give the set of $|F(hkl)|^2$ for all h, k, l to a maximum value of Q, from which it is possible to obtain the magnitudes $|F(hkl)|$. However, the structure factor has a phase, whether positive or negative for centrosymmetric crystals or a general phase angle for the complex structure factor in acentric crystals. The Fourier transform of the set of $F(hkl)$ required to obtain $\rho(x, y, z)$ presupposes that the phase is incorporated into the structure factor.

It is interesting to consider how important the phase is. Figure 6.19 shows two two-dimensional objects that have been digitized to give two-dimensional functions. Below each function is the diffraction pattern obtained as the square of the modulus of the Fourier transform. Then we show the inverse Fourier

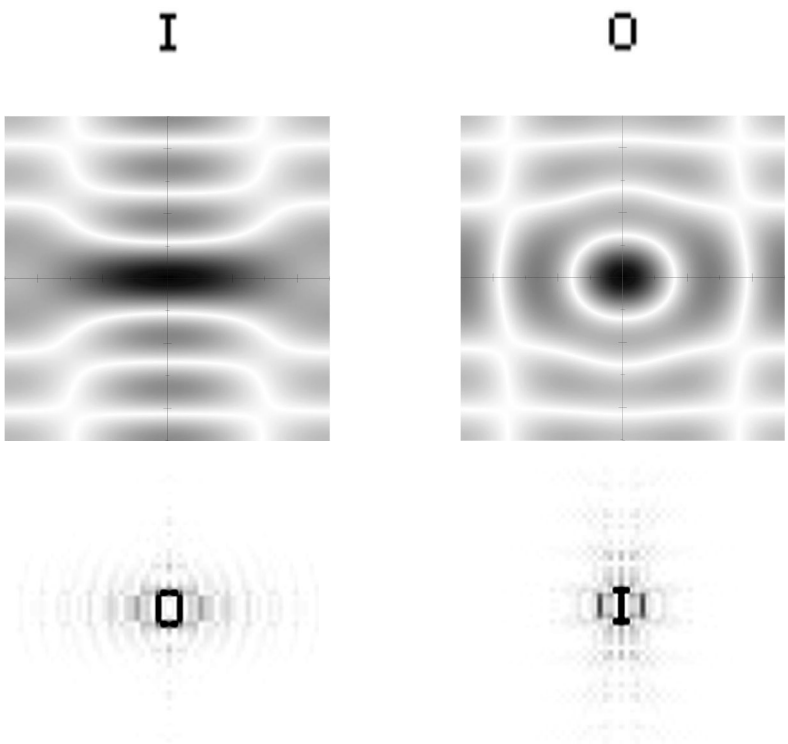


Fig. 6.19 Example of the importance of the phase problem. The example begins with digitized images of the letters I and O (top). The figures immediately below show in the inner portions the squared modulus of the Fourier transforms, corresponding to what would be measured as $|F(\mathbf{Q})|^2$. The bottom figures show the result of reconstructing the images using the inverse Fourier transform of the functions constructed by combining the values of $|F(\mathbf{Q})|$ for the images shown at the top of the column and the phases of the calculated $F(\mathbf{Q})$ for the image in the other column. What is noted is that the reconstructed images are primarily determined by the phases, and the effects of using different amplitudes is simply to generate noise in the image.

transform constructed using the magnitudes of the Fourier transform of each object but by swapping the phases of the two Fourier transforms. The interesting point is that the result in each case is an object that is not very different from the object from which the phases were derived. This is striking – taken at face value it suggests that most of the information about the object is contained within the phase of the Fourier transform, and the phase is the one part of the structure factor that cannot be measured. It appears that the magnitudes only give a bit of fine detail once the object has been reconstructed.

Given the demonstrated importance of the phases, it is not clear how the magnitudes of the structure factors could ever be used to reconstruct the crystal structure. It is the issue of how this problem can be solved that occupies the rest of this chapter. But before specifically dealing with this problem it is useful to consider how the symmetry of the crystal (Section 3.5) can make the problem rather easier. In a complicated structure the formal use of symmetry shows how the positions of some atoms are related. This means that the number of unknown parameters in the crystal structure can be reduced.

6.8 Effects of symmetry on diffraction patterns

6.8.1 Friedel's law

From the equation for $F(hkl)$ we note that

$$F(\bar{h}\bar{k}\bar{l}) = F^*(hkl) \tag{6.42}$$

where F^* is the complex conjugate of F. This then gives

$$|F(hkl)|^2 = F(hkl) \times F^*(hkl) = |F(\bar{h}\bar{k}\bar{l})|^2 \tag{6.43}$$

This relation is known as **Friedel's law**, and predicts that the distribution of intensities in reciprocal space has a centre of symmetry.

Friedel's law can be violated in crystals without a centre of symmetry, when the energy of the X-ray beam is close to an absorption energy of an atom. In this case the atomic scattering factor contains an imaginary component:

$$f = f' + if'' \tag{6.44}$$

The real component, f', corresponds to the amplitude of the scattering process, and the imaginary component, f'', corresponds to the amplitude of the absorption process. The absorption process requires that the energy of the X-ray photon is close to an absorption resonance of the atom, so f'', unlike f', is very sensitive to the X-ray wavelength.

Consider a structure factor of a crystal containing N atoms, and with one atom that has a complex scattering factor. Its structure factor can be written as

$$\begin{aligned} F(hkl) = &\sum_{j=1}^{N-1} f_j \exp\left[2\pi i(hx_j + ky_j + lz_j)\right] \\ &+ f_N \exp[2\pi i(hx_N + ky_N + lz_N)] \end{aligned} \tag{6.45}$$

where atom N is the atom with the complex scattering factor. We denote hkl as $\mathbf{h}$, and rewrite the structure factor as

$$F(\mathbf{h}) = F_{N-1}(\mathbf{h}) + F_N(\mathbf{h}) \tag{6.46}$$

where $F_{N-1}(\mathbf{h})$ denotes the component of the structure factor due to all the atoms, apart from the one with the complex scattering factor, and $F_N(\mathbf{h})$ denotes the component from this atom. For normal atoms, we have $F_{N-1}(-\mathbf{h}) = F^*_{N-1}(\mathbf{h})$. However, the relation is not so simple for anomalous scattering. We write

$$\begin{aligned} F_N(\mathbf{h}) &= (f' + if'')(a(\mathbf{h}) + ib(\mathbf{h})) \\ &= (f'a(\mathbf{h}) - f''b(\mathbf{h})) + i(f'b(\mathbf{h}) + f''a(\mathbf{h})) \end{aligned} \tag{6.47}$$

where $a(\mathbf{h})$ and $b(\mathbf{h})$ are the real and imaginary parts of the phase factor. For $F_N(-\mathbf{h})$ we have

$$\begin{aligned} F_N(-\mathbf{h}) &= (f' + if'')(a(\mathbf{h}) - ib(\mathbf{h})) \\ &= (f'a(\mathbf{h}) + f''b(\mathbf{h})) + i(-f'b(\mathbf{h}) + f''a(\mathbf{h})) \end{aligned} \tag{6.48}$$

Clearly $F_N(-\mathbf{h}) \neq F^*_N(\mathbf{h})$, so it follows that $F(-\mathbf{h}) \neq F^*(\mathbf{h})$. This is illustrated in Fig. 6.20.

Whilst anomalous scattering does not often occur naturally with standard laboratory X-ray sources, X-rays generated by synchrotron sources can be tuned to specific wavelengths close to absorption wavelengths in order to produce a violation of Friedel's law. This can be a useful indication of whether or not a crystal has a centre of symmetry, in which case $b(\mathbf{h})$ would be zero and Friedel's law would still be obeyed. It can also be useful for resolving problems in the determination of the crystal structure, particularly when two different atoms have similar scattering factors.

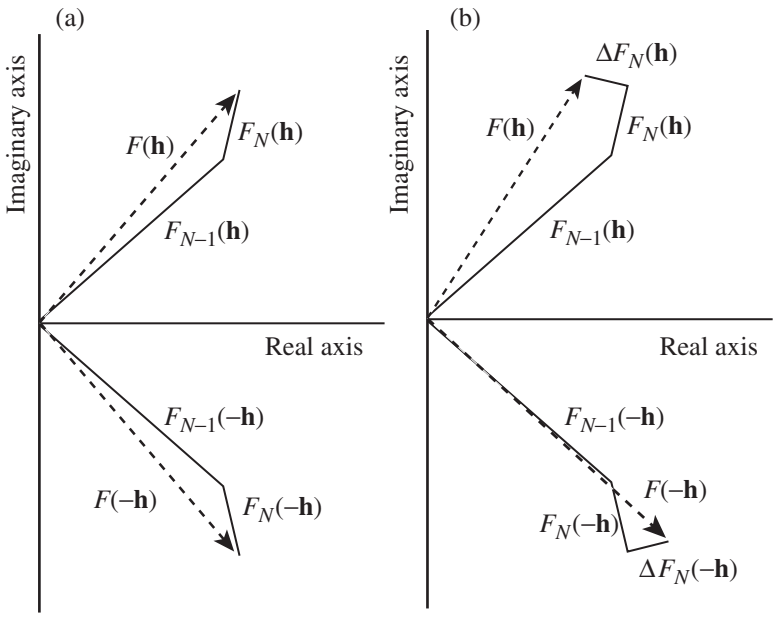


Fig. 6.20 Illustration of the effect of anomalous scattering on the structure factors of a pair of $\mathbf{h} = hkl$ and $-\mathbf{h} = \overline{hkl}$ reflections: (a) represents the case where there is no anomalous scattering, and (b) represents the case of anomalous scattering.

6.8.2 Point symmetry of diffraction patterns

The structure factors are subject to the same point symmetry of the crystal. For example, in a crystal with mirror symmetry that causes an atom with coordinates x, y, z to be replicated at $x, \bar{y}, z$, and N atoms in the unit cell, the structure factor (without the temperature factor) is given as

$$F(hkl) = \sum_{j=1}^{N/2} f_j[\exp(2\pi i(hx_j + ky_j + lz_j)) + \exp(2\pi i(hx_j - ky_j + lz_j))] \tag{6.49}$$

By inspection it can be seen that $F(hkl) = F(h\bar{k}l)$, so that the structure factors reflect the same mirror symmetry. Similarly, if the crystal has a 2-fold rotation axis parallel to [001] that causes an atom with coordinates x, y, z to be replicated at $\bar{x}, \bar{y}, z$, the structure factor (without the temperature factor) is given as

$$F(hkl) = \sum_{j=1}^{N/2} f_j[\exp(2\pi i(hx_j + ky_j + lz_j)) + \exp(2\pi i(-hx_j - ky_j + lz_j))] \tag{6.50}$$

In this case it can be seen that $F(hkl) = F(\bar{h}\bar{k}l)$.

The general result is that the point group symmetry operations of the crystal are also reflected in the diffraction pattern. However, Friedel's law means that the diffraction pattern has a centre of symmetry, even when the symmetry of the crystal does not contain a centre of symmetry. As a result, the point symmetry of the diffraction pattern is the point symmetry of the crystal plus a centre of

symmetry. This gives the **Laue symmetry**:

System	Laue class	Point groups of the Laue class			
Triclinic	$\bar{1}$	1	$\bar{1}$		
Monoclinic	$2/m$	2	m	$2/m$	
Orthorhombic	mmm	222	$mm2$	mmm	
Trigonal	$\bar{3}$	3	$\bar{3}$		
	$\bar{3}m$	32	$3m$	$\bar{3}m$	
Tetragonal	$4/m$	4	$\bar{4}$	$4/m$	
	$4/mmm$	422	$4mm$	$\bar{4}2m$	$4/mmm$
Hexagonal	$6/m$	6	$\bar{6}$	$6/m$	
	$6/mmm$	622	$6mm$	$\bar{6}2m$	$6/mmm$
Cubic	$m3$	23	$3m$		
	$m\bar{3}m$	432	$\bar{4}3m$	$m\bar{3}m$	

It is not possible to tell from the symmetry of the diffraction pattern whether or not the crystal has a centre of symmetry. We will revisit this issue shortly, and will show that this information is in fact buried in the intensity distribution of the diffraction pattern.

6.8.3 Centre of symmetry

If there is a centre of symmetry in a crystal with N atoms per unit cell, the structure factor can be written as

$$\begin{aligned} F(hkl) &= \sum_{j=1}^{N/2} f_j \{\exp[2\pi i(hx_j + ky_j + lz_j)] \\ &\quad + \exp[-2\pi i(hx_j + ky_j + lz_j)]\} \\ &= \sum_{j=1}^{N} f_j \cos[2\pi i(hx_j + ky_j + lz_j)] \end{aligned} \tag{6.51}$$

In this case the phase of $F(hkl)$ is either 0 or π, which means that $F(hkl)$ is a real quantity with either positive or negative sign.

6.8.4 Systematic absences

There are some symmetry operations that have the effect of making some structure factors have zero value in a systematic way. These reflections are called **systematic absences**. These arise from the presence of translational symmetry in either of two ways. The first type of absences are those due to a non-primitive lattice, and these affect reflections across the whole of the reciprocal lattice. These essentially arise as a consequence of the conventional way of describing the unit cell, as was described in Section 4.3. The second type arises from the presence of screw axes and glide planes, and their effects are restricted to lines or planes in reciprocal space respectively. The observation of systematic

absences provides unambiguous evidence that the particular type of symmetry is present, which helps in determining the space group.

There are no systematic absences associated with the other types of symmetry. These could be deduced if the point symmetry of the crystal is known, but the problem is that the presence of a centre of symmetry in the diffraction pattern means that there is always an ambiguity about the point group of the crystal unless there is a glide plane on each of the main lattice planes. There is no way of telling whether the crystal has a centre of symmetry directly from a first inspection of the diffraction pattern. However, we will discover later that the presence or absence of a centre of symmetry does have an affect on the statistical distribution of structure factor values.

Non-primitive lattices

In earlier chapters (2–4) we discussed how the description of a non-primitive lattice could make thinking about the lattice more intuitive. The volume of the conventional (non-primitive) unit cell will be larger than that of the primitive unit cell by some factor (2, 3, or 4, depending on the lattice). This means that the volume of the reciprocal unit cell of the conventional lattice will be smaller than that of the primitive lattice by the same factor, which means that only the corresponding fraction of reciprocal lattice vectors implied by the conventional description are true reciprocal lattice vectors, and when using the conventional description, the structure factors for a systematic set of reciprocal lattice vectors will have zero value. The systematic absences for the different non-primitive lattice types are

Lattice type	Systematic absences
P	no systematic absences
I	absences when h, k, l are not all even or not all odd
F	absences when $h + k + l$ is an odd number
A	absences when $k + l$ is an odd number
B	absences when $h + l$ is an odd number
C	absences when $h + k$ is an odd number
R	absences when $-h + k + l$ is not equal to $3 \times$ an integer

These results can be obtained using the transformation matrices of Chapter 4 (Section 4.3.2, Problems 4.6 and 4.7). The observation of any of these systematic absences in a diffraction pattern will give unambiguous identification of the Bravais lattice.

Glide planes

Consider a glide plane on (010) with translations along $\mathbf{c}/2$ such that an atom at x, y, z is replicated at $x, \overline{y}, \frac{1}{2} + z$. The structure factor is then given as

$$F(hkl) = \sum_{j=1}^{N/2} f_j \left\{ \exp[2\pi i (hx_j + ky_j + lz_j)] + \exp\left[2\pi i \left(hx_j + -ky_j + l\left(\tfrac{1}{2} + z_j\right)\right)\right]\right\} \tag{6.52}$$

For $h0l$ reflections we have

$$\begin{aligned} F(h0l) &= \sum_{j=1}^{N/2} f_j \left\{ \exp(2\pi i(hx_j + lz_j)) + \exp\left[2\pi i \left(hx_j + l\left(\tfrac{1}{2} + z_j\right)\right)\right]\right\} \\ &= [1 + (-1)^l] \sum_j f_j \exp[2\pi i(hx_j + lz_j)] \\ &= 0 \quad \text{if } l = 2n + 1 \quad \text{(odd)} \end{aligned} \tag{6.53}$$

where n is an integer. Similar systematic absences arise from glide planes involving other directions.

Screw axes

Consider a 2_1 screw axis along **b** so that an atom at x, y, z is replicated at $\overline{x}, \frac{1}{2} + y, \overline{z}$. The structure factor is then given as

$$\begin{aligned} F(hkl) = \sum_{j=1}^{N/2} f_j \Big\{ &\exp[2\pi i(hx_j + ky_j + lz_j)] \\ &+ \exp\left[2\pi i \left(-hx_j + k\left(\tfrac{1}{2} + y_j\right) - lz_j\right)\right]\Big\} \end{aligned} \tag{6.54}$$

For $0k0$ reflections we have

$$\begin{aligned} F(0k0) &= \sum_{j=1}^{N/2} f_j \left\{ \exp(2\pi iky_j) + \exp\left[2\pi ik\left(\tfrac{1}{2} + y_j\right)\right]\right\} \\ &= [1 + (-1)^k] \sum_j f_j \exp(2\pi iky_j) \\ &= 0 \quad \text{if } k = 2n + 1 \quad \text{(odd)} \end{aligned} \tag{6.55}$$

where n is an integer. Similar systematic absences arise from screw axes along other directions.

Some readers may note that $0k0$ reflections with odd values of k will be absent if there is a c or n glide on (100) through the conditions on $0kl$ reflections with odd l or odd $h + l$ respectively, or if there is an a or n glide on (001) through the conditions on $hk0$ reflections with odd h or odd $h + k$ respectively. This does indeed give an ambiguity, and in all these cases it is not possible to determine directly whether or not there is a 2_1 axis along [010].

6.8.5 Determination of space-group symmetry

Clearly from the systematic absences it is possible to determine the type of lattice and the presence of glide planes and screw axes. This may not give an ambiguous identification of the space group, but it leads to a very restricted choice.

To illustrate the point, consider an orthorhombic crystal. Its Laue symmetry is necessarily of point group mmm. Suppose the systematic absences are $0kl$ reflections absent if k is an odd number, and $h0l$ reflections absent if l is an odd number. The first set of absences indicates the presence of a b glide plane

on (100), and the second set indicates the presence of a *c* glide plane on (010). We cannot determine anything about the symmetry on the z-axis from the systematic absences – recall that there are no systematic absences associated with mirror planes and 2-fold rotation axes, and this is an example of where the existence of a 2_1 axis along the third axis will be masked by the systematic absences from one of the glide planes. However, the presence of two glide planes has at least reduced the possible choices of point group to $mm2$ or mmm. If the crystal has point symmetry $mm2$, we need to determine whether the space group is either $Pbc2$ or $Pbc2_1$. We can deduce which is possible from the overall symmetry. If we start with a point in the unit cell at x, y, z, the *b*-glide on (100) will generate a symmetrically related point at $\bar{x}, \frac{1}{2} + y, z$. The *c*-glide on (010) will then generate the following two new symmetrically related points:

$$x, y, z \;\rightarrow\; x, \bar{y}, \frac{1}{2} + z$$

and

$$\bar{x}, \frac{1}{2} + y, z \;\rightarrow\; \bar{x}, \frac{1}{2} - y, \frac{1}{2} + z$$

The last position is equivalent to that generated by the operation on the point x, y, z by a 2_1 axis parallel to z. Thus we conclude that the $mm2$ space group possibility is $Pbc2_1$. On the other hand, it is possible that the point group is mmm, in which case the missing symmetry operation involving the z-axis is a mirror plane parallel to (001), and the space group will be $Pbcm$. For the moment we are unable to demonstrate which of the two space groups will be the correct one for the crystal, but at least we have managed to reduce the choice to only two out of the complete set of 230 space groups.

It would be possible to proceed now with this limited uncertainty, but there are ways to resolve some of these ambiguities. Measurements of physical properties, such as of piezoelectric coefficients (Chapter 7), may give an independent determination of the presence or absence of a centre of symmetry. If appropriate, it will also be possible to use anomalous scattering (Section 6.8.1) to determine whether or not the crystal has a centre of symmetry. However, it turns out that within the diffraction data there is an internal check on the presence of a centre of symmetry. In order to make use of such a check, it is necessary to attempt to convert the measured structure factors to an absolute scale. Clearly the measured intensity is partly determined by the size of the crystal and the length of time of the measurement (there are other factors such as absorbsion to take into account, but these are details beyond our interest here). Formally we can write the intensity as $I(\mathbf{h}) = s|F(\mathbf{h})|^2$, where s is a scale factor. A method for estimating both the value of s *and* the overall temperature factor is described in Appendix I.

It proves useful to define the **normalized structure factors**

$$|E(\mathbf{h})|^2 = \frac{|F(\mathbf{h})|^2 \exp(2B \sin\theta/\lambda)}{\sum_j f_j^2} \tag{6.56}$$

The normalized structure factors approximate the structure factors of the crystal containing point atoms at rest, each containing the equivalent of one electron. We recall that the crystal structure is defined as the convolution of the

δ-functions defining the positions of the atoms with the shapes of the atoms. By dividing out the temperature factor and the atomic scattering factor we are effectively deconvoluting out the size of the atom (that is, the spread of electron distributions about each atom position, and the distribution of atom positions due to thermal motions) from the Fourier transform of the normalized structure factors. Statistical analysis, called **Wilson statistics**, shows that when the normalized structure factors are averaged over all reflections,

$$\langle E^2 \rangle = 1 \tag{6.57}$$

For centrosymmetric structures, Wilson statistics give

$$\langle E \rangle = \sqrt{2/\pi} = 0.798 \tag{6.58}$$

$$\langle |E^2 - 1| \rangle = 0.968 \tag{6.59}$$

For acentric crystals, the same analysis gives

$$\langle E \rangle = \sqrt{\pi}/2 = 0.886 \tag{6.60}$$

$$\langle |E^2 - 1| \rangle = 0.736 \tag{6.61}$$

In fact, the distribution of the values of E has been explicitly given for both the centrosymmetric and acentric cases. With good data, checked by forming the average $\langle E^2 \rangle$, calculations of the averages $\langle E \rangle$ and $\langle |E^2 - 1| \rangle$ can help decide whether the crystal has a centre of symmetry.

6.9 Solution of the phase problem and determination of crystal structure

6.9.1 The origin of the phase problem

We noted the core of the phase problem earlier in this chapter: the experimental measurements of the diffracted intensities yield the magnitude of the structure factor but not its phase. It is not possible to measure the phase of the structure factor, but the phase is essential if the structure factors are to be used to deduce the crystal structure. Now we turn to address this problem head on.

6.9.2 Historical review of attempts to bypass the phase problem

The traditional idea has been that if the phases of the most intense reflections could be deduced somehow, then these could be used in a Fourier synthesis of the structure. These phases could only be deduced by estimating the positions of some of the atoms, particularly those that scatter most strongly. One technique was to form a three-dimensional representation of $g(\mathbf{r})$ (Sections 2.7.2 and 6.6) from the measured intensities, the so-called **Patterson method**. In this case $g(\mathbf{r})$ gives a set of peaks at positions which correspond to the interatomic vectors. By careful manipulation of $g(\mathbf{r})$, the coordinates of the heaviest atoms can be deduced, particularly when taking account of symmetry. The structure factors can be calculated using only these few atoms, and the phases of the strongest

reflections can be taken to apply to the measured structure factors. The structure (electron density) that is obtained by the initial Fourier synthesis may not be complete, but it may show some new atoms. These can then be used to improve the calculation of the phases, bringing in some of the weaker reflections, and the procedure repeated until the positions of all the atoms have been deduced.

6.9.3 Direct methods to overcome the phase problem

Although we cannot measure the phases of the structure reflections, it transpires that some information on the phases is contained in the magnitudes of the structure factors. The development of methods to extract the phases from the magnitudes was one of the great recent breakthroughs in modern crystallography, allowing the solution of crystal structures to become less of an art and more of a routine exercise.

First we consider a simple demonstration. For a centrosymmetric crystal subject to the (trivial) condition that the electron density is positive everywhere, it can be shown that pairs of reflections are bound by the inequality

$$2\left|\frac{F(\mathbf{h})}{F(000)}\right|^2 \leq 1 + \frac{F(2\mathbf{h})}{F(000)} \tag{6.62}$$

It should be noted that in all cases $F(hkl) \leq F(000)$. If $|F(\mathbf{h})/F(000)|^2 > 1/2$, the phase of $F(2\mathbf{h})$ must necessarily be positive. Other inequalities can also be defined arising from other symmetry operations. In practice the number of strong reflections will be small, so that these inequalities are of limited practical usefulness. Nevertheless, the recognition of the existence of these inequalities gave the first indication that some useful information about phase values is contained within the magnitudes of the structure factors.

One of the important results to follow was **Sayre's equation**. In its simplest form for centrosymmetric crystals, if $s(\mathbf{Q})$ is the phase of a given reflection ($= \pm 1$), the phases of three reflections are related by the simple product

$$s(\mathbf{h}_1) \times s(\mathbf{h}_2) \times s(\mathbf{h}_3) \approx +1 \tag{6.63}$$

where the three scattering vectors are related by $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$ (i.e. $h_1 + h_2 = -h_3$ etc.), and the symbol $\approx$ denotes 'probably equal to'. This probability is a function of the magnitude of the E-values of the three reflections, and is higher for larger values of E. Thus Sayre's equation properly holds only for the strongest reflections. This may be enough to generate the phases of a useful number of reflections, at which stage a stronger version of Sayre's equation can be used:

$$s(\mathbf{h}) \approx \sum_{\mathbf{h}'} s(\mathbf{h}') \times s(\mathbf{h} - \mathbf{h}') \tag{6.64}$$

The probability of this relationship is higher than when using a single triplet, so that some of the less intense reflections can be used.

In practice, if a small number of initial phases can be estimated, it is possible to use Sayre's equation to generate the phases of further reflections. By repeated iteration, the phases of a sufficient number of reflections can be generated in order to produce a Fourier synthesis that contains all the atoms. This method

has been developed so that computer programs can assign phases more-or-less automatically in straightforward cases, even for acentric crystal structures.

Once the phases of an adequate number of reflections have been determined, an approximate electron density map can be produced by Fourier synthesis (eqn 6.40):

$$\rho(x, y, z) = \sum_{h,k,l} F(hkl)\exp[-2\pi i(hx + ky + lz)] \tag{6.65}$$

where the range of x, y, and z cover the unit cell. The peaks in $\rho(x, y, z)$ will correspond to the atoms in the unit cell. If the summation is not complete, the positions of only some of the atoms may be determined this way, but these may be enough to calculate the phases of other reflections, which can then be added to the Fourier synthesis. An example is shown in Fig. 6.21.

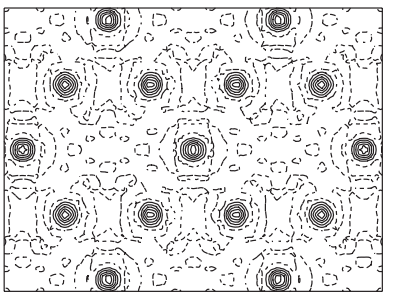


Fig. 6.21 Electron density map of the mineral bobkingite, $Cu_5Cl_2(OH)_8{\cdot}2H_2O$, in the **a**–**b** plane. The peaks in the contour map correspond to two of the symmetrically distinct copper atoms. (Picture kindly provided by Mark Cooper, University of Manitoba, and Mark Welch, Natural History Museum, London. The structure has been reported by Hawthorne *et al.*, *Min. Mag.* **66**, 301, 2002.)

The end result of this process is a set of atomic coordinates. No attempt has been made to pinpoint the positions with any precision, and this is the task we now address. Precise coordinates are necessary for the determination of bond lengths. We may also be interested in the amplitudes of thermal vibrations.

6.9.4 Refinement of the crystal structure

When the approximate positions of the atoms in a crystal structure have been deduced, the final stage is to obtain greater accuracy on the results, together with good values for the temperature factors. It is common to use a least-squares method to improve on the accuracy of a structure once the main features have been correctly deduced.

The quality of a structure determination is quantified by comparing the calculated and measured structure factors. Crystallographers commonly use a quantity called the R-factor, defined as

$$R = 100 \times \frac{\sum_{hkl} ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)||}{\sum_{hkl} |F_{\text{obs}}(hkl)|}\% \tag{6.66}$$

By comparing the moduli of the structure factors any problem with the phases is avoided: the comparison is directly with the experimental quantities. Typically the value of R for a refined structure will be of order 2–4% for X-ray single-crystal data.

In order to refine the structure, the structural variables (atomic fractional coordinates and temperature factors) are adjusted to minimize the quantity

$$M = \sum_{hkl} ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)||\,/\sigma^2(hkl) \tag{6.67}$$

where $\sigma(hkl)$ is the estimated error on $F_{\text{obs}}(hkl)$. Computer algorithms are available to change the values of the structural parameters in order to bring the values of $F_{\text{calc}}(hkl)$ closer to the values of $F_{\text{obs}}(hkl)$. The temperature factors can be treated as anisotropic. It is also possible to incorporate effects that cause modification of the measurements of the structure factors, such as absorption corrections, which can also be adjusted in the refinement procedure.

The quality of the final result always depends on the quality of the data, but typically absolute atomic positions can be refined to within 0.01 Å or better. Thus bond lengths can be determined with reasonable precision.

Summary of chapter

- **Bragg's law** relates the angle through which beams of radiation are diffracted to the interplanar spacing.
- The main types of radiation used for the determination of crystal structures are **X-rays** and **neutrons**. Beams of X-rays can be produced in the laboratory from tubes in which electrons strike metal targets. Beams of electromagnetic radiation within the X-ray spectrum can also be produced as **synchrotron radiation** in an electron accelerator. Beams of neutrons are produced in nuclear reactors or as a result of a beam of protons striking a heavy metal target.
- It is also possible to use beams of electrons in diffraction. This is a widespread technique for imaging with a resolution at a near-atomic scale, but is rarely used for structure determination. Instead it is mostly used for imaging lattice defects.
- Diffraction can be performed using single-crystal or powdered samples.
- The intensity of a diffracted beam is determined by the interference caused by scattering from individual atoms. The intensity provides information about the separations of all pairs of atoms. The formalism is developed in terms of the structure factor, with the intensity being given by the square of the modulus of the structure factor. The **structure factor** is the **Fourier transform** of the crystal structure, and therefore contains information about the positions of atoms rather than the separations of pairs of atoms. The reverse Fourier transform of the complete set of structure factors gives the distribution of electron density of the crystal for X-ray diffraction, and the distribution of nuclear positions for neutron diffraction. However, the phase of the structure factor cannot be measured directly in a diffraction experiment.
- The **scattering vector** $\mathbf{Q}$ is defined as the change in wave vector of the beam of radiation as a result of diffraction (or any scattering process). The formalism for scattering of beams of radiation can be completely described in terms of $\mathbf{Q}$ rather than the separate wave vectors of the incident and scattered beams. Bragg scattering only occurs when $\mathbf{Q}$ is equal to a reciprocal lattice vector.
- The scattering of X-rays from a single atom is described by the Fourier transform of the electron density. This function is called the **atomic scattering factor**. The atomic scattering factor for $Q = 0$ is equal to the number of electrons. The width of the atomic scattering factor in Q is inversely proportional to the size of the atom.
- The scattering of neutrons from atomic nuclei is virtually independent of the scattering vector because the size of the nucleus is around 10^5 times smaller than the wavelength used for diffraction. The amplitude is called the **scattering length**.
- The crystal structure can be described as the convolution of the lattice with the basis, the electrons, and the thermal motion of the atoms. The Fourier transform is therefore given as the product of the Fourier transforms of the individual components of the convolution.

- The distribution of atomic displacements due to harmonic lattice vibrations is described by a Gaussian function. The Fourier transform is also a Gaussian, with width determined by the size of the mean-square atomic displacement. This is described empirically by the **temperature factor**.
- Some aspects of the symmetry of the crystal structure can be determined by the symmetry of the diffraction pattern. The point symmetry of the diffraction pattern is given by the point symmetry of the structure plus a centre of symmetry, the **Laue symmetry**. Thus it is possible to obtain the Bravais lattice symmetry from the diffraction pattern, and some limited information about the possible point group symmetry. **Systematic absences** provide direct information on the presence of screw axes and glide planes.
- The rule that diffraction patterns have a centre of symmetry can be broken for non-centrosymmetric crystals if the energy of the X-ray beam is close to an absorption resonance energy of one of the electrons.
- The presence of a centre of symmetry can be obtained by statistical analysis of measured structure factors when scaled by the atomic scattering factors and temperature factors. These scaled structure factors represent the structure factors for point atoms of equal weights. This analysis is called **Wilson statistics**.
- Further statistical analysis of the distribution of scaled structure factors can provide information about the phases of the structure factors. This set of procedures is called **direct methods**.
- The crystal structure deduced using direct methods can be refined by adjusting the atomic fractional coordinates and temperature factors to give best agreement between measured and calculated structure factors.

Further reading

Diffraction is discussed in many textbooks, because it is a technique that has a central role in the investigation of crystalline materials. Both Kittel (1996) and Aschcroft and Mermin (1976) give a good introduction to the topic. General texts, which include determination of crystal structure, are Azároff (1974), Clegg (1998), Cullity (2001), Giacovazzo *et al.* (1992), Glusker and Trueblood (1985), Hammond (2001), Kelly *et al.* (2000), Ladd and Palmer (1977), and Woolfson (1997). Direct methods are the primary focus of Schenk (1984). Experimental techniques are discussed in detail by Aslanov *et al.* (1998) and McKie and McKie (1986). Crystallography using synchroton radiation is presented by Coppens (1992), and neutron diffraction is described by Dobryznski and Blinowski (1994), Bacon (1973, 1987), Newport *et al.* (1988), Nield and Keen (2001), and Squires (1978). The Rietveld method, which is used for refinement of crystal structures from powder diffraction data, is discussed is detail in a collection of papers edited by Young (1993). The use of X-ray diffraction to obtain information about the distribution of charge in crystalline materials is described by Coppens (1997). Electron diffraction, which could not be developed within this book, is described by Cowley (1992).

Exercises

(6.1) In a time-of-flight neutron powder diffractometer, a sharp pulse of neutrons with a range of wavelengths is fired at the sample. The diffracted signal is measured at a fixed scattering angle 2θ, and the diffraction pattern comes from measurements of the time taken for the neutrons in a single pulse to travel the distance l from the source to the detector. Using the de Broglie relationship, eqn 6.1 together with the Bragg relationship, eqn 6.2, show that the time t taken by the neutrons to travel the distance l is related to the size of the d-spacing of a given reflection by

$$d = \frac{ht}{2ml \sin\theta}$$

where m is the mass of the neutron.

No quantities are known exactly. For an uncertainty of $\Delta\theta$ on the Bragg angle, error analysis gives the corresponding uncertainty on d:

$$\Delta d = \frac{\partial d}{\partial \theta}\Delta\theta$$

Show by differentiation of the Bragg law that for a given uncertainty $\Delta\theta$ this leads to a maximum resolution of

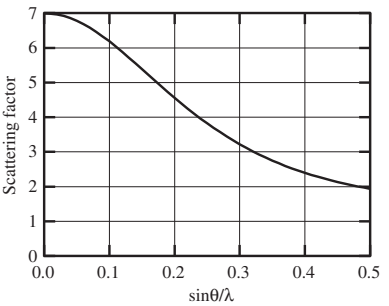
$$\left|\frac{\Delta d}{d}\right| = \cot\theta\,\Delta\theta$$

In order to optimize the resolution of a time-of-flight powder diffractometer, the scattering angle 2θ is chosen to be as close to 180° as possible. In this case a significant source of uncertainty is in the distance travelled by the neutron beam. Show that the uncertainty in the flight path l of Δl leads to a resolution limit of

$$\frac{\Delta d}{d} = \frac{\Delta l}{l}$$

If the uncertainty in the value of l comes from the width of the neutron moderator of 2 cm, calculate the flight paths necessary to achieve (a) a moderate resolution of $\Delta d/d = 10^{-3}$, and (b) a high resolution of $\Delta d/d = 2 \times 10^{-4}$. Comment on your answers, and check out the real situation at the ISIS spallation neutron source by looking at the instruments POLARIS and HRPD from http://www.isis.rl.ac.uk.

(6.2) The atomic scattering factor for nitrogen is shown in the figure below:

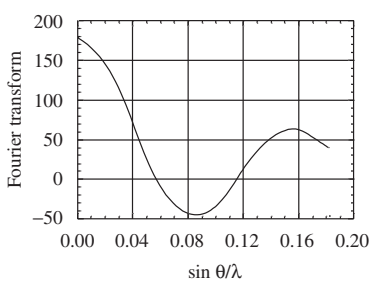


Consider two nitrogen atoms at positions $\mathbf{r} = \pm 0.5$, 0, 0 Å, with an incident beam of Cu Kα X-rays along the vector direction [1, 0, 0]. For scattering of the beam in the x–y plane, calculate the modulus of the scattering vector Q, and hence form the vector $\mathbf{Q}$ for the beams scattered at angles of (a) 40° and (b) 80° to in the x–y plane. Use the atomic scattering factors for these two values of Q together with a convolution approach to calculate the scattering function $F(\mathbf{Q})$ for the pair of atoms for each of the two scattering vectors, by first calculating $F(\mathbf{Q})$ for the positions of the atoms and then incorporating the scattering from the individual atoms.

(6.3) Sketch a two-dimensional section of reciprocal space showing the area that would be swept out by the Ewald sphere in a powder diffraction experiment. (See Appendix H.)

(6.4) An orthorhombic crystal of lattice parameters $a = 6.3$, $b = 4.7$, $c = 3.5$ Å is rotated about [001] during an X-ray a diffraction experiment performed using Cu $K\alpha$ radiation. Use the Ewald sphere to show which $h00$ and which $h02$ reflections can be measured in such an experiment. (See Appendix H.)

(6.5) Pyrophyllite and muscovite are closely related layer structures in which sheets of composition $Al_2(Si,Al)_4O_{10}(OH)_2$ lie parallel to (001). The figure below gives the Fourier transform of a pyrophyllite-type layer along a line in reciprocal space normal to (001):



In pyrophyllite these sheets are stacked to give a repeat of 9.18 Å along the normal to (001). From the convolution theorem we know that the structure factor of an $00l$ reflection will be equal to the Fourier transform of a layer evaluated at the appropriate value of $\sin\theta/\lambda$. Calculate the structure factors for the first three $00l$ reflections.

In muscovite there are potassium ions (one per formula unit) that lie exactly half-way between two layers, which now stack every 9.99 Å. Deduce the values of the structure factors of the first three $00l$ reflections for both pyrophyllite and muscovite. (Assume the potassium atom to be a point atom containing 18 electrons.)

(6.6) Measurements for copper metal give the mean-squared displacements of the atoms from their average positions as 0.021 $Å^2$ at 293 K, and 0.096 $Å^2$ at 1273 K. Copper has a cubic close-packed structure with $a = 3.615$ Å. Calculate the ratios of the intensities at the two temperatures of (a) the lowest-angle and (b) highest-angle reflections that can be measured with Cu Kα radiation.

(6.7) From the following sets of systematic absences in the diffraction patterns of orthorhombic crystals, determine the possible space groups (noting that in some cases there may be alternative possibilities, and also noting that in some cases one set of systematic absences implies another set):

	(a)	(b)	(c)	(d)
$0kl$:	k odd	l odd	—	—
$h0l$:	h odd	l odd	$h+l$ odd	h odd
$hk0$:	—	$h+k$ odd	—	—
$h00$:	h odd	h odd	h odd	h odd
$0k0$:	k odd	k odd	—	—
$00l$:	—	l odd	l odd	—

(6.8) List the systematic absences for the following orthorhombic space groups, and in each case identify any other space groups that might have the same set of systematic absences:

(a) $Pcca$; (b) $Pmmn$; (c) $Pnn2$; (d) $Pmn2_1$

7 Physical properties

7.1 Overview

7.1.1 Crystal anisotropy

We are used to thinking about the physical properties of materials in terms of scalar quantities, such as volume and energy. These will vary as the external conditions are changed. Simple examples are the coefficient of volume expansion,

$$\beta = \frac{1}{V}\frac{\partial V}{\partial T} \tag{7.1}$$

or the compressibility,

$$K = \frac{1}{V}\frac{\partial V}{\partial P} \tag{7.2}$$

We also understand that some properties can have a vector form. For example, the magnetization of a ferromagnetic material is a vector quantity. Moreover, some external stimuli have the nature as vectors and give a response that can be described as a vector. An example is an electric field, **E**, applied to a dielectric material giving an induced polarization field, **P**. For an **isotropic** material (quite what we mean by 'isotropic' can remain ill-defined for the moment – it is sufficient to take the word at face value as implying no dependence of properties on direction) the direction of the response vector will be the same as that of the stimulus. Thus we might define a susceptibility χ as a simple scalar:

$$\frac{1}{\epsilon_0}\mathbf{P} = \chi\mathbf{E} \tag{7.3}$$

For many technological applications it is necessary to generalize this trivial approach to more complicated properties and to consider the effects of **anisotropy**, namely the dependence of properties on direction. Following our example, in an anisotropic material the application of a field **E** may result in a polarization **P** that is not parallel to **E**. Thus we might write the components of the vector $\mathbf{P} = (P_1, P_2, P_3)$ piecewise in terms of the components of $\mathbf{E} = (E_1, E_2, E_3)$:

$$\frac{1}{\varepsilon_0}\mathbf{P} = \boldsymbol{\chi}\cdot\mathbf{E} \;\Rightarrow\; \frac{1}{\epsilon_0}P_i = \sum_{i,j}\chi_{ij}E_j \tag{7.4}$$

The susceptibility χ now has the form of a 3×3 matrix. Clearly now the response **P** depends on the direction of **E** in both its direction and magnitude, and the matrix of numbers that form the susceptibility determine the form of this dependence on direction.

So far the extension from the isotropic case seems relatively easy. Too easy in fact, because if we are not careful we will quickly end up in the wrong direction. An example of thermal expansion will demonstrate where we are heading. Consider a hexagonal material that expands along the [001] direction on heating but shrinks along the ⟨100⟩ directions – this is not as unlikely as it might appear. You can imagine that it should be possible to cut the crystal into the shape of a plate with an orientation relative to the crystal axes such that there is an exact balance between the positive and negative contributions of the thermal expansion to give zero thermal expansion in the direction normal to the plate. This situation is illustrated in Fig. 7.1. The crystal axes, which define the main symmetry directions, are shown as the main axes, and the plate with zero thermal expansion normal to its face is shown in its appropriate orientation. To calculate the orientation of the cut faces of the plate relative to the crystal axes is not quite as trivial as you might first think. The proper description of thermal expansion is in terms of the concept of **strain**, which describes the change in both the size and shape of an object. To calculate how to cut a plate that does not expand on heating, we need to rotate the components of the thermal expansion, which means rotating the full mathematical form of the strain. This operation is not as simple as the rotation of a vector, although there are some common elements. In fact this type of problem is quite common, and the purpose of this chapter is to introduce the mathematical tools that have been developed to handle it. We will see that these tools are fairly powerful, and can be extended to determine the effects of symmetry on the physical properties of a crystal.

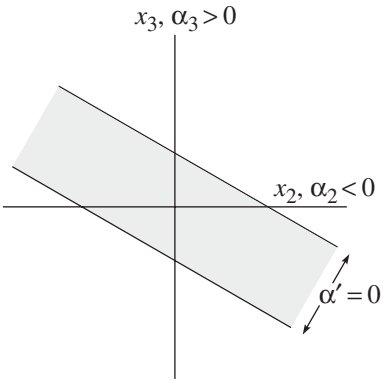


Fig. 7.1 Orientation of crystal plate with negative thermal expansion in one direction chosen so that the expansion of the plate is exactly zero.

7.1.2 An introduction to tensors

The mathematical form of the anisotropic properties of crystals is described using **tensors**. Vectors, such as those used to describe the dielectric polarization, and matrices, such as those used to describe the susceptibility, are special forms of tensors, and we will quickly meet higher-order versions. Tensors are rather more than mere collections of numbers. Integral to the definition of a tensor is the ability to change the definition of the coordinate system, the operation that we encountered in the example of the crystal plate with zero net thermal expansion. Since the physical properties of a crystal do not depend on our choice of axes, we expect the transformation tools for the rotation of the axes to leave the physics intact. A significant part of this chapter is to understand how to apply such transformations.

The vector tensors such as polarization are called **first-rank tensors**, and the matrix tensors such as suceptibility are called **second-rank tensors**. We will show later that strain is also described as a second-rank tensor. Higher-rank tensors will be used to describe how the action of an applied stimulus will produce a change in the physical properties of a crystal. The susceptibility tensor, eqn 7.4, is a simple example.

7.1.3 Field and matter tensors

Implicit in the discussion so far is the fact that some tensor quantities, such as the applied field, are controlled by the person setting up an experiment and are independent of the crystal. It is convenient to treat these tensors as special quantities, which are called **field tensors**. On the other hand, there are some tensor quantities that are properties of the material, such as the coefficients of thermal expansion and dielectric susceptibility, and these are called **matter tensors**. In the formalism it often appears as if the matter tensors are constants of the material. However, we will meet examples of how the values of the matter tensors can be modified by application of a field tensor quantity (Section 7.6). This will be particularly noticeable if the application of an external field changes the symmetry of the crystal, because the form of the matter tensors are determined by symmetry.

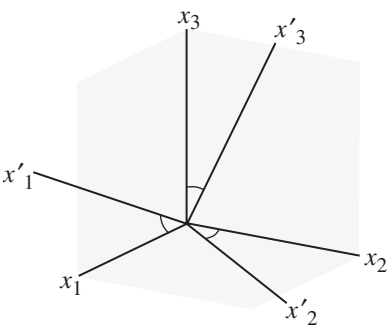


Fig. 7.2 Two sets of orthogonal axes oriented with respect to each other, as used in the transformation equations.

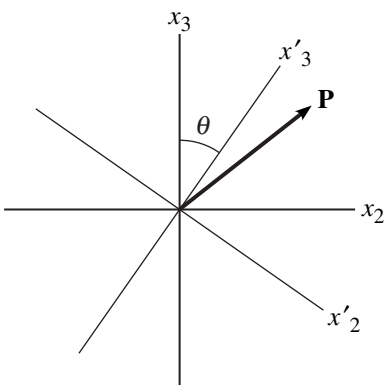


Fig. 7.3 Schematic rotation of axes in two dimensions. θ is the angle between the x_3 and x'_3 axes.

7.2 First-rank tensors

A first-rank tensor transforms as a vector. Consider the two sets of axes shown in Fig. 7.2, where for the convenience of notation later we label the x, y, z axes of one set of axes as x_1, x_2, x_3, and of the second set of axes as x'_1, x'_2, x'_3. We define the matrix of direction cosines as $a_{ij} = \cos(x'_i : x_j)$. This is more easily pictured for the two-dimensional case shown in Fig. 7.3. We define $\theta = x'_3 : x_3 = x'_2 : x_2$, so that $x'_2 : x_3 = 90° + \theta$ and $x'_3 : x_2 = 90° - \theta$. Thus the **matrix of direction cosines** can be written as

$$\mathbf{a} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \tag{7.5}$$

Note that $a_{ij} = -a_{ji}$ for $i \neq j$, which will be true in general for rotation matrices. It is left as a problem at the end of the chapter (Problem 7.1) to show that for the general case $\mathbf{a}^{-1} = \mathbf{a}^{\mathrm{T}}$; eqn 7.5 is clearly consistent with this.

If we have a vector **P** in the first set of axes, this is defined as **P′** in the second set of axes, and we have the normal vector transformations:

$$\begin{aligned} &\mathbf{P}' = \mathbf{a} \cdot \mathbf{P}; \quad \mathbf{P} = \mathbf{a}^{-1} \cdot \mathbf{P}' = \mathbf{a}^{\mathrm{T}} \cdot \mathbf{P}' \\ &\Rightarrow P'_i = \sum_j a_{ij} P_j; \quad P_i = \sum_j a_{ji} P'_j \end{aligned} \tag{7.6}$$

We have already met two first-rank tensors, namely the applied electric field **E** and the polarization **P**. We also have the magnetic analogues **B** and **M** respectively. Other first-rank tensors are heat flow, temperature gradient, and electric current.

7.3 Second-rank tensors

7.3.1 Basic ideas

We met the susceptibility tensor as a second-rank tensor in eqn 7.4. In fact whenever one quantity described by a first-rank tensor depends on another

first-rank tensor quantity, such as when **P** is obtained by application of an external **E**, the link between the two is described by a second-rank tensor. Before we explore the inner wonders of second-rank tensors in more detail, we deal first with the methods used to rotate a second-rank tensor. We start by considering the following tensor product involving **P** and **Q** as first-rank tensors and **T** as a second-rank tensor:

$$\mathbf{P} = \mathbf{T} \cdot \mathbf{Q} \Rightarrow P_i = \sum_j T_{ij} Q_j \tag{7.7}$$

If we have the same tensors in a different orientation of the axes, we have a corresponding equation, $\mathbf{P}' = \mathbf{T}' \cdot \mathbf{Q}'$. From eqn 7.6, $\mathbf{P}' = \mathbf{a} \cdot \mathbf{P}$ and $\mathbf{Q}' = \mathbf{a} \cdot \mathbf{Q}$. The equation that links $\mathbf{T}'$ to **T** is obtained from

$$\begin{aligned} &\mathbf{P}' = \mathbf{a} \cdot \mathbf{P} = \mathbf{a} \cdot (\mathbf{T} \cdot \mathbf{Q}) \\ &\mathbf{P}' = \mathbf{T}' \cdot \mathbf{Q}' = \mathbf{T}' \cdot (\mathbf{a} \cdot \mathbf{Q}) \\ &\quad \Rightarrow \mathbf{a} \cdot \mathbf{T} = \mathbf{T}' \cdot \mathbf{a} \\ &\quad \Rightarrow \mathbf{T}' = \mathbf{a} \cdot \mathbf{T} \cdot \mathbf{a}^{\mathrm{T}} \end{aligned} \tag{7.8}$$

Written out long-hand, this gives

$$T'_{ij} = \sum_{k,l} a_{ik} a_{jl} T_{kl} \tag{7.9}$$

The reverse transformation gives

$$\begin{aligned} &\mathbf{T} = \mathbf{a}^{\mathrm{T}} \cdot \mathbf{T}' \cdot \mathbf{a} \\ &\Rightarrow T_{ij} = \sum_{k,l} a_{ki} a_{lj} T'_{kl} \end{aligned} \tag{7.10}$$

As an example, consider the tetragonal ferroelectric phase of $BaTiO_3$. The susceptibility tensor at room temperature has $\chi_{11} = \chi_{22} = 70$, $\chi_{33} = 250$ (see Fig. 2.31). Suppose a crystal is cut in the form of a plate such that the normal to the plate is in the x–z plane at an angle of θ to [001], as shown in Fig. 7.4, and suppose that an electric field $\mathbf{E}'$ is applied normal to the plate. We can determine the resultant polarization field $\mathbf{P}'$ in either of two ways. The first is to transform $\mathbf{E}'$ into the frame of the principal axes by the following transformation:

$$\mathbf{E} = \begin{pmatrix} a_{11} E'_1 + a_{13} E'_3 \\ E'_2 \\ a_{31} E'_1 + a_{33} E'_3 \end{pmatrix} \tag{7.11}$$

We can then obtain $\mathbf{P} = \epsilon_0 \boldsymbol{\chi} \cdot \mathbf{E}$ directly, and then calculate the polarization in the axes of the crystal plate, $\mathbf{P}'$, by the reverse transformation. However, we could do the whole calculation in the axes of the crystal plate by transforming $\boldsymbol{\chi}$. Application of the transformation equation gives

$$\boldsymbol{\chi}' = \begin{pmatrix} a_{11}^2 \chi_{11} + a_{13}^2 \chi_{33} & 0 & a_{11} a_{31} \chi_{11} + a_{13} a_{33} \chi_{33} \\ 0 & \chi_{22} & 0 \\ a_{31} a_{11} \chi_{11} + a_{33} a_{13} \chi_{33} & 0 & a_{31}^2 \chi_{11} + a_{33}^2 \chi_{33} \end{pmatrix} \tag{7.12}$$

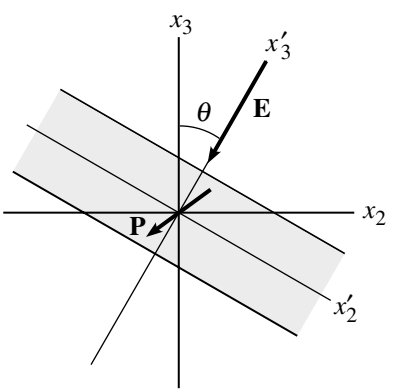


Fig. 7.4 Schematic illustration of an anisotropic dielectric plate, showing how the orientation of the induced **P** is not parallel to the applied **E** if x_2 and x_3 are the principal crystal axes.

Since the applied electric field is of the form

$$\mathbf{E}' = \begin{pmatrix} 0 \\ 0 \\ E'_3 \end{pmatrix} \tag{7.13}$$

we have

$$\mathbf{P}' = \epsilon_0 \boldsymbol{\chi}' \cdot \mathbf{E}' = \epsilon_0 \begin{pmatrix} (a_{11}a_{31}\chi_{11} + a_{13}a_{33}\chi_{33})E'_3 \\ 0 \\ (a_{31}^2\chi_{11} + a_{33}^2\chi_{33})E'_3 \end{pmatrix} \tag{7.14}$$

Clearly there is now a component of the polarization that is normal to the direction of the applied electric field (Problem 7.11).

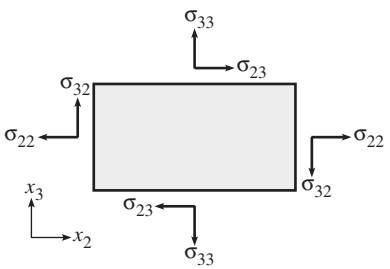


Fig. 7.5 Schematic definition of the full stress tensor. The arrows indicate the direction of the applied force relative to the orientation of the surface.

7.3.2 Stress as a second-rank tensor

Stress arises from applying a force to a surface, and its units are force/area. Force is a vector quantity, and hence is a first-rank tensor. But when you apply force to a surface you have to consider also the relative orientations of the force vector and the normal to the surface. This is illustrated in Fig. 7.5. In the case of an isotropic material, if the force is applied normal to the surface, the result will be a compression of the crystal in the direction of the force. But if the force is applied in a direction parallel to the face of the crystal, the result is a shear of the crystal. To distinguish between these two outcomes, we need a definition of stress that includes both the direction of the force and the direction of the normal to the face the force is acting on. If the two directions are parallel, we have a tensile stress, and if the two directions are orthogonal we have a shear stress. When a force is applied at an arbitrary direction to a surface, we have a mixture of the two types of stress, which can be obtained by separating the force vector into the components normal and parallel to the face. The **stress tensor** is given the symbol $\boldsymbol{\sigma}$.

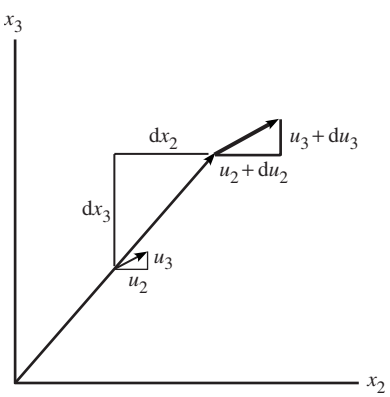


Fig. 7.6 Schematic two-dimensional representation of how a displacement vector (u_2, u_3) for a small element of a body under strain deformation will vary depending on position (x_2, x_3). How the changes in displacements $(\mathrm{d}u_2, \mathrm{d}u_3)$ depend on (x_2, x_3) is described by the strain tensor **e**.

7.3.3 Strain as a second-rank tensor

If stress is a second-rank tensor, we might expect the response, namely the strain, also to be a second-rank tensor. We noted before that strain includes effects on both the size and shape of an object: formally this is both the tensile and shear strain.

Strain is formally defined by the relationship between the displacement of a small element of a solid and its position. This is illustrated for two dimensions in Fig. 7.6, which shows how a general displacement induced by the strain varies according to position. The tensile strain relates the displacement along one direction to the position in the same direction. A small element of the material at initial position x_2, x_3 will be displaced by u_2, u_3, and the element at position $x_2 + \mathrm{d}x_2, x_3 + \mathrm{d}x_3$ will be displaced by $u_2 + \mathrm{d}u_2, u_3 + \mathrm{d}u_3$. The dependence of u_2 on u_3, x_2, and x_3 respectively is described by the general equation for tensile strain:

$$e_{ii} = \frac{\mathrm{d}u_i}{\mathrm{d}x_i} \tag{7.15}$$

Now we consider the shear strain, which links the dependence of u_2 on x_3 (and similarly for other combinations). We write the shear strain as

$$e_{23} = \frac{\mathrm{d}u_2}{\mathrm{d}x_3} \tag{7.16}$$

Clearly we can establish a general definition that encompasses both types of strain:

$$e_{ij} = \frac{\mathrm{d}u_i}{\mathrm{d}x_j} \tag{7.17}$$

There is, however, an additional complication with the shear strains that did not arise with the definition of stress. The problem with the general definition is that a pure rotation of an object will also give rise to a value of e_{ij}. This is illustrated in Fig. 7.7. For a general deformation we have different magnitudes of e_{ij} and e_{ji}, representing a mixture of shear strain and rotation. This problem is resolved by defining a new strain tensor $\boldsymbol{\epsilon}$ as the symmetric part of the tensor $\mathbf{e}$:

$$\epsilon_{ij} = \frac{1}{2}(e_{ij} + e_{ji}) \tag{7.18}$$

and a rotation tensor $\boldsymbol{\omega}$ as the antisymmetric part of $\mathbf{e}$:

$$\omega_{ij} = \frac{1}{2}(e_{ij} - e_{ji}) = -\omega_{ji} \tag{7.19}$$

These two combinations are illustrated in Fig. 7.7. The rotation matrix $\boldsymbol{\omega}$ will not interest us further; the pure shear strain tensor $\boldsymbol{\epsilon}$ contains all the important information.

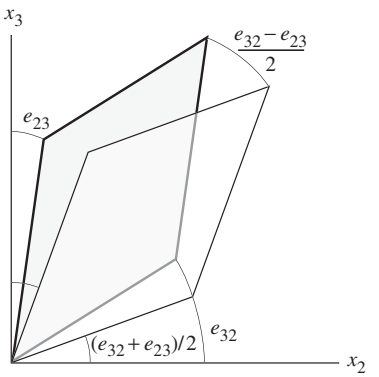


Fig. 7.7 Schematic representation of the relationship between shear strain and rotation. The initial object is a rectangle with edges parallel to the axes. $(e_{32} + e_{23})/2$ defines the shear strain, and $(e_{32} - e_{23})/2$ defines the rotation.

7.3.4 45° rotation of the strain tensor and the conversion between tensile and shear strain

The fact that we can link both tensile and shear strain into a single tensor suggests that the distinctions of the two types of strain may depend on the definition of the coordinate system. This point is illustrated for a 45° rotation of the coordinate system about the x_1-axis, as shown in Fig. 7.8. We have the following transformation matrix:

$$\mathbf{a} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \tag{7.20}$$

A tensor that describes a pure shear strain in the x_2–x_3 plane is

$$\boldsymbol{\epsilon} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \epsilon \\ 0 & \epsilon & 0 \end{pmatrix} \tag{7.21}$$

When we transform the axes, we obtain the new strain tensor:

$$\epsilon'_{22} = a_{22}a_{23}\epsilon_{23} + a_{23}a_{22}\epsilon_{32} \tag{7.22}$$

$$\epsilon'_{33} = a_{32}a_{33}\epsilon_{23} + a_{33}a_{32}\epsilon_{32} = -\epsilon'_{22} \tag{7.23}$$

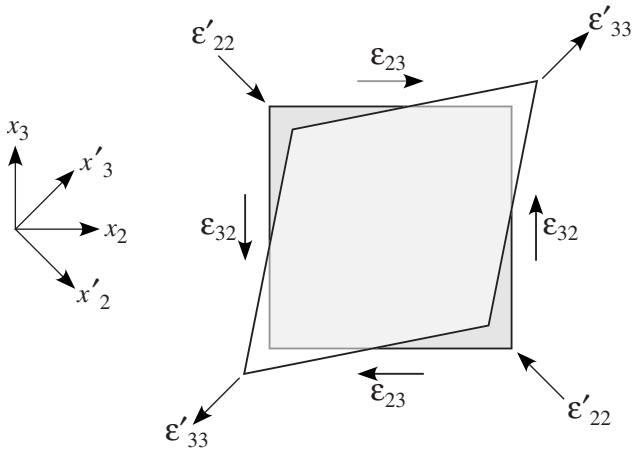


Fig. 7.8 Rotation of a shear strain tensor $\boldsymbol{\varepsilon}$ leads to a tensile strain tensor $\boldsymbol{\varepsilon}'$.

with all other components equal to zero. Thus

$$\boldsymbol{\epsilon}' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & -\epsilon \end{pmatrix} \tag{7.24}$$

What is clear from this result is that the rotation of the axes by 45° has led to the conversion of the definition of the type of strain from pure shear to pure tensile.

We can also remark that if the initial crystal has volume V, the new volume following application of the strain tensor described by eqn 7.24 will be

$$V(1+\varepsilon)(1-\varepsilon) = V(1-\varepsilon^2).$$

This demonstrates that a pure shear strain gives only a quadratic change in volume.

7.3.5 Voigt notation

It will prove to be useful to simplify the notation used in the indices. In what is called the **Voigt notation**, the pairs of indices 11, 22, 33 are replaced by the single indices 1, 2, and 3 respectively, and the pairs of indices 23, 13, and 12 are replaced by the single indices 4, 5, and 6 respectively. Thus the strain tensor can be represented by

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{pmatrix} = \begin{pmatrix} \epsilon_1 & \epsilon_6 & \epsilon_5 \\ \epsilon_6 & \epsilon_2 & \epsilon_4 \\ \epsilon_5 & \epsilon_4 & \epsilon_3 \end{pmatrix} \equiv \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix} \tag{7.25}$$

This notation makes the labelling easier, and makes the representation of higher-rank tensors considerably easier. For example, the tensor linking strain and

stress is the fourth-rank **elastic constant tensor**. The Voigt notation reduces the number of elements from $3^4 = 81$ to $6 \times 6 = 36$. We will shortly meet some complicating rules that are implied by the use of the Voigt notation (Section 7.4.2), and we have to concede that it is much easier not to use the Voigt notation when rotating tensors (Section 7.4.3), but the ease of representation has led to the near-universal use of the Voigt notation.

7.3.6 Principal axes

Implicit in what we have said so far is that the second-rank matter tensors are symmetric. This arises from the fact that there is for any crystal a set of orthogonal axes for which a matter tensor will be diagonal. These are called the **principal axes**. A vector field applied along one of the principal axes will produce a vector field along the same direction. In tetragonal, hexagonal, and trigonal crystals, one of the principal axes is along [001] (the perpendicular plane is isotropic, so any orientation of the other two axes is equivalent to any other). In orthorhombic systems the principal axes are parallel to the three lattice vectors. For monoclinic crystals one of the principal axes is parallel to [010], with the other two axes lying somewhere in the **a**–**c** plane.

Now we consider the rotation of a diagonal second-rank tensor. We take the following two components:

$$T'_{ij} = \sum_{kl} a_{ik} a_{jl} T_{kl} = \sum_{k} a_{ik} a_{jk} T_{kk} \tag{7.26}$$

$$T'_{ji} = \sum_{kl} a_{jk} a_{il} T_{kl} = \sum_{k} a_{jk} a_{ik} T_{kk} \tag{7.27}$$

Clearly $T'_{ij} = T'_{ji}$, which demonstrates that the second-rank matter tensor will be symmetric for all orientations of the axes.

It is useful to be able to obtain the orientation of the principal axes of a second-rank tensor. We illustrate the rotation involved with an example that will require rotation about only one axis:

$$\mathbf{T} = \begin{pmatrix} T_{11} & 0 & T_{13} \\ 0 & T_{22} & 0 \\ T_{31} & 0 & T_{33} \end{pmatrix} \tag{7.28}$$

To reduce this tensor to its principal axes requires rotation about y. The transformation matrix is

$$\mathbf{a} = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \tag{7.29}$$

When the tensor has been transformed to its principal axes, $T'_{13} = 0$. This allows us to write (noting that $T_{31} = T_{13}$)

$$\begin{aligned} T'_{13} = 0 &= a_{11}a_{31}T_{11} + a_{11}a_{33}T_{13} + a_{13}a_{31}T_{31} + a_{13}a_{33}T_{33} \\ &= \cos\theta \sin\theta (T_{33} - T_{11}) + (\cos^2\theta - \sin^2\theta) T_{13} \\ &= \frac{1}{2} \sin 2\theta (T_{33} - T_{11}) + \cos 2\theta T_{13} \\ \Rightarrow \tan 2\theta &= \frac{2T_{13}}{T_{11} - T_{33}} \end{aligned} \tag{7.30}$$

From this equation, the rotation angle θ can be obtained, from which the full transformation matrix and hence the principal components of the tensor can be determined. In the case where we need to find three angles, the geometric method is not practical. In this case, the principal components are equivalent to the eigenvalues of the second-rank tensor, and the transformation matrix is the corresponding set of eigenvectors.

7.3.7 Symmetry and second-rank matter tensors

It is straightforward to appreciate that symmetry will lead to some components of second-rank matter tensors being equal when in the setting of the principal axes. For example, for cubic symmetry the three diagonal components will be equal. For tetragonal, hexagonal, and trigonal symmetry, the matter tensor components have $T_{11} = T_{22} \neq T_{33}$.

Symmetry plays an important role in determining whether some of the components of a matter tensor should be zero. We consider two examples. The first is the case of a 2-fold rotation axis along y. The relationship between atomic coordinate $x, y, z \rightarrow \overline{x}, y, \overline{z}$ has the corresponding transformation matrix:

$$\mathbf{a} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{7.31}$$

The symmetry requires that the transformed tensor $\mathbf{T}'$ must be identical to the initial tensor $\mathbf{T}$. The transformed tensor is obtained as

$$\mathbf{T}' = \begin{pmatrix} T_{11} & -T_{12} & T_{13} \\ -T_{21} & T_{22} & -T_{23} \\ T_{31} & -T_{23} & T_{33} \end{pmatrix} \tag{7.32}$$

The solution to $T'_{12} = T_{12} = -T_{12}$ is $T_{12} = 0$. Similarly, $T_{23} = 0$. On the other hand, since $T'_{13} = T_{13}$, this component of the tensor can have a non-zero value. The same result will follow for the transformation of a mirror plane normal to [010]. This simple analysis shows how a symmetry transformation can be used to determine which coefficients of a matter tensor must have a value of zero and which are allowed to have non-zero values.

7.3.8 Example of zero thermal expansion

We now return to our initial example of a material that has both positive and negative values of the coefficient of the thermal expansion tensor. Since thermal expansion involves the strain tensor, the thermal expansion tensor α is second-rank, and for a scalar change in temperature ΔT we have

$$\boldsymbol{\epsilon} = \boldsymbol{\alpha} \cdot \Delta T \tag{7.33}$$

Now we consider a uniaxial material in which $\alpha_{11} = \alpha_{22} < 0$ and $\alpha_{33} > 0$, the example shown in Fig. 7.1. If we define the direction normal to the plate to be x'_3, we have

$$\alpha'_{33} = a_{32}^2 \alpha_{22} + a_{33}^2 \alpha_{33} = 0 \tag{7.34}$$

We now need to find a value for the rotation angle such that $\alpha'_{33} = 0$. Substituting the trigonometric values of the rotation matrix elements a_{ij}, we obtain

$$\tan^2\theta = -\alpha_{33}/\alpha_{22} \tag{7.35}$$

Now we consider the other strain components. These are given as

$$\alpha'_{22} = a_{22}^2\alpha_{22} + a_{23}^2\alpha_{33} \tag{7.36}$$
$$\alpha'_{23} = a_{22}a_{32}\alpha_{22} + a_{23}a_{33}\alpha_{33} \tag{7.37}$$

Not only is there an expansion of the plate in the plane of the plate, but there is also a shear strain in the plane of the plate induced by the temperature change.

7.4 Third-rank tensors

7.4.1 Piezoelectricity

The definition of the piezoelectric tensor

Arguably one of the most important third-rank tensors for 21st-century applications is the **piezoelectric tensor**, d_{ijk}. This links an induced polarization to an applied stress (the **direct** piezoelectric effect), or an induced strain to an applied electric field (the **indirect** piezoelectric effect):

$$P_i = \sum_{jk} d_{ijk}\sigma_{jk} \tag{7.38}$$
$$\epsilon_{jk} = \sum_{k} d_{ijk}E_i \tag{7.39}$$

The equivalence of these forms can be obtained by a thermodynamic argument. We write the internal energy as

$$\mathrm{d}U = \sum_{jk}\sigma_{jk}\mathrm{d}\epsilon_{jk} + \sum_{i} E_i\mathrm{d}P_i \tag{7.40}$$

The free energy function, which we denote as Φ, can be written as

$$\Phi = U - \sum_{jk}\sigma_{jk}\epsilon_{jk} - \sum_{i} E_i P_i \tag{7.41}$$

This can be written in differential form as

$$\mathrm{d}\Phi = -\sum_{jk}\epsilon_{jk}\mathrm{d}\sigma_{jk} - \sum_{i} P_i\mathrm{d}E_i \tag{7.42}$$

Thus we have the following two differentials:

$$\left.\frac{\partial\Phi}{\partial\sigma_{jk}}\right|_E = -\epsilon_{jk} \tag{7.43}$$
$$\left.\frac{\partial\Phi}{\partial E_i}\right|_\sigma = -P_i \tag{7.44}$$

From the second differential we have

$$-\frac{\partial^2 \Phi}{\partial E_i \partial \sigma_{jk}} = \frac{\partial \epsilon_{jk}}{\partial E_i} = \frac{\partial P_i}{\partial \sigma_{jk}} = d_{ijk} \tag{7.45}$$

Both first differentials relate to the piezoelectric effect, as the indirect and direct effect respectively. This demonstrates that the same coefficient applies to both effects.

Symmetry and piezoelectricity: example of the tetragonal ferroelectric phase of perovskite

Symmetry plays an important role in determining the piezoelectric properties of a material. Consider the direct piezoelectric effect. The simplest example is the role of a centre of symmetry. An applied stress will not remove the centre (stress always has a symmetry corresponding to a centrosymmetric point group), and so application of stress by itself cannot create an electrical polarization in a centrosymmetric crystal. Therefore the first symmetry condition that must be met for a piezoelectric material is that it will not have a centre of symmetry. There are further symmetry conditions on the type of stress that can induce an electrical polarization, and on the direction of that polarization.

The role of symmetry can be illustrated by the example of the piezoelectric effect given by the ferroelectric tetragonal phase of a perovskite material such as $BaTiO_3$ or $PbTiO_3$, which is shown schematically in Fig. 7.9. The symmetry of the crystal is point group $4mm$, that is, there is a 4-fold rotation axis along [001], and mirror planes perpendicular to the equivalent $\langle 100 \rangle$ and $\langle 110 \rangle$ directions. Because of the symmetry of the ferroelectric phase there is already an electrical polarization along [001] in the unstressed crystal, together with an elongation of the crystal in this direction. Application of the stress σ_3, that is a force along [001] applied on the (001) face, will reduce the extent of the elongation along the [001] direction. This will squash the TiO_6 octahedra and force the Ti^{4+} cations back towards the centre of the octahedra. Similarly, there will be a compression of the oxygen atoms around the Ba/Pb sites, forcing the cation to move towards the centre of the 12 oxygens. The effects of these induced cation displacements will be to reduce the overall electrical polarization along $x_3 = [001]$.

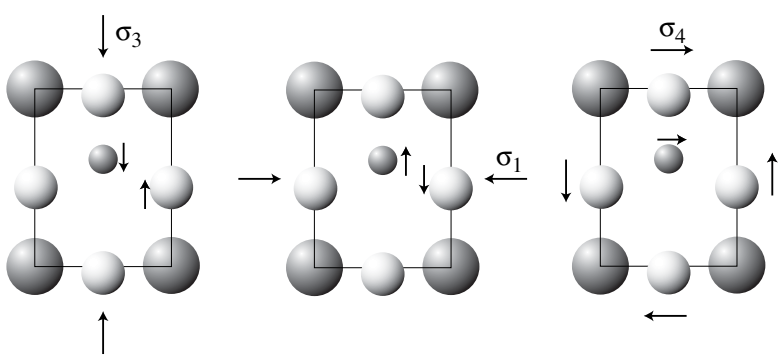


Fig. 7.9 Schematic representation of the piezoelectric effect in a ferroelectric perovskite of point group symmetry $4mm$. [001] (x_3) is vertical, and [100] (x_1) is horizontal.

Now consider application of the stress σ_1 (because of the 4-fold rotational symmetry the same arguments here will also apply to σ_2). There is a mirror plane normal to the face on which the force is applied, and the stress is not able to change this symmetry. Therefore there will be no induced electrical polarization along the [100] direction. However, by squashing together the oxygen atoms in the TiO_6 octahedra, the Ti^{4+} cation will be further displaced along [001], and so an additional contribution to the electrical polarization along [001] will be induced.

We now consider the effects of an applied shear stress. First we consider σ_6, which will cause a shear of the **a**–**b** plane (not illustrated in Fig. 7.9). This will have the effect of moving two of the oxygens in the TiO_6 octahedra closer together, and of moving two others further apart. This shear does not destroy the orthogonal mirror planes on the {110} planes, so there can be no polarization in the **a**–**b** plane. To first order, the changes in the distances between the two pairs of oxygen atoms in the **a**–**b** plane along the ⟨110⟩ directions will be of equal size but opposite sign, so there will no net effect on the position of the Ti^{4+} cation to first order (although there may be a second-order effect that would be represented by a tensor of higher order). Thus σ_6 is not able to produce an induced electrical polarization. On the other hand σ_5, which causes a shear on the **a**–**c** plane, removes all mirror planes other than the (010) mirror plane, and therefore a polarization is allowed on a general direction in the **a**–**c** plane.

Writing everything out full, we can state the results of this discussion in matrix form:

$$\begin{pmatrix} P_1 \\ P_2 \\ \Delta P_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} \tag{7.46}$$

where we have noted that $d_{32} = d_{31}$ and $d_{24} = d_{15}$, and we have explicitly noted that P_1 and P_2 are new, but that P_3 changes by ΔP_3.

Piezoelectricity of non-polar systems: example of a tetrahedron

The example of the tetragonal ferroelectric phase of perovskite is a case where the existing polarization of the crystal is changed by the application of stress. Piezoelectricity is also found in crystals without a centre of symmetry even when there is no polarization in the stress-free crystal. We explore this point by the example of a single tetrahedral group of ions, as illustrated in Fig. 7.10. In this case we assume that the vertices of the tetrahedra are anions, of charge $-Q$, and the centre of the cluster is a cation, of charge $+4Q$. Note that the point symmetry of the tetrahedral cluster is $\bar{4}3m$ – there is no centre of symmetry. We will assume for the sake of the discussion that the anion–cation bond is strong, and remains of constant length when a small stress is applied. This can be achieved if the anion–cation–anion bonds can flex with the applied stress to give an angle different to the ideal tetrahedral angle of 109.47°. In Fig. 7.10 the axes are defined along the $\bar{4}$ axes, with the central cation at the origin.

We first consider the application of the stress σ_1. This has the effect of squashing the tetrahedral cluster along the x_1 direction. This causes the 3-fold

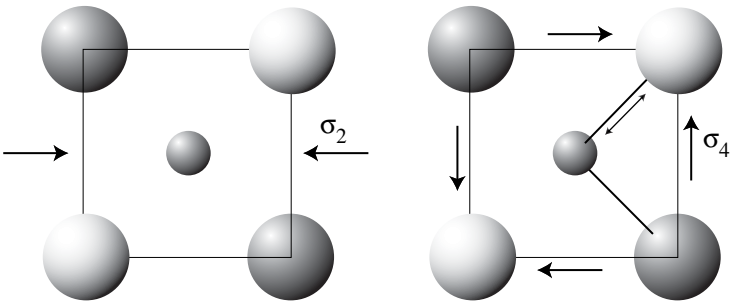


Fig. 7.10 Schematic representation of the piezoelectric effect in a tetrahedron of initial point group symmetry $\bar{4}3m$. The x_1 axis is horizontal, and the x_2 axis is vertical. The lightly-shaded anions (large spheres) are above the plane of the diagram, and the darkly-shaded anions are below the plane of the diagram; the central cation (small sphere) is in the plane of the diagram.

rotational symmetry to be lost along each of the axes, and the $\bar{4}$ rotoinversion axes along both the x_2 and x_3 axes are reduced to 2-fold rotation axes along these directions. However, the $\bar{4}$ rotoinversion axis along x_1 is retained. With the presence of these three orthogonal rotation and rotoinversion axes, there can be no polarization.

Now we consider the application of a shear stress, σ_6. This has the effect of moving the top two anions away from each other, and the lower two anions apart. Since the system tries to keep the cation–anion distances constant, the cation is moved upwards. It can be shown (Problem 7.4) that the displacement of the cation along the x_3 direction, $u^+_{(z)}$, is related to the displacements of the anions in the $x_1 - x_2(x - y)$ direction, $\pm u^-_{(xy)}$, by

$$|u^+_{(z)}| = |u^-_{(xy)}|/\sqrt{2} \tag{7.47}$$

where we use the modulus because the different directions of motion all cause the same displacement. The induced polarization is simply equal to $4Qu^+_{(z)}$.

The final result can be summarized by the matrix equation:

$$\begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & d_{13} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{13} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{13} \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} \tag{7.48}$$

where we have used the fact that $d_{13} = d_{25} = d_{36}$ by symmetry.

Given that a perfect tetrahedral arrangement of cations and anions has a non-zero piezoelectric tensor, we might expect crystals that contain tetrahedral arrangements of atoms to also have non-zero piezoelectric tensor. This applies to the semiconducting cubic ZnS and related materials, where each cation is in a tetrahedral coordination with neighbouring anions (and *vice versa*) and to systems such as quartz that consist of networks of linked SiO_4 tetrahedra. Of course, the arrangements of the tetrahedral groups may give additional symmetry that changes the piezoelectric properties – for example, the crystal structure of the high-temperature cristobalite phase of silica has a centre of symmetry – but the point is that a crystal without a centre of symmetry can have

a piezoelectric effect even without there being a polarization in the stress-free crystal.

7.4.2 Use of Voigt notation for third-rank tensors

We have made use of the Voigt notation in the example above, but to perform operations such as the rotation of axes we need to convert to the long-handed three-index mode. In long hand, we can write the equation for the direct piezoelectric effect as

$$\begin{aligned} P_1 = {} & d_{111}\sigma_{11} + d_{112}\sigma_{12} + d_{113}\sigma_{13} \\ & + d_{121}\sigma_{21} + d_{122}\sigma_{22} + d_{123}\sigma_{23} \\ & + d_{131}\sigma_{31} + d_{132}\sigma_{32} + d_{133}\sigma_{33} \end{aligned}$$

and similarly for P_2 and P_3. When we take account of the symmetry of the σ tensor by using Voigt notation for this tensor, i.e. $\sigma_{12} = \sigma_{21} = \sigma_6$, etc., we can rewrite this equation as

$$\begin{aligned} P_1 = {} & d_{111}\sigma_1 + (d_{112} + d_{121})\sigma_6 + (d_{113} + d_{131})\sigma_5 \\ & + d_{122}\sigma_2 + (d_{123} + d_{132})\sigma_4 + d_{133}\sigma_3 \end{aligned}$$

By comparison with the full Voigt expansion:

$$P_1 = d_{11}\sigma_1 + d_{12}\sigma_2 + d_{13}\sigma_3 + d_{14}\sigma_4 + d_{15}\sigma_5 + d_{16}\sigma_6 \tag{7.49}$$

we see that $d_{11} = d_{111}$, $d_{12} = d_{122}$, and $d_{13} = d_{133}$, whereas $d_{14} = 2d_{123}$, $d_{15} = 2d_{113}$, and $d_{16} = 2d_{112}$. Similar results follow from an analysis of P_2 and P_3. The new factor of 2 is important!

7.4.3 Transformations of third-rank tensors

Transformation equations

Earlier we used physical arguments to obtain the non-zero components of the piezoelectric tensor for a tetragonal crystal and a tetrahedron. The symmetry properties can be obtained by using symmetry transformations, but for this we need to see how to transform a third-rank tensor by a change in the definitions of the axes. This will follow the approach taken in transformation of second-rank tensors. We define a first-rank tensor $\mathbf{P}$, a second-rank tensor $\mathbf{Q}$, and a third-rank tensor $\mathbf{T}$. These are linked by $\mathbf{P} = \mathbf{T} \cdot \mathbf{Q}$. Similarly we can define a set of transformed tensors: $\mathbf{P}' = \mathbf{T}' \cdot \mathbf{Q}'$. Following the earlier procedure, we obtain

$$\begin{aligned} \mathbf{P}' &= \mathbf{a} \cdot \mathbf{P} = \mathbf{a} \cdot (\mathbf{T} \cdot \mathbf{Q}) \\ \mathbf{P}' &= \mathbf{T}' \cdot \mathbf{Q}' = \mathbf{T}' \cdot (\mathbf{a} \cdot \mathbf{Q} \cdot \mathbf{a}^{\mathrm{T}}) \\ \Rightarrow \mathbf{a} \cdot \mathbf{T} \cdot \mathbf{Q} &= \mathbf{T}' \cdot \mathbf{a} \cdot \mathbf{Q} \cdot \mathbf{a}^{\mathrm{T}} = \mathbf{T}' \cdot \mathbf{a} \cdot \mathbf{a} \cdot \mathbf{Q} \\ \Rightarrow \mathbf{T}' &= \mathbf{a} \cdot \mathbf{T} \cdot \mathbf{a}^{\mathrm{T}} \cdot \mathbf{a}^{\mathrm{T}} = \mathbf{a} \cdot \mathbf{a} \cdot \mathbf{a} \cdot \mathbf{T} \end{aligned} \tag{7.50}$$

where we have made use of the fact that $\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}$ and $\mathbf{T}^{\mathrm{T}} = \mathbf{T}$, for example in writing $\mathbf{Q} \cdot \mathbf{a}^{\mathrm{T}} = \mathbf{a} \cdot \mathbf{Q}$. In long hand, this can be written as

$$T'_{ijk} = \sum_{lmn} a_{il} a_{jm} a_{kn} T_{lmn} \tag{7.51}$$

with the reverse transformation

$$T_{ijk} = \sum_{lmn} a_{li} a_{mj} a_{nk} T'_{lmn} \tag{7.52}$$

These equations can be readily extended for higher-rank tensor transformations.

Effect of symmetry

We now use the transformation equation to demonstrate how symmetry can determine which elements of the piezoelectric tensor are zero. We will consider the effect of 4-fold rotation axis along x_3. The transformation matrix for a 90° rotation is

$$\mathbf{a} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7.53}$$

First we consider d_{11} and d_{22}. The transformation equations give

$$\left.\begin{aligned} d'_{111} &= a_{12}a_{12}a_{12}d_{222} = d_{222} \\ d'_{222} &= a_{21}a_{21}a_{21}d_{111} = -d_{111} \end{aligned}\right\} \;\Rightarrow\; d_{11} = -d_{22} = 0$$

For d_{33} we have

$$d'_{333} = a_{33}a_{33}a_{33}d_{333} = d_{333} \;\Rightarrow\; d_{33} \neq 0$$

We now consider the pairs involving the tensile stresses, using the fact that we expect pairs to be equal by symmetry:

$$\left.\begin{aligned} d'_{122} &= a_{12}a_{21}a_{21}d_{211} = d_{211} \\ d'_{211} &= a_{21}a_{12}a_{12}d_{112} = -d_{112} \end{aligned}\right\} \;\Rightarrow\; d_{12} = -d_{21} = 0$$

$$\left.\begin{aligned} d'_{133} &= a_{12}a_{33}a_{33}d_{233} = d_{233} \\ d'_{233} &= a_{21}a_{33}a_{33}d_{133} = -d_{133} \end{aligned}\right\} \;\Rightarrow\; d_{13} = -d_{23} = 0$$

$$\left.\begin{aligned} d'_{311} &= a_{33}a_{12}a_{12}d_{322} = d_{322} \\ d'_{322} &= a_{33}a_{21}a_{21}d_{311} = d_{311} \end{aligned}\right\} \;\Rightarrow\; d_{31} = d_{32} \neq 0$$

Now we consider the piezoelectric coefficients involving the shear stresses:

$$\begin{aligned}
&\left.\begin{aligned} d'_{123} &= a_{12}a_{21}a_{33}d_{213} = -d_{213} \\ d'_{213} &= a_{21}a_{12}a_{33}d_{123} = -d_{123} \end{aligned}\right\} \;\Rightarrow\; d_{14} = -d_{25} \neq 0 \\
&\left.\begin{aligned} d'_{223} &= a_{21}a_{21}a_{33}d_{113} = d_{113} \\ d'_{113} &= a_{12}a_{12}a_{33}d_{223} = d_{223} \end{aligned}\right\} \;\Rightarrow\; d_{15} = d_{24} \neq 0 \\
&\left.\begin{aligned} d'_{323} &= a_{33}a_{21}a_{33}d_{313} = -d_{313} \\ d'_{313} &= a_{33}a_{12}a_{33}d_{323} = d_{323} \end{aligned}\right\} \;\Rightarrow\; d_{35} = -d_{34} = 0 \\
&\left.\begin{aligned} d'_{121} &= a_{12}a_{21}a_{12}d_{212} = -d_{212} \\ d'_{212} &= a_{21}a_{12}a_{21}d_{121} = d_{121} \end{aligned}\right\} \;\Rightarrow\; d_{26} = -d_{16} = 0 \\
&d'_{312} = a_{33}a_{12}a_{21}d_{321} = -d_{312} \;\Rightarrow\; d_{36} = 0
\end{aligned} \tag{7.54}$$

This gives the piezoelectric tensor for a material with point group 4 of

$$\mathbf{d} = \begin{pmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & -d_{14} & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{pmatrix} \tag{7.55}$$

As a result of symmetry, the number of non-zero coefficients of the piezoelectric coefficient will be reduced to a relatively small number in most crystal classes.

7.5 Fourth-rank tensors

7.5.1 A hierachy of higher-order tensors

The principles developed in this chapter, including the methods for transforming tensors when there is a rotation of the frame of reference, and the role of symmetry in determining which coefficients of a matter tensor are allowed to have non-zero values, will apply in the same way (albeit with rather more tedius algebra) to higher-rank tensors. Therefore we will mention only one fourth-rank tensor in closing this chapter, namely the elasticity tensors that link stress and strain.

Since all the field tensors first or are second-rank, the higher-order tensors will correspond to cases where the application of one field tensor is able to change the values of the matter tensors. We will consider briefly examples of this kind in the next section, which will be third-rank and fourth-rank effects.

7.5.2 The elasticity tensors

There are two important tensors that link stress to strain, namely the **elastic compliance tensor** that links an induced strain to an applied stress,

$$\epsilon_{ik} = \sum_{jl} s_{ijkl}\sigma_{jl} \tag{7.56}$$

and the **elastic constant tensor** that gives the value of an applied stress required to produce a given set of values of the strains:

$$\sigma_{ik} = \sum_{jl} c_{ijkl}\epsilon_{jl} \tag{7.57}$$

The energy that accompanies a set of strains is

$$E = \frac{1}{2}\sum_{ijkl} c_{ijkl}\epsilon_{ij}\epsilon_{kl} \tag{7.58}$$

Although there are 81 ($=3^4$) terms in the full elastic constant tensor, the number of independent terms is smaller. Use of the symmetry implicit in the Voigt notation, namely that $c_{ijkl} = c_{jikl} = c_{ijlk} = c_{jilk}$ reduces the number of independent coefficients to $6 \times 6 = 36$. Within the Voigt notation, we also have the symmetry that $c_{mn} = c_{nm}$, so the maximum number of independent

coefficients is 27. Symmetry of the crystal will reduce this further. For example, in the case of a cubic material the elastic constant tensor can be written as

$$\mathbf{c} = \begin{pmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{pmatrix} \tag{7.59}$$

In an isotropic material, $c_{44} = (c_{11} - c_{12})/2$. Also, in cubic crystals in which all the interatomic interactions can be realistically described by pair potentials, $c_{44} = c_{12}$. In NaCl, $c_{12} = 1.24\,\text{GPa}$ and $c_{44} = 1.26\,\text{GPa}$. This result holds to a similar or slightly lesser extent in other alkali halide crystals. It is not so appropriate for cubic metals. For example, in Cu, $c_{12} = 12.1\,\text{GPa}$ and $c_{44} = 7.5\,\text{GPa}$, although in Ni the agreement is rather closer, $c_{12} = 14.0\,\text{GPa}$ and $c_{44} = 12.5\,\text{GPa}$.

7.6 Induced changes in matter tensors

7.6.1 Basic ideas

When we separated the types of tensors into matter tensors and field tensors, we implied that the field tensors can be varied (by turning a dial on a piece of equipment, say), whereas the matter tensors were constants of the material. However, the values of matter tensors are subject to change if there are changes in the external conditions of the material. For example, changing temperature may give rise to a phase transition, which may have a dramatic effect on the matter tensors, such as the dielectric susceptibility as seen earlier in Figs 1.8 and 2.31. In the case of the ferroelectric perovskites, the high-temperature cubic phase has a centre of symmetry and no non-zero piezoelectric coefficients, but some of these coefficients (e.g. d_{33}, as seen above) will become non-zero on cooling through the phase transition, and further cooling will lead to more distortion of the crystal structure and increases in the sizes of the piezoelectric coefficients. A similar example for quartz will be shown later in Section 12.2.4. Other matter tensors can be changed by application of higher-order field tensors. This is exemplified by the effects of external conditions on the optical properties of materials.

7.6.2 Refractive index, the electro-optic effect, and the photoelastic effect

The refractive index of a material is given as the square root of the high-frequency dielectric constant. In principle, this is also a tensor property. For an anisotropic crystal, the refractive indices along the principle directions will therefore be equal to $n_{ii} = \varepsilon_{ii}^{1/2}$. For many applications, the primary tensor used is $\mathbf{B}$, where the principal coefficients are related to the refractive index by $B_{ii} = 1/n_{ii}^2$.

The tensor $\mathbf{B}$ is a matter tensor, but this can be affected by application of an electric field or elastic stress. Specifically, these effects are written as the

change in the coefficients of the **B** tensor:

$$\Delta B_{ij} = \sum_k z_{ijk} E_k + \sum_{kl} \pi_{ijkl} \sigma_{kl} \tag{7.60}$$

The third-rank tensor **z** is called the **linear electro-optic** tensor, and the fourth-rank tensor $\boldsymbol{\pi}$ is the **piezo-elastic** tensor. Clearly the latter term could also be written using the strain tensor instead.

Finally, the first term in eqn (7.60) describes the linear dependence of the refractive index on applied electric field. There are many cases where the linear dependence is zero, and instead a strong second-order dependence is seen. This is described by a fourth-rank tensor:

$$\Delta B_{ij} = \sum_{kl} \alpha_{ijkl} E_k E_l \tag{7.61}$$

When allowing for higher-order effects, it is clear that these will require higher-rank tensors. The extension of the analysis for these higher-rank tensors follows from the main results presented here.

Summary of chapter

- Physical properties in anisotropic materials are formally described using the language of **tensors**.
- The important feature of tensors is that they can be transformed following a transformation of the system of axes.
- Each crystal has a set of axes – the **principal axes** – for which its properties are described by a tensor in which only the diagonal components have non-zero values.
- The non-zero values for the tensor properties of a crystal are determined by symmetry.

Further reading

The main source texts for tensor properties of crystals are Lovett (1989) and Nye (1985). The mathematics of tensors are described by Sands (1982, 1995).

Exercises

(7.1) Show that the transformation matrix **a** of eqn 7.5 has the condition $\mathbf{a}^{-1} = \mathbf{a}^{\mathrm{T}}$.

(7.2) The coefficients of the thermal expansion tensor of calcite (point group $\bar{3}m$) are given in Chapter 1 (page 7). Calculate the orientation of the normal of a plate that can be cut so that it does not change its thickness on heating. Calculate the shear strain induced within the plane if there is a change in temperature of 100 K.

(7.3) A cubic crystal is cut in the form of a plate parallel to (111). Denoting the axis normal to the plate as x'_3, if a

stress σ_3' is applied to the plate, what are the components of the stress tensor described in the reference frame of the principal axes?

(7.4) Derive eqn 7.47 for the constraint that the bond lengths do not change.

(7.5) The non-zero components of the piezoelectric tensor for a cubic crystal of point group $\bar{4}3m$ are $d_{14} = d_{25} = d_{36}$. Deduce expressions for the components of the dielectric polarization referred to the principal axes if a stress σ_3' is applied normal to the (111) plate of the previous question. The cubic phase of ZnS (point group $\bar{4}3m$) has $d_{14} = 3.1 \times 10^{-12}\,\mathrm{C\,N^{-1}}$. Calculate the charges induced on a (111) plate of area $30\,\mathrm{mm}^2$. (*Note*: the obvious way to proceed is to transform the polarization tensor back into the reference axes of the plate, but a more direct method would be to transform the piezoelectric tensor **d** into the axes of the plate. It would be worth using both methods to give a check on the answers.)

(7.6) The non-zero piezoelectric coefficients are quartz are $d_{11} = -d_{12}$, $d_{26} = -2d_{11}$, and $d_{14} = -d_{25}$. Data are given later in the book in Fig. 12.11. Calculate the strains produced when an electric field of strength $10^6\,\mathrm{V\,m^{-1}}$ is applied along the [100] direction, which corresponds to the x_1 axis, at room temperature.

(7.7) Write the transformation matrix for a mirror plane, and use this to determine the non-zero components of the piezoelectric tensor d_{ijk}. Repeat the exercise for the case of a mirror plane at right angles to the original mirror plane and hence determine the non-zero components for a crystal of point group $mm2$. Show that the addition of a third mirror plane causes all components of the piezoelectric tensor to be zero.

(7.8) Write the transformation matrix analogous to eqn 7.53 for the operation of a $\bar{4}$ axis. Use the transformation matrix to obtain the non-zero values of the piezoelectric coefficients, and compare with those allowed in the presence of a 4-fold rotation axis.

(7.9) Obtain the non-zero piezoelectric coefficients of a tetrahedron of point group symmetry $\bar{4}3m$, eqn 7.49, by consecutive application of the transformation matrices for the $\bar{4}$ along the three orthogonal axes.

(7.10) The elastic energy is given by eqn 7.58. Show that for a cubic crystal, for which the only non-zero values of the elastic constant tensor are c_{11}, c_{12}, and c_{44} and their symmetrically equivalent components, stability is only achieved if $c_{44} > 0$, $c_{11} > |c_{12}|$ and $c_{11} + 2c_{12} > 0$. (*Hint*: for the second two conditions, one approach would be to write the equation for the energy as $E = \mathbf{e}^{\mathrm{T}} \cdot \mathbf{c} \cdot \mathbf{e}$, where **e** is the column matrix with elements e_1, e_2, e_3 and **c** is a 3×3 matrix with the diagonal elements equal to c_{11} and the other elements equal to c_{12}. The condition for stability $E > 0$ is given by the condition that the eigenvalues of **c** are greater than 0.)

(7.11) Show that the transformation of $\mathbf{P} \rightarrow \mathbf{P}'$ obtained from the applied electric field of eqn 7.11 is consistent with eqn 7.14.

Lattice dynamics

8

8.1 Why do we need to consider dynamics?

Up to this point in the book we have paid no attention to the motions of atoms in solids. Without so doing we have achieved a considerable level of understanding of the properties and behaviour of solids. There are, however, a number of issues that we cannot begin to tackle without considering dynamics. These include the whole area of thermodynamics, including phase stability. For example, the stable phase of any material is that for which the free energy is a minimum. The free energy G has two parts, due to the enthalpy H (assuming constant pressure) and entropy S:

$$G = H - TS \tag{8.1}$$

For many solids the primary origin of entropy is from thermal vibrations. Neglect of dynamics is tantamount to asserting that temperature is zero, and hence that entropy is also zero. It is only by including dynamics that we are able to study issues associated with phase stability under changing temperature, such as the ferroelectric phase transitions we have met in the earlier parts of the book.

Thermal expansion and pyroelectricity are two important issues that depend directly on dynamics through their dependence on temperature. The piezoelectric properties of ferroelectric crystals also have a strong dependence on temperature, which can only be understood through the development of a model for the dynamics. In effect, without an understanding of the dynamics of crystals we are unable to develop quantitative models for crystal behaviour that incorporate temperature as a variable.

Our route into the subject is through a few simple models. The simple models can be generalized for application to real materials (and we do so in this chapter), but in keeping them simple we will be able to deduce the important general principles with minimum of effort.

8.2 The harmonic approximation

The ideas developed in this chapter are based on a simple approximation called the harmonic approximation. In principle, if $u_{j,\alpha}$ describes a component of the vector displacement ($\alpha = x, y, z$) of the atom labelled j from its equilibrium

position, the energy can be expressed as a Taylor expansion:

$$E = E_0 + \frac{1}{2}\sum_{\substack{j,j'\\ \alpha,\alpha'}} \frac{\partial^2 E}{\partial u_{\alpha,j}\partial u_{\alpha',j'}} u_{\alpha,j}u_{\alpha',j'} + \cdots + \frac{1}{n!}\sum_{\substack{j^{(1)},\ldots,j^{(n)}\\ \alpha^{(1)},\ldots,\alpha^{(n)}}} \frac{\partial^n E}{\partial u_{\alpha^{(1)},j^{(1)}}\cdots\partial u_{\alpha^{(n)},j^{(n)}}} u_{\alpha^{(1)},j^{(1)}}\cdots u_{\alpha^{(n)},j^{(n)}} + \cdots \tag{8.2}$$

E_0 is the equilibrium lattice energy (Chapter 5), which we stress again is the energy with all atoms at rest. There is no first-order term in the expansion because by the definition of equilibrium all residual forces are zero. The second-order term is the **harmonic energy**. All higher-order terms are classed together as the **anharmonic energy**. The essence of **harmonic approximation** is to neglect all the anharmonic terms.

The harmonic approximation may appear to be quite crude (brutal even). Its merit in our immediate context is that it gives a mathematical model that has exact solutions. The anharmonic model, even if truncated at the first additional term, cannot be solved exactly (and to truncate after an odd-order term allows the energy to be lowered to a value of $-\infty$). Where the harmonic model has a particular merit is that the anharmonic terms can be introduced as a modification to the harmonic model, provided that the atomic displacements are not large. This condition is often fulfilled. The harmonic model gives many of the features of lattice dynamics, such as the dependence of frequency on wave vector, and the features that are not explained by the harmonic model (such as thermal expansion) can be incorporated into a model that looks like a harmonic model. Thus the harmonic approximation does rather more than simply provide a model that can be solved – it provides the whole framework for the study of lattice dynamics into which improvements can be incorporated with little damage to the framework.

8.3 Lattice vibrations of one-dimensional monatomic crystals

8.3.1 The linear chain model

The simplest illustrative model is a chain of atoms, each of mass m, and separated by the unit cell length a, as shown in Fig. 8.1. Initially we consider that each atom only feels the force of its immediate neighbour, which we will assume to be harmonic. Writing u_n as the displacement of the n-th atom from its equilibrium position, the total energy of a chain of N atoms is

$$E = NE_{\text{bond}} + \frac{1}{2}J\sum_{i=1}^{N}(u_n - u_{n+1})^2 \tag{8.3}$$

where the force constant J is given by

$$J = \left.\frac{\partial^2 E}{\partial u^2}\right|_{u=0} \tag{8.4}$$

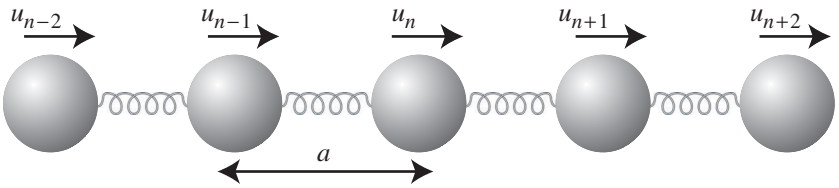


Fig. 8.1 Linear chain model. J is the harmonic force constant (the interactions are represented by springs), and the atomic displacements are represented by u.

The force constant is identical for each pair of atoms in the chain. In forming the differential we note that the distance between two atoms n and $n+1$ is $r = a + (u_n - u_{n+1})$. Thus the derivatives of E with respect to u are equivalent to the derivatives with respect to r. Since a is the equilibrium unit cell length, the first derivative of E at $r = a$ is zero.

8.3.2 Sound waves – vibrations with long wavelengths

Vibrations with long wavelengths, and hence small wave vectors, will correspond to sound waves. There is a simple relationship between the speed of sound, v, and the isotropic elastic stiffness c:

$$v^2 = c/\rho \tag{8.5}$$

where ρ is the density, which in our one-dimensional model is equal to $\rho = m/a$. The one-dimensional stiffness in the model is easy to compute: a strain ϵ gives a displacement $u_n = \epsilon n a$ to atom n, so the strain energy *per unit length* is given as

$$\begin{aligned} E_{\text{strain}} &= \frac{1}{2}\frac{J(u_n - u_{n+1})^2}{a} \\ &= \frac{1}{2}J\epsilon^2 a \equiv \frac{1}{2}c\epsilon^2 \end{aligned} \tag{8.6}$$

The last part of this equation defines the stiffness c:

$$c = Ja \tag{8.7}$$

We therefore have

$$v^2 = Ja^2/m \tag{8.8}$$

The velocity of the sound wave is related to the angular frequency and wave vector through

$$\omega = vk \tag{8.9}$$

Thus we have an equation for the dependence of angular frequency on wave vector:

$$\omega = \left(a\sqrt{J/m}\right)k \tag{8.10}$$

We have therefore related the vibration frequency to the second derivative of the interatomic potential, and obtained the dependence of the frequency on wave vector in the limit of small k.

8.3.3 Vibrations with shorter wavelengths: general features

We start from the energy of atom n:

$$E_n = \frac{1}{2}J(u_{n+1} - u_n)^2 + \frac{1}{2}J(u_n - u_{n-1})^2 \tag{8.11}$$

The force on the n-th atom (which is given by the differential of the energy) is linked to its acceleration by the classical Newton equation of motion, *force* = *mass* × *acceleration*:

$$\begin{aligned} F_n &= -\frac{\partial E_n}{\partial u_n} = J(u_{n+1} - u_n) - J(u_n - u_{n-1}) \\ &= -J(2u_n - u_{n+1} - u_{n-1}) = m\frac{\partial^2 u_n}{\partial t^2} \end{aligned} \tag{8.12}$$

The solution of the harmonic equation of motion is a sinusoidal wave, so the motion of the whole system will correspond to a set of travelling waves. Our aim then is to find the set of frequencies of these waves. A wave of a given wave vector k and angular frequency ω_k will cause atom n to be displaced according to the travelling wave equation

$$u_{n,k} = \tilde{u}_k \exp[i(kx_n - \omega t)] \tag{8.13}$$

where k is any wave vector ($= 2\pi/\text{wavelength}$), ω_k is the corresponding angular frequency, $\tilde{u}_k$ is the amplitude, and $x_n = na$ is the position of the n-th atom. From the outset we use the complex exponential form instead of the simpler cosine or sine function, because it makes the manipulation of the differential equations more transparent. We note for future reference that the displacements of the atom's neighbours are given by

$$u_{n\pm1,k} = u_{n,k} \exp(\pm ika) \tag{8.14}$$

We will find later that there is a discrete set of allowed values of k. We expect the time-dependent motion of the n-th atom to be a linear superposition of each of the travelling waves allowed along the chain; the mathematical representation is

$$u_n(t) = \sum_k \tilde{u}_k \exp[i(kna - \omega t)] \tag{8.15}$$

8.3.4 Vibrations with shorter wavelengths: the special case of $\lambda = 2a$

We consider first the special case of a vibration where the wavelength is equal to twice the atom repeat distance, which means that $k = \pi/a$. In this case, neighbouring atoms will move by equal amounts in opposite directions, i.e. $u_n = -u_{n+1}$ (consistent with eqn 8.14). The force on atom n will now be given by

$$F_n = -J(2u_n - u_{n+1} - u_{n-1}) = -4Ju_n = m\frac{\partial^2 u_n}{\partial t^2} \tag{8.16}$$

We can write the displacement of atom n for this wave as

$$u_n = \tilde{u} \exp(i\omega t) \tag{8.17}$$

Substitution of this equation into both sides of eqn 8.16 gives

$$-4Ju_n = -m\omega^2 u_n$$
$$\Rightarrow \omega = \left(\frac{4J}{m}\right)^{1/2} \tag{8.18}$$

The result for $k = \pi/a$ is compared with the result for $k \to 0$ in Fig. 8.2. It should be noted that we expect that the vibration with $\lambda = 2a$ will be a standing wave, which will have the group velocity $\partial\omega/\partial k = 0$. The interpolation between the results for $k \to 0$ and $k = \pi/a$ will now be developed exactly using the same methods.

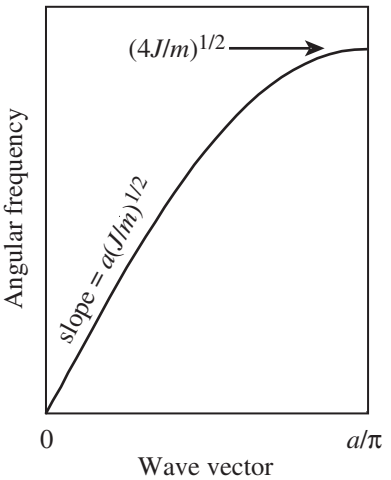


Fig. 8.2 Interpolation of the dependence of frequency on wave vector for the one-dimensional harmonic chain.

8.3.5 Vibrations with shorter wavelengths: the general case

If we consider each wave vector separately, we can solve the equation of motion for each individual wave propagating along the one-dimensional chain of atoms. We substitute the general equations for u_n and $u_{n\pm1}$, eqns 8.13 and 8.14, into the force component of eqn 8.12 to obtain the force acting on atom n:

$$\begin{aligned} -\frac{\partial E_n}{\partial u_n} &= -J(2u_n - u_{n+1} - u_{n-1}) \\ &= -Ju_n[2 - \exp(ika) - \exp(-ika)] \\ &= -2Ju_n[1 - \cos(ka)] \\ &= -4Ju_n \sin^2(ka/2) \end{aligned} \tag{8.19}$$

The *mass* × *acceleration* term is given as

$$m\frac{\partial^2 u_n}{\partial t^2} = -m\omega_k^2 u_n \tag{8.20}$$

When we equate the two expressions, and divide out u_n from both sides of the equation, we obtain the following expression for the angular frequency as a function of wave vector, ω_k:

$$\omega_k^2 = \left(\frac{4J}{m}\right)\sin^2(ka/2)$$
$$\Rightarrow \omega_k = \left(\frac{4J}{m}\right)^{1/2} |\sin(ka/2)| \tag{8.21}$$

By taking the limit $k \to 0$ we recover eqn 8.10. By taking only the positive roots we obtain the behaviour of the angular frequency shown in Fig. 8.3. The graph of ω_k is known as a **dispersion curve**. It should be noted that there is a periodicity of the results in k. For every k equal to a reciprocal lattice vector, i.e. $k = 2n\pi/a$, where n is any integer, the angular frequency falls to zero. This result was anticipated in Section 4.5, where we noted that any two waves with wave vectors that differ by a reciprocal lattice vector will be equivalent. It is a rather elegant point to note that we have deduced the existence of the reciprocal

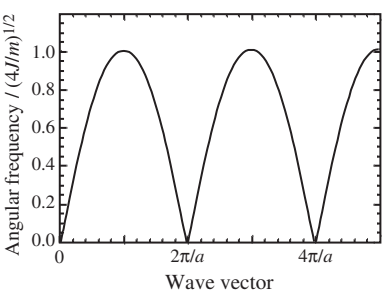


Fig. 8.3 Acoustic mode dispersion curve for the one-dimensional chain.

lattice (at least in one dimension, although the result can be generalized to three dimensions) without having introduced the idea of the reciprocal lattice at any stage in the development of the theory. It persuades this author that reciprocal space is as real a concept as real space, and not, as supposed by some, a creation of crystallographers to make their tasks easier but which is best avoided by everyone else whenever possible!

We note at this point that vibrations with dispersion curves of the form of Fig. 8.3, namely with $\omega \to 0$ as $k \to 0$, are called **acoustic modes**. This name highlights the link to sound waves in the limit $k \to 0$. In Section 8.5 we will introduce another type of vibration, called an **optic mode**, which has a quite different behaviour in the limit $k \to 0$.

8.3.6 Extension of model of monatomic chain to include distant neighbours

The analysis of the lattice vibrations of the monatomic chain can be extended to include more interactions beyond the nearest neighbour interaction. If $\phi_p(r_p)$ is the energy between the n-th atom and its p-th neighbour, which have equilibrium separation $r_p = pa$, the energy of the chain is given by

$$E = N \sum_p \phi_p(pa) + \frac{1}{2} \sum_{n,p} J_p (u_n - u_{n+p})^2 \tag{8.22}$$

where

$$J_p = \frac{\partial^2 \phi_p}{\partial r_p^2} \tag{8.23}$$

The equation of motion for any atom given by eqn 8.12 is now generalized to

$$m \frac{\partial^2 u_n}{\partial t^2} = -\frac{\partial E}{\partial u_n} = -\sum_p (2u_n - u_{n+p} - u_{n-p}) \tag{8.24}$$

We again substitute the travelling wave solution for each wave vector, eqn 8.13 (and the generalization of eqn 8.14):

$$-m\omega_k^2 u_n = u_n \sum_p J_p [2 - \exp(ikpa) - \exp(-ikpa)] \tag{8.25}$$

As before, this simply reduces to

$$\begin{aligned} m\omega_k^2 &= 2 \sum_p J_p [1 - \cos(kpa)] \\ \Rightarrow \omega_k^2 &= \frac{4}{m} \sum_p J_p \sin^2(kpa/2) \end{aligned} \tag{8.26}$$

This result represents only a minor modification to the result for nearest neighbours. In particular, it has the same behaviour for $k \to 0$ and at $k = \pi/a$ as described earlier. The main difference is that the additional terms give the possibility that the dispersion curve has a more detailed shape than the simpler nearest-neighbour model.

8.3.7 Reciprocal lattice, the Brillouin zone, and allowed wave vectors

We now recall the general result obtained in Section 4.5 that two waves with wave vectors k and k' that differ by a reciprocal lattice vector G, i.e. $k' = k+G$, have the same wave properties. This was seen in the calculation of the dispersion curves, where the plot of ω_k is periodic in $2\pi/a$. To reiterate the point, we compare the displacements u_{nk} and $u_{nk'}$, eqn (8.13). Noting that $\omega_k = \omega_{k'}$, we rewrite $u_{nk'}$ as

$$\begin{aligned} u(x) &= \tilde{u}\exp(i[k'na - \omega_{k'}t]) \\ &= \tilde{u}\exp(i[(k+G)na - \omega_k t]) \\ &= \tilde{u}\exp(iGna)\exp(i[kna - \omega_k t]) \end{aligned} \tag{8.27}$$

Since $\exp(iGa) = 1$, we see that $u_{nk'} = u_{nk}$, i.e. the two travelling waves have identical effects on the atoms in the chain. This is illustrated in Fig. 8.4, where we see that two waves with wave vectors differing by a reciprocal lattice vector give identical displacements at each lattice position. This result can be generalized to three dimensions, and to unit cells containing more than one atom.

Since the equations of motion for waves with wave vectors that differ by a reciprocal lattice vector are equivalent, the only unique information contained in the dispersion curves is for the range of wave vectors lying between the limits $-\pi/a < k < \pi/a$. This range of wave vectors is the one-dimensional equivalent of the Brillouin zone introduced in Section 4.5. The wave vectors $k = \pm\pi/a$ define special points in reciprocal space, called **Brillouin zone boundaries**, which lie half-way between reciprocal lattice points. The group velocity, $\partial\omega/\partial k$, is equal to zero at these points, meaning that the wave with this wave vector is a standing wave. It will turn out that this is a general result in three dimensions.

It is common practice to represent the wave vector as a normalized quantity by dividing its value by the magnitude of the first reciprocal lattice vector lying along the direction of the wave vector. This gives what is called the **reduced wave vector**. For our one-dimensional example, the wave vector at the Brillouin zone boundary, $k = \pi/a$, is represented by the reduced wave vector of value $1/2$, obtained by dividing the wave vector $a^*/2$ by the reciprocal lattice vector a^*. This is really no different to the normal use of Miller indices.

Although the dispersion curve is a continuous curve, a finite chain will only permit a discrete set of wave vectors. This set is determined by noting that in practice we never define an origin of the crystal, and our equations do

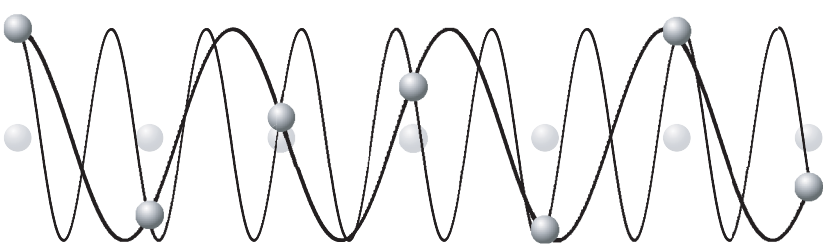


Fig. 8.4 The atomic motions associated with two waves whose wave vectors differ by a reciprocal lattice vector. We see that the two waves have identical effects on the atoms.

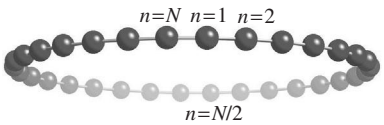


Fig. 8.5 Joined chain to show operation of periodic boundary conditions.

not suppose the existence of any boundaries. In effect, the two ends of the one-dimensional chain of atoms are effectively joined together. This state of having removed the free boundaries by assuming that the periodicity of the structure spans the boundaries in both directions is known as **periodic boundary conditions**. This construction of the periodic boundary conditions is illustrated in Fig. 8.5. We briefly met this construction in Section 5.5.5.

With periodic boundary conditions, the atom at one end of the chain, labelled by the integer $n = N$ with position $x = Na$, is identical to the atom at the origin, which is labelled as $n = 0$ and position $x = 0$. Thus the wave equation gives terms of the form

$$\exp(ikNa) = \exp(0) = 1 \tag{8.28}$$

Therefore the allowed values of k are given by the discrete set:

$$k = \frac{2\pi n}{Na} \tag{8.29}$$

where n is an integer. This shows that there are N allowed wave vectors within one Brillouin zone. In the general case, including the generalization to three dimensions, the number of allowed wave vectors within one Brillouin zone is equal to the number of primitive unit cells in the crystal.

8.3.8 Three-dimensional monatomic crystals: general principles

The one-dimensional model can be easily extended to three dimensions if, in the equations of motions, the variables u refer not to displacements of atoms but of planes of atoms, and appropriate force constants are defined. We also need to include motions that are perpendicular to the wave vector. These follow as a simple extension of our one-dimensional model. The perpendicular displacements will have different force constants from the longitudinal force constants, but they will be expressed in an identical manner so that the results for the different types of motion will follow identical equations. This situation is illustrated in Fig. 8.6.

As there are two orthogonal directions perpendicular to the wave vector, there will be two more acoustic branches with different dispersion curves determined by the new force constants. The corresponding vibrations are called **transverse acoustic modes** (TA), whereas the vibrations where the atoms move in directions parallel to the wave vector are called **longitudinal acoustic modes** (LA). In high symmetry situations, where there are three-, four- or six-fold rotation axes along the direction of the wave vector, the force constants will be the same for any two transverse displacements, so that the two transverse modes will have degenerate (i.e., equal) frequency dispersion.

Experimental data of dispersion curves are commonly represented using the **reduced wave vector scheme** introduced in the previous section. In three dimensions, any general wave vector can be written as

$$\mathbf{k} = \xi\mathbf{a}^* + \zeta\mathbf{b}^* + \varsigma\mathbf{c}^* \tag{8.30}$$

where ξ, ζ, ς typically have values between $\pm 1/2$ (except for non-primitive lattices). The set (ξ, ζ, ς) is directly analogous to the set (h, k, l) used to

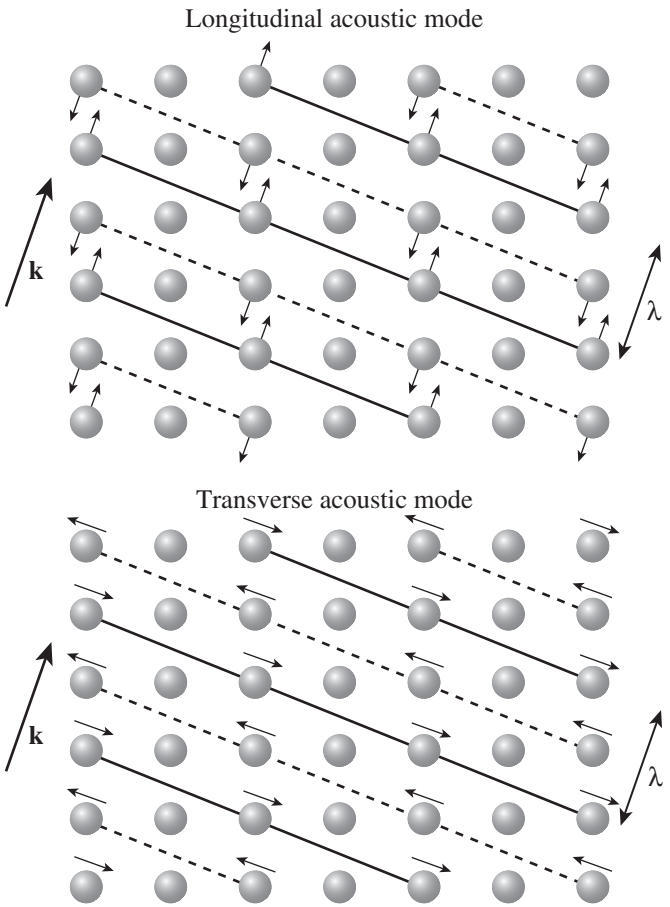


Fig. 8.6 Travelling waves with longitudinal and transverse atomic displacements in a two-dimensional section of a monatomic crystal.

represent lattice planes (Section 4.2), except that the values of ξ, ζ, ς are not confined to integers. The reduced wave vector representation (ξ, ζ, ς) is widely used for representing wave vectors in diagrams of dispersion curves.

8.4 Dispersion curves in face-centred cubic materials

8.4.1 Dispersion curves of neon

The simplest face-centred cubic materials are crystals of the rare gases. The bonding in these materials is primarily of van der Waal's type (Section 5.4.4), and the interactions between second-nearest neighbour atoms are much weaker that those between nearest neighbours. The dispersion curves of crystalline neon are shown in Fig. 8.7. The data are shown using the reduced zone scheme, restricting the range of wave vectors to within the first Brillouin zone. The data for the three symmetry directions are plotted back-to-back to highlight the fact that the Brillouin zone boundaries along the directions [100] and

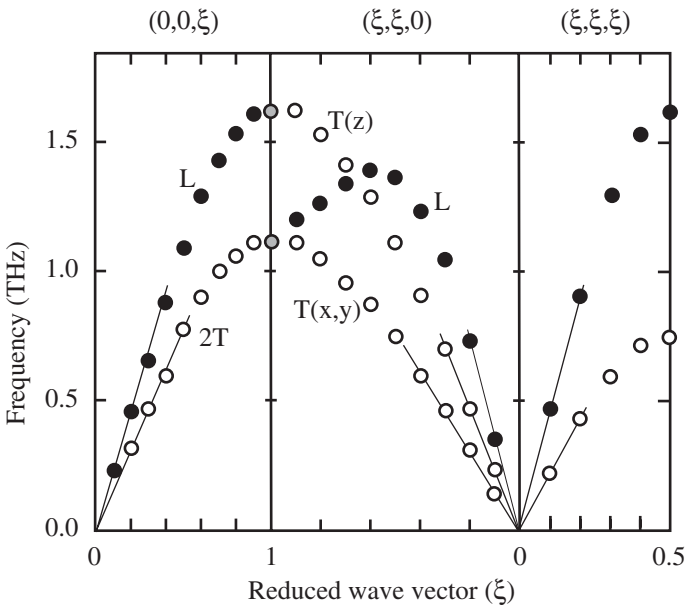


Fig. 8.7 Acoustic mode dispersion curves for neon (ccp structure) measured by inelastic neutron scattering. L and T denote longitudinal and transverse modes respectively. For the [001] and [111] directions the two transverse modes are degenerate (as indicated by the label 2T). Along the [110] direction, the (x,y) and (z) labels indicate the directions of the atomic motions. (Data taken from Endoh *et al.*, *Phys. Rev. B* **11**, 1681, 1975.)

[110] are symmetrically equivalent points in reciprocal space (see below, and Problem 4.5).

There are a number of overall features that should be noted at the outset. The transverse acoustic modes for wave vectors along [100] and [111] are degenerate as a result of the 4-fold and 3-fold rotational symmetry along these directions respectively. On the other hand, the [110] direction contains only a 2-fold rotation axis, so the two transverse modes have different frequencies. The dispersion curves along the [100] and [111] directions closely follow the sinusoidal relationship developed for the simple monatomic chain with nearest-neighbour interactions. The more complex shapes of the dispersion curves along [110] arise from second-neighbour interactions between planes of atoms which arise even when there are only interactions between nearest-neighbour atoms. This will be discussed in more detail below.

Another point that should be noted is that the [100] and [110] points in reciprocal space are symmetrically equivalent (see Problem 4.5), and this means that the dispersion curves along the two directions converge to common frequencies at this point. In particular, it is necessary for two of the modes along [110] to become degenerate at the Brillouin zone boundary. The reason why it is the longitudinal and one of the transverse modes that become degenerate is explored in Problem 8.1.

The data in Fig. 8.7, as in other plots of dispersion curves that will be given later in this chapter, were obtained by inelastic neutron scattering experiments. The basic theory of the method of inelastic neutron scattering will be discussed in Chapter 10, but for the moment we can outline the basic idea. Neutrons are scattered by quanta of lattice vibrations (to be discussed in Chapter 9), and the frequency and wave vector of any vibration can be obtained by measuring the change in energy and wave vector of the scattered neutron beam. The change

in wave vector corresponds to the wave vector of the lattice vibration, and the change in energy is equal to the frequency of the vibration multiplied by Planck's constant.

The dispersion curves of the other rare-gas crystals are very similar to those of neon (these are gathered together in Dove 1993), except that there is a scaling of the frequency range due to differences in mass and in the values of nearest-neighbour force constants.

Dispersion curves of neon along [100]

We can interpret the dispersion curves of neon along [100] by mapping the interatomic interactions onto the interactions between planes. This mapping is shown in Fig. 8.8. The atomic force constants act along the direction of the distance between atoms, and are given by $\partial^2\phi/\partial r^2$, where r is the interatomic distance. The interatomic force constants need to be mapped onto the force constant between planes for both longitudinal and transverse displacements of planes. It can be shown (Problem 8.2) that if we write the interatomic vector as $\mathbf{r} = (x, y, z)$,

$$\frac{\partial^2\phi}{\partial x^2} = \frac{x^2}{r^2}\frac{\partial^2\phi}{\partial r^2} \tag{8.31}$$

Similar results hold for the y and z directions. To obtain the force constants for the three relative displacements of neighbouring planes we need to sum over all interatomic interactions between the planes. If we denote $K = \partial^2\phi/\partial r^2$ we have the following three force constants (Problem 8.3).

$$J_x = \sum_j \frac{\partial^2\phi}{\partial x_j^2} = 2K \quad J_y = \sum_j \frac{\partial^2\phi}{\partial y_j^2} = K \quad J_z = \sum_j \frac{\partial^2\phi}{\partial z_j^2} = K \tag{8.32}$$

J_x gives the force constant for the longitudinal modes, and J_y and J_z give the force constants for the transverse modes. The four nearest-neighbour interatomic vectors between the planes shown in Fig. 8.8 are

$$\frac{a}{2}[1, \pm 1, 0] \qquad \frac{a}{2}[1, 0, \pm 1]$$

There are no interactions between second-neighbour planes. We therefore expect both the dispersion curves for the longitudinal and transverse modes

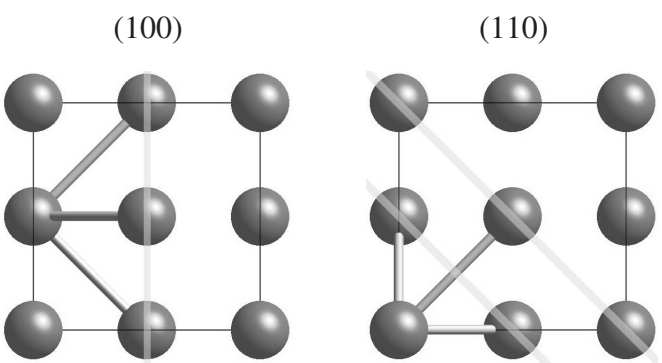


Fig. 8.8 Mapping of interatomic interactions onto the (100) and (110) planes of the ccp crystal, with the planes marked by the thick light lines. The relevant nearest-neighbour distances are represented by the tie lines between atoms. For neighbouring (100) planes there are four interatomic interactions (noting that one lies directly above another in the view of the diagram). For nearest-neighbour (110) planes there are again four interatomic interactions (two being masked out), and there is one interatomic interaction between second-neighbour planes.

along [100] to have the sinusoidal form from before, with the frequency of the longitudinal mode at each wave vector to be $\sqrt{2}$ times larger than the frequencies of the two transverse modes at the same wave vector. Moreover, the frequency of the longitudinal mode should be the same as the LA mode along [100]. Both these features are confirmed by the data shown in Fig. 8.7.

Dispersion curves of neon along [111]

The planes in this case are the close-packed planes, and each atom in one plane interacts with three atoms in the neighbouring plane. It is convenient to define a new coordinate system with $x \parallel [111]$ and $y \parallel [1\bar{1}0]$. This gives the interplanar force constants (Problem 8.3)

$$J_x = 2K \qquad J_y = J_z = K/2 \tag{8.33}$$

Again, there are no interactions between second-neighbour planes. The ratio of the two force constants suggest that the frequency of the longitudinal mode should be twice as large as the frequencies of the corresponding transverse modes for wave vectors along [110], and this is confirmed by the data in Fig. 8.7.

Dispersion curves of neon along [110]

The mapping of the atomic model onto the (110) planes is shown in Fig. 8.8. It is now convenient to define a coordinate system such that $x \parallel [110]$, $y \parallel [1\bar{1}0]$, and $z \parallel [001]$. In this case, the mapping shows that there is an interaction between second-nearest planes. The mapping gives the following force constants for nearest-neighbour planes (Problem 8.3):

$$J_x^{(1)} = J_y^{(1)} = K \qquad J_z^{(1)} = 2\frac{\partial^2 \phi}{\partial z^2} = 2K \tag{8.34}$$

and the following force constants for second-neighbour planes:

$$J_x^{(2)} = K \qquad J_y^{(2)} = J_z^{(2)} = 0$$

In this case the dispersion curves for the two transverse modes will be the standard sinusoidal form, with one of the modes having frequencies that are $\sqrt{2}$ times larger than the frequencies of the second TA mode. The LA mode will have the extended form given by eqn 8.26, involving the sum of two sinusoidal contributions to ω^2. The second contribution will have a maximum at the wave vector half-way to the Brillouin zone boundary, and zero contribution at the zone boundary. Therefore we expect the LA mode to have the same value of the frequency at the zone boundary as the TA(y) mode. These features are confirmed in the data of Fig. 8.7.

We have therefore been able to rationalize the form of the dispersion curves of crystalline neon for the three principal symmetry directions in reciprocal space. The good agreement between the simple theory and the experimental data shows that taking only nearest-neighbour interatomic interactions is a very good approximation. The shape of the LA curve along [110] shows that mapping a real three-dimensional crystal onto the linear chain model may lead to second-neighbour interplanar interactions even though the interatomic interactions are only taken to nearest-neighbour pairs.

8.4.2 Dispersion curves of lead

Crystalline lead also has a fcc structure. The measured dispersion curves are shown in Fig. 8.9. These show many of the features seen in the dispersion curves of the crystalline rare gas elements. However, it is clear from the shapes of the curves along [100] and [110] that there are important interatomic interactions that extend beyond nearest-neighbour pairs of atoms. There are also gliches in the continuous dispersion curves, most obviously in the LA mode along [110] at the wave vector around [0.3, 0.3, 0], which arise when the wave vectors match the wave vectors of the electronic Fermi surface (Section 5.4.7).

8.4.3 Dispersion curves of potassium

Crystalline potassium has a body-centred cubic structure. The dispersion curves are shown in Fig. 8.10. In the bcc lattice the Brillouin zone boundaries along the [100] and [111] directions are symmetrically equivalent points in reciprocal space, and this is highlighted in the scheme chosen to represent the dispersion curves.

The forms of the LA and TA dispersion curves along [100] and of the TA mode along [110] show that the nearest-neighbour interatomic interactions dominate. The mapping of the interatomic interactions onto the interactions between planes is shown in Fig. 8.11. The vectors between an atom in one plane and its four neighbours in the first-nearest (100) plane are

$$\frac{a}{2}(1, \pm 1, \pm 1)$$

It can be shown (Problem 8.6) that all force constants are equal:

$$J_x = \sum_j \frac{\partial^2 \phi}{\partial x_j^2} = 2K \qquad J_y = \sum_j \frac{\partial^2 \phi}{\partial y_j^2} = 2K \qquad J_z = \sum_j \frac{\partial^2 \phi}{\partial z_j^2} = 2K$$

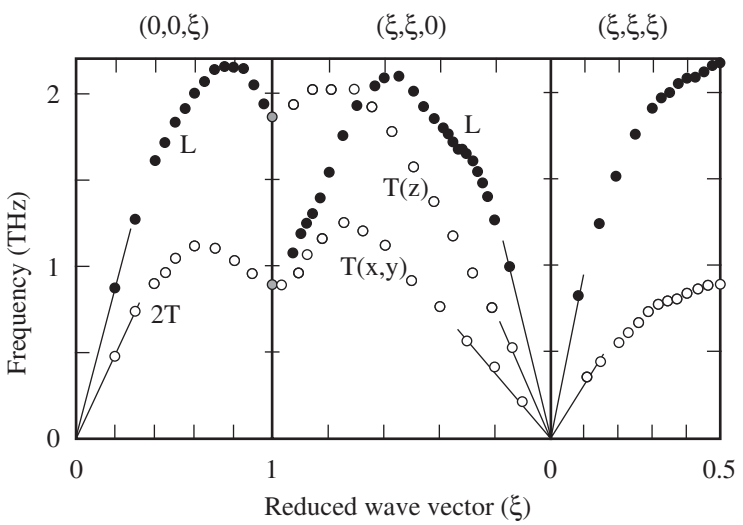


Fig. 8.9 Acoustic mode dispersion curves for lead (fcc) measured by inelastic neutron scattering. (Data taken from Brockhouse *et al.*, *Phys. Rev.* **128**, 1099, 1962.)

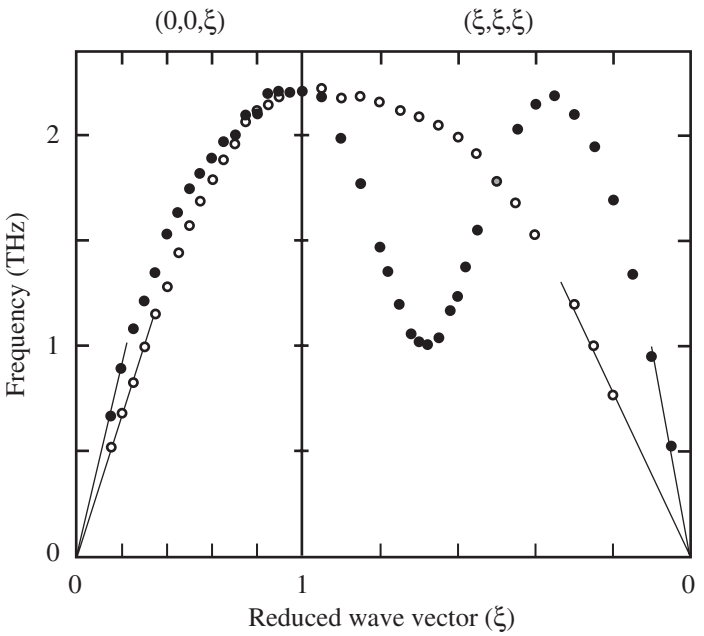


Fig. 8.10 Acoustic mode dispersion curves for potassium (bcc) measured by inelastic neutron scattering. The right-hand plot shows the relevant portion of reciprocal space. (Data taken from Cowley *et al.*, *Phys. Rev.* **150**, 487, 1966.)

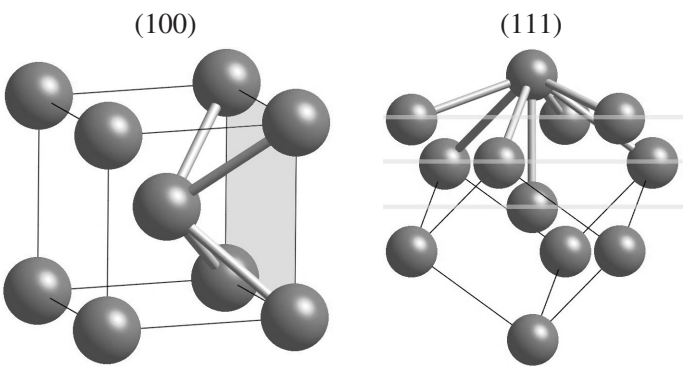


Fig. 8.11 Mapping of interatomic interactions onto the (100) and (111) planes of the bcc crystal, with the planes marked by the light shading. The nearest-neighbour interactions are represented by the tie lines between atoms. There are four interatomic interactions contributing to the (100) interplanar force constants. On the other hand, for the (111) planes, the nearest-neighbour interatomic interactions contribute to interplanar interactions up to third-distant neighbouring planes.

This shows why, if the nearest-neighbour interatomic interactions dominate, the LA and TA modes have very similar frequencies for wave vectors along [100], as seen in Fig. 8.10. The second-neighbour interatomic interaction involves the separation of one unit cell length, and this will contribute to only the LA interplanar force constant. The small contribution of this second-neighbour interaction can be seen in the small increase in frequency of the LA mode over the TA mode that reaches a maximum at the reduced wave vector $(\frac{1}{2}, 0, 0)$.

The interactions between (111) planes are a little more complicated, as seen in Fig. 8.11. The interactions involved in the nearest-neighbour planes are the nearest-neighbour interatomic distances, but the projection is at a wide angle, meaning that the contribution to the longitudinal force constants is not large. The interaction between second-neighbour planes involves second-neighbour interatomic interactions, and these are expected to be somewhat weaker. The

third-neighbour interplanar force constants also involve the nearest-neighbour interatomic interaction. In this case the interatomic vector is normal to the planes, so there will be a large contribution to the longitudinal force constant but no interaction to the transverse force constants. If we define K as the first-neighbour interatomic force constant as in our analysis of the ccp structures, and set the second-neighbour interatomic force constant to zero, it can be shown that the interplanar force constants are

$$J_{\mathrm{L}}^{(1)} = K/3 \qquad J_{\mathrm{T}}^{(1)} = 4K/3 \qquad J_{\mathrm{L}}^{(3)} = K$$

$J_L^{(2)}$ is only sensitive to second-neighbour interactions, and is therefore likely to be lower in value than $J_{\mathrm{L}}^{(1)}$. Clearly $J_{\mathrm{L}}^{(3)}$ is the largest of the longitudinal force constants. Writing d as the (111) interplanar distance, and projecting the wave vector onto the [111] direction, the LA dispersion curve has the form

$$\omega_{\mathrm{LA}}^2 = \frac{4}{m}\left(J_{\mathrm{L}}^{(1)} \sin^2(kd/2) + J_{\mathrm{L}}^{(2)} \sin^2(kd) + J_{\mathrm{L}}^{(3)} \sin^2(3kd/2)\right) \qquad (8.35)$$

As we noted, the strongest force constant is $J_{\mathrm{L}}^{(3)}$, and it can be seen that the contribution from this force constant vanishes when $k = 2\pi/3d$, which corresponds to the three-dimensional reduced wave vector $[\frac{2}{3}, \frac{2}{3}, \frac{2}{3}]$. This is the origin of the large dip in the dispersion curve at this wave vector seen in the data of Fig. 8.10 (see Problem 8.7).

8.5 Lattice vibrations of crystals with several atoms in the unit cell

8.5.1 The basic model

The main features of the lattice dynamics of crystals with several atoms in the unit cell can be represented by a simple model that is an extension of the one-dimensional chain. This is shown in Fig. 8.12. Two types of atoms of masses M and m interact with force constant J, and their displacements U and u are defined in this figure. All atoms are separated by $a/2$ when at rest. This model represents crystals that have alternating planes of two types of atoms, as for the (111) planes of atoms in the NaCl structure. We will consider only the longitudinal motions, as the extension to include transverse motions is the same as for the monatomic case. We will also only consider nearest-neighbour interactions. This model will give us some new important characteristics, but despite its simplicity it proves to be about as far as we can go before the

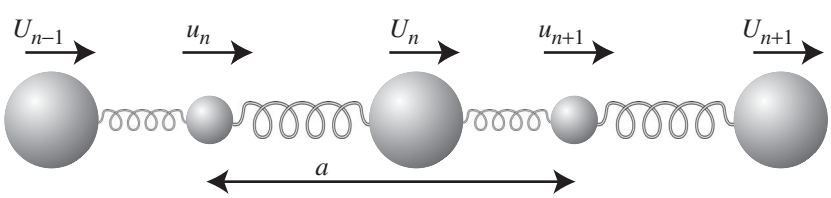


Fig. 8.12 Diatomic linear harmonic chain. The model is shown here with two distinct types of atoms and two distinct harmonic interactions. In the main text we treat the two types of interactions as equivalent, and Problem 8.11 treats the case of distinct interactions and both atoms having the same mass.

arithmetic becomes too complicated. For anything more complex we will need to resort to the computer!

The harmonic energy of the model is given as

$$E = \frac{1}{2} J \sum_n (U_n - u_n)^2 + (u_{n+1} - U_n)^2 \tag{8.36}$$

Following the procedure used for the analysis of the monatomic chain, we can write down the equations for the forces acting on both types of atoms. For the big atom labelled n, the energy is given as

$$E(U_n) = \frac{1}{2} J[(U_n - u_n)^2 + (u_{n+1} - U_n)^2] \tag{8.37}$$

and the force on it is given by

$$\begin{aligned} F_n = -\frac{\partial E(U_n)}{U_n} &= -J[(U_n - u_n) - (u_{n+1} - U_n)] \\ &= -J(2U_n - u_n - u_{n+1}) \\ &= M\frac{\partial^2 U_n}{\partial t^2} \end{aligned} \tag{8.38}$$

Similarly, for the small atom labelled n, the energy is given as

$$E(u_n) = \frac{1}{2} J[(u_n - U_{n-1})^2 + (U_n - u_n)^2] \tag{8.39}$$

and the force on it is given by

$$\begin{aligned} f_n = -\frac{\partial E(u_n)}{u_n} &= -J[(u_n - U_{n-1}) - (U_n - u_n)] \\ &= -J(2u_n - U_n - U_{n-1}) \\ &= m\frac{\partial^2 u_n}{\partial t^2} \end{aligned} \tag{8.40}$$

We can assume the same general solutions as used for the monatomic chain, one for each atom type:

$$\begin{aligned} U_n &= \sum_k \tilde{U}_k \exp(i[kna - \omega_k t]) \\ u_n &= \sum_k \tilde{u}_k \exp(i[kna - \omega_k t]) \end{aligned} \tag{8.41}$$

where k is the wave vector, ω_k is the angular frequency of a given mode, and $\tilde{U}_k$ and $\tilde{u}_k$ are the two amplitudes for a single given mode (the two atoms each have a different amplitude for any one mode).

8.5.2 Solution for small wave vector

Consider the two types of waves illustrated in Fig. 8.13. The first wave is a sound wave, where all atoms move in phase. An extension of the earlier model leads to the result for the acoustic mode frequency for small k:

$$\omega = a\left(\frac{J}{2(M+m)}\right)^{1/2} k \tag{8.42}$$

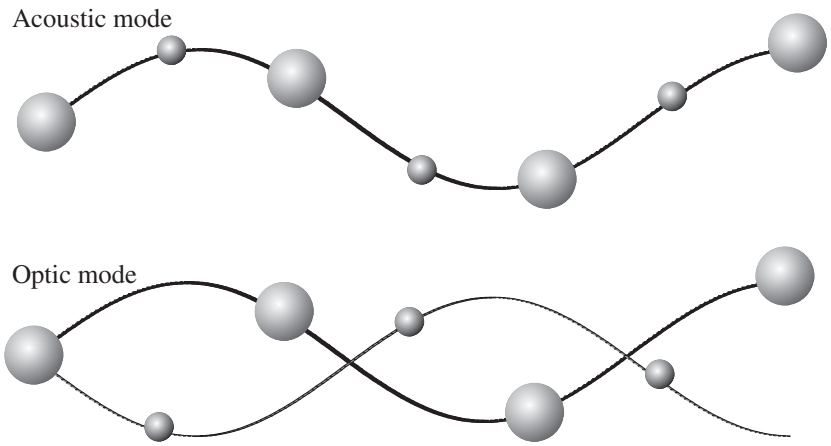


Fig. 8.13 Exaggerated transverse atomic motions in a one-dimensional diatomic crystal for an acoustic mode (motions of different types of atoms are in phase) and an optic mode (motions of different types of atoms are out of phase).

The second wave has neighbouring atoms (of different types) moving out of phase. For ionic crystals, this wave will have the cations moving in opposite directions to neighbouring anions. For $k = 0$ (infinite wavelength) the condition that the centre of mass doesn't move means that

$$MU_n = -mu_n \tag{8.43}$$

As a result we can write

$$u_n = \frac{A}{m}\exp(i\omega t) \qquad U_n = -\frac{A}{M}\exp(i\omega t) \tag{8.44}$$

Note that at $k = 0$, $u_n = u_{n+1}$ and $U_n = U_{n+1}$ for all n. We substitute these solutions into the equation of motion for the big atom:

$$\begin{aligned} M\frac{\partial^2 U_n}{\partial t^2} &= -A\omega^2\exp(i\omega t) = -\frac{\partial E}{\partial U_n} \\ &= -J(U_n - u_{n+1} - u_n) \\ &= -J\left(2\frac{A}{M} + \frac{A}{m} + \frac{A}{m}\right)\exp(i\omega t) \\ &= -2AJ\left(\frac{1}{M} + \frac{1}{m}\right)\exp(i\omega t) \end{aligned} \tag{8.45}$$

This leads to the result for the angular frequency:

$$\omega^2 = 2J\left(\frac{1}{M} + \frac{1}{m}\right) \tag{8.46}$$

The same result is obtained when we repeat the procedure for the small atom (Problem 8.9). Unlike the acoustic mode, this vibration has a non-zero frequency at $k = 0$ which is, to first order, independent of k. This is a *standing wave*, which is properly defined as such through the condition $\partial\omega/\partial k = 0$. This is called the **optic mode**, because if the two atoms are of opposite charge their relative motions correspond to the motions that would be induced by the passage of an electromagnetic wave of optical frequencies. Actually the frequencies of optic modes are usually within the infrared region of the electromagnetic spectrum, as we will see in the examples later in this chapter.

8.5.3 General result

The general solution for all wave vectors can be obtained with a little more algebra. Using eqns 8.37–8.41, we have

$$\begin{aligned} -M\omega^2\tilde{U} &= -J\{2\tilde{U} - \tilde{u}[1 + \exp(ika)]\} \\ -m\omega^2\tilde{u} &= -J\{2\tilde{u} - \tilde{U}[1 + \exp(-ika)]\} \end{aligned} \tag{8.47}$$

These equations can be rewritten in matrix form:

$$\begin{pmatrix} M\omega^2 - 2J & J[1 + \exp(ika)] \\ J[1 + \exp(-ika)] & m\omega^2 - 2J \end{pmatrix} \begin{pmatrix} \tilde{U} \\ \tilde{u} \end{pmatrix} = 0 \tag{8.48}$$

This solution of this matrix equation is that for which the determinant of the matrix is zero. This gives

$$\begin{aligned} &[M\omega^2 - 2J][m\omega^2 - 2J] = J^2[1 + \exp(ika)][1 + \exp(-ika)] \\ &\Rightarrow \quad Mm\omega^4 - 2J(M + m)\omega^2 + 4J^2\sin^2(ka/2) = 0 \\ &\Rightarrow \quad \omega^2 = J\frac{M + m}{Mm} \pm \frac{J}{Mn}\sqrt{(M + m)^2 - 4Mm\sin^2(ka/2)} \end{aligned} \tag{8.49}$$

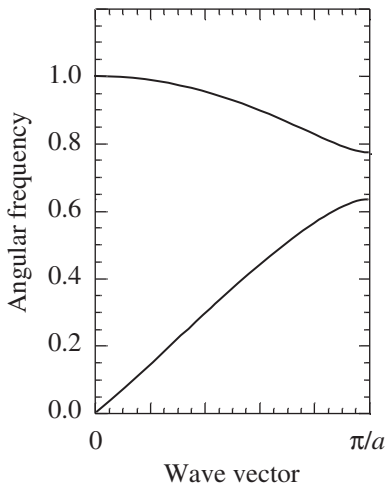


Fig. 8.14 Dispersion curves for the one-dimensional diatomic chain obtained from eqn 8.49.

This general result is shown for one set of masses in Fig. 8.14, which shows both the acoustic and optic mode frequencies as functions of the wave vector. It can readily be shown (Problem 8.10) that the results given earlier for $k \to 0$ (eqns 8.42 and 8.46) are obtained from this general result. The general result also allows us to solve for the atomic motions. For the two solutions for $k \to 0$ it is straightforward to show that the acoustic mode solution for the frequency gives $\tilde{U} = \tilde{u}$, and that the optic model solution gives $M\tilde{U} = -m\tilde{u}$.

It is interesting to consider the behaviour at the Brillouin zone boundary, $k = \pi/a$. In this case the two frequencies can be shown to be (Problem 8.10)

$$\omega^2 = \frac{2J}{M} \qquad \omega^2 = \frac{2J}{m} \tag{8.50}$$

The matrix equation (eqn 8.48) then simplifies to

$$\begin{pmatrix} M\omega^2 - 2J & 0 \\ 0 & m\omega^2 - 2J \end{pmatrix} \begin{pmatrix} \tilde{U} \\ \tilde{u} \end{pmatrix} = 0 \tag{8.51}$$

For the first solution, $\omega^2 = 2J/M$, $\tilde{u} = 0$ and the value of $\tilde{U}$ is undetermined. For the second solution, $\omega^2 = 2J/m$, we now have $\tilde{U} = 0$ and the value of $\tilde{u}$ is undetermined. So for both modes, one atom moves while the other stands still.

The value of the amplitude of the atom that moves cannot be determined by the dynamic equations developed so far. It will depend on temperature, as will be shown in the next chapter. The fact that in both modes only one atom moves suggests that at the Brillouin zone boundary we have lost the distinction between acoustic and optic modes. This is not always the case (for example, a similar model with two equal masses but two different force constants will retain the difference between acoustic and optic modes in the displacements for all wave vectors, Problem 8.11), but in many crystals the distinction between acoustic and optic modes is only properly maintained in the limit $k \to 0$. This point is not always appreciated!

8.5.4 Generalization for more complex cases: atomic motions

We reinforce our descriptions of acoustic and optic modes by considering the case of four atoms in a unit cell, all moving along the x direction, in either a positive or negative sense, and with $k \to 0$. These four motions are shown in Fig. 8.15 (a similar picture is given in Fig. 3.13). There are four combinations of motions of the four atoms, $++++$, $++--$, $+-+-$, and $+--+$, where the signs denote the relative signs of the displacements of the four atoms. These are the only four independent linear combinations of positive or negative displacements, which were labelled according to their irreducible representations in Section 3.4. For the last three combinations, the sums of the displacements are zero, which will therefore correspond to the optic modes. The first mode, for which all the atoms move in phase, is the acoustic mode. All other possible displacement patterns correspond to linear combinations of these basic modes.

It is easy to generalize this argument for Z atoms per cell, showing that for displacements along one direction there will always be one acoustic mode and $(Z-1)$ optic modes. If we now allow motions along the y and z directions as well, it is clear that we will have three acoustic modes and $(3Z-3)$ optic modes. We can conclude that the principal difference between acoustic and optic modes is that in the limit $k \to 0$ the acoustic modes have all the atoms moving in phase, whereas the optic modes have the atoms moving out of phase. This distinction is not relevant for larger wave vectors, as we have seen in one case for our simple one-dimensional model. (Note that if the atoms have different masses, the displacement patterns described above will describe mass-weighted displacements.)

In high-symmetry situations, it is possible to separate **longitudinal optic** (LO) modes from **transverse optic** (TO) modes, just as we noted the separation of LA and TA modes in Section 8.3.8. The distinction is easy to appreciate in simple crystals such as NaCl (Section 8.5.6), but it may not be appropriate in more complex systems.

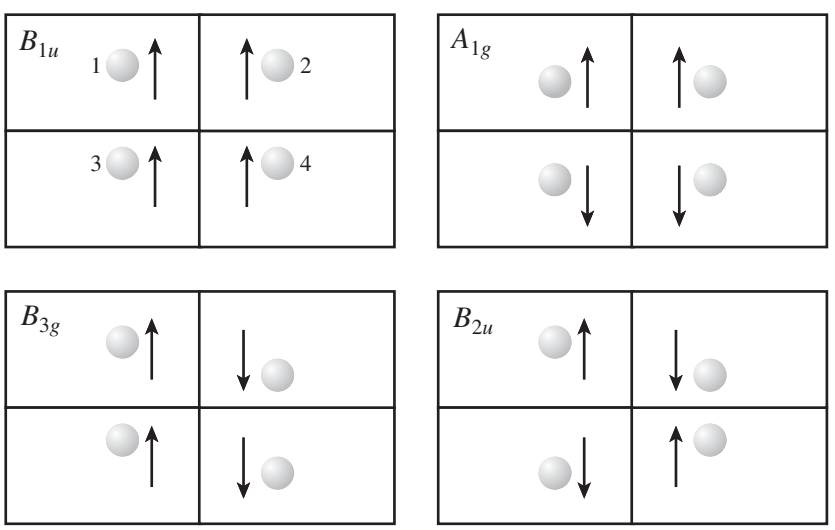


Fig. 8.15 Examples of four modes of motion along x in a two-dimensional crystal of point group symmetry $mm2$ (similar to Fig. 3.13). The model labelled B_{1u} is the acoustic mode; the three other modes are optic modes. For the A_{1g} mode, in which the atomic displacements do not break the symmetry, the atoms are drawn in their mean positions.

8.5.5 Generalization for more complex cases: the dynamical matrix

The mathematical formalism developed for the diatomic chain can be generalized for more complex cases, but only by imposing a certain order on the procedure we used. We will consider another one-dimensional example where we allow interactions to extend beyond nearest neighbour, and we will allow an arbitrary number of atoms in the unit cell. Consider each atom in the unit cell to be labelled j. The energy is now written as

$$E_j = \frac{1}{2}\sum_{n,j'} J_{j,j',n}(u_{j,0} - u_{j,n})^2 \tag{8.52}$$

where the sum over j' runs over all atoms in other unit cells, and n labels the unit cell such that $n = 0$ is the reference unit cell, and the sum over n also includes the $n = 0$ components. The force on the atom is given by

$$\begin{aligned} F_j &= -\frac{\partial E_j}{\partial u_{j,0}} = -\sum_{n,j'} J_{j,j',n}(u_{j,0} - u_{j',n}) \\ &= m_j \frac{\partial^2 u_{j,0}}{\partial t^2} \end{aligned} \tag{8.53}$$

As before, we assume that any wave will affect all atoms in the unit cell. We have a general equation for the displacement of an atom that has the form we have met earlier, but we now need to include the sum over all the modes. Strictly the sum should include all wave vectors, but since this sum will occur in both sides of the force equation we can account for the different wave vectors separately. It will also become convenient to express the atomic displacement equation in a way that separates out a mass term, which may seem to be a little odd for the moment:

$$\begin{aligned} u_{j,n} &= \sum_{\nu} \tilde{u}_{j,\nu} \exp(i[kna - \omega_\nu t]) \\ &= \sum_{\nu} A_\nu m_j^{-1/2} e_{j,\nu} \exp(i[kna - \omega_\nu t]) \end{aligned} \tag{8.54}$$

where ν labels the mode (in a one-dimensional system there are as many modes as atoms in the unit cell). A_ν is a scale factor that will drop out of the equations – this is the part that is dependent on temperature. The new variables $e_{j,\nu}$ are normalized such that

$$\sum_j |e_{j,\nu}|^2 = 1 \tag{8.55}$$

for each value of ν. These normalized variables contain the information about the *relative* displacements of the atoms due to the wave.

Substituting the general wave equation into the force equation gives, for each value of ν:

$$-m_j^{1/2} e_{j,\nu}\omega_\nu^2 = -\sum_{n,j'} J_{j,j',n}\left(m_j^{1/2} e_{j,\nu} - m_{j'}^{1/2} e_{j',\nu}\exp(ikna)\right) \tag{8.56}$$

This equation can be written in matrix form. If we define the $N \times 1$ matrix $\mathbf{e}_\nu$ as formed from the different components $e_{j,\nu}$, we can write

$$\mathbf{e}_\nu \omega_\nu^2 = \mathbf{D} \times \mathbf{e} \tag{8.57}$$

where the $N \times N$ matrix $\mathbf{D}$, which is called the **dynamical matrix**, has components

$$D_{j,j'} = (m_j m_{j'})^{-1/2} \sum_n J_{j,j',n}(\delta_{j,j'} - \exp(ikna)) \tag{8.58}$$

We can pack together the individual $N \times 1$ matrices $\mathbf{e}_\nu$ to form the $N \times N$ matrix $\mathbf{e}$. This matrix has the property that $\mathbf{e}^{-1} = (\mathbf{e}^*)^{\mathrm{T}}$. We can also form the matrix $\mathbf{\Omega}$ with components $\Omega_{\nu,\nu'} = \omega_\nu^2 \delta_{\nu,\nu'}$ (i.e. $\mathbf{\Omega}$ is a diagonal matrix, whose components are the squares of the vibrational angular frequencies). By using these new matrices, we can write

$$\mathbf{e} \times \mathbf{\Omega} = \mathbf{D} \times \mathbf{e} \Rightarrow \mathbf{\Omega} = \mathbf{e}^{-1} \times \mathbf{D} \times \mathbf{e} \tag{8.59}$$

This is now in the form of an **eigenproblem**: the components of the matrix $\mathbf{\Omega}$ are the eigenvalues of $\mathbf{D}$, and the components of $\mathbf{e}$ contain the corresponding eigenvectors. For this reason, the vectors $\mathbf{e}$ are known as the **mode eigenvectors**. Finding the eigenvalues and eigenvectors of a matrix is a task that has been optimized for calculations on a computer. The task then is to set up the dynamical matrix using the individual interatomic force constants, and then find its eigenvalues to get the mode frequencies. Subsuming the masses into the definitions of the $\mathbf{D}$ and $\mathbf{e}$ matrices has allowed the writing of the dynamical equations in the form of an eigenvalue problem.

The dynamical matrix formalism can readily be extended for calculations on three-dimensional crystals. The $\exp(ikna)$ phase factor of the one-dimensional chain can be replaced by $\exp(i\mathbf{k} \cdot \mathbf{r})$, where $\mathbf{r}$ is the vector between the reference unit cell and the neighbouring unit cell, and $\mathbf{k}$ is the three-dimensional wave vector. The other change is that account has to be taken of the three components of the displacement vectors, which means that each component of the dynamical matrix has to be replaced by a 3×3 array representing the combinations of the components of each pair of displacements. Similarly the **frequency matrix** $\mathbf{\Omega}$ and the eigenvector matrix $\mathbf{e}$ both become $3N \times 3N$ matrices.

Finally, we remark that the form of the dynamical matrix will be controlled by the symmetry of the crystal. If we recall the example earlier of four atoms in an orthorhombic crystal of symmetry $mm2$, we can remark that the symmetry of the crystal will ensure that the diagonalization of the dynamical matrix automatically produces the combinations of symmetry that were drawn earlier. We illustrate this by considering the crystal of $mm2$ symmetry shown in Fig. 8.15. We will assume that the atoms have interactions only with nearest neighbours, but a more complicated model will not actually modify the ideas in the analysis. It can be shown (Problem 8.13) that the dynamical matrix at $\mathbf{k} = 0$ has the form

$$\mathbf{D}(\mathbf{k} = 0) = \begin{pmatrix} a & b & c & d \\ b & a & d & c \\ c & d & a & b \\ d & c & b & a \end{pmatrix} \tag{8.60}$$

where b, c, and d involve the second differentials of the interactions between atoms 1 and 2, 1 and 3, and 1 and 4 respectively, and will have negative values. It can be shown that $a = -(b+c+d)$. We consider in turn the four eigenvectors, labelled according to their irreducible representations:

$$\mathbf{e}_1(B_{1u}) = \begin{pmatrix} +1 \\ +1 \\ +1 \\ +1 \end{pmatrix} \qquad \mathbf{e}_2(A_{1g}) = \begin{pmatrix} +1 \\ +1 \\ -1 \\ -1 \end{pmatrix}$$

$$\mathbf{e}_3(B_{3g}) = \begin{pmatrix} +1 \\ -1 \\ +1 \\ -1 \end{pmatrix} \qquad \mathbf{e}_4(B_{2u}) = \begin{pmatrix} +1 \\ -1 \\ -1 \\ +1 \end{pmatrix}$$

We now construct the equations of the form $\mathbf{D} \cdot \mathbf{e}_j$, and if the proposed vectors $\mathbf{e}_j$ are true eigenvectors of $\mathbf{D}$ we expect to find $\mathbf{D} \cdot \mathbf{e}_j = \omega_j^2 \mathbf{e}_j$, where ω_j^2 will be a function of the components of $\mathbf{D}$. Indeed, by forming the matrix product $\mathbf{D} \cdot \mathbf{e}_j$ for each j we find that the vectors $\mathbf{e}_j$ are all eigenvectors of $\mathbf{D}$, with

$$\omega_1^2 = a + b + c + d \qquad \omega_2^2 = a + b - c - d$$
$$\omega_3^2 = a - b + c - d \qquad \omega_4^2 = a - b - c + d$$

From the condition $a + b + c + d = 0$ noted earlier, we have the result that $\omega_1 = 0$, as expected for the acoustic mode eigenvector when $k \to 0$.

8.5.6 Lattice dynamics of ionic crystals

The dynamical matrix has two parts. The first is the long-range Coulomb part. This extends to infinite range, and needs to be handled with an Ewald summation as briefly discussed in Section 5.4.1. The second part consists of the short-range interactions. This part can be treated in two ways. On one hand, the individual values of the $J_{j,j',n}$ parameters can be taken as adjustable parameters, tuned to give the best agreement with the experimental measurements of the dispersion curves. On the other hand, the $J_{j,j',n}$ parameters can be written as the differentials of an assumed potential energy function, with a form such as the Buckingham potential (eqn 5.23). In this case the same function can be used to determine values of $J_{j,j',n}$ for different neighbouring unit cells. This has the advantage of tying the force constants to a physical model, which has fewer adjustable parameters which can also be optimized against structural parameters.

Early in the development of lattice dynamics models that were being used to describe the first experimental data on lattice dynamics (the early 1960s) it was realized that it is necessary to take account of the ionic polarizability. This is highlighted by a relationship for simple cubic ionic materials called the **Lyddane–Sachs–Teller** (LST) relation. This relates the limiting values of the LO and TO modes as $k \to 0$ to the dielectric constants for electric fields of zero and infinite frequency through the simple equality of ratios

$$\frac{\epsilon(\omega = 0)}{\epsilon(\omega = \infty)} = \frac{\omega_{\mathrm{LO}}^2}{\omega_{\mathrm{TO}}^2} \tag{8.61}$$

The frequencies in the ratio are for the LO and TO modes with similar atomic motions. For example, for alkali halides, the LST relation might apply to vibrations with wave vectors of the form $\mathbf{k} = [\eta, 0, 0]$ as $\eta \rightarrow 0$. In this case the LO modes have the cations and anions moving in opposite directions parallel to [100], and the TO modes have the cations and anions moving in opposite directions perpendicular to [100]. Since both types of motion are along symmetrically equivalent directions, when $k = 0$ the types of motion become exactly equivalent. In this case one might have expected that the LO and TO frequencies would be equal. But for wave vectors close to but not exactly at $k = 0$, the fact that the long-wavelength optic modes generate electric fields that are either parallel or perpendicular to the direction of propogation of the optic mode will have a significant effect on the frequency of the mode. Since $\epsilon(\omega = 0) > \epsilon(\omega = \infty)$, it follows that $\omega_{\mathrm{LO}} > \omega_{\mathrm{TO}}$. This effect is known as **LO/TO splitting**, reflecting the fact that at exactly zero wave vector the frequencies are the same. We will see examples of the size of the LO/TO splitting in data for alkali halides below.

The importance of ionic polarizability is that it determines the size of $\epsilon(\omega = \infty)$, which would have a value of 1 without polarizability. It also makes a significant contribution to $\epsilon(\omega = 0)$, which would otherwise only be determined by the polarizability of the crystal structure through displacements of the ions. The shell model introduced in Section 5.4.5 works reasonably well in reproducing experimental values of the two dielectric constants.

The shell model can be incorporated into the formalism of the dynamical matrix by writing separate equations of motion for the core and the shell. Because the mass of the shell is zero in the shell model, there is no separate $m\omega^2$ for the shell. Instead, the shell instantaneously relaxes to its equilibrium position so that there is no restoring force on the shell. Thus the displacement of the shell is completely tied to the displacements of the cores, and therefore there is not a separate solution for the shells. The contribution of the shell folds into an expanded dynamical matrix for the ionic cores, giving the same number of solutions as for models without ionic shells.

8.5.7 The lattice dynamics of the alkali halides

The alkali halides were among the first crystals whose dispersion curves were measured by the emerging technique of inelastic neutron scattering. Measured dispersion curves for NaCl are shown in Fig. 8.16, using the same reduced scheme as shown earlier for the dispersion curves of fcc monatomic crystals. There are two ions in the primitive cell, so there are three acoustic and three optic modes. For wave vectors along the [001] and [111] symmetry directions the TO and TA modes are both degenerate, similar to the degeneracy of the TA modes in argon, lead and potassium (Figs 8.7, 8.9 and 8.10 respectively).

The dispersion curves for the alkali halides show two important features of the lattice dynamics of ionic crystals. First, the data highlight the significance of the LO/TO splitting. In NaCl, the LO and TO modes have frequencies of around 8 and 5 THz respectively, which are significantly different. From eqn 8.61, the ratio of the squares of the LO and TO frequencies suggests that the static dielectric constant is 2.5 times larger than the high-frequency dielectric constant. In fact the experimental values are $\epsilon(\omega = 0) = 5.90$ and $\epsilon(\omega = \infty) = 2.34$, giving, from eqn 8.61, the ratio $\omega^2(\mathrm{LO})/\omega^2(\mathrm{TO}) = 2.52$.

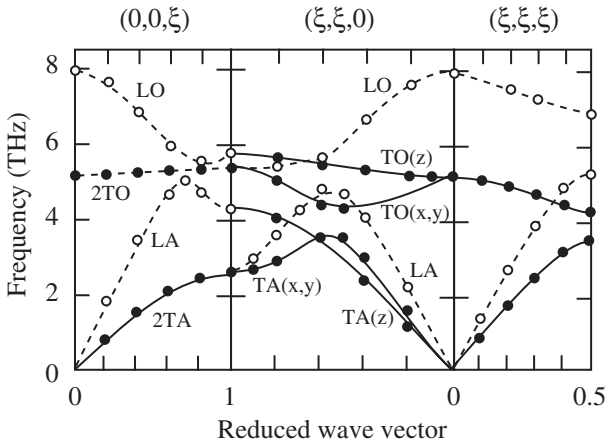


Fig. 8.16 Acoustic and optic mode dispersion curves for NaCl along the main symmetry directions, as measured by inelastic neutron scattering. (Data taken from Raunio *et al.*, *Phys. Rev.* **178**, 1496, 1969.)

A second feature to note is that there is not a separation of optic and acoustic mode frequencies across the range of wave vectors. The simple model of the diatomic chain given earlier showed a separation of the optic and acoustic mode frequencies, but this is an artefact of the simple nature of the model. Consider the two sets of modes for wave vectors along [111]. In this case both the transverse and longitudinal modes show a clear separation of the optic and acoustic modes. In this case there are alternating planes of cations and anions normal to [111] as in the simple diatomic chain, so we might expect the behaviour to resemble the simple model. However, for wave vectors along [100] the planes of atoms normal to this direction contain both cations and anions, so there is much less of a match with the diatomic chain model. The transverse modes do indeed show a separation of the optic and acoustic modes, but there is not a separation of the LO and LA modes. Indeed, for NaCl the dispersion curves for these two types of modes appear to cross. In fact symmetry plays an important role in this case, because the two types of vibration have the same symmetry. Modes with the same symmetry are not permitted to cross in the dispersion diagram, but will appear to repel each other as they start to get close on changing wave vector. The apparent near-crossing, called **anti-crossing**, of the LO and TO modes, and of the LA and TA modes, is also seen for the dispersion curves along [110].

8.5.8 The lattice dynamics of quartz

We choose the example of quartz to show the dispersion curves of a more complex insulating material. The bonds in quartz show elements of both covalent and ionic bonding. The SiO_4 tetrahedra are held together by strong covalent bonds, but the ions do not lose all their charge, so there are important electrostatic interactions. The dispersion curves have been measured by inelastic neutron scattering, and are shown in Fig. 8.17. Quartz is a good example because it is rather more complex than the alkali halides without being too complicated. There are nine atoms in the unit cell, so there are 27 dispersion curves (three acoustic, 24 optic).

Quartz has trigonal symmetry. For wave vectors along [001] there is a three-fold rotation axis, with three types of symmetry for the vibrations. In Fig. 8.17

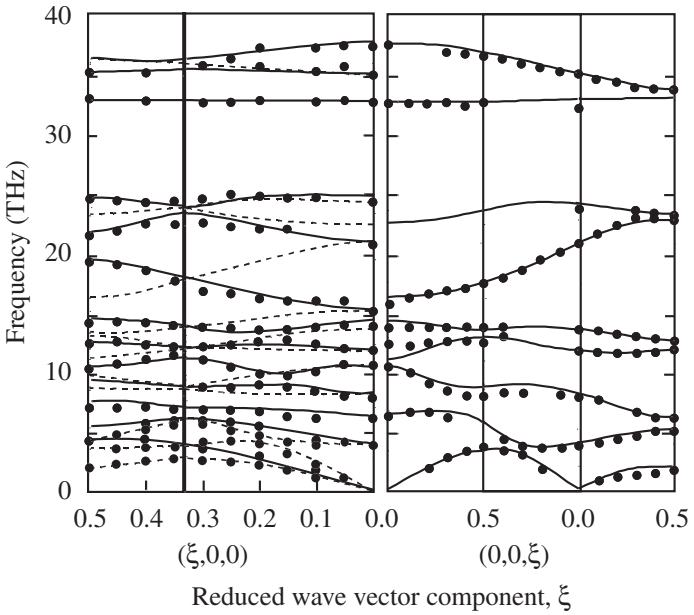


Fig. 8.17 Dispersion curves for quartz along two symmetry directions, noting the extended mode scheme for the [001] data, and the presence of two symmetry representations for the [100] data. The curves represent the results of a harmonic lattice dynamics calculation using force constants whose values have been fitted against the data. (Data are taken from Strauch and Dorner, *J. Phys.: Cond. Matter*, **5**, 6149, 1993.)

the modes of different symmetry for the [001] direction are plotted on separate diagrams, called an **extended zone scheme**. The order of these diagrams reflects the cases where modes become degenerate for the wave vectors exactly at the $k = 0$ or at the Brillouin zone boundary. For vibrations along [100] there are only two types of symmetry, and the dispersion curves are plotted on the same diagram.

The data for quartz show how complex the dispersion curves can be for complicated crystal structures. Some of the branches have quite a lot of variation of their frequency, whereas others are more constant. The frequency range extends nearly to 40 THz. This is five times larger than in NaCl, and since ω^2 is directly proportional to a force constant, it can be deduced that there are forces that are 25 times stronger in quartz than in NaCl. This is a reflection of the covalent bonds in quartz, which are particularly strong for vibrations involving stretching of the Si–O bonds.

The measurements of the dispersion curves in quartz represent a significant achievement. As will be shown in Chapter 10, neutron scattering methods are much easier for low-frequency modes, and often it is not possible to measure dispersion curves across the whole range of frequencies.

Summary of chapter

- The key assumption in the development of the theory of lattice dynamics is that atomic displacements are small so that the energy can be described as a **harmonic** function of the relative displacements of neighbouring atoms. Furthermore, it is assumed that if this approximation needs to be relaxed, it will be possible to modify the results of the harmonic model to account for the higher-order contributions to the energy.

- Within the **harmonic approximation**, equations of motion for all atoms are solved assuming that the atomic displacements can be described by **harmonic travelling waves**. The solutions of the equations give the frequency as a function of wave vector, $\omega(\mathbf{k})$. Plots of $\omega(\mathbf{k})$ are called **dispersion curves**. For systems with more than one atom in the unit cell, the solutions of the equations of motion will also give the relative displacement of each atom.
- Models of crystals with one atom in the unit cell give dispersion curves for which $\omega(\mathbf{k}) \rightarrow 0$ as $\mathbf{k} \rightarrow 0$. These solutions correspond to sound waves, and are called **acoustic modes**. Models with d dimensions have d acoustic modes.
- Models for three-dimensional monatomic crystals are solved and compared with experimental measurements of dispersion curves. The main features of the dispersion curves can be understood from the dominance of the **nearest-neighbour interatomic interactions**.
- Models with Z atoms and d dimensions have dZ modes of vibration for each wave vector, of which d are the acoustic modes. The acoustic modes are the vibrations for small $\mathbf{k}$ in which all the atoms in the unit cell move in phase, and the other modes represent all other combinations of the phase of atomic displacements. The vibrations in which atoms in the unit cell move with some atoms out of phase with others are called the **optic modes**.
- In general $\partial\omega^2/\partial\mathbf{k} = 0$ for $\mathbf{k} = 0$ and $\mathbf{k}$ half-way between reciprocal lattice vectors. The exception is where two modes become degenerate only at the latter point, in which case the sum of $\partial\omega/\partial\mathbf{k}$ for the two modes is equal to zero.
- The equations of motion can be conveniently represented by the formalism of the **dynamical matrix**, $\mathbf{D}(\mathbf{k})$. The eigenvalues of $\mathbf{D}(\mathbf{k})$ give $\omega^2(\mathbf{k})$, and the eigenvectors give the relative atomic displacements of each mode.

Further reading

Introductions of the subject of lattice dynamics are given by Kittel (1996) and Ashcroft and Mermin (1976). More detailed treatments of the topic are given by Barron and White (1999), Böttger (1983), Di Bartolo and Powell (1976), Dove (1993), Elliot (1998), Hardy and Karo (1979), Kosevich (1990), and Ziman (2000). Ferroelectric materials are discussed in detail by Lines and Glass (1977) and Mitsui and Tatsuzaki (1976).

Exercises

(8.1) From a sketch of the atomic displacements associated with the [110] and [1$\bar{1}$0] zero boundary acoustic modes in a ccp crystal, demonstrate that the LA and one of the TA modes are equivalent at the $\langle 110 \rangle$ zone boundary wave vector.

(8.2) Derive eqn 8.31 for the nearest-neighbour interaction. (*Hint*: start with $\partial r/\partial x = x/r$, which comes from $r^2 = x^2 + y^2 + z^2$, and note that for models with only nearest-neighbour interactions $\partial\phi/\partial r = 0$.)

(8.3) Derive the expressions for the transverse and longitudinal interplanar force constants for neon for wave vectors along the (a) [100], (b) [111], and (c) [110] directions as given in eqns 8.32, 8.33–8.4.1. (*Hint*: for each direction of the wave vector use a set of axes in which the longitudinal direction is along one of the new axes, and express the vectors between an atom and its neighbours as a vector in this new set of axes.)

(8.4) From the dispersion curves for neon (Fig. 8.7) obtain an estimate for the nearest-neighbour interatomic force constant K (Section 8.4.1). Assuming that the interaction between atoms can be represented using a Lennard-Jones potential 5.23, show that

$$K = \frac{72\epsilon}{2^{1/3}\sigma^2}$$

Compare your estimate of the value of K with that calculated from values of ϵ and σ obtained in Problem 5.1.

(8.5) The model of the monatomic chain (Section 8.3) can be extended by including a local interaction for each atom of the form

$$E_{\text{local}}(u) = \frac{1}{2}\alpha u^2$$

Show that the effect of an additional local potential energy is to shift the dispersion curve for the case with nearest-neighbour interactions to

$$m\omega^2 = \alpha + 4J\sin^2 ka/2 \qquad (8.62)$$

(8.6) Explain why the LA(100) and TA(100) modes for potassium (Fig. 8.10) have very similar dispersion curves, by deriving the expressions for the interplanar force constants based on nearest-neighbour interatomic interactions given in Section 8.4.3.

(8.7) Estimate the relative sizes of the three force constants that describe the dispersion curve of the LA mode along [111] in potassium (Fig. 8.10, Section 8.4.3). Explain this result.

(8.8) The crystal structure of the linear molecule HCN is I-centred tetragonal, with one molecule per lattice point oriented along [001]. The two-dimensional **a**–**b** section of the crystal is shown below. If there is a single force constant between nearest-nieghbour molecules for motions in the **a**–**b** plane as shown in the structure plan, show that (a) the frequencies of the LA and TA(y) modes along [100] are identical; (b) the frequency of the LA mode along [110] reaches the same maximum value as the LA model along [100]; (c) the TA(x–y) model for all wave vectors along [110] has zero frequency within the approximation of a single force constant. The latter result gives a ferroelastic phase transition in this material at a temperature of 170 K (Section 12.1.2).

(8.9) Repeat the analysis leading to eqn 8.46 from the equation of motion for the small atom in the model for the diatomic chain (Section 8.5), showing that the same result is obtained.

(8.10) From the general solution for the harmonic diatomic chain with unequal masses, eqn 8.4.9:

$$\omega^2 = \frac{J(M+m)}{Mn} \pm \frac{J}{Mn}\sqrt{(M+m)^2 - 4Mm\sin^2(ka/2)}$$

obtain the acoustic and optic mode solutions in the limit $k \to 0$ and at $k = \pi/a$.

(8.11) Recast the harmonic model for the diatomic chain (Section 8.5) for the case where the atoms are equivalent, but with different force constants between the two neighbours of each atom, and obtain the solution analogous to eqn 8.49. Such a model could represent the lattice dynamics of the diamond or hcp structures. Show that the atomic displacements of the two modes at $k = \pi/a$ are different from the case of unequal masses but identical force constants (eqn 8.51).

(8.12) Compare the measured dispersion curves of NaCl (Fig. 8.16) for a wave vector along [111] with the prediction of the simple diatomic chain (eqn 8.49).

(8.13) Show that the one-dimensional dynamical matrix for the crystal of Fig. 8.15 has the form given by eqn 8.60.

9 Thermodynamics and lattice dynamics

9.1 The quantization of lattice vibrations

9.1.1 Phonons: the quanta of harmonic lattice vibrations

The harmonic theory of lattice dynamics developed in the previous chapter has allowed us to determine the distribution of vibrational frequencies in a crystal structure, but it has given no information about the amplitudes of the vibrations, or equivalently about the sizes of the atomic displacements. The vibrations will be excited by the effects of temperature, with higher temperatures leading to large amplitudes, but the theory does not yet include temperature. The theory has been firmly based in classical mechanics, but we might expect that a proper treatment that includes temperature will necessarily incorporate quantum mechanics. This is the main point of this chapter.

The harmonic theory of lattice dynamics lends itself to an analogy with the theory of electromagnetic radiation. The classical theory of electromagnetism leads to a classical wave equation for electromagnetic waves that is not modified in substance when quantum mechanics is brought into the picture. Instead, the effect of quantum mechanics is to quantize the harmonic electromagnetic field without changing properties such as frequency. The quanta of the electromagnetic field are photons, which have energy $\hbar\omega$, where $\hbar$ $(= h/2\pi)$ is Planck's constant. This leads to the wave–particle duality of the electromagnetic field seen, for example, in the photoelectric effect. All harmonic vibrations can be quantized in the same way. The fundamental quanta of lattice vibrations are called **phonons**, by analogy with the quanta of light (the link of this name to sound is clear), and like photons these quanta have energy $\hbar\omega$. The values of ω are unchanged from the classical values that can be calculated using the methods of the previous chapter.

One of the properties of harmonic vibrations that arises purely from the quantization is that the lowest energy state (**ground state**) is not zero energy as it would be in a classical picture, but is equal to

$$E_0 = \frac{1}{2}\hbar\omega \tag{9.1}$$

This is called the **zero-point energy**, and the corresponding motions at $T = 0\,\mathrm{K}$ are called the **zero-point motions**. The point is that in the quantum picture, the vibration cannot stop, and that the last bit of energy cannot be removed from the vibration. The zero-point motions can have significant consequences for low-temperature behaviour, as we will see later in this chapter and in Chapter 12.

The energy of the vibration can be changed by integral units of the phonon energy, $\hbar\omega$. The mean energy of each vibrational mode is then given as:

$$E = \hbar\omega\left[\frac{1}{2} + n(\omega, T)\right] \tag{9.2}$$

where $n(\omega, T)$ is the number of phonons with frequency ω at temperature T, and is called the **phonon number**.

9.1.2 The Bose–Einstein relation, $n(\omega, T)$

An equation for $n(\omega, T)$ was obtained by Bose and Einstein, after whom it is named. The **Bose–Einstein relation** can be obtained using statistical mechanics. If a system has a variable q, and the energy $E(q)$ can be written as a function of q, the equilibrium value of q can be obtained from

$$\langle q\rangle = \frac{\sum_q q\exp(-\beta E(q))}{\sum_q \exp(-\beta E(q))} \tag{9.3}$$

where we have written $\beta = 1/k_B T$ as a convenient shorthand, and the sums are over all possible values of q. The denominator is given a special status in statistical mechanics, and is called the **partition function**, Z. The reader is referred to Appendix L for a general discussion on the partition function, with examples of how it can be used to calculate equilibrium values of properties.

Now we consider a system with a vibration of angular frequency ω, and write the quantum of this vibration as $\epsilon = \hbar\omega$. The energy of this vibration is equal to $n\epsilon$, so the partition function can be written as

$$Z = \sum_n \exp(-\beta n\epsilon) \tag{9.4}$$

Using the standard approach in statistical mechanics, the average value of n is given by

$$\begin{aligned}\langle n\rangle &= \frac{1}{Z}\sum_n n\exp(-\beta n\epsilon)\\ &= -\frac{1}{\beta Z}\frac{\partial Z}{\partial \epsilon}\end{aligned} \tag{9.5}$$

We now use a standard result to evaluate the partition function:

$$\sum_n x^n = \frac{1}{1-x} \tag{9.6}$$

Using the substitution $x = \exp(-\beta\epsilon)$, so that $x^n = \exp(-\beta n\epsilon)$, we obtain

$$Z = \frac{1}{1-\exp(-\beta\epsilon)} \tag{9.7}$$

Evaluation of eqn 9.5 with this form of the partition function gives

$$\langle n\rangle = Z\exp(-\beta\epsilon) = \frac{1}{\exp(\beta\epsilon)-1} \tag{9.8}$$

This result is the Bose–Einstein distribution, which we write in long-hand as

$$n(\omega, T) = \frac{1}{\exp(\hbar\omega/k_B T)-1} \tag{9.9}$$

This function is shown in Fig. 9.1.

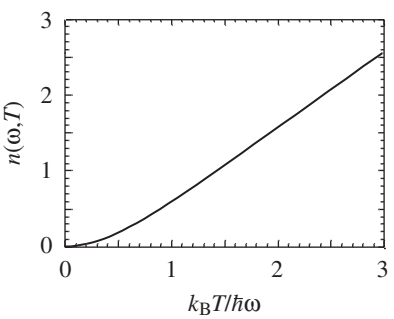


Fig. 9.1 The Bose–Einstein distribution $n(\omega, T)$ as a function of $k_B T/\hbar\omega$.

9.1.3 High-temperature behaviour

At high temperatures, defined as $k_B T > \hbar\omega$, we can expand the energy function:

$$\begin{aligned}\frac{1}{2} + n(\omega, T) = \frac{1}{2} + \frac{1}{\exp(\beta\epsilon) - 1} &= \frac{\exp(\beta\epsilon) + 1}{2[\exp(\beta\epsilon) - 1]} \\ &= \frac{(1 + \beta\epsilon + \cdots) + 1}{2[(1 + \beta\epsilon + \beta^2\epsilon^2/2 \cdots) - 1]} \\ &= \frac{2 + \beta\epsilon + \cdots}{\beta\epsilon(2 + \beta\epsilon + \cdots)} \\ &\approx \frac{1}{\beta\epsilon} = \frac{k_B T}{\hbar\omega} \qquad (9.10)\end{aligned}$$

Thus in the high-temperature limit, the energy of a single vibration is given as

$$E = \hbar\omega\left[\frac{1}{2} + n\right] = k_B T \qquad (9.11)$$

Since there are three vibrations for each atom, the vibrational energy of a mole of atoms at high temperatures is given as

$$E = \sum \hbar\omega\left[\frac{1}{2} + n\right] \rightarrow 3RT \qquad (9.12)$$

This is equivalent to the internal energy.

9.1.4 Heat capacity

The heat capacity at constant volume, c_V, is equal to

$$\begin{aligned}c_V &= \left(\frac{\partial E}{\partial T}\right)_V \\ &= \sum \hbar\omega \frac{\partial n}{\partial T} \qquad (9.13) \\ &= \sum k_B \left(\frac{\hbar\omega}{k_B T}\right)^2 \frac{\exp(\hbar\omega/k_B T)}{[\exp(\hbar\omega/k_B T) - 1]^2} \qquad (9.14)\end{aligned}$$

(Note that the harmonic model of lattice dynamics does not allow for thermal expansion, so all thermodynamic functions are for constant volume.) In the high-temperature limit, this expression reduces to $3R\ \mathrm{J\,K^{-1}}$ per mole of atoms, consistent with the expression for the energy in the high-temperature limit. This is the classical **Dulong–Petit** result, noted empirically in the early 19th century. However, it is known that, contrary to the classical result, the heat capacity falls to zero on cooling to zero temperature. An example of the heat capacity for the mineral andalusite, Al_2SiO_5, is shown in Fig. 9.2, and compared with calculations using eqn 9.14 and vibrational frequencies obtained using the empirical models discussed in Chapter 5. It can be seen that at high temperature the heat capacity is close to $24R\ \mathrm{J\,K^{-1}}$ per mole of formula units (eight atoms in the formula unit), consistent with the Dulong–Petit result. This example also shows the decrease in the heat capacity on cooling. In fact, the heat capacity in this material falls significantly below the Dulong–Petit value at quite high temperatures. This is correlated with the bonds in this material being quite strong (specifically the Si–O and Al–O bonds), leading to high vibrational frequencies.

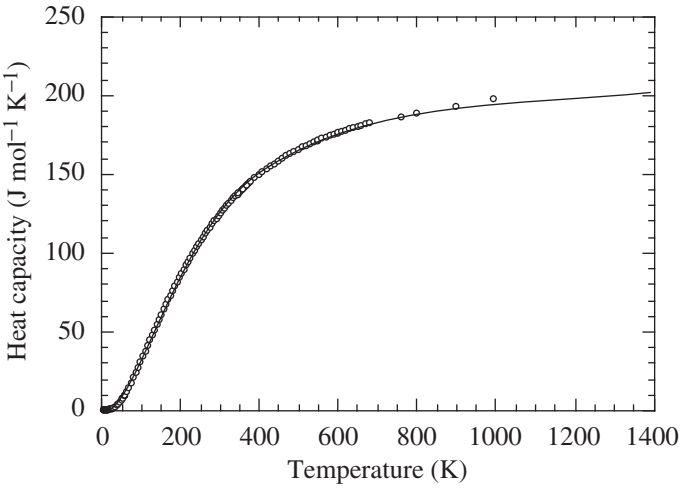


Fig. 9.2 Comparison of calculated (line) and experimental (points) heat capacity of the mineral andalusite, Al_2SiO_5.

9.1.5 Phonon free energy and entropy

The partition function for phonons, which was used to obtain the Bose–Einstein relation, also gives the free energy of the phonons. In Appendix L it is shown that the free energy is given by

$$F = -\frac{1}{\beta} \ln Z \tag{9.15}$$

When we use eqn 9.7 for Z for phonons, the resultant free energy does not include the contribution from the zero-point vibrations, and this must be added to the phonon free energy. Using the phonon partition function, eqn 9.7 and the zero-point energy, eqn 9.1, we obtain the full free energy for a single vibration:

$$F = \frac{1}{2}\epsilon + \frac{1}{\beta} \ln[1 - \exp(-\beta\epsilon)] \tag{9.16}$$

$$= \frac{1}{\beta} \ln[2 \sinh(\beta\epsilon/2)] \tag{9.17}$$

The second form of the free energy function will be used in Chapter 12 in the study of phase transitions. The derivation is kept for Problem 9.2. In the high-temperature limit, $\beta\epsilon < 1$, we use the small-argument limit of the sinh function to obtain

$$F = \frac{1}{\beta} \ln(\beta\epsilon) \tag{9.18}$$

The entropy, S, is given as $S = -\partial F/\partial T$. In the high-temperature limit, we have

$$S = k_B[1 - \ln(\beta\varepsilon)]$$

9.2 Thermodynamic functions for crystals

9.2.1 Thermodynamic functions

The thermodynamic functions, eqns 9.2–9.17, have been defined for a single vibration that is quantized with energy levels that differ by multiples of $\epsilon = \hbar\omega$. For a crystal the thermodynamic functions have to be summed over all vibrations. Thus, for example, the free energy of a crystal will be written as

$$F = \frac{1}{\beta} \sum_{\mathbf{k},\nu} \ln\{2 \sinh[\beta\hbar\omega(\mathbf{k}, \nu)/2]\} \tag{9.19}$$

Using a model interatomic potential, or a model for the interatomic force constants fitted to vibrational data, it is straightforward to compute a long list of vibrational frequencies for a grid of wave vectors in reciprocal space, and use these in the summation as an approximation to the summation over all normal modes of a crystal containing of order N_A atoms. This procedure will give reasonably good results, as shown in the comparison of the calculated heat capacity of andalusite with experimental data in Fig. 9.2.

9.2.2 The Einstein model

Long lists of frequencies are fine for calculations of values of the thermodynamic functions, but they are not readily ported to general theories. As a result it would be ideal to be able to replace the long list by an average value. This approach was used by Einstein in his pioneering studies of the thermodynamics of quantum crystals (which appeared two years after he proposed his theory of special relativity – although $E = mc^2$ has more popular appeal than the equation for the heat capacity of a solid, this particular achievement of Einstein should not be unappreciated). If we denote the average angular frequency as $\langle\omega\rangle$, the free energy and heat capacity for a crystal containing one mole of atoms can be written as

$$F = \frac{3N_A}{\beta} \ln[2 \sinh(\beta\hbar\langle\omega\rangle/2)] \tag{9.20}$$

$$c_V = 3R(\beta\hbar\langle\omega\rangle)^2 \frac{\exp(\beta\hbar\langle\omega\rangle)}{[\exp(\beta\hbar\langle\omega\rangle) - 1]^2} \tag{9.21}$$

The factor of 3 arises from the fact that there are three degrees of freedom associated with each atom (Chapter 8). We will make use of the **Einstein model** in Chapter 12 to develop of thermodynamic models of phase transitions.

The Einstein model provides a reasonable qualitative description of the heat capacity of a crystal, but has not been adopted for regular use in this regard because it fails to reproduce the behaviour at very low temperature. This is because at low temperature the energy of the Einstein phonon, $\hbar\langle\omega\rangle$, is significantly above the thermal energy, $k_B T$, and is barely excited. On the other hand, in a real crystal there will always be some acoustic modes, namely those with small wave vectors, whose phonon energies are lower than the thermal energy. It is this failure of the Einstein model to represent the contribution of the acoustic modes to the thermodynamic functions, particularly at low temperature when the contribution of the acoustic modes dominates the thermodynamic

functions, that has limited the use of the model to cases where only the behaviour of the optic modes is important. The **Debye model**, which we will shortly discuss, overcomes this failure.

9.2.3 Density of states

One common approach is to replace the long list of frequencies by a distribution of frequencies, known as the **density of states**. Specifically, the density of states, $g(\omega)$, is defined such that the number of vibrations with angular frequencies between ω and $\omega + \mathrm{d}\omega$ is given by $g(\omega)\,\mathrm{d}\omega$. The density of states can readily be obtained from a long list of computed frequencies by standard histogram methods.

Within the formalism of the density of states, the internal energy and heat capacity can be written in an integral form:

$$U = \int \left(n(\omega) + \frac{1}{2} \right) \hbar\omega g(\omega)\,\mathrm{d}\omega \tag{9.22}$$

$$c_V = \int \frac{\partial n}{\partial T} \hbar\omega g(\omega)\,\mathrm{d}\omega \tag{9.23}$$

Similar equations can be written for the free energy or other thermodynamic functions.

Although for many purposes the density of states has few advantages over the long list of frequency values, there are cases where the density of states can be obtained directly without first obtaining the list of frequency values. First, neutron scattering methods (Chapter 10) can be used to measure density of states functions weighted by the neutron atomic scattering factors. Second, the computer simulation method called **molecular dynamics simulation** generates the time-varying positions and velocities of a sample of atoms using a numerical integration of Newtonian mechanics, and the vibrational density of states can be computed from the analysis of the velocities with much less effort and more accuracy than for the calculation of individual frequency values (see Appendix F of Dove, 1993, for example). From the perspective of providing a concise description of the properties of a material, the density of states provides a nice representation of the spread of frequency values in a crystal. Perhaps the most important application of the density of states is in determining the form of the heat capacity of a crystal at low temperatures.

9.2.4 Density of states for acoustic modes

The density of states for the optic modes depends only on the force constants and has no particular shape. However, there is a particular shape for the density of states for the acoustic modes because there is an exact dependence of frequency on wave vector as $k \to 0$, namely

$$\omega = ck \tag{9.24}$$

where c is an average velocity of sound, and k is the modulus of the wave vector (Chapter 8). To obtain the frequency density of states we make use of the distribution of wave vectors. The number of wave vectors with values

between k and $k + \mathrm{d}k$ is given by $g(k)\,\mathrm{d}k$, and by relating the two distribution functions we have

$$g(\omega)\,\mathrm{d}\omega = g(k)\,\mathrm{d}k \tag{9.25}$$

for corresponding values of ω and k. $g(k)$ is relatively straightforward to evaluate since the individual wave vectors are distributed across a lattice grid in reciprocal space. In a crystal containing N unit cells and with total volume V, the volume of the unit cell is equal to V/N and the volume of the Brillouin zone (equal to that of the reciprocal unit cell) is $V_{\mathrm{BZ}} = (2\pi)^3 N/V$ (the same grid was used to count electron states in Section 5.4.7). The crystal is defined as containing N unit cells, so there will be N wave vectors in one Brillouin zone, and the number of wave vectors per unit volume of reciprocal space will be $N/V_{\mathrm{BZ}} = V/(2\pi)^3$. The number of wave vectors in a shell of radius k and thickness $\mathrm{d}k$ is therefore equal to

$$g(k)\,\mathrm{d}k = \frac{V}{(2\pi)^3}\,4\pi k^2\,\mathrm{d}k \tag{9.26}$$

Substituting $k = \omega/c$ and $\mathrm{d}k = \mathrm{d}\omega/c$, we have

$$\begin{aligned} g(\omega)\,\mathrm{d}\omega &= 3\frac{V}{(2\pi)^3}4\pi\left(\frac{\omega}{c}\right)^2\frac{\mathrm{d}\omega}{c} \\ &= \frac{3V\omega^2}{2\pi^2 c^3}\,\mathrm{d}\omega \end{aligned} \tag{9.27}$$

The factor of 3 comes from the fact that there are three acoustic modes for each wave vector. The important feature of this result is that $g(\omega) \propto \omega^2$. This will always be the limiting form of $g(\omega)$ as $\omega \to 0$ for all crystals that have normal sound waves, for frequencies up to the lowest optical mode frequencies.

9.2.5 Debye model of heat capacity

The **Debye model** of the heat capacity is developed from the functional form $g(\omega) \propto \omega^2$ for acoustic modes. In this form, the phonon energy is given as

$$E = \int_0^{\omega_{\mathrm{D}}} \frac{3V\omega^2}{2\pi^2 c^3}\hbar\omega\frac{1}{\exp(\hbar\omega/k_{\mathrm{B}}T) - 1}\,\mathrm{d}\omega \tag{9.28}$$

where ω_{D} is an effective cut-off frequency called the **Debye frequency**. It is possible to treat ω_{D} as a free variable whose value can be adjusted to give the best agreement with experimental data.

The heat capacity follows as

$$c_V = \frac{\partial E}{\partial T} = \int_0^{\omega_{\mathrm{D}}} \frac{3V\omega^2}{2\pi^2 c^3}\hbar\omega\frac{\partial n}{\partial T}\,\mathrm{d}\omega \tag{9.29}$$

It is common to represent the heat capacity of a solid using this function, adjusting ω_{D} to give the best agreement with experiment, or indeed, obtaining a value of ω_{D} for each temperature.

At low temperature, only low-frequency phonons can be excited, since the Bose–Einstein function falls to zero for values of ω that are a long way short of

ω_D. Thus the value of the upper limit on the integral is not important (so long as it is larger than the highest frequency that can be excited at the temperature), so it can be taken to be infinity. We make the substitution $x = \hbar\omega/k_B T$. Thus $\omega^3\,d\omega = (k_B T/\hbar)^4 x^3\,dx$, and

$$E = \frac{3V\hbar}{2\pi^2 c^3}\left(\frac{k_B T}{\hbar}\right)^4 \int_0^\infty x^3 (e^x - 1)^{-1}\,dx \tag{9.30}$$

The integral is of a standard form, and has the result $\pi^4/15$. Thus the energy can be written as

$$E = \frac{V\pi^2 (k_B T)^4}{10(c\hbar)^3} \tag{9.31}$$

The important point is that $E \propto T^4$, so that $c_V \propto T^3$. Specifically,

$$c_V = \frac{2V\pi^2 k_B}{5(c\hbar/k_B)^3} T^3 \tag{9.32}$$

A commonly used parameter is the **Debye Temperature**, θ_D. This is defined as $\theta_D = \hbar\omega_D/k_B$. It can be shown (Problem 9.3) that for a crystal of **N** atoms and volume **V**,

$$\theta_D = \frac{\hbar c}{k_B}\left(\frac{6\pi N^2}{V}\right)^{1/3}. \tag{9.33}$$

In this case, the heat capacity can be written as

$$c_V = \frac{12\pi^4 N k_B}{S}\left(\frac{T}{\theta_D}\right)^3 \tag{9.34}$$

9.2.6 Example of thermodynamic functions of fluorite, CaF_2

The main conclusions of the preceding discussion can be illustrated by calculations of the thermodynamic properties of fluorite, CaF_2. Phonon frequencies were calculated using a model empirical potential energy function of the type described in Chapter 5. This included a Buckingham F. . .F potential, a Born–Mayer Ca. . .F potential, and a shell model for the F^- anions, together with formal charges for the cation and anion. The parameters in the model were optimized to give best agreement with the crystal structure and phonon dispersion curves.

The density of states calculated from this model is shown in Fig. 9.3. The ω^2 dependence is clear for a wide range of frequencies above $\omega = 0$. The various peaks arise from specific optic modes.

The heat capacity constructed from the phonon calculations is shown in Fig. 9.4. At high temperatures the heat capacity tends towards the classical result $9R = 74.83\,J\,K^{-1}\,mol^{-1}$. At low temperatures the heat capacity shows the characteristic T^3 form from the acoustic modes. The calculated internal energy, free energy, and entropy are shown in Fig. 9.5.

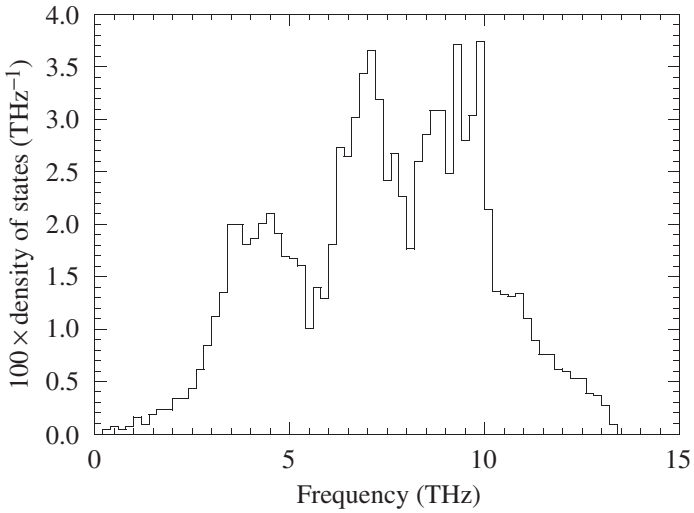


Fig. 9.3 Density of states calculated for fluorite, CaF_2. Note that at low frequencies there is the characteristic ω^2 form arising from the acoustic modes.

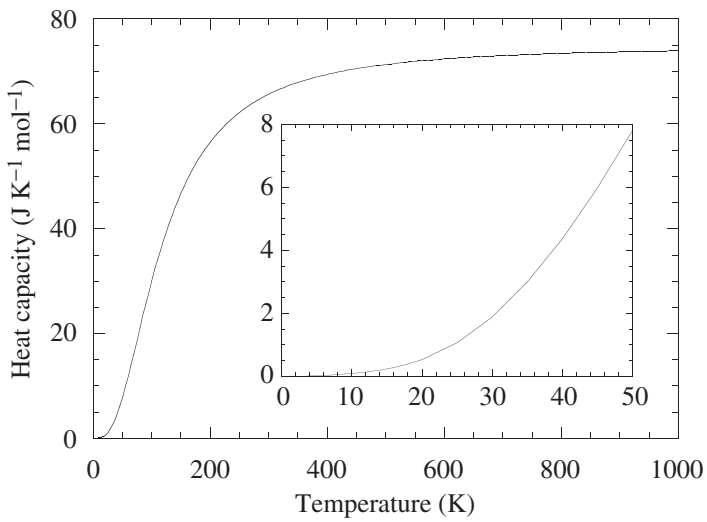


Fig. 9.4 Heat capacity of fluorite, CaF_2, calculated from the phonon frequencies. Note that at low frequencies there is the characteristic T^3 form arising from the acoustic modes.

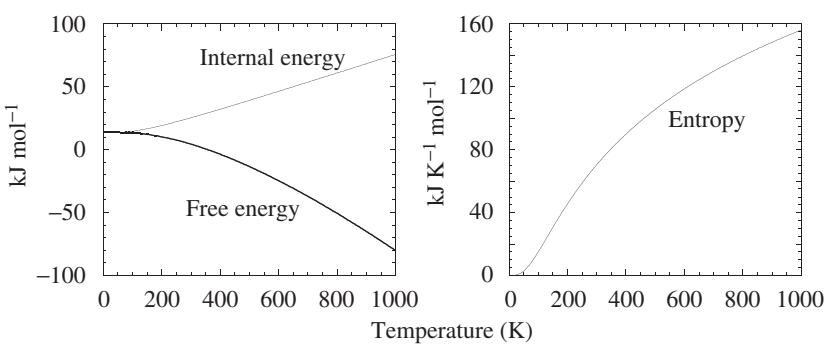


Fig. 9.5 Free energy, internal energy, and entropy of fluorite, CaF_2, calculated from the phonon frequencies.

9.3 Atomic displacements

9.3.1 Normal mode coordinates

Any wave causes all atoms in the unit cell to be displaced by relative amounts given by the mode eigenvector $\mathbf{e}(\mathbf{k}, \nu)$, where we label the eigenvector with the wave vector $\mathbf{k}$ and the mode label ν. The eigenvectors are normalized, and within the harmonic approximation will not depend on temperature. However, we expect the resultant atomic displacement to depend on temperature, and we now consider the mode amplitudes. We therefore write the contribution of all the normal modes to the displacement of the atom labelled j in the unit cell labelled l:

$$\mathbf{u}(jl) = \frac{1}{\sqrt{Nm_j}} \sum_{\mathbf{k},\nu} \mathbf{e}(j, \mathbf{k}, \nu) \exp(i\mathbf{k} \cdot \mathbf{r}(jl)) Q(\mathbf{k}, \nu) \tag{9.35}$$

The position of the atom is written as $\mathbf{r}(jl)$. The important new quantity is $Q(\mathbf{k}, \nu)$, which subsumes the time dependence of the wave as well as the temperature dependence. The time derivative can be written as

$$\dot{Q}(\mathbf{k}, \nu) = -i\omega(\mathbf{k}, \nu) Q(\mathbf{k}, \nu) \tag{9.36}$$

The quantity $Q(\mathbf{k}, \nu)$ is called the **normal mode coordinate**. At the end of this chapter we will show that it can play the role of a coordinate variable. Moreover, $\dot{Q}(\mathbf{k}, \nu)$ can play the role of a velocity variable, as we will now demonstrate.

9.3.2 Vibrational energy and amplitude

The velocity of an atom is given as

$$\begin{aligned} \dot{\mathbf{u}}(jl) &= \frac{1}{\sqrt{Nm_j}} \sum_{\mathbf{k},\nu} \mathbf{e}(j, \mathbf{k}, \nu) \exp(i\mathbf{k} \cdot \mathbf{r}(jl)) \dot{Q}(\mathbf{k}, \nu) \\ &= \frac{-i}{\sqrt{Nm_j}} \sum_{\mathbf{k},\nu} \omega(\mathbf{k}, \nu) \mathbf{e}(j, \mathbf{k}, \nu) \exp(i\mathbf{k} \cdot \mathbf{r}(jl)) Q(\mathbf{k}, \nu) \end{aligned} \tag{9.37}$$

The kinetic energy of the atom is then

$$\begin{aligned} \frac{1}{2} m_j |\dot{\mathbf{u}}(jl)|^2 = \frac{1}{2N} \sum_{\substack{\mathbf{k},\mathbf{k}' \\ \nu,\nu'}} &\omega(\mathbf{k}, \nu)\omega(\mathbf{k}', \nu') \mathbf{e}(j, \mathbf{k}, \nu) \cdot \mathbf{e}^*(j, \mathbf{k}', \nu') \\ &\times \exp(i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}(jl)) Q(\mathbf{k}, \nu) Q^*(\mathbf{k}', \nu') \end{aligned} \tag{9.38}$$

The kinetic energy of the crystal can be obtained by summing over all atoms:

$$\begin{aligned} \frac{1}{2} \sum_{j,l} m_j |\dot{\mathbf{u}}(jl)|^2 = \frac{1}{2N} \sum_{j,l} \sum_{\substack{\mathbf{k},\mathbf{k}' \\ \nu,\nu'}} &\omega(\mathbf{k}, \nu)\omega(\mathbf{k}', \nu') \mathbf{e}(j, \mathbf{k}, \nu) \cdot \mathbf{e}^*(j, \mathbf{k}', \nu') \\ &\times \exp(i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}(jl)) Q(\mathbf{k}, \nu) Q^*(\mathbf{k}', \nu') \end{aligned} \tag{9.39}$$

This equation can be drastically simplified. We use eqn M.1 for lattice sums (see Appendix M). This gives the result that the sum over l is only non-zero

if $\mathbf{k}' = \mathbf{k}$. As a result, the sum over l is replaced by a factor of N, and the dependence on $\mathbf{r}(jl)$ vanishes:

$$\frac{1}{2}\sum_{j,l} m_j |\dot{\mathbf{u}}(jl)|^2 = \frac{1}{2}\sum_{j}\sum_{\substack{\mathbf{k}\\ \nu,\nu'}} \omega(\mathbf{k},\nu)\omega(\mathbf{k},\nu')\mathbf{e}(j,\mathbf{k},\nu)\cdot\mathbf{e}^*(j,\mathbf{k},\nu') \times Q(\mathbf{k},\nu)Q^*(\mathbf{k},\nu') \tag{9.40}$$

We now make use of the normalization condition

$$\sum_j \mathbf{e}(j,\mathbf{k},\nu)\cdot\mathbf{e}^*(j,\mathbf{k}',\nu') = \delta_{\nu,\nu'} \tag{9.41}$$

This allows us to further simplify the equation for the kinetic energy of the crystal, now accounting for the sum over j:

$$\frac{1}{2}\sum_{j,l} m_j |\mathbf{u}(jl)|^2 = \frac{1}{2}\sum_{\mathbf{k},\nu} \omega^2(\mathbf{k},\nu)|Q(\mathbf{k},\nu)|^2 \tag{9.42}$$

The total phonon energy is the sum of the contribution of the kinetic energy and the contribution of the vibrational potential energy. Rather than derive an expression for the latter term, which would involve the use of the dynamical matrix formalism (Section 8.5.5), we simply note that, for a harmonic oscillation, *the average vibrational potential energy is equal to the average kinetic energy*. Therefore the total phonon energy is simply equal to *twice* the phonon kinetic energy. We can now equate the vibrational energy calculated here to the phonon energy given in eqn 9.2:

$$E = \sum_{\mathbf{k},\nu} \omega^2(\mathbf{k},\nu)\langle|Q(\mathbf{k},\nu)|^2\rangle = \sum_{\mathbf{k},\nu} \hbar\omega(\mathbf{k},\nu)\left(n(\omega(\mathbf{k},\nu),T)+\frac{1}{2}\right) \tag{9.43}$$

This immediately gives an equation for the mean-square value of the normal mode coordinate for each value of $\mathbf{k}$ and ν:

$$\langle|Q(\mathbf{k},\nu)|^2\rangle = \frac{\hbar}{\omega(\mathbf{k},\nu)}\left(n(\omega(\mathbf{k},\nu),T)+\frac{1}{2}\right) \tag{9.44}$$

In the high-temperature limit, $k_\mathrm{B}T > \hbar\omega$, we have

$$\langle|Q(\mathbf{k},\nu)|^2\rangle = \frac{\hbar}{\omega(\mathbf{k},\nu)} \times \frac{k_\mathrm{B}T}{\hbar\omega(\mathbf{k},\nu)} = \frac{k_\mathrm{B}T}{\omega^2(\mathbf{k},\nu)} \tag{9.45}$$

This implies that at high temperatures, the mean-square atomic displacements will be proportional to temperature, as anticipated in Section 6.7.1. This is discussed further in Appendix N.

9.3.3 Recasting the crystal Hamiltonian

One of the merits of the formalism of the normal mode coordinates is that it is possible to rework the Hamiltonian of the crystal in terms of the individual vibrations. The Hamiltonian of the harmonic crystal is given by

$$\mathcal{H} = \frac{1}{2}\sum_{jl} m_j|\dot{\mathbf{u}}(jl)|^2 + \frac{1}{2}\sum_{\substack{j,l\\ j',l'}} \frac{\partial^2 E}{\partial\mathbf{u}(jl)\cdot\partial\mathbf{u}(j'l')}\mathbf{u}(jl)\cdot\mathbf{u}(j'l') \tag{9.46}$$

The first term, the kinetic energy, follows from the preceding discussion. However, we now rework the equations in terms of $\dot{Q}(\mathbf{k}, \nu)$ to obtain (Problem 9.6)

$$\frac{1}{2}\sum_{j,l} m_j |\dot{\mathbf{u}}(jl)|^2 = \frac{1}{2}\sum_{\mathbf{k},\nu} \dot{Q}(\mathbf{k}, \nu)\dot{Q}(-\mathbf{k}, \nu) \tag{9.47}$$

The second term in eqn 9.46 is the potential energy. Following the previous analysis, this term can be written as

$$\begin{aligned}&\frac{1}{2}\sum_{\substack{j,l\\ j',l'}} \frac{\partial^2 E}{\partial \mathbf{u}(jl)\cdot\partial \mathbf{u}(j'l')}\mathbf{u}(jl)\cdot\mathbf{u}(j'l')\\ &\quad = \frac{1}{2}\sum_{\mathbf{k},\nu} \omega^2(\mathbf{k}, \nu)Q(\mathbf{k}, \nu)Q(-\mathbf{k}, \nu)\end{aligned} \tag{9.48}$$

This result is anticipated by our earlier analysis, but a detailed derivation requires a more thorough treatment of the dynamical matrix than is appropriate here (the main result is derived in Dove 1993).

Following the representation of the separate terms in terms of the normal mode coordinates, the Hamiltonian of the harmonic crystal can be written as

$$\mathcal{H} = \frac{1}{2}\sum_{\mathbf{k},\nu} \dot{Q}(\mathbf{k}, \nu)\dot{Q}(-\mathbf{k}, \nu) + \frac{1}{2}\sum_{\mathbf{k},\nu} \omega^2(\mathbf{k}, \nu)Q(\mathbf{k}, \nu)Q(-\mathbf{k}, \nu) \tag{9.49}$$

This equation has a certain elegance – we have replaced pairs of variables related to individual atoms with pairs of variables related to individual normal modes. In effect, this Hamiltonian is the Fourier transform equivalent of the harmonic atomic Hamiltonian. This new version of the Hamiltonian will prove to be useful in the development of the theory of anharmonic interactions, and for the study of displacive phase transitions. It is also easy to quantize, if the variables Q and $\dot{Q}$ are replaced by appropriate operators (but this is beyond the scope of this book; see Chapter 11 of Dove 1993).

Summary of chapter

- The energy of a lattice vibration is quantized, with each quantum having energy $\hbar\omega$. The quanta are called **phonons**.
- The ground state of a harmonic oscillator has energy $\frac{1}{2}\hbar\omega$. The motions of the atoms associated with this energy are called **zero point motions**.
- The number of phonons excited in thermal equilibrium at any temperature is given by the **Bose–Einstein distribution**. This is derived from the partition function for phonons. Expressions for the phonon contributions to other thermodynamic functions, including the free energy, internal energy, and their derivatives, are derived from the phonon partition function.
- The **high-temperature limit** is defined as the regime for which $k_B T > \hbar\omega$. In this limit it is possible to expand the thermodynamic functions as power series of $\hbar\omega/k_B T$, leading to classical thermodynamic results (such as

for internal energy and heat capacity). The high-temperature limit will be useful for ideas developed in later chapters.

- The **Einstein model** gives a representation of the phonon thermodynamic functions by assigning a single average frequency for all modes. This model captures the behaviour of the thermodynamic functions over a wide range of temperatures if the average frequency can be treated as an adjustable phenomenological parameter, but the model does not work at low temperatures where the important vibrations are the acoustic modes.
- The **Debye model** represents all vibrations as a linear function of wave vectors up to an appropriate cut-off. This model also gives a good representation of the thermodynamic functions if the cut-off frequency, ω_{D}, can be treated as a phenomenological parameter, particularly at low temperatures where the model does not suffer the same deficiencies as the Einstein model.
- The distribution of vibration frequencies is given by the **density of states**, $g(\omega)$. For a three-dimensional crystal, the density of states at low ω is determined by the acoustic modes, and $g(\omega) \propto \omega^2$. The thermodynamic functions can be written as integrals over the density of states, and the form of $g(\omega)$ at low ω leads to the prediction that the heat capacity at low temperatures varies as T^3.
- The amplitude of each vibration can be written in terms of the **normal mode coordinate** $Q(k)$, and the atomic displacements can therefore be written in terms of the normal mode coordinates of all vibrations. The kinetic energy can be written in terms of the time derivatives of the normal mode coordinates, $\dot{Q}(k)$.
- The mean square value of the normal mode coordinate in the high-temperature limit is simply given as $\langle |Q^2| \rangle = k_{\mathrm{B}}T/\omega^2$.
- The Hamiltonian describing the harmonic atomic motions of the crystal can be written as a simple function of $Q(k)$ and $\dot{Q}(k)$, with a sum over all vibrations.

Further reading

All textbooks on lattice dynamics will describe the thermodynamic aspects. Introductions of the subject of lattice dynamics are given by Kittel (1996) and Ashcroft and Mermin (1976). More detailed treatments of the topic are given by Barron and White (1999), Böttger (1983), Di Bartolo and Powell (1976), Dove (1993), Elliot (1998), Hardy and Karo (1979), Kosevich (1990), and Ziman (2000).

Exercises

(9.1) When making an expansion of any function $f(x)$ in terms of x it is usually the case that the expansion is only valid when $x \ll 1$. For $x = \hbar\omega/k_{\mathrm{B}}T$, calculate the relative error in the internal energy in the high-temperature approximation when $\hbar\omega = k_{\mathrm{B}}T$ (i.e., the difference between $n(x) + 1/2$ and x when $x = 1$).

(9.2) Obtain the second form of the free energy in eqn 9.17.

(9.3) Derive eqns 9.33 and 9.34.

(9.4) Show that for a d-dimensional crystal, the density of states varies as $g(\omega) \propto \omega^{d-1}$, and that the heat capacity at low temperatures varies as $c_V \propto T^d$.

(9.5) Show that the mean-square displacement of an atom j in a unit cell is given by

$$\langle \mathbf{u}(j) \cdot \mathbf{u}(j) \rangle = \frac{1}{Nm_j} \sum_{\mathbf{k},\nu} \mathbf{e}(j, \mathbf{k}, \nu) \cdot \mathbf{e}(j, -\mathbf{k}, \nu) \times \langle Q(\mathbf{k}, \nu) Q(-\mathbf{k}, \nu) \rangle$$

where the average is over all unit cells in the crystal. Hence show that in the high-temperature limit, the mean-square displacement of at atom is proportional to temperature (Appendix N).

(9.6) Show that the mean kinetic energy of all atoms in a crystal is given by eqn 9.47.

(9.7) Show that the mean-square displacements for all atoms in the crystal can be written in the high-temperature limit as

$$\langle \mathbf{u} \cdot \mathbf{u} \rangle = \frac{k_B T}{Nm} \int \frac{1}{\omega^2} g(\omega) \, \mathrm{d}\omega$$

Use the result from Problem 9.4 to show that the mean-square displacement is infinite for $d < 3$. This has important implications for the possible existence of flatworld.

10 Experimental methods for measurements of vibrational frequencies

10.1 Introduction

The main experimental approach to study the structure of materials is through scattering of radiation, as was discussed in some depth in Chapter 6. It was noted that several different types of radiation beams can be used in diffraction, including electromagnetic radiation and beams of neutrons, each with its own relative merits. It is not surprising that scattering of radiation also proves to be one of the main techniques for studying dynamics. Nor is it surprising that, like in structural studies, there are several types of radiation that can be used, with each radiation type giving complementary information. Our purpose in this chapter is to review the various spectroscopic techniques that are in common use for the study of dynamic processes in solids.

As in diffraction, spectroscopy exploits the wave–particle duality of quantum mechanics, in part because of the quantization of the radiation beam (photons or neutrons, each with particle energy and momentum, and also with wave vector/wavelength), and also because of the quantization of the lattice waves giving phonons with definite energy and wave vector. The basic idea in vibrational spectroscopy is that a particle in the radiation beam will interact with the lattice vibrations, either through a scattering process or, specifically in the case of electromagnetic radiation, through an absorption process. These processes can involve one or more phonons in the crystal. In a scattering process, this interaction will give rise to a change in the wave vector and energy of the scattered beam with respect to that of the incident beam.

Like diffraction (Chapter 6), spectroscopy is another technique that exploits the duality between waves and particles, but in this case the duality is evident in both the radiation beam and the lattice waves. The interactions between the radiation beams and the vibrations involve individual particles within the radiation beam, whether neutrons or phonons, absorbing or losing energy as phonons, the quanta of the lattice vibrations (Chapter 9). However, many aspects of the formalism, including the use of wave vectors of the radiation beam and the lattice vibrations, and the frequencies of the vibrations, retain firm roots in the wave picture of both the radiation beam and the vibrations. The analysis of diffraction, Chapter 6, was firmly based on the wave picture of the radiation beam. Although we do not develop the theory of spectroscopy in the same detail here, it is nevertheless firmly grounded in the wave theory.

Before we discuss scattering processes in detail, we need to make some additional comments on wave–particle duality. The starting point is the link

between the particle momentum, **p** and its wave vector **k** and wavelength λ, which is described in eqn 6.1. The energy of a particle with rest mass m is given by

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \tag{10.1}$$

On the other hand, for electromagnetic radiation there is a linear relationship between energy and wave vector:

$$E = \hbar\omega = \hbar ck \tag{10.2}$$

To appreciate how the relationships between the energy and wave vector for the different types of radiation are linked, it is useful to note the relationship between the **phase velocity** v_{ph}, the velocity of the *wave front*, and the **group velocity** v_{g}, the velocity of the *energy flow*:

$$v_{\mathrm{ph}} = \frac{\omega}{k} \qquad v_{\mathrm{g}} = \frac{\partial\omega}{\partial k} \tag{10.3}$$

Since for electromagnetic radiation $\omega = ck$, $v_{\mathrm{ph}} = v_{\mathrm{g}} = c$. However, for beams of massive particles, the group velocity is the particle velocity, since the particle carries all the energy. We can identify an angular frequency with the wave carried along with the neutron. If we write

$$\hbar\omega = \frac{\hbar^2 k^2}{2m} \tag{10.4}$$

we can write the group velocity as

$$v_{\mathrm{g}} = \frac{\hbar k}{m} \tag{10.5}$$

This result gives

$$p = m v_{\mathrm{g}} = \hbar k \tag{10.6}$$

which is consistent with v_{g} being assigned with the neutron velocity, since the neutron is the particle that carries the energy of the associated wave. The phase velocity is equal to $v_{\mathrm{ph}} = \hbar k/2m = v_{\mathrm{g}}/2$.

10.2 Basic ideas of spectroscopy

There are two basic approaches to spectroscopy. In the first, a beam of radiation can be scattering by interactions with the phonons in the crystal. The changes in energy and wave vector of the beam will give direct information about the energies and wave vectors of the phonons involved in the scattering process. The second approach involves the absorption of electromagnetic radiation in the crystal by the creation of phonons. If this is a resonance process, the resonance energies provide direct information about the energies of the phonons.

Scattering processes are subject to constraints from the conservation of energy and momentum. In a scattering process, a beam of particles (whether

neutrons, photons, or some other form of radiation) incident on the sample with wave vector $\mathbf{k}_\mathrm{i}$ and energy E_i is scattered with wave vector $\mathbf{k}_\mathrm{f}$ and energy E_f (the subscript i stands for *initial*, and the subscript f stands for *final*, both of which denote the quantum state of the particle in the radiation beam). The scattering process can involve absorption or creation of a single phonon of energy. In this case there are constraints imposed by the conservation of energy and crystal momentum respectively:

$$\pm\hbar\omega = E_\mathrm{i} - E_\mathrm{f} \tag{10.7}$$

$$\pm\mathbf{k} = \mathbf{k}_\mathrm{i} - \mathbf{k}_\mathrm{f} - \mathbf{G} \tag{10.8}$$

where ω and $\mathbf{k}$ are the angular frequency and wave vector of the phonon, and $\mathbf{G}$ is any reciprocal lattice vector. $E = E_\mathrm{i} - E_\mathrm{f}$ is called the **energy transfer** (you might think that ΔE would be a better symbol, but E is what is used in practice). As in Chapter 6, we define $\mathbf{Q} = \mathbf{k}_\mathrm{i} - \mathbf{k}_\mathrm{f}$ as the **scattering vector**, and $\hbar\mathbf{Q}$ can be called the **momentum transfer**. The one-phonon scattering process is illustrated in Fig. 10.1. Clearly a knowledge of both E_i and E_f, and of both $\mathbf{k}_\mathrm{i}$ and $\mathbf{k}_\mathrm{f}$, which are obtained from the settings on the instruments, will provide direct information about ω and $\mathbf{k}$. The conservation laws give constraints between the energy transfer and scattering vectors. We write

$$E = \frac{\hbar^2}{2m}(k_\mathrm{i}^2 - k_\mathrm{f}^2) \tag{10.9}$$

In some experiments, the instrument (which will be discussed below) is set to a constant value of k_f^2. We can therefore substitute for $k_\mathrm{f}^2 = |\mathbf{k}_\mathrm{i} - \mathbf{Q}|^2$ to obtain

$$E = \frac{\hbar^2}{2m}(2\mathbf{k}_\mathrm{i}\cdot\mathbf{Q} - Q^2) \tag{10.10}$$

In many experiments, one crystal axis is held normal to the plane containing both $\mathbf{k}_\mathrm{i}$ and $\mathbf{k}_\mathrm{f}$, and the crystal is rotated about this axis to measure the phonons in the zone of reciprocal space. In this case, if it is intended to perform measurements at one particular value of $\mathbf{Q}$, the conservation laws can be satisfied by adjusting both the values of k_i^2 and the angle subtended by $\mathbf{Q}$ and $\mathbf{k}_\mathrm{i}$. We will discuss the experimental realization of these conditions in Section 10.3.

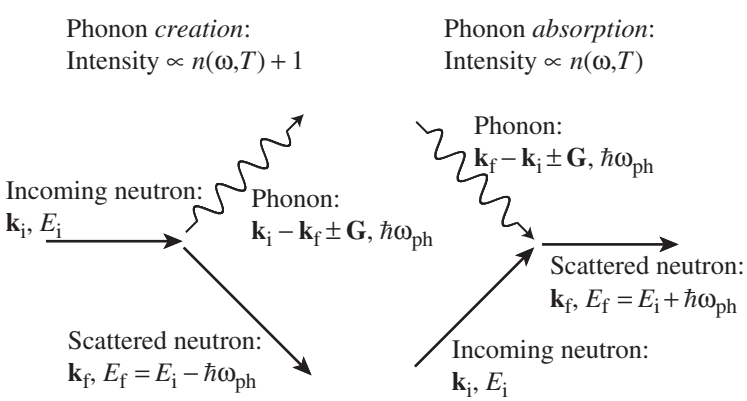


Fig. 10.1 Representation of the two types of one-phonon scattering process, drawn here for neutron scattering, but the same diagrams are also applicable to photon scattering (X-ray or light).

Scattering processes can also involve more than one phonon. In these cases, the conservation laws for crystal momentum and energy are less stringent, since these quantities are shared by more particles. As a result, it is very difficult to extract information about phonon energies and wave vectors from these multiphonon processes. However, the looser conservation laws mean that multiphonon processes tend to feature in neutron scattering experiments as a smoothly varying background. As a result, they are usually no more than a slight nuisance, so we will have little to say about them.

In absorption spectroscopy, the incoming beam of radiation is absorbed at vibrational frequencies as a resonance process, and its energy converted to phonons. This means that there must be a match between the energy and wave vector of the radiation beam and of the phonon.

10.3 Neutron scattering techniques

10.3.1 Neutrons for spectroscopic measurements

The properties of neutrons were described in some detail in Section 6.2.4. Many of the properties of neutron beams that ensure they are suitable for diffraction also give neutron beams a special role in the study of lattice dynamics. In particular, neutron beams with thermal energies will have ideal energies for studies of dynamics in addition to having ideal wavelengths. For example, the beam of neutrons with energies corresponding to room temperature will have an energy of 25 meV, which corresponds to $E/2\pi\hbar \equiv 6.05$ THz. This is within the range of phonon frequencies. The fact that both energy and wavelength are matched to structural and dynamic processes in materials means that neutron scattering methods are capable of probing lattice dynamics across the ranges of energy and wave vector of phonons in materials.

10.3.2 Neutron scattering experimental methods: the *triple-axis spectrometer*

As outlined in Chapter 6, there are two types of neutron sources, the nuclear reactor and the spallation source. These two sources give neutron beams that have such different characteristics that the experimental methods for both spectroscopy and diffraction will be radically different. Historically the most important instrument for measuring phonon dispersion curves has been the **triple-axis spectrometer**, which won for its developer, Bertram Brockhouse, a share of the 1994 Nobel prize. This instrument is optimized for operation on a reactor source of neutrons, on which it is possible to use a relatively intense monochromatic beam of radiation. We will restrict ourselves to a discussion of this type of instrument – it illustrates most of the basic points of neutron spectroscopy reasonably easily, although with the advent of spallation neutron sources other techniques are becoming increasingly important.

The basic design of a triple-axis spectrometer is shown in Fig. 10.2. The basic point made earlier is that it is required to measure a change in energy for a given value of $\mathbf{Q}$. The measurement can be performed by setting both the magnitudes and orientations of the two vectors $\mathbf{k}_\mathrm{i}$ and $\mathbf{k}_\mathrm{f}$. The neutron beam that emerges from the reactor will contain a broad spectrum of wavelengths, and

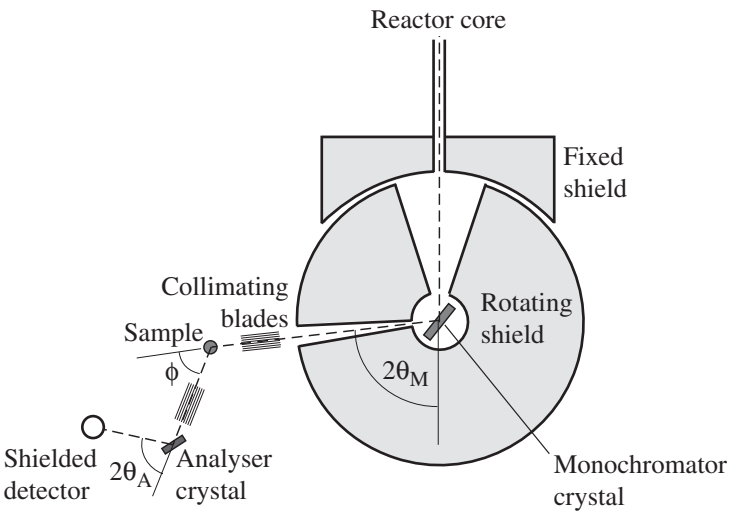


Fig. 10.2 Schematic representation of the triple axis spectrometer. The angles $2\theta_M$, ϕ and $2\theta_A$ can all be adjusted in routine operation.

a monochromatic beam is produced by reflection through the Bragg angle θ_M from an aligned monochromator crystal. Bragg's law ensures that the beam scattered at one angle will contain the fundamental wavelength of choice, namely $2\pi/k_i$ together with higher-order wavelengths that can be removed by appropriate filters or simply by the settings of the instrument. Similarly, a **monochromator** crystal can be used to set the value of k_f – this crystal is called the **analyser**, and the relevant Bragg angle is labelled θ_A. The most common crystals for the monochromator and analyser are graphite, scattering from the (001) planes. Silicon and germanium crystals can also be used: these do not have as high a reflecting power, but can give 'cleaner' beams without contamination from higher-order reflections.

Having the monochromator and analyser set the magnitudes k_i and k_f, the relative directions of the two beams will then define the scattering vector $\mathbf{Q}$. The angle between the two beams is labelled ϕ in Fig. 10.2. The crystal then needs to be oriented with respect to $\mathbf{k}_i$, $\mathbf{k}_f$, and $\mathbf{Q}$; the angle of the crystal is usually labelled ψ.

The triple-axis spectrometer is often run with a fixed value of θ_A to give a constant value of k_f. The reason for wanting to be able to change the value of θ_A is that it controls the resolution of the instrument, and different phonon measurements, even in one crystal, may require different settings of the resolution. For example, the frequencies of the low-energy phonons often have more variation with wave vector than the high-energy modes, and thus better resolution is required, particularly if measurements are performed using a constant value of $\mathbf{Q}$. On the other hand, the higher-energy modes have less variation with wave vector, but, as we will see, the intensity of the scattering from high-energy modes is lower (we will show below that the intensity scales as ω^{-2}). For these modes, lowering the resolution will lead to a higher count-rate of scattered neutrons, and the cost in terms of resolution need not be significant.

When working with fixed values of θ_A, it is necessary to constantly change the value of θ_M. The difficulty is that the movement involves components close to the face of the reactor, which means that shielding has to be moved as well. Typically the monochromator crystal is enclosed in a large drum, the

Fig. 10.3 Photograph of a working triple axis spectrometer. The large drum on the far right contains the monochromator crystal, the sample is contained in the cryostat on the table to the left of the monochromator assembly, and the analyser stage and detector are on the left side of the instrument. The scientists operating the instrument give an idea of the scale. (Photograph courtesy of the Institute Laue-Langevin.)

proportions of which can be seen in Fig. 10.3. Moreover, since the orientation of the beam coming from the monochromator is constantly changing, the whole spectrometer, including crystal and analyser assembly, must also be able to rotate. This is typically accomplished by moving the whole assembly on air pads, which can be seen in Fig. 10.3.

With control of the instrument being determined by the settings of the angles θ_{M}, ϕ, and ψ, θ_{A}, which in turn determine the values of E and $\mathbf{Q}$ and their resolution, the task is to choose the most appropriate strategy for running an experiment. As noted above, it is common to work with a fixed value of θ_{A}. In addition, it is common to perform measurements with the angles set to maintain a fixed value of $\mathbf{Q}$, and to then measure the intensity of the scattering beam as a function of E. This strategy is illustrated in Fig. 10.4. If there is no phonon with wave vector $\mathbf{k} = \mathbf{Q} \pm \mathbf{G}$ (where $\mathbf{G}$ is a reciprocal lattice vector – recall from Chapter 8 that phonon frequencies remain unchanged when a reciprocal lattice vector is added to $\mathbf{k}$) and phonon energy $\hbar\omega = E$, there will be no signal in the measurement. But when there is a phonon with $\mathbf{k} = \mathbf{Q} \pm \mathbf{G}$ and $\hbar\omega = E$, the scattered intensity will be measured. The combination of the finite resolution of the instrument and the non-zero width of the phonon peak (the former usually being the dominant factor unless there are strong anharmonic processes, Chapter 11) will mean that the phonon will be measured as a peak with a width that covers several incremental settings of E.

As we have remarked, the triple-axis spectrometer is specifically designed for operation on a reactor source. It requires a polychromatic beam of radiation in order to adjust the value of E_{i}, but works by selecting a single value of E_{i} for a single measurement. The measurements as a function of ($\mathbf{Q}$, E) proceed following a point-by-point trajectory in the space spanned by $\mathbf{Q}$–E. This is because there is no way to keep track of individual neutrons in a continuous source.

On the other hand, most spallation neutron sources produce pulses of neutrons with a wide spread of energies, but because the pulses are produced at times that are known, it is possible to use a clock (albeit a fast clock) to keep track of individual neutrons within each pulse. This allows the whole beam to be used, with all incident energies, at the same time, and a number of strategies have been developed that exploit the time-keeping capabilities. As a result, a much

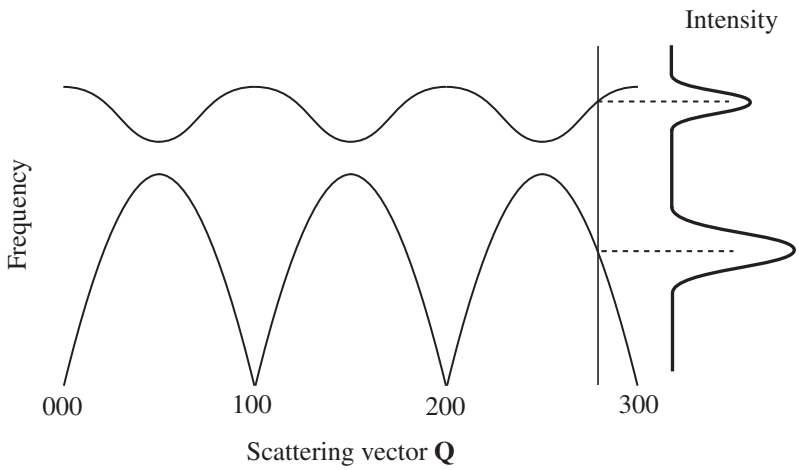


Fig. 10.4 Illustration of the constant-**Q** mode of operation of a triple-axis spectrometer. The acoustic and optic mode dispersion curves are shown extending across several Brillouin zones. The vertical line close to the 300 reciprocal lattice point represents the value of **Q** used in the illustrated experiment. The intensity of the measured signal peaks at the frequencies of the vibrations for this value of **Q**.

larger fraction of the neutron beam can be used for simultaneous measurements, which is, at least in principle, a big improvement over the point-to-point mode of the triple-axis spectrometer. The most important advantage of this is that it helps overcome one of the main problems of spallation sources, namely that the neutron flux of a spallation source when integrated over time is much weaker than the corresponding flux from a reactor source. There is, however, some degree of cost in this more efficient style of working, which is that part of the flexibility of the triple-axis spectrometer is lost.

10.3.3 General formalism of neutron scattering

General approach and the one-phonon scattering function

The analysis of Chapter 6, leading to eqn 6.22 which describes the intensity of the scattering of X-rays or neutrons from an ensemble of atoms, has been based on the implicit assumption that the coherence effects arise from scattering from pairs of atoms at the same time. This assumption is not necessary, but it leads to correct expressions for the diffraction of beams of radiation from a crystal. However, a full treatment can also take account of the contributions to interference effects from components of the beam that are scattered at different times. This means that we can include terms of the form

$$b_j b_k \exp\{i\mathbf{Q} \cdot [\mathbf{r}_j(t) - \mathbf{r}_k(0)]\}$$

in the scattering equation. In practice we need to integrate over all times, but care is needed because although we retain the concept of the change in scattering vector, **Q**, we need to include the fact that scattering from atoms at different times brings in a change in the energy $E = \hbar\omega$ of the scattered beam. This is accounted for by noting that we have to perform a time Fourier transform of the scattered beam; the measurement is performed in the energy (frequency) domain rather than the time domain. Thus the scattering function of eqn 6.22 is modified to

$$S(\mathbf{Q}, E) = \sum_{i,j} b_i b_j \int \langle \exp(i\mathbf{Q} \cdot [\mathbf{r}_i(t) - \mathbf{r}_j(0)]) \rangle \exp(-i\omega t)\, \mathrm{d}t \qquad (10.11)$$

where E is the change in energy of the neutron beam on scattering. This is the pivotal equation for neutron scattering. It can be shown (Problem 10.1) that the special case $E = 0$, which is called **elastic scattering**, corresponds to the Bragg scattering of Chapter 6. It can also be shown that if an experiment is designed to integrate over all energies, giving a measurement of $\int S(\mathbf{Q}, E)\,\mathrm{d}E$, the result is a function that contains information about the instantaneous distances between particles (Problem 10.4). For the purposes of this chapter, we will consider the full function $S(\mathbf{Q}, E)$. It should be noted that a full derivation of eqn 10.11 is somewhat beyond the scope of this book – the full derivation can only be obtained within the formalism of quantum mechanics, whereas we have restricted ourselves to arguments based on classical mechanics.

Equation 10.11 covers all scattering processes, which need to be separated out of the equation. We first assume that we can write the instantaneous position of an atom labelled i as

$$\mathbf{r}_i(t) = \mathbf{R}_i + \mathbf{u}_i(t) \tag{10.12}$$

where $\mathbf{R}_i$ is the average position of the atom, and $\mathbf{u}_i(t)$ is the instantaneous displacement relative to the average position. Thus we have

$$S(\mathbf{Q}, E) = \sum_{i,j} b_i b_j \exp(i\mathbf{Q} \cdot (\mathbf{R}_i - \mathbf{R}_j) \times \int \langle \exp(i\mathbf{Q} \cdot (\mathbf{u}_i(t) - \mathbf{u}_j(0))) \rangle \exp(-i\omega t)\,\mathrm{d}t \tag{10.13}$$

We now make use of a result for two variables X and Y whose distributions are characteristics of harmonic motions:

$$\langle \exp(i(X + Y)) \rangle = \exp(\langle (X + Y)^2 \rangle / 2) \tag{10.14}$$

Thus we can write

$$\langle \exp(i\mathbf{Q} \cdot (\mathbf{u}_i(t) - \mathbf{u}_j(0))) \rangle = \exp(-\langle (\mathbf{Q} \cdot \mathbf{u}_i)^2 / 2 \rangle) \times \exp(-\langle (\mathbf{Q} \cdot \mathbf{u}_j)^2 / 2 \rangle) \times \exp(\langle [\mathbf{Q} \cdot \mathbf{u}_i(t)]\,[\mathbf{Q} \cdot \mathbf{u}_j(0)] \rangle) \tag{10.15}$$

The first two terms in the product are the temperature factors of diffraction (Chapter 6, Appendix N), which are denoted as $T_i(\mathbf{Q})$ and $T_j(\mathbf{Q})$ *respectively*. To analyse the third term, we expand it as a power series:

$$\exp(\langle [\mathbf{Q} \cdot \mathbf{u}_i(t)]\,[\mathbf{Q} \cdot \mathbf{u}_j(0)] \rangle) = \sum_{m=0}^{\infty} \frac{1}{m!} \langle [\mathbf{Q} \cdot \mathbf{u}_i(t)]\,[\mathbf{Q} \cdot \mathbf{u}_j(0)] \rangle^m \tag{10.16}$$

The term $m = 0$ corresponds to Bragg diffraction (Problem 10.2). The term $m = 1$ corresponds to scattering processes that involve single phonons; these are the interesting terms for spectroscopy. The terms $m > 1$ correspond to scattering processes involving more than one phonon. As we have argued above, these terms only contribute to the background scattering in a neutron scattering measurement, and we will not consider these further.

The scattering function with only the $m = 1$ term is

$$S(\mathbf{Q}, E) = \sum_{i,j} b_i b_j \exp(i\mathbf{Q} \cdot (\mathbf{R}_i - \mathbf{R}_j) T_i(\mathbf{Q}) T_j(\mathbf{Q})) \times \int \langle [\mathbf{Q} \cdot \mathbf{u}_i(t)]\,[\mathbf{Q} \cdot \mathbf{u}_j(0)] \rangle \exp(-i\omega t)\,\mathrm{d}t \tag{10.17}$$

The next stage is to expand the displacements in terms of the component normal modes, eqn 9.35 – but be careful not to confuse the symbol **Q** for scattering vector with that for the normal mode coordinate, $Q(\mathbf{k}, \nu, t)$. Thus we have

$$\mathbf{Q} \cdot \mathbf{u}_i(t) = \frac{1}{\sqrt{N m_i}} \sum_{\mathbf{k},\nu} \mathbf{Q} \cdot \mathbf{e}(j, \mathbf{k}, \nu) \exp(i\mathbf{k} \cdot \mathbf{R}_i) Q(\mathbf{k}, \nu, t) \quad (10.18)$$

Denoting k for $\mathbf{k}$, ν (and $-k$ for $-\mathbf{k}$, ν), we can put all the components back together to obtain

$$\begin{aligned} S(\mathbf{Q}, E) = \sum_{i,j} \frac{1}{N\sqrt{m_i m_j}} b_i b_j \exp(i\mathbf{Q} \cdot (\mathbf{R}_i - \mathbf{R}_j) T_i(\mathbf{Q}) T_j(\mathbf{Q})) \\ \times \int \sum_{k,k'} [\mathbf{Q} \cdot \mathbf{e}_i(k)] [\mathbf{Q} \cdot \mathbf{e}_j(k')] \exp(i\mathbf{k} \cdot \mathbf{R}_i) \exp(i\mathbf{k}' \cdot \mathbf{R}_j) \\ \times \langle Q(k, t) Q(k', 0) \rangle \exp(-i\omega t)\, \mathrm{d}t \end{aligned} \quad (10.19)$$

It can be shown (Problem 10.3) that the double summations over atoms and wave vectors give rise to cancellations of terms, so that the one-phonon scattering function reduces to

$$\begin{aligned} S(\mathbf{Q}, E) = N \sum_{i,j}^{\text{basis}} \frac{1}{\sqrt{m_i m_j}} b_i b_j \exp(i\mathbf{Q} \cdot (\mathbf{R}_i - \mathbf{R}_j) T_i(\mathbf{Q}) T_j(\mathbf{Q})) \\ \times \int \sum_{\nu} [\mathbf{Q} \cdot \mathbf{e}_i(\mathbf{Q}, \nu)] [\mathbf{Q} \cdot \mathbf{e}_j(-\mathbf{Q}, \nu)] \\ \times \langle Q(\mathbf{Q}, \nu, t) Q(-\mathbf{Q}, \nu, 0) \rangle \exp(-i\omega t)\, \mathrm{d}t \end{aligned} \quad (10.20)$$

Note that **Q** is equal to the phonon wave vector in any Brillouin zone, i.e., **k** plus a reciprocal lattice vector. The double sum over atoms is now over atoms in the unit cell basis rather than over the whole crystal.

We now note that the correlation function can be written as

$$\langle Q(\mathbf{Q}, \nu, t) Q(-\mathbf{Q}, \nu, 0) \rangle = \langle Q(\mathbf{Q}, \nu) Q(-\mathbf{Q}, \nu) \rangle \cos \omega(\mathbf{Q}, \nu) t \quad (10.21)$$

The thermal average in the high-temperature limit is given by eqn 9.45. The time Fourier transform of the cosine term is simply equal to the Dirac delta function $\delta(E - \hbar\omega_k)$. We can now put everything back together in the high-temperature limit to give

$$\begin{aligned} S(\mathbf{Q}, E) = N k_{\mathrm{B}} T \sum_{i,j}^{\text{basis}} \frac{1}{\sqrt{m_i m_j}} b_i b_j \exp(i\mathbf{Q} \cdot (\mathbf{R}_i - \mathbf{R}_j) T_i(\mathbf{Q}) T_j(\mathbf{Q})) \\ \times \sum_{\nu} \omega^{-2}(\mathbf{Q}, \nu) \delta(E \pm \hbar\omega(\mathbf{Q}, \nu)) \end{aligned} \quad (10.22)$$

The use of the high-temperature limit means that this is a classical result. A full quantum-mechanical analysis is beyond the scope of this chapter; the result that does not rely on the high-temperature approximation is quoted below.

Unpacking the one-phonon scattering function

The general equation for the one-phonon $S(\mathbf{Q}, E)$ contains a number of key features that we now explore separately. The key point is that when an instrument such as a triple-axis spectrometer is set to record the scattering as a function of both $\mathbf{Q}$ and E; when the values of both parameters match those for a phonon there will be a signal to be measured. This is given by the factor $\delta(E \pm \hbar\omega(\mathbf{Q}, \nu))$ in eqn 10.22. The actual intensity of the signal is determined by the other factors in the equation.

It is possible to think of an inelastic scattering process as a Doppler process. The beam is reflected from a moving plane of atoms, and this gives a Doppler shift to the neutron beam. Accordingly, it would be expected that there will be more intensity if the motion of the plane is in a direction parallel to $\mathbf{Q}$. The relative motion of each atom is given by the normal mode eigenvector $\mathbf{e}$, and we therefore expect that the intensity will be determined in part by factors of the form $m_j^{-1/2}\mathbf{Q} \cdot \mathbf{e}_j$, where we consider each atom individually (see Sections 8.5 and 9.3 for an explanation of the factor of $m_j^{-1/2}$). We also expect the basic reflecting power of the particular plane of atoms to be important, and this is controlled by the structure factor. Combining this with the $\mathbf{Q} \cdot \mathbf{e}$ factor gives an overall factor for the amplitude of the inelastic scattering process of

$$\sum_j \frac{b_j}{m_j^{1/2}}[\mathbf{Q} \cdot \mathbf{e}_j] \exp(i\mathbf{Q} \cdot \mathbf{R}_j) T_j(\mathbf{Q})$$

This factor is clearly at the heart of eqn 10.22. It is actually very important for measurements of phonon dispersion curves. The point is that this factor will give a different intensity for a mode of given $\mathbf{k}$ with different values of $\mathbf{Q} = \mathbf{k} + \mathbf{G}$. Thus one mode of a given $\mathbf{k}$ may be particularly strong at one value of $\mathbf{Q}$ relative to the other modes of the same $\mathbf{k}$, but at another value of $\mathbf{Q}$ corresponding to the same value of $\mathbf{k}$ another mode may be strongest. Because of this, it is possible to develop a strategy for performing measurements at different values of $\mathbf{Q}$ corresponding to the same value of $\mathbf{k}$ in order to pick up as many modes as possible. The measurements can then be checked against calculations to confirm that different branches are being picked up in different measurements. It should also be remarked that the fact that the $\mathbf{Q} \cdot \mathbf{e}$ factor is squared in the equation for the intensity means that the intensity of neutron scattering will scale as Q^2, so that it is invariably favourable to perform measurements at higher values of Q. The use of the $\mathbf{Q} \cdot \mathbf{e}$ factor is explored in Problem 10.5.

The $\mathbf{Q} \cdot \mathbf{e}$ factor takes account of the relative displacement of atoms, but not of the absolute amplitude of motion. In the classical approach, this is given by the factor T/ω^2 in eqn 10.22. This result is the same for both processes of phonon absorption (energy gain) and phonon creation (energy loss). Within a complete quantum mechanical treatment we expect these two processes to give different intensities. At low temperatures there are very few phonons to be absorbed by the scattering beam, but it will always be possible to create phonons. We therefore expect the phonon creation (energy loss) processes to give a higher intensity of the scattered beam than the phonon absorption (energy gain) processes. The result is that for absorption processes the intensity will be proportional to

$$\frac{\hbar}{2\omega} n(\omega, T)\delta(E - \hbar\omega)$$

and for creation processes the intensity will be proportional to

$$\frac{\hbar}{2\omega}[n(\omega, T) + 1]\delta(E + \hbar\omega)$$

Thus the one-phonon scattering function in the quantum mechanical limit is given by

$$S(\mathbf{Q}, E) = \frac{N\hbar}{2}\sum_{\nu}\frac{1}{\omega_\nu}\left|\sum_j \frac{b_j}{m_j^{1/2}}[\mathbf{Q}\cdot\mathbf{e}_j(\mathbf{k}, \nu)]\exp(i\mathbf{Q}\cdot\mathbf{r}_j)T_j(\mathbf{Q})\right|^2 \times ([n(\omega, T) + 1]\delta(E + \hbar\omega_\nu) + n(\omega, T)\delta(E - \hbar\omega_\nu)) \quad (10.23)$$

The factor $\hbar n(\omega, T)/\omega$, which reduces to T/ω^2 at high temperatures (eqn 10.22), means that phonons with higher values of ω will be harder to measure than phonons with lower values of ω. For this reason, many neutron phonon measurements are restricted to lower frequencies. To obtain data for dispersion curves such as that for quartz shown in Fig. 8.17, it is necessary to work very hard to obtain the high-frequency components. The same factor also suggests that the intensity is better at higher values of temperature. However, this may be offset by anharmonic broadening of the phonon peaks (discussed in the next chapter) and increased multiphonon scattering in the background.

10.3.4 Applications of neutron inelastic scattering

Single-crystal studies

When neutron inelastic scattering was developed as a tool in the early 1960s, it was primarily used for the determination of phonon dispersion curves. Most initial studies focused on relatively simple systems such as the elements and simple ionic solids – the results shown in Chapter 8, namely Figs 8.7 for neon, 8.9 for lead, 8.10 for potassium, 8.16 for NaCl, and 8.17 for quartz, came from this initiative. The results of these studies provided new information about the forces between atoms. On a simple level the process of fitting force constants to give best agreement between calculated and measured dispersion curves was able to provide information about the size and range of the forces. This fitting was also able to provide new insights. It was found that to obtain the best agreement between experiment and calculation it was necessary to incorporate the effects of ionic polarizability through the shell model discussed in Section 5.4.5 – this was not the first use of the shell model, but the analysis of phonon dispersion curves in terms of the shell model showed how important it is. Detailed phonon dispersion curves for a wide variety of materials have been reported, some of which have been quite complicated, and all of which have been used to provide information about interatomic forces.

One of the important developments that followed from the early experimental work on lattice dynamics was the recognition that displacive phase transitions can be understood from a lattice dynamical perspective. This is discussed in some detail in Chapter 12. As a result, it became clear that there is often an advantage in using tools such as neutron inelastic scattering to study a specific part of the phonon spectrum in some detail, for example by studying the

frequency of a particular phonon as a function of temperature. Some examples will be shown in Chapter 12. Similar measurements can be carried out using Raman and infrared spectroscopy (described below in Sections 10.5.1 and 10.6). In contrast with these techniques, neutron scattering has lower resolution and lower intensity. However, the significant advantage of neutron scattering is that it covers a wide range of scattering vectors, whereas (as will be discussed in Section 10.5.1) Raman and infrared spectroscopy are restricted to measurements of phonons with $k \sim 0$. Thus studies of phonons with wave vectors other than $k \sim 0$ necessarily require the use of neutron scattering techniques.

Studies with polycrystalline samples

In addition to studies of single crystals for the measurements of phonon dispersion curves or for studies of individual phonons, it can often be useful to perform studies using polycrystalline samples. When polycrystalline studies are performed, data for a given modulus of scattering vector Q are averaged over all orientations of the corresponding set of scattering vectors **Q**. In some studies the aim is to average over a wide range of values of Q to obtain a measurement of the phonon density of states (albeit one that is weighted by the scattering lengths of the individual atoms).

It is also possible to make use of the information contained in the dependence of the neutron scattering on Q for polycrystalline samples. Figure 10.5 shows inelastic neutron scattering data obtained on two phases of cristobalite (see Sections 2.3.3, 2.6.3 and 12.1) and silica glass (Section 2.7). The data are plotted as maps across both energy transfer, $E = \hbar\omega$, and Q, showing how the intensity of the scattering varies across the maps. It is interesting to note how similar the three maps are. The maps for α-cristobalite and silica glass both show a band of scattering at 5 meV across a wide range of values of Q, which follow a similar variation in intensity across Q. In particular, the regions of large intensity in the data for α-cristobalite, which are seen as larger light patches in the maps, occur at the same positions as strong Bragg peaks (which are seen in the band at zero energy). The same band of scattering at 5 meV is also seen in the data for β-cristobalite, but in this case there is additional scattering at lower energies in comparison with α-cristobalite which appears stronger than the band at 5 meV. The data for silica glass appear to lie between these two extremes, with additional scattering at low energies compared to the data for α-cristobalite.

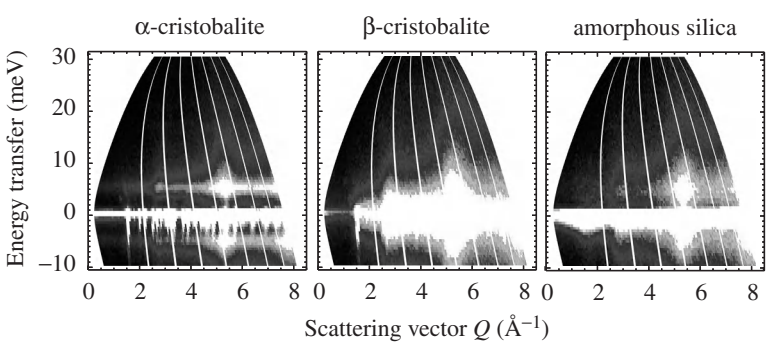


Fig. 10.5 Inelastic neutron scattering data for polycrystalline samples of the two phases of cristobalite and silica glass obtained using a chopper spectrometer at the ISIS spallation neutron source (Figure 6.9). The α and β phases of cristobalite are the ordered and disordered phases respectively.

10.4 Inelastic X-ray scattering

In principle the processes involved in inelastic neutron scattering can be applied with minor modifications to inelastic scattering of X-rays. In this case, the wavelengths of the X-ray photons are similar to those used in a neutron scattering experiment. The big difference is the energy scale. We can contrast a 25 meV neutron (wavelength 1.8 Å, corresponding frequency 6 THz) with an X-ray photon of the same wavelength. This will have a frequency of 1.67×10^6 THz. Clearly to resolve energy changes of the order of 1 THz or less is a significant challenge, and until recently this was too difficult to be contemplated. The big breakthrough came with the intense beams of X-rays that are generated by the current new generation of synchrotron sources. The high intensity enables equipment to be built to the necessarily high resolution (note that high resolution is always obtained by a trade-off with intensity). The equivalent of changing angles θ_A or θ_M to high resolution is accomplished by changing the d-spacing of the reflections rather than the angle, which can be achieved using an electric field with a piezoelectric crystal or by changing temperature.

The formalism for X-ray inelastic scattering is similar to that of neutron inelastic scattering, but there is one inherent difference. This is that the atomic scattering factors fall to zero on increasing Q – this point was discussed in Chapter 6 – and too fast to be balanced by the Q^2 factor in the equation for the inelastic scattering intensity that we noted above. This means that X-ray inelastic scattering methods work better at lower values of Q, in contrast to neutron inelastic scattering methods, which work better at higher values of Q. Thus X-ray inelastic scattering is complementary to neutron inelastic scattering, and finds particular applications in studies where the interest is in the low-Q regime of the dynamics. Typically this area is particularly important in the study of the dynamics of liquids and glasses.

Inelastic X-ray spectroscopy can still be said to be in its infancy compared to the more mature technique of inelastic neutron scattering. However, considerable progress has been made in using the technique for the study of phonon dispersion curves. Data for the dispersion curves of quartz for wave vectors along $\mathbf{a}^*$ are shown in Fig. 10.6. The data are in close agreement with the neutron scattering data shown in Fig. 8.17. The experimental effort required to perform these measurements was similar to that required for the neutron scattering measurements.

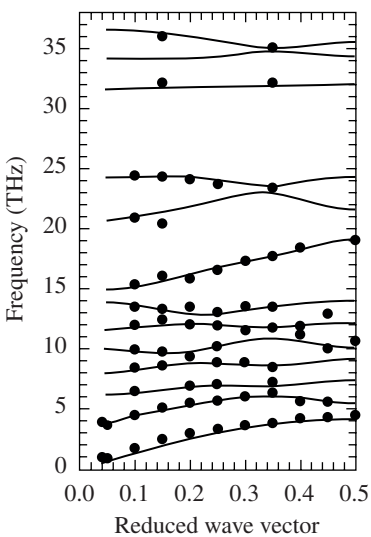


Fig. 10.6 Phonon dispersion curves of quartz for wave vectors along $\mathbf{a}^*$, obtained by inelastic X-ray scattering. The curves are the results of the shell-model calculations shown in Fig. 8.17. The experimental data should be compared with the inelastic neutron scattering data in that figure. (Original data from Ch. Halcoussis, cited in Burkel, *J. Phys.: Cond. Matter* **13**, 7627, 2001.)

10.5 Light scattering

10.5.1 Basic idea of Raman scattering

Whereas neutron beams have the property that both the values of energy and wave vector are close to those of phonons, electromagnetic radiation either has a similar wave vector but much higher energy, as in X-ray beams, or similar energy but much smaller wave vector, as in light. When matching either condition, it is possible to reproduce the main processes in neutron scattering, but there are differences owing to the mismatch on some of the conditions.

In fact the use of visible light for spectroscopy came several decades before inelastic X-ray scattering.

The discovery of the inelastic scattering of photons with wavelengths within the visible spectrum was made by Raman in 1928, and the process is accordingly known as the **Raman effect**. The main applications in the early years was to investigate vibrations of molecules, but in the 1940s the technique was applied to quartz. A significant boost to Raman spectroscopy came with the development of lasers, which are able to deliver an incident light beam to a sample with high intensity and high precision wavelength, which permit spectroscopic measurements of reasonably high resolution.

Like inelastic neutron scattering, Raman scattering involves the scattering of photons from phonons with change in energy and wave vector. However the fact that the wave vector of light is small means that the changes in the wave vector through the scattering process must also be small. In effect, this means that measurements are restricted to phonons with wave vectors that are barely different from zero when compared with the range of wave vectors within the Brillouin zone. To illustrate this, consider an incident light beam from an Ar laser with wavelength 5145 Å, and hence a frequency of 583 THz. With the precision available with optics, changes in the frequency as small as 0.01 THz can be measured (although most phonon peaks are intrinsically much broader than this, and this will limit the resolution of a measurement). The wave vector of the incident laser beam is 1.22×10^{-3} Å^{-1}. We now calculate the scattering vector for a one-phonon interaction that changes the frequency by 5 THz. This will lead to a change in the wave vector of around 1%. If this beam is scattered through 90°, the scattering vector Q will be close to $\sqrt{2} \times k_i = 1.7 \times 10^{-3}$ Å^{-1}. When we compare this value with the typical reciprocal lattice vector, which is of order 1 Å^{-1}, we can see that the scattering vectors for Raman scattering are very close to $Q \sim 0$. In effect, Raman scattering measures optic phonons with $k = 0$ (the corresponding processes that lead to scattering from acoustic modes give rise to **Brillouin scattering**, and are briefly discussed in Section 10.5.4 below). This is strikingly different from neutron scattering, which can measure phonons with wave vectors across a wide range of reciprocal space. On the other hand, Raman scattering can achieve a resolution that is virtually unattainable by single-crystal neutron inelastic scattering, whilst covering a wide range of energies that is hard to cover by neutron scattering, and on a time scale that is much faster (as well as being a much cheaper technique).

10.5.2 Mechanism of Raman scattering

The mechanism of Raman scattering arises from the fact that electromagnetic radiation is scattering through the re-radiation of the light beam. This means that the actual process must involve the polarization of the material. For a vibration to scatter a phonon by the Raman effect it has to be one whose atomic displacements change the crystal polarizability. Such a mode is said to be **Raman active**. The modes that are Raman active are a subset of all normal modes at $k = 0$. This point can be seen by the simple example of the CO_2 molecule. Three modes of molecular vibration are shown in Fig. 10.7. The totally symmetric mode, labelled A_{1g}, involves stretching of the bonds, which will change the polarizabilities, and so we expect this mode to be Raman active.

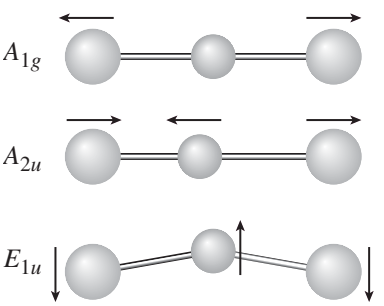


Fig. 10.7 The three normal modes of the CO_2 molecule.

The second mode, labelled A_{2u}, involves stretching and compression of the two bonds in equal measure. Thus there will be an increase in one bond polarizability and an equal decrease in the other bond polarizability. Thus the overall polarizability will remain constant to first order, and this mode is not Raman active. The third mode, labelled E_{1u} involves equal stretching of both bonds, but only to second order, and is therefore not Raman active.

The conditions that determine whether or not a normal mode is Raman active are controlled by the symmetry, and can be analysed using group theory. The resultant rules are called **selection rules**. The fact that Raman scattering is only able to select $k = 0$ modes is another form of a selection rule.

10.5.3 Applications of Raman spectroscopy

The applications of Raman scattering reflect the three most important features of the technique, namely its high resolution, the existence of the selection rules, and the relative ease of use. An example of a Raman spectrum is shown in Fig. 10.8. This plot shows the spectrum for calcite at room temperature. The corresponding infrared absorption spectrum of calcite was previously shown in Fig. 1.10, and some points of comparison will be discussed below (Section 10.6). Several important points about the spectrum in Fig. 10.8 should be noted. The first is that all peaks are sharp, which reflects the high resolution of the technique (the reader who has made the comparison with the infrared spectrum in Fig. 1.10 will note that some of the peaks are considerably sharper in the Raman spectrum – this is actually not an issue of resolution but a consequence of the LO/TO splitting described in Section 8.5.6, and this point will be discussed below). The high resolution means that it is possible to study small changes of phonon frequencies with external conditions such as temperature or pressure with sufficient accuracy that permits quantitative analysis. Applications of Raman spectroscopy in the study of temperature-induced phase transitions will be given in Chapter 12.

The second point is that the Raman spectrum contains peaks for a wide range of frequency, with the intensities of the high-frequency modes being strong enough to permit quantitative analysis. The third point, which emerges from a comparison with Fig. 1.10, is that only two peaks are also seen in the infrared spectrum. Thus the two types of spectroscopy provide complementary

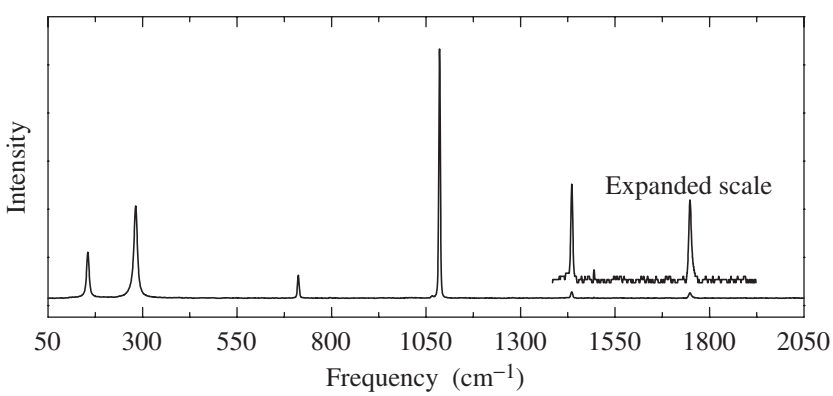


Fig. 10.8 The Raman scattering spectrum of calcite at room temperature, which can be compared with the infrared closeup absorption spectrum shown in Fig. 1.10 (data courtesy of Ming Zhang, University of Cambridge).

information, and the matching of selection rules for the two types of radiation can facilitate a clearer understanding of the dynamics of a material.

Raman scattering experiments can be performed on single crystal or polycrystalline samples. The data shown in Fig. 10.8 were obtained on a powder sample. The main applications for single-crystal Raman spectroscopy arise from the fact that the selection rules have a dependence on the alignment of the sample, so comparison of spectra for different crystal alignments can provide new information about the symmetries of the vibrations.

10.5.4 Brillouin scattering

The Raman process as described above occurs on an energy scale that corresponds to optic vibrations. On the other hand, acoustic modes with small wave vectors have much lower energy, and it is possible to fulfil the conservation constraints that apply to neutron scattering in order to measure the acoustic mode dispersion curves over a very limited range of wave vector and energy. The scattering of photons from acoustic modes is called **Brillouin scattering**. From Brillouin scattering data it is possible to determine the elastic constants with reasonably high accuracy. Certainly it is possible to obtain rather better precision on the acoustic modes at low wave vector than is possible with neutron scattering. Moreover, it is also possible to perform measurements as a detailed function of external conditions, which means that Brillouin scattering can be used to obtain information about the variation of elastic constants with changes in temperature or pressure. Unlike Raman spectroscopy, but following inelastic neutron scattering, Brillouin scattering experiments need to be performed using single-crystal samples.

10.6 Infrared absorption spectroscopy

Infrared spectroscopy is unlike the other spectroscopy techniques we have discussed in this chapter in that it involves an absorption process rather than a phonon scattering interaction. An incident polychromatic beam of infrared radiation is fed onto a sample, and the transmission spectrum is measured. The radiation is absorbed at frequencies that correspond to $k = 0$ phonon modes with appropriate atomic motions.

Like Raman spectroscopy, infrared absorption spectroscopy has selection rules, but they have a different origin. When a photon travels through a crystal, it will generate motions of the cations and anions that follow the time-varying electric field component of the electromagnetic radiation. The absorption arises as a resonance process when there is a match between the atomic motions induced by the radiation and those associated with one of the phonons in the crystal. Specifically, the infrared vibration will generate an oscillation of the dipole moment of the crystal in the transverse direction. Because the wavelength of the infrared radiation is much larger than the repeat distance of the crystal lattice, this induced oscillation will effectively correspond to a $k \sim 0$ vibration. When the induced oscillation is close to that of a phonon, in terms of both motion and frequency, there will be resonance absorption of the radiation. The selection rule is that the vibration should involve a change in the dipole moment of the

crystal. This point is illustrated by the example of CO_2 in Fig. 10.7. The two modes labelled A_{2u} and E_{1u} both involve changes in the dipole moment, whereas the A_{1g} mode retains the centre of symmetry. As in Raman scattering, the selection rules for infrared spectroscopy are determined by the symmetry of the vibration.

The main applications of infrared spectroscopy are very similar to those for Raman scattering. The technique is capable of achieving the same precision as Raman scattering. Because the selection rules for the two types of spectroscopy are different, they are complementary, and the use of both techniques can give a very balanced view of the $k = 0$ dynamics of a crystal. This complementarity is seen by comparing the infrared absorption spectrum for calcite that was shown in Fig. 1.10 with the corresponding Raman spectrum shown in Fig. 10.8.

We noted above that some of the peaks in the infrared absorption spectrum of calcite are much broader that the peaks in the Raman spectrum. This arises from the LO/TO splitting (Section 8.5.6). The modes that are susceptible to LO/TO splitting are those that give rise to a dipole moment, which include the modes that are detected in an infrared absorption spectroscopy experiment. In a transmission experiment, the significant absorption process is for the TO modes, with absorption of the transmitted beam. On the other hand, it is possible for processes to also include absorption of beams *reflected* from the sample. Reflection spectra contain both LO and TO components, and will be broadened because there is a range of frequencies between the frequencies of the pure LO and TO components. Analysis of reflection spectra is therefore somewhat more complicated, but with a proper reflection experiment the LO and TO mode frequencies can be separately recovered from such an analysis. In a transmission experiment, the contribution of reflection will be simply to broaden the peaks, and the TO mode frequencies are obtained as the peak positions.

Two of the strengths of Raman spectroscopy and infrared absorption spectroscopy are that both techniques have high resolution and both techniques have relatively short measurement times for covering a wide range of frequency. This means that it is possible to perform detailed measurements of the spectra of a material for a wide range of temperatures or pressures. This is particularly useful when studying phase transitions (Chapter 12), since the behaviour of the phonons through the phase transition can give information about the phase transition (this point will be made explicit in Section 12.5.3). The changes in the spectra can be analysed with high precision to give quantitative information about the changes in the structure through the phase transition. Some phonon peaks will change frequency in a way that reflects the phase transition, and some phonon peaks will split into two as the change in symmetry causes two degenerate phonon frequencies to differ. Moreover, some peaks will appear in the low-temperature phase due to changes in the selection rules following changes in crystal symmetry, and the intensity of such a peak will follow changes in the crystal structure. The example of the use of infrared spectroscopy for the study of the phase transition in $LiKSO_4$ is shown in Fig. 10.9. The phase transition occurs at a temperature of 110 K. It is accompanied by the appearance of new phonon peaks in the spectra at lower frequencies, the splitting of peaks (particularly clear for the peak at 470 cm^{-1}), changes in the peak frequencies, and broadening of the phonon peaks. An example of the anharmonic broadening of phonon frequencies measured by infrared spectroscopy will be shown in

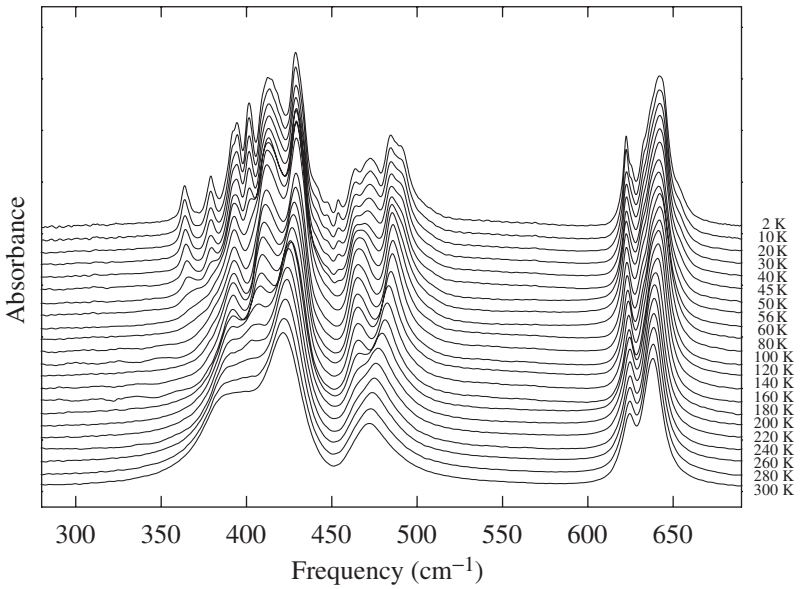


Fig. 10.9 The infrared absorption spectra of $LiKSO_4$ at several temperatures, showing the effects of a phase transition at 110 K (data courtesy of Ming Zhang, University of Cambridge).

Fig. 11.2. Following the discussion in Section 11.1, the width of a phonon peak is associated with the mean lifetime of the phonon due to anharmonic interactions. It is commonly found that there is structural disorder in the high-temperature phases of materials which contributes significantly to the anharmonic processes.

Summary of chapter

- Spectroscopic measurements of vibrational frequencies involve processes in which beam energies change by the energies of single phonons.
- **Inelastic neutron scattering** methods, particularly using the **triple axis spectrometer**, allow measurements of phonon frequencies for any phonon wave vectors and across a wide range of scattering vectors. These techniques enable the experimental determination of phonon dispersion curves.
- **Inelastic X-ray scattering** methods can now be used to obtain measurements similar to those obtained using inelastic neutron scattering. The main difference between the two types of measurements are those associated with the differences between the X-ray atomic scattering factor and the neutron scattering length discussed in Chapter 6. The high energies of X-ray photons relative to phonon energies means that extremely high resolution is required.
- **Raman scattering** is the equivalent process involving inelastic scattering of photons whose energies are in the optical region of the electromagnetic spectrum. Because the wavelength of light is much longer than interatomic spacings, Raman scattering can only measure phonons with wave vectors close to zero. Tight selection rules mean that not all phonons can be measured. Raman scattering measurements have much higher

resolution than inelastic neutron or X-ray scattering, and measurements can be performed much more quickly, which means that Raman scattering is particularly good for studies as functions of external variables (such as temperature or pressure).

- Phonons of appropriate symmetry can absorb photons of an incident beam of infrared radiation. Measurements of the absorption spectrum provide information on phonon frequencies that is complementary to the information obtained from Raman spectroscopy because different selection rules operate.

Further reading

Neutron spectroscopic methods are described by Ashcroft and Mermin (1976), Dobrzynski and Blinowski (1994), Bacon (1973, 1987), Bée (1988), Dorner (1982), Dove (1993), Hardy and Karo (1979), Lovesey (1984), Newport *et al.* (1988), Squires *et al.* (1978), and Sköld and Price (1986). Vibrational Raman and infrared spectroscopy are discussed by Diem (1993), Dove (1993), Elliott (1998), and Iqbal and Owens (1984).

Exercises

(10.1) Show that eqn 10.11 for $S(\mathbf{Q}, E)$ with energy transfer $E = \hbar\omega = 0$ is formally equivalent to Bragg diffraction.

(10.2) Show that the term with $m = 0$ in eqn 10.16 is consistent with Bragg diffraction.

(10.3) Equation 10.19 for the one-phonon scattering function contains double sums over all atoms in the crystal and over all wave vectors. Use the lattice sums of Appendix M together with the fact that each sum over all atoms can be reduced to a sum over all lattice vectors followed by a sum over all atoms in each unit cell, to show that $\mathbf{k} = \mathbf{k}' = \mathbf{Q}$, and that the summations over all atoms reduce to summations over the atoms in the unit cell, eqn 10.20. (*Hint*: treat the two sums over atoms separately, noting that the sum over i brings in $\mathbf{k}$ and that the sum over j brings in $\mathbf{k}'$). Argue that the only non-zero terms in the double summation over ν and ν' will be those for which $\nu = \nu'$.

(10.4) Show that the information contained in a measurement of $S(\mathbf{Q}, E)$ (eqn 10.11) integrated over all values of energy transfer, that is, $\int S(\mathbf{Q}, E)\,\mathrm{d}E$, will provide information about the instantaneous separations of atoms – this is the measurement that provides information about the short-range structures of glasses and other disordered materials.

(10.5) If a neutron scattering experiment is performed on a triple axis spectrometer to determine the phonon dispersion curves of a ccp metal, what information will be provided by a constant-$\mathbf{Q}$ experiment performed at the following values of $\mathbf{Q}$:

(a) (2.2, 0, 0), (d) (3, 1, 0),
(b) (2.2, 0, 4), (e) (3.2, 2.2, 4),
(c) (1.7, 1.7, 2.3), (f) (2.6, 3.4, −2).

(10.6) For each of the four vibrations shown in Fig. 8.15 determine whether the vibration could be measured in a Raman scattering or infrared absorption experiment.

(10.7) The formalism for $S(\mathbf{Q}, E)$ developed in the chapter has assumed that all atoms of the same type scatter the same way. This is not always true for neutron scattering. One exception to this assumption is when two different isotopes of one type of atom have different neutron scattering lengths. Another exception is for nuclei with spin. One important example is hydrogen, where there is a significant difference in the scattering length when the spins of the neutron and proton point along the same or opposite directions. In this case, we must write $S(\mathbf{Q}, E)$ as

$$S(\mathbf{Q}, E) = \sum_{j,k} \langle b_j b_k \exp[i\mathbf{Q} \cdot (\mathbf{r}_j(t) - \mathbf{r}_k(0))] \rangle$$

where we have included the scattering lengths of the individual atoms within the averages. Explain why the equation for $S(\mathbf{Q}, E)$ can be rewritten as

$$S(\mathbf{Q}, E) = \sum_{j,k} \langle b_j b_k \rangle \langle \exp[i\mathbf{Q} \cdot (\mathbf{r}_j(t) - \mathbf{r}_k(0))] \rangle$$

where the average for any site is over all possible values for the scattering length of the atom of that type.

The general equation for the scattering function can be written in terms of two contributions:

$$S(\mathbf{Q}, E) = S_{\text{coherent}}(\mathbf{Q}, E) + S_{\text{incoherent}}(\mathbf{Q}, E)$$

where

$$S_{\text{coherent}}(\mathbf{Q}, E) = \sum_{j,k} \overline{b}_j \overline{b}_k \langle \exp[i\mathbf{Q} \cdot (\mathbf{r}_j(t) - \mathbf{r}_k(0))] \rangle$$

and

$$S_{\text{incoherent}}(\mathbf{Q}, E) = \sum_{j,k} (\overline{b_j b_k} - \overline{b}_j \overline{b}_k) \times \langle \exp[i\mathbf{Q} \cdot (\mathbf{r}_j(t) - \mathbf{r}_k(0))] \rangle$$

where $\overline{b} = \langle b \rangle$. $S_{\text{coherent}}(\mathbf{Q}, E)$ is the **coherent scattering function**, and corresponds to the type of scattering discussed in this chapter. The second term is the **incoherent scattering function**, because it is sensitive to the variations of the scattering lengths for any atom type. Show that the value of $\overline{b_j b_k} - \overline{b}_j \overline{b}_k$ is only non-zero if $j = k$, that is, the average is zero when we consider different atoms. In this case, if we write $\sigma_j = \overline{b_j^2} - (\overline{b}_j)^2$, show that

$$S_{\text{incoherent}}(\mathbf{Q}, E) = \sum_j \sigma_j \langle \exp[i\mathbf{Q} \cdot (\mathbf{r}_j(t) - \mathbf{r}_j(0))] \rangle$$

What information is contained within the incoherent scattering function?

(10.8) Suppose an experiment is performed to measure the phonon dispersion curves of potassium on a triple-axis spectrometer, when the energy of the beam scattered into the analyser is held fixed at 3.5 THz. For a measurement of the LA mode at $\mathbf{Q} = [2.5, 0, 0]$, what will be the energy of the incident beam for an experiment in which the neutron beam loses energy in the creation of a phonon. The phonon dispersion curves of potassium are given in Fig. 8.10, and the lattice parameter of potassium is 5.23 Å.

(10.9) Contrast the techniques of neutron scattering, Raman scattering and infrared absorption. What are the relative merits or disadvantages of each technique?

(10.10) Explain how measurements of the Raman scattering signal for both cases of energy being absorbed from the sample and energy lost to the sample in a single experiment can provide a measurement of the sample temperature.

11 Anharmonic interactions

11.1 Introduction

The basic harmonic model theory of lattice dynamics developed in Chapters 8 and 9 has proven to be extremely powerful. It has enabled an understanding of a wide range of basic phenomena, such as the propogation of sound waves, thermodynamic properties, and interaction of radiation such as light, X-rays, and neutron beams with crystalline matter (Chapter 10). It has provided the link between the microscopic interatomic interactions introduced in Chapter 5, as represented by the force constants, and these phenomena. The harmonic model is also very useful as a predictive tool, for example facilitating the prediction of thermodynamic properties. However, there are some phenomena that cannot be explained within the limits of the harmonic approximation. These include thermal expansion, thermal conductivity, the temperature dependence of phonon frequencies, and corresponding macroscopic properties (such as dielectric and elastic constants), the observation that phonons appear to have finite lifetimes, and the existence of displacive phase transitions. It is the purpose of this chapter to look at how the basic harmonic model can be modified to account for these phenomena.

In Chapter 8 (eqn 8.2) we expanded the energy of the crystal in terms of atomic displacements as

$$E = E_0 + \frac{1}{2}\sum_{\substack{j,j'\\ \alpha,\alpha'}} \frac{\partial^2 E}{\partial u_{j,\alpha}\partial u_{j',\alpha'}} u_{j,\alpha}u_{j',\alpha'} + \cdots + \frac{1}{n!}\sum_{\substack{j^{(1)},\ldots,j^{(n)}\\ \alpha^{(1)},\ldots,\alpha^{(n)}}} \frac{\partial^n E}{\partial u_{j^{(1)},\alpha^{(1)}}\cdots\partial u_{j^{(n)},\alpha^{(n)}}} u_{j^{(1)},\alpha^{(1)}}\cdots u_{j^{(n)},\alpha^{(n)}} + \cdots \tag{11.1}$$

The harmonic approximation consisted in taking account only of the term that is quadratic in the atomic displacements (the linear term sums to zero in the equilibrium structure). All the higher-order terms are the **anharmonic** terms. We noted that the virtue of the harmonic approximation is that the equations can be solved exactly – the corollary of this is that the anharmonic model does not have an exact solution. The easiest way forward is to consider that the anharmonic interactions simply act to modify the harmonic picture. Provided that the anharmonic terms are not too large, this is a reasonable approach, and

we will restrict ourselves to this limiting case. An example of this approach is the **quasiharmonic** approximation, in which the anharmonic interactions are assumed to give rise to changes in the phonon frequencies through their dependence on the structure of the material, and these phonon frequencies are then used within the harmonic model for calculations of quantities such as the phonon thermodynamic functions. The phonon frequencies can be modified through indirect effects, such as their dependence on crystal volume (the usual understanding of the **quasiharmonic** approximation), or by direct interactions between different phonons (through what is usually called **renormalized phonon theory**).

We will find that a great deal of progress can be obtained by retaining the concept of the normal mode coordinate, and expanding the full anharmonic Hamiltonian in terms of the normal mode coordinates in the way that we did for the harmonic Hamiltonian at the end of Chapter 9 (eqn 9.49):

$$\begin{aligned}\mathcal{H} = {} & \frac{1}{2}\sum_{\mathbf{k},\nu}\dot{Q}(\mathbf{k},\nu)\dot{Q}(-\mathbf{k},\nu) + \frac{1}{2}\sum_{\mathbf{k},\nu}\omega^2(\mathbf{k},\nu)Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu) \\ & + \frac{1}{3!}\sum_{\substack{\mathbf{k},\mathbf{k}',\mathbf{k}'' \\ \nu,\nu',\nu''}}\alpha^{(3)}_{\substack{\mathbf{k},\mathbf{k}',\mathbf{k}'' \\ \nu,\nu',\nu''}}Q(\mathbf{k},\nu)Q(\mathbf{k}',\nu')Q(\mathbf{k}'',\nu'')\Delta(\mathbf{k}+\mathbf{k}'+\mathbf{k}'') \\ & + \frac{1}{4!}\sum_{\substack{\mathbf{k},\mathbf{k}',\mathbf{k}'',\mathbf{k}''' \\ \nu,\nu',\nu'',\nu'''}}\alpha^{(4)}_{\substack{\mathbf{k},\mathbf{k}',\mathbf{k}'',\mathbf{k}''' \\ \nu,\nu',\nu'',\nu'''}}Q(\mathbf{k},\nu)Q(\mathbf{k}',\nu')Q(\mathbf{k}'',\nu'')Q(\mathbf{k}''',\nu''') \\ & \qquad \times \Delta(\mathbf{k}+\mathbf{k}'+\mathbf{k}''+\mathbf{k}''') \\ & + \cdots \end{aligned} \tag{11.2}$$

The $\alpha^{(n)}$ coefficients are related to the n-th derivative of the energy analogous to ω^2 for the harmonic coefficient. The important point in the equation, which is implicit in the harmonic term, is that there is a conservation law operating for the wave vectors of the normal mode coordinates in each term. This is represented by the functions $\Delta(\mathbf{K})$, where $\mathbf{K}$ is any combination of wave vectors, such that

$$\Delta(\mathbf{K}) = \begin{cases} 1 & \text{if } \mathbf{K} = \mathbf{G} \\ 0 & \text{if } \mathbf{K} \neq \mathbf{G} \end{cases} \tag{11.3}$$

where $\mathbf{G}$ is a reciprocal lattice vector.

When this Hamiltonian is treated properly within the framework of quantum mechanics, specifically within the formalism of what is called **second quantization**, it can be shown that the anharmonic terms have a physical interpretation in terms of collisions between phonons that lead to changes in the frequencies and wave vectors, and even creation or annihilation of phonons. Consider the third-order anharmonic term. This corresponds to events such as those shown in Fig. 11.1. In one case, one phonon spontaneously decays into two of lower energy (this rather stretches the meaning of a collision!). In the other case, two phonons merge to form a third, with an energy equal to the sum of the energies of the initial two phonons. The fourth-order anharmonic interactions are also shown in Fig. 11.1. These are more complicated than the third-order interactions. In one case, two phonons interact to form two different phonons, with changes in energy and wave vector. In the other two cases, one phonon

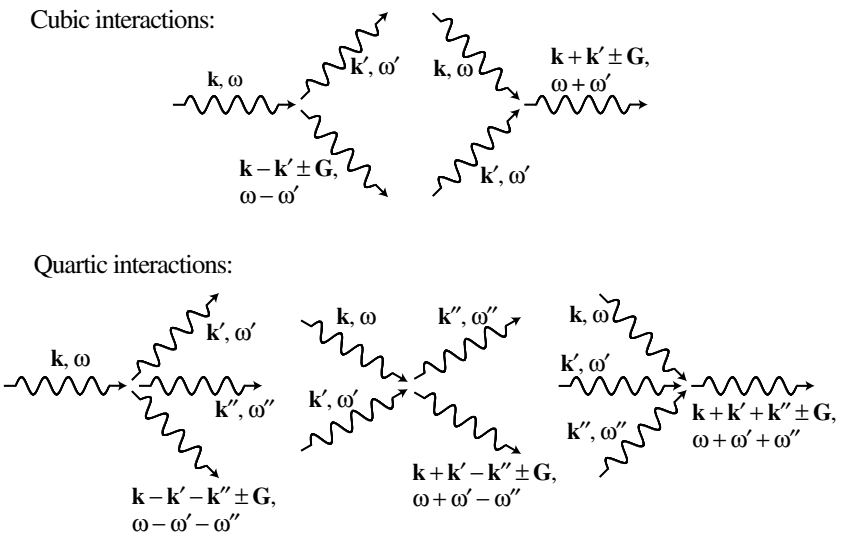


Fig. 11.1 Third- and fourth-order anharmonic phonon interactions represented as collisions between phonons with conservation restrictions. The phonons are represented by waves with arrows.

spontaneously decays into three of lower energy, or three phonons merge to form a single phonon.

In the collision processes shown in Fig. 11.1, the phonons are sketched in a way that suggests that the motions either side of the interactions are in the same direction. However, the law guiding conservation of wave vector allows a change in the overall wave vector by addition or subtraction of a reciprocal lattice vector. As a result, the flow of energy can be changed, even reversed, by the anharmonic interactions.

It is worth remarking that although the third-order interactions are expected to be larger than the fourth-order interactions, the constraints on the possible interactions due to the conditions of conservation of energy and wave vector, eqn 11.3, are more restrictive for the third-order terms than the fourth-order terms. Because of this, the fourth-order terms make a significant contribution to the anharmonic energy of a crystal.

The fact that we can picture anharmonic processes as events that change the wave vectors and energies (and even causing creation and annihilation) of phonons, means that we can assign a mean lifetime to a phonon. One can think of the lattice vibration being described as a sinusoidal function multiplied by an exponential decay to represent the finite lifetime. The Fourier transform, which gives the description of the wave in the frequency domain rather than the time domain, will be described by the convolution of a Dirac δ-function at the vibrational frequency and the Fourier transform of the exponential function, which is a Lorentzian function – this is demonstrated in Appendix B, Section B.3.3. Thus a measurement of an anharmonic phonon peak in a high-resolution spectroscopy experiment (Chapter 10) will give a peak with a width that is equal to the inverse of the phonon lifetime. The width of this peak will increase on heating, since with more phonons in the system the probability per unit time of a phonon being involved in a collision with other phonons increases. An example of a spectroscopic measurement of a phonon peak on increasing temperature is shown in Fig. 11.2, with the variation of its width with temperature shown in Fig. 11.3.

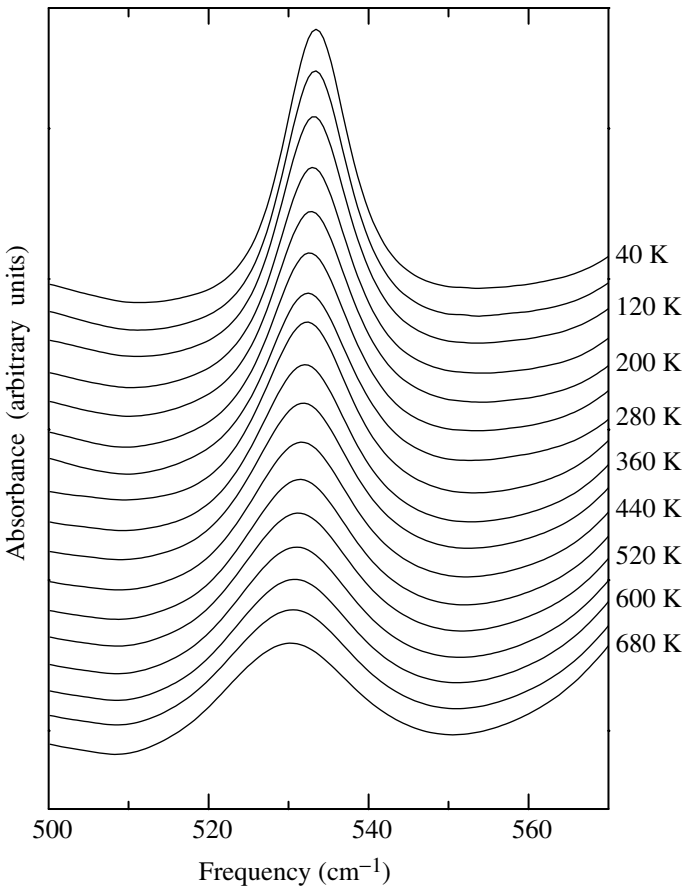


Fig. 11.2 Example of temperature dependence of one of the phonon spectra in the mineral albite, $NaAlSi_3O_8$, measured in an infrared spectroscopy experiment (Chapter 10). (Data courtesy of Ming Zhang, University of Cambridge.)

11.2 Thermal conductivity

Thermal conductivity gives the ease with which energy in the form of heat can pass through a material. In the harmonic approximation, a heat source will create lattice vibrations in the form of phonons, the phonons will move through the sample, and will deposit their energy at the other end of the material. Since there is nothing within the harmonic model to inhibit the flow of phonons, the thermal conductivity would be infinite in value (which we will demonstrate below). Experimentally it is found that the thermal conductivity decreases on heating, which implies that the phonons must be inhibiting the flow of heat.

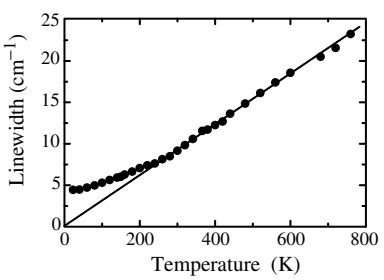


Fig. 11.3 Temperature dependence of the width of the phonon peak shown in Fig. 11.2. Note that quantum effects come into play for temperatures below 250 K.

The situation we need to analyse is shown schematically in Fig. 11.4. This shows a rod of unit area, along which is a temperature gradient. For there to be a true flow of heat energy, the heat energy needs to be carried by phonons and deposited some distance away. Moreover, in any region of the rod it is necessary for the phonon distribution to attain equilibrium. The first condition could be initiated by the existence of defects (Appendix A), in which case the length λ_x would correspond to the mean distance along the x direction that a phonon can travel before being scattered by a defect. However, the existence of defects will not enable different regions of the crystal to reach thermal equilibrium: this has

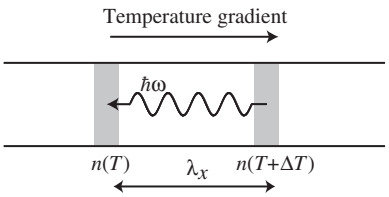


Fig. 11.4 Phonon drift in a rod leading to a finite value of the thermal conductivity.

to be due to anharmonic interactions between phonons. Thus we will identify λ_x as the x component of the mean-square distance a phonon can travel before being scattered by other phonons.

The flow of phonons between two regions of different temperature occurs because the number of phonons in each region is different. Thus we have a diffusion of phonons between the hotter and cooler regions. The temperature difference between two regions separated by distance λ_x is trivially given as

$$\Delta T = \lambda_x \frac{\mathrm{d}T}{\mathrm{d}x} \tag{11.4}$$

If λ_x is determined by anharmonic interactions, we can write

$$\lambda_x = v_x \tau \tag{11.5}$$

where v_x is the component of the phonon velocity along the x direction, and τ is the mean time between collisions. The difference in phonon numbers between the two regions is given as

$$\Delta n = n(T + \Delta T) - n(T) = \frac{\partial n}{\partial T}\Delta T = v_x \tau \frac{\partial n}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}x} \tag{11.6}$$

The flux of phonons crossing a point between the two regions is given by

$$J_{\mathrm{ph}} = -\Delta n \frac{3N}{V} v_x \tag{11.7}$$

where $3N/V$ is the number of phonon modes per unit volume (equal to 3 times the number of atoms per unit volume), and the negative sign accounts for the fact that the heat flow is in the opposite direction to the temperature gradient. Thus we can write

$$J_{\mathrm{ph}} = -\frac{3N}{V} v_x^2 \tau \frac{\partial n}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}x} \tag{11.8}$$

To convert to a flux of energy we assume each phonon carries an average energy $\hbar\omega$. The heat energy flux is therefore given as $J = \hbar\omega J_{\mathrm{ph}}$. We note that we can write $v_x^2 = \overline{v^2}/3$, where $\overline{v^2}$ is the mean-square phonon velocity (we are now averaging over all directions of phonon motion). The quantity

$$\frac{3N}{V}\hbar\omega\frac{\partial n}{\partial T} = c_{\mathrm{V}}$$

is simply the heat capacity per unit volume. The heat flux is given as

$$J = -K\frac{\mathrm{d}T}{\mathrm{d}x} \tag{11.9}$$

where K is the coefficient of thermal conductivity. Pulling together eqns 11.8–11.9, we obtain

$$K = \frac{1}{3}\overline{v^2}\tau c_{\mathrm{V}} \tag{11.10}$$

In the absence of anharmonic interactions, and if the material is free of defects, $\lambda = \infty$, and hence the thermal conductivity will be infinite. However,

it is found experimentally that the value of K falls on increasing temperature. At high temperature, the heat capacity is constant (at its maximum value), and the phonon velocity is roughly constant with temperature (in contrast to classical particles). The temperature dependence of K is primarily due to the temperature dependence of τ, which typically varies as $\tau \propto 1/T$. This arises from the anharmonic interactions. As a result, at high temperature $K \propto 1/T$. On the other hand, at low temperature the mean free phonon path length λ is determined either by the mean distance between defects or by the dimensions of the sample. As a result, τ has a constant value, and the temperature dependence of K is determined by the temperature dependence of c_V. Thus K also falls at low temperatures. The overall temperature dependence of K is shown schematically in Fig. 11.5.

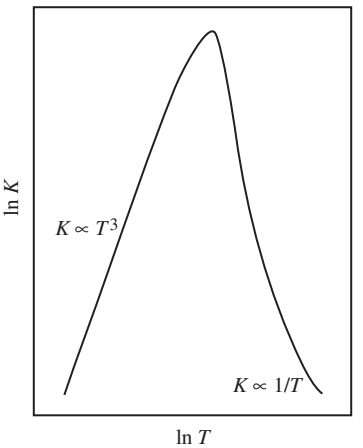


Fig. 11.5 Schematic representation of the temperature dependence of the thermal conductivity.

Implicit in the analysis developed above is that when the phonon deposits its energy, it does so in the process of establishing thermodynamic equilibrium in the local environment. The state of thermodynamic equilibrium is that for which all phonon modes are excited in accordance with the Bose–Einstein distribution for that temperature. This includes the modes of equivalent frequency for wave vectors $\mathbf{k}$ and $-\mathbf{k}$. Thus in the equilibrium state, the sum of wave vectors of all phonons is zero. In the process of the phonon flow that transfers heat, the group of phonons that carries the heat has a net component of wave vector in the direction of the flow of heat. When the phonons undergo anharmonic interactions, the conservation law implied by eqn 11.3 acts as a constraint on the allowed collision processes. For an anharmonic interaction involving n initial phonons of wave vector $\mathbf{k}_i$ and n' final phonons of wave vector $\mathbf{k}'_j$, the conservation law can be rewritten as

$$\sum_{i=1}^{n} \mathbf{k}_i = \sum_{j=1}^{n'} \mathbf{k}'_j + \mathbf{G} \tag{11.11}$$

Clearly if $\mathbf{G} = 0$ there is no change in the overall wave vector. Such a process is called a **normal** process, and cannot assist in the development of thermal equilibrium for the pool of phonons. In order for the flow of phonons to be converted to one with a net wave-vector value of zero, it is necessary for many of the anharmonic interactions to have $\mathbf{G} \neq 0$. These processes are called **Umklapp** processes, and it is the existence of the Umklapp processes that gives the possibility for the phonons to deposit their heat energy as the local environment establishes thermodynamic equilibrium.

11.3 Thermal expansion

11.3.1 Theory

The coefficient of volume thermal expansion, β, can be written as

$$\beta = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P = K_T \left(\frac{\partial P}{\partial T}\right)_V \tag{11.12}$$

where K_T is the isothermal compressibility:

$$K_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T = \sum_{i,j=1}^{3} S_{ij} \tag{11.13}$$

The latter part of the equation relates the compressibility to the coefficients of the inverse of the elastic constant tensor, called the **elastic compliance** tensor (Section 7.5.2).

Pressure is given by the differential of the free energy:

$$P = -\left(\frac{\partial F}{\partial V}\right)_T \tag{11.14}$$

For greater simplicity and transparency we will work within the high-temperature limit, and will cite the results for the general case at the end. The free energy of the crystal is given from Chapter 9 as

$$F = E + k_B T \sum_j \ln(\hbar\omega_j / k_B T) \tag{11.15}$$

where E is the lattice energy. The pressure is obtained as

$$P = -\frac{\partial E}{\partial V} - k_B T \sum_j \frac{1}{\omega_j}\frac{\partial \omega_j}{\partial V} \tag{11.16}$$

In general frequencies will increase on decreasing volume since the force constants will be larger as atoms are forced closer together. Thus the differential of frequency with volume will be negative, and the second term will give a positive contribution to the pressure.

We now obtain the differential of pressure with respect to temperature:

$$\frac{\partial P}{\partial T} = -k_B \sum_j \frac{1}{\omega_j}\frac{\partial \omega_j}{\partial V} \tag{11.17}$$

where we have assumed that the phonon frequencies do not have a direct dependence on temperature, but change with temperature only through their dependence on volume. This assumption is not appropriate when we consider phase transitions (Section 11.4, Chapter 12), but in the general case it can be reasonable.

When we put everything together we obtain

$$\beta = -K_T k_B \sum_j \frac{1}{\omega_j}\frac{\partial \omega_j}{\partial V} \tag{11.18}$$

It is common to rewrite this equation using a dimensionless parameter called the **Grüneisen parameter**, defined as

$$\gamma_j = -\frac{V}{\omega_j}\frac{\partial \omega_j}{\partial V} = -\frac{\partial \ln \omega_j}{\partial \ln V} \tag{11.19}$$

The Grüneisen parameter will be positive for the usual case where a mode frequency increases as the volume decreases. By working with one mole of atoms, we can re-express the equation for thermal expansion as

$$\beta = \frac{K_T k_B}{V} \sum_j \gamma_j = \frac{3K_T R\gamma}{V} \tag{11.20}$$

where we have defined an overall γ as the sum over the $3N_A$ individual Grüneisen parameters, and V is the volume occupied by one mole of atoms.

If we use the complete quantum mechanical free energy, we obtain

$$P = -\frac{\partial E}{\partial V} - \frac{1}{2}\sum_j \hbar \frac{\partial \omega_j}{\partial V} - \sum_j n(\omega_j, T)\frac{\partial \omega_j}{\partial V} \tag{11.21}$$

The coefficient of thermal expansion is then obtained as

$$\beta = -K_T \hbar \sum_j \frac{\partial \omega_j}{\partial V}\frac{\partial n(\omega_j, T)}{\partial T} \tag{11.22}$$

The term $\partial n/\partial T$ reminds us of the heat capacity. We can define the **mean Grüneisen parameter** as

$$\gamma = \frac{1}{C_V}\sum_j \gamma_j \hbar \omega_j \frac{\partial n(\omega_j, T)}{\partial T} \tag{11.23}$$

This then gives

$$\beta = \frac{K_T \gamma C_V}{V} \tag{11.24}$$

11.3.2 Example: calculation of thermal expansion in fluorite

To illustrate the discussion, the overall Grüneisen parameter and coefficient of volume thermal expansion for fluorite, CaF_2, have been calculated, and the results are shown in Fig. 11.6. The calculations were carried out using the same model as used for the calculation of thermodynamic properties in Chapter 9. The Grüneisen parameter has a value of around 1.5 from a temperature of 200 K upwards. The peak at lower temperatures is not uncommon, and the value of the Grüneisen parameter almost always decreases at low temperature.

The temperature dependence of the coefficient of thermal expansion nearly reflects the temperature dependence of the heat capacity, since the temperature dependence of the volume, compressibility, and Grüneisen parameter is much weaker. In particular, the coefficient of thermal expansion falls to zero on cooling to 0 K. Typical values for the coefficient of thermal expansion are usually of order 10^{-5} K^{-1}.

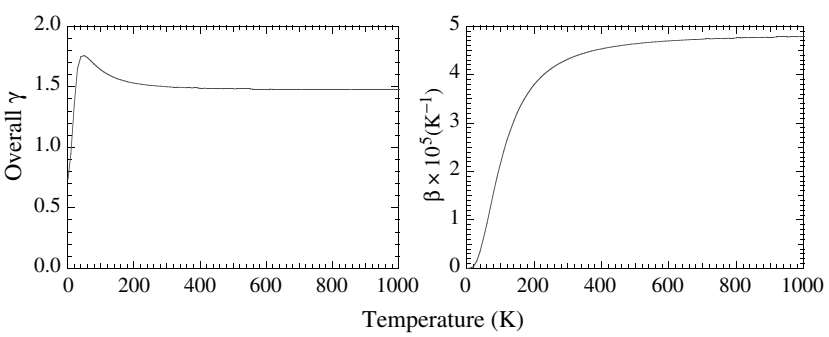


Fig. 11.6 Temperature dependence of the overall Grüneisen parameter (left) and coefficient of volume thermal expansion (right) for fluorite, CaF_2, calculated using a simple model set of interatomic interactions (Chapter 5).

11.4 Temperature dependence of phonon frequencies

We cannot solve the anharmonic Hamiltonian, eqn 11.2 exactly. Instead we will look for an approximate solution that retains the main features of the harmonic model. The quasiharmonic approach of the preceding section assumes that the temperature dependence of the phonon frequencies is only due to the dependence on force constants with volume, leading to a temperature dependence of the phonon frequencies that is due only to the temperature dependence of the crystal volume. Our task now is to develop the anharmonic theory to obtain an estimate of the intrinsic temperature dependence of phonon frequencies.

Our approach will eventually require us to have pairs of normal mode coordinates, which means that we need to work with even-order terms in the anharmonic expansion. We focus on the quartic Hamiltonian, neglecting the kinetic energy term:

$$\begin{aligned}\mathcal{H} = &\frac{1}{2}\sum_{\mathbf{k},\nu}\omega^2(\mathbf{k},\nu)Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu)\\ &+\frac{1}{4!}\sum_{\substack{\mathbf{k},\mathbf{k}',\mathbf{k}'',\mathbf{k}'''\\ \nu,\nu',\nu'',\nu'''}}\alpha^{(4)}_{\substack{\mathbf{k},\mathbf{k}',\mathbf{k}'',\mathbf{k}'''\\ \nu,\nu',\nu'',\nu'''}}Q(\mathbf{k},\nu)Q(\mathbf{k}',\nu')Q(\mathbf{k}'',\nu'')Q(\mathbf{k}''',\nu''')\\ &\qquad\times\Delta(\mathbf{k}+\mathbf{k}'+\mathbf{k}''+\mathbf{k}''')\end{aligned}\tag{11.25}$$

We apply what is called a **mean-field** approach. This is a common approximation in condensed matter physics by which an averaging is carried out that, in effect, neglects localized fluctuations. Mean-field approximations are common in the study of order/disorder processes, such as ferromagnetism. Often the central assumption causes grave errors in the calculations of quantities such as phase transition temperatures, but the approximation can be useful to make progress when a more exact treatment is difficult – the mean-field approximation gives a certain level of transparency to any theory. In the treatment of eqn 11.25, the mean-field approximation involves replacing pairs of normal mode coordinates by their thermal averages. We proceed by assuming that we can also work within the high-temperature limit:

$$\begin{aligned}Q(\mathbf{k}'',\nu'')Q(\mathbf{k}''',\nu''') &\rightarrow \langle Q(\mathbf{k}'',\nu'')Q(\mathbf{k}''',\nu''')\rangle\\ &\simeq \frac{k_\mathrm{B}T}{\omega^2(\mathbf{k}'',\nu'')}\delta_{\mathbf{k}'',-\mathbf{k}'''}\delta_{\nu'',\nu'''}\end{aligned}\tag{11.26}$$

where the average is only non-zero if $\mathbf{k}'' = -\mathbf{k}'''$ by conservation of wave vector, and $\nu'' = \nu'''$ by the orthogonality condition. Thus the anharmonic Hamiltonian can be written as

$$\begin{aligned}\mathcal{H} = &\frac{1}{2}\sum_{\mathbf{k},\nu}\omega^2(\mathbf{k},\nu)Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu)\\ &+\frac{1}{4}\sum_{\substack{\mathbf{k},\mathbf{k}''\\ \nu,\nu''}}\alpha^{(4)}_{\substack{\mathbf{k},-\mathbf{k},\mathbf{k}'',-\mathbf{k}''\\ \nu,\nu,\nu'',\nu''}}Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu)\langle Q(\mathbf{k}'',\nu'')Q(-\mathbf{k}'',\nu''')\rangle\end{aligned}$$

$$= \frac{1}{2}\sum_{\mathbf{k},\nu}\omega^2(\mathbf{k},\nu)Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu) + \frac{k_\mathrm{B}T}{4}\sum_{\substack{\mathbf{k},\mathbf{k}''\\ \nu,\nu''}}\alpha^{(4)}_{\substack{\mathbf{k},-\mathbf{k},\mathbf{k}'',-\mathbf{k}''\\ \nu,\nu,\nu'',\nu''}}Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu)/\omega^2(\mathbf{k}'',\nu'') \tag{11.27}$$

There are six ways of selecting independent pairs from the fourth-order term in the anharmonic Hamiltonian, which is why the numerical prefactor has been multiplied by a factor of 6. This equation now looks like a harmonic Hamiltonian (eqn 9.49), since we have pairs of normal mode coordinates in both the last two terms. We can therefore merge these terms to obtain

$$\mathcal{H} = \frac{1}{2}\sum_{\mathbf{k},\nu}\left(\omega^2(\mathbf{k},\nu) + \frac{k_\mathrm{B}T}{2}\sum_{\mathbf{k}'',\nu''}\alpha^{(4)}_{\substack{\mathbf{k},-\mathbf{k},\mathbf{k}'',-\mathbf{k}''\\ \nu,\nu,\nu'',\nu''}}/\omega^2(\mathbf{k}',\nu')\right) \times Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu) \tag{11.28}$$

$$= \frac{1}{2}\sum_{\mathbf{k},\nu}\tilde{\omega}^2(\mathbf{k},\nu)Q(\mathbf{k},\nu)Q(-\mathbf{k},\nu) \tag{11.29}$$

where we have a new set of **renormalized phonon frequencies:**

$$\tilde{\omega}^2(\mathbf{k},\nu) = \omega^2(\mathbf{k},\nu) + \frac{k_\mathrm{B}T}{2}\sum_{\mathbf{k}'',\nu''}\alpha^{(4)}_{\substack{\mathbf{k},-\mathbf{k},\mathbf{k}'',-\mathbf{k}''\\ \nu,\nu,\nu'',\nu''}}/\omega^2(\mathbf{k}',\nu') \tag{11.30}$$

This applies to all phonons in the crystal: the frequency of each phonon is modified by its anharmonic interaction will all other phonons in the crystal. This mechanism gives a temperature dependence to all phonons over and above that due to thermal expansion. If the coefficients are positive, the phonon frequencies will increase with temperature. This is opposite to the usual effects of thermal expansion, which cause a decrease in the phonon frequency since the force constants decrease as bonds increase in length. In fact, the effects of thermal expansion usually dominate, and hence mask, the intrinsic anharmonic effect. The increase of the phonon frequency due to the intrinsic anharmonic interactions is of major importance in giving the microscopic origin of displacive phase transitions, as discussed in Chapter 12.

Summary of chapter

- **Anharmonic processes** can be represented as processes involving interactions between phonons, such as one phonon decaying to form two or more new phonons, two or more phonons merging to form a new phonon, or two or more phonons combining to form different phonons. These interactions are subject to the constraint that energy must be conserved, but the wave vector is only conserved to within an additive reciprocal lattice vector. The interactions between phonons give rise to a finite lifetime of each phonon, which is measured as the inverse of the width of a phonon peak in a spectroscopy experiment.

- Thermal conductivity is one example of a process that can only be explained by anharmonic phonon theory. The important phonon interactions are those in which the sum of the wave vectors changes by a reciprocal lattice vector: such an interaction is known as an **Umklapp process**.
- Thermal expansion can be described using harmonic phonon theory in which the harmonic frequencies are assumed to be functions of volume V. The effect is described by the dimensionless **Grüneisen parameter**: $\gamma = -\partial \ln \omega / \partial \ln V$. The resultant thermal expansion is a simple function of γ, and can readily be calculated from standard lattice dynamics models.
- Anharmonic processes give a temperature dependence of phonon frequencies that is independent of the effect of thermal expansion.

Further reading

Both the standard texts of Kittel (1996) and Ashcroft and Mermin (1976) contain good introductory descriptions of anharmonic processes. More advanced discussions of a range of anharmonic effects are given by Baron and White (1999), Krishnan *et al.* (1979), and Srivastava (1990).

Exercises

(11.1) Show that the Grüneisen parameter for a ccp crystal with nearest-neighbour Lennard-Jones interactions (eqn 5.23) has value 3.5. (*Hint*: the values of ω^2 are determined by the second differential of the energy – express the Grüinesen parameter in terms of ω^2 rather than ω, and replace the differential with respect to volume by the differential with respect to nearest-neighbour distance through $\partial/\partial V = (\partial V/\partial r)^{-1}\partial/\partial r$.)

(11.2) Obtain an expression for the coefficient of thermal expansion of the same crystal at high temperature, using the expression for the bulk modulus of the same crystal obtained in Problem 5.2 and the result of the previous question, together with the fact that at high temperature the heat capacity per mole of atoms is equal to $3R$. Use values of the parameters in the Lennard-Jones potential for neon (Problem 5.1) to obtain the coefficient of thermal expansion.

Displacive phase transitions

12

12.1 Introduction to displacive phase transitions

Many applications of the solid state sciences exploit the electronic or mechanical properties of solids. The important materials are metals or semiconducting materials. These materials tend to have relatively simple crystal structures, such as the cubic and hexagonal structures we met in Chapter 2 (fcc, hcp, diamond, NaCl, and ZnS structures), and interest is focused more on the behaviour of the electrons than on the behaviour of the atoms *per se*. On the other hand, there are many properties of materials that depend on some degree of complexity of the crystal structure, and on its anisotropy (Chapter 7), and these can be significantly enhanced if the material undergoes a phase transition at some temperature. This is particularly true for dielectric, pyroelectric, and piezoelectric properties. Moreover, displacive phase transitions are associated with important phenomena such as high-temperature superconductivity in cuprate ceramics.

Not only are phase transitions of considerable technological importance, it is also true that the study of phase transitions is particularly enjoyable because it brings together many of the topics we have discussed as separate topics in earlier chapters. In particular, the study of phase transitions brings together issues of crystal structures, symmetry, bonding, thermodynamics, and lattice dynamics, including the need to include anharmonic interactions. Because of this, it is appropriate to close this book with a survey of phase displacive transitions in crystalline materials.

We met a number of phase transitions in Chapter 2. It can be argued that the most important examples are the perovskites, which have displacive phase transitions that involve ferroelectric distortions and rotations of structural polyhedra. These types of phase transitions are not restricted to perovskites, and we will also discuss phases of silica that have similar phase transitions. We also met phase transitions involving rotations of molecules (C_{60} and N_2), and phase transitions involving disordering of positions of atoms, as in the fast-ion conducting phases such as AgI.

In this chapter we are primarily concerned with phase transitions in which the crystal structure of one phase is a distorted form of the crystal structure of the other phase, and where the transition is accomplished by small displacements of the atoms. These are called displacive phase transitions. This point is illustrated in Fig. 12.1, in which we show the crystal structures of the phases of the silica (SiO_2) polymorphs quartz and cristobalite. Both polymorphs undergo displacive phase transitions, in which the symmetry of one phase is lowered by small displacements of some of the atoms. In quartz, the change in symmetry

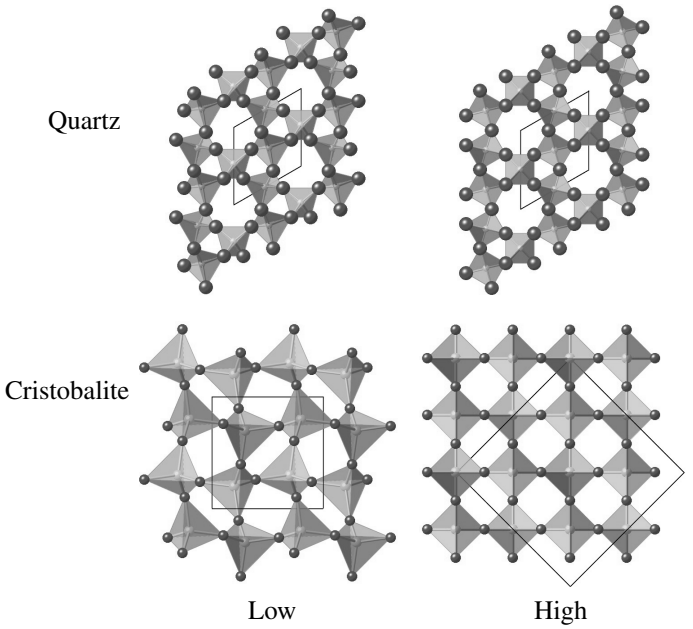


Fig. 12.1 Crystal structures of the two phases of quartz and cristobalite.

is produced by small rotations of the SiO_4 tetrahedra, as a result of which the 6-fold rotation symmetry is replaced by 3-fold rotation symmetry. Thus the symmetry of the lattice is lowered from hexagonal to trigonal, with a change in the point group from 622 to 32 (the space group changes from $P6_222$ to $P3_221$). In the case of cristobalite, there are similar rotations of the SiO_4 tetrahedra, and in this case the changes in symmetry are from face-centred cubic (point group $m\bar{3}m$, space group $Fd3m$) to tetragonal (point group 422, space group $P4_12_12$), with the unit cell of the tetragonal phase encompassing two primitive unit cells of the cubic phase. The phase transitions in the perovskite structures we met earlier are other examples of displacive phase transitions.

The defining feature of a displacive phase transition may be understood to be a change in symmetry. There are cases of phase transitions where it appears that the change in symmetry is not a defining feature, but these cases are rare, and when reported they may need to be treated with some caution. The point about symmetry is that it is either present or absent, there is no intermediate state. The deformation of the structure may be reduced continuously on heating up to a phase transition, but whilst there is even an infinitesimal distortion of the structure, the symmetry is lost. The changes of symmetry associated with phase transitions in crystalline silica polymorphs have been highlighted above. In the case of perovskites, the ferroelectric distortion on transforming from the cubic phase of point group $m\bar{3}m$ to $4mm$ involves loss of the 3-fold axes and the centre of symmetry. It is important to note that the mere transition from cubic to tetragonal, as would be achieved by a simple expansion of the lattice along [001], does not describe the complete symmetry change; this is a point that will be important later when we discuss couplings between strain and atomic displacements. The rotational phase transitions in perovskites involve a

change in the description of the unit cell. The unit cell of the lower-symmetry phase encompasses two unit cells of the cubic phase in each direction. The change in symmetry in this case includes translational symmetry.

A phrase often used to describe the change of symmetry at a phase transition is **broken symmetry**. An example of broken symmetry is shown in Fig. 12.2. The medieval cathedral of Ely once had a symmetrical west end, with a large central tower and two side towers. The defining symmetry was a vertical mirror plane. Sadly, one tower collapsed and was never rebuilt. The loss of symmetry is quite clear.

The starting point to understanding the origin of phase transitions is with thermodynamics. We know that the equilibrium state of matter is that for which the free energy is a global minimum. This is an incredibly powerful statement. If we have some means to write down the free energy as a function of the critical variables (such as volume as a general variable, or some measure of the distortion of the structure as a specific variable), minimization of the free energy will yield an equation for the temperature or pressure dependence of these variables. It will also then become possible to obtain properties such as the susceptibility or heat capacity, quantities that are known to have an anomalous behaviour associated with a phase transition.

Fig. 12.2 A real case of broken symmetry: the mirror symmetry of the west front to Ely cathedral was lost when the north tower collapsed.

12.1.1 Importance of thermodynamic analysis

We recall from Chapter 5 that the free energy can be written as

$$G = U - TS + PV \tag{12.1}$$

It is nearly always the case that the lower-symmetry phase will be stable at lower temperatures and higher pressures, and the higher-symmetry phase will be stable at higher temperatures and lower pressures. We can rationalize this from the general form of the free energy in a way that shows the power of this concept. The internal energy U is primarily associated with the bond energy and the energy stored in the phonons. Clearly at low temperature the free energy will be dominated by the enthalpy, $H = U + PV$. Typically the distortion of the structure will lead to an increase in its density, and the lower volume will lead to a lower energy U and a lower value of PV, the latter term becoming more important at higher pressures. Hence we are able to understand why the distorted phase will be the stable phase at low pressure and low temperature. On heating, the entropy makes a growing contribution to the free energy. We can make a reasonable presupposition about the form of this contribution. The distorted phase has a more compact structure than the high-symmetry phase, and this will result in slightly higher phonon frequencies and hence a lower entropy (Section 9.1.5). Thus the $-TS$ term will favour the high-symmetry structure over the distorted one when T is sufficiently large.

In principle exact functions can be obtained for the different terms in the free energy, and these can be related to microscopic interactions. Later in this chapter (Section 12.5) we will derive this link, but first we describe several different types of displacive phase transitions.

12.1.2 Various types of displacive phase transitions

Ferroelectric phase transitions

In Section 2.4 we outlined the main features of ferroelectric phase transitions, namely that the symmetry-breaking displacements of the atoms allowed the formation of a dielectric polarization of the crystal. We also noted that the phase transition is accompanied by a divergence of the dielectric susceptibility.

One of the important characteristics of ferroelectric phase transitions is that the dielectric polarization can be reversed by the application of an electric field. This is not surprising, since the field only has to reverse the directions of the relatively small static atomic displacements. How this reversal takes place is illustrated in Fig. 12.3. The field is first applied along the direction of the polarization. This increases the polarization by an amount determined by the susceptibility, $\Delta P = \chi E$. The field is then reduced to zero, so that the polarization returns to its equilibrium value P_0. The field is then increased in the opposite direction. The polarization is then reduced, instead of immediately switching direction. It is only when the reverse field reaches a certain strength, known as the **coercive field strength**, labelled E_c in Fig. 12.3, that the polarization suddenly switches directions. Then the polarization increases in magnitude along the opposite direction following the susceptibility. The reverse process then happens when the field is reduced and then turned on in the opposite direction. The resultant plot of polarization against electric field has the shape of a loop, which is called the **hysteresis loop**, as shown in Fig. 12.3. The same type of hysteresis loops are also found in ferromagnets, and it is the close analogy between the ferromagnetic and ferroelectric phase transitions that led to the naming of the latter.

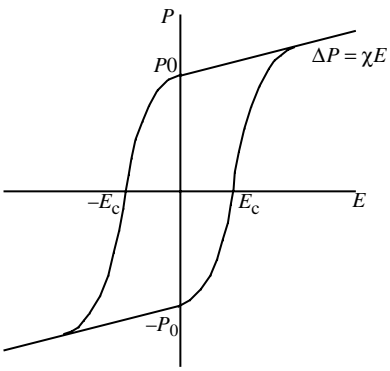


Fig. 12.3 Ferroelectic hysteresis loops.

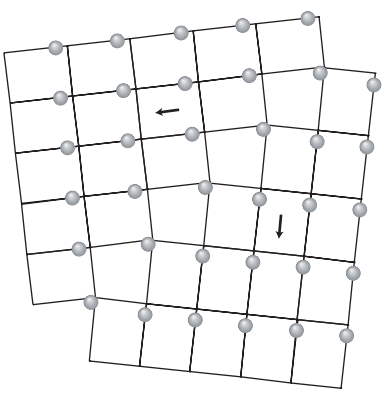


Fig. 12.4 Ferroelectic domain wall, showing motion of domain wall under application of an electric field.

In principle, the size of the coercive field could be related to the size of the potential energy barriers against movements of individual atoms, or of a free energy barrier. However, the most important factor concerns the mechanism by which the polarization can be reversed. When a material undergoes a ferroelectric phase transition, it can form small regions of the crystal in which the polarization points along different directions. In $PbTiO_3$, the polarization can point along any of the ⟨001⟩ directions of the parent cubic phase, which means that there are six possible polarization states of the ferroelectric field. In practice, all six orientations can be found in different places in one crystal. These regions are called **domains**, and the boundaries between the regions are called **domain walls**. A schematic illustration of a domain wall is shown in Fig. 12.4. Switching the polarization direction of any domain will take place at the domain wall, since the atoms bounding the domain walls do not feel the effects of so many aligned neighbours. This is illustrated in Fig. 12.4. The effect of switching the polarization of the unit cells close to the domain wall has the appearance of the domain wall moving through the material. It is the force required to move the domain wall that determines the size of the coercive field, and clearly this may be affected by the presence of defects that can act to pin domain walls. Domains and domain walls are discussed in Appendix A.

Antiferroic (unit-cell combining) displacive phase transitions

In a ferromagnet, the magnetic moments on the atoms in each unit cell align in the same direction. However, there is a large class of magnetically-ordered

materials in which the magnetic moments of the atoms in adjacent unit cells point along opposite directions – a simple example of this ordering is shown in Fig. 12.5 (examples are oxides with the NaCl structure, such as NiO and MnO). These are called **antiferromagnetic** materials. There are many displacive phase transitions that are analogous to this, and these are called **antiferroelectric** phase transitions, if there are similarities to the displacements found in ferroelectric phase transitions, or just **antiferroic** displacive phase transitions in more general cases. The point is that there are many phase transitions in which displacements of the atoms in one unit cell are balanced by opposite displacements of the atoms in neighbouring unit cells. Thus the unit cell of the low-symmetry phase will be a combination of two, four or eight unit cells of the high-symmetry phase.

A simple example is the rotational phase transition in $SrTiO_3$, which was discussed in Chapter 2 and shown in Fig. 12.6. The oxygen atoms are displaced in a way that corresponds to the rotations of the TiO_6 octahedra. The resultant change in symmetry does not allow for the development of a dielectric polarization. Most displacive phase transitions are of this type, including quartz and cristobalite shown in Fig. 12.1.

Antiferroic phases will also show formation of domains, and examples are given in Appendix A. The number of types of domains that can be formed will depend on symmetry.

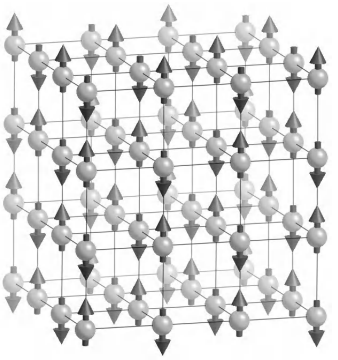


Fig. 12.5 Antiferromagnetic ordering in a simple cubic lattice.

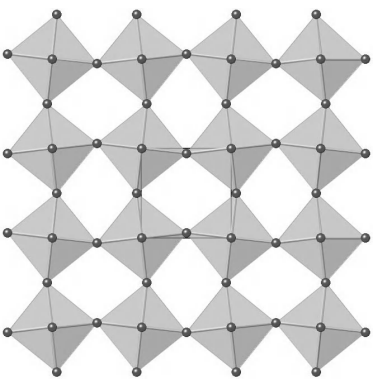


Fig. 12.6 Octahedra rotation phase transition in perovskite showing opposite rotations of oactahedra leading to a doubling of the unit cell in each of the two directions in the plane of the diagram.

Ferroelastic phase transitions

Most displacive phase transitions will involve a change in the size and shape of the unit cell. For example, as a result of the ferroelectric phase transition in the perovskite $PbTiO_3$, the c axis will expand relative to the a and b axes as the lattice type changes from cubic to tetragonal. This strain deformation of the lattice is called **spontaneous strain**. In the case of the perovskite ferroelectric, the size and sign of the spontaneous strain are independent of whether the polarization points along $\langle 001 \rangle$ or in the opposite direction, $\langle 00\overline{1} \rangle$. This is because the structures of the two domains are simply mirror images of each other, and thus the symmetry requires the spontaneous strains to be identical in both cases. This is achieved if the relationship between the spontaneous strain and polarization does not involve the sign of the polarization, which means that, to lowest order, the spontaneous strain $\epsilon_s \propto P^2$. When generalized to any phase transition, we often find that ϵ_s is proportional to the square of the atomic displacements.

However, there is a class of phase transitions in which the spontaneous strain can change sign. These are called **ferroelastic phase transitions**. In a manner similar to ferromagnetic and ferroelectric phase transitions, the strain can be reversed by application of a suitable external stress. The analogy with the ferromagnetic and ferroelectric phase transitions also extends to the point that the quantity analogous to the magnetic or dielectric susceptibility, namely the elastic compliance (the inverse of the elastic constant), will diverge at a ferroelastic phase transition. This is seen as a softening of one or more of the elastic constants of the material.

One example of a ferroelastic phase transition is found in Na_2CO_3, which is shown in Fig. 12.7. In this case the hexagonal high-temperature phase transforms to a monoclinic phase by a shear in the **a**–**b** plane. This shear is accompanied by the relative shifts of the columns of NaO_6 octahedra, which

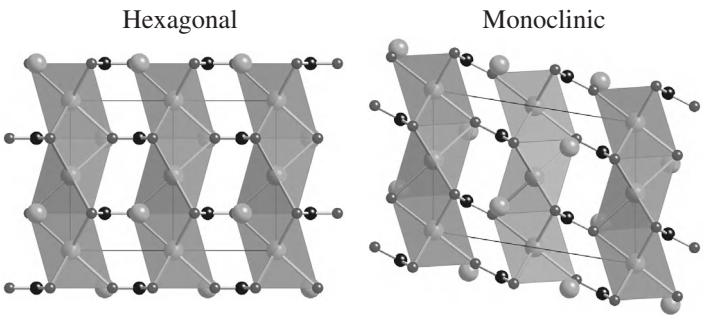


Fig. 12.7 Crystal structures of the two phases of Na_2CO_3, showing how the low-temperature monoclinic phase can be obtained by a shear strain of the high-temperature hexagonal phase.

are hinged by the rotating planar carbonate molecular ions. The shear strain is proportional to the size of the rotation, and can be reversed by application of the corresponding shear stress.

The reader may permit the author to interject with some additional detail on the ferroelastic phase transition Na_2CO_3. The shear instability involves the value of the c_{44} elastic constant in the hexagonal phase vanishing at the phase transition. Because of the hexagonal symmetry, $c_{55} = c_{44}$, and the shear elastic instability occurs for any direction in the **a**–**b** plane. It has been shown that this type of instability gives rise to a divergence of the atomic mean-square displacements, and these in turn cause the Bragg peaks with $l \neq 0$ to vanish at the phase transition. At this point, there is only true long-range order in the [001] direction, and the material has order that resembles a two-dimensional liquid in all directions normal to [001]. More details are given in Harris *et al.*, *J. Phys.: Cond. Matter* **8**, 7073, 1996, and references cited therein.

12.2 Quantitative description of displacive phase transitions: the concept of the order parameter

12.2.1 The general definition of the order parameter

The descriptions of the displacive phase transitions given above have been in terms of symmetry-breaking atomic displacements that give rise to a change in symmetry. For many systems there will be several atoms moving as a result of the phase transition. For example, in ferroelectric perovskites such as $PbTiO_3$, the phase transition involves displacements of both cations and anions. When several atoms are involved in the symmetry-breaking deformations associated with the phase transition, it is usually possible to identify an overall pattern of relative atomic displacements, and it is the amplitude of the overall pattern of displacements that changes with temperature. In principle one might want to generate a thermodynamic theory that takes account of all symmetry-breaking displacements individually, but this could be rather complicated. Instead it is common to treat the pattern of atomic displacements as a fixed pattern, and the overall amplitude is the quantity that would be used in any theoretical treatment. This quantity is what is called the **order parameter**. This definition is analogous

to the concept of the normal mode coordinate in the theory of lattice dynamics. The pattern of displacements associated with any normal mode of the crystal is defined by the pattern of force constants, and is independent of the amplitude of the mode. The normal mode coordinate gives the overall amplitude of the displacement at any time, and this is the quantity that contains the dependence on temperature. Following the analogy further, the low-symmetry phase can be described as arising from a freezing of one of the normal modes of the high-symmetry phase into the structure as a static deformation, and in this case the order parameter is then the static equivalent of the normal mode coordinate. This suggests that the lattice dynamics of a crystal are implicated in the origin of a phase transitions, a point that we will develop later in this chapter (Section 12.4).

The power behind the use of the order parameter is that with a single parameter (or a small number of parameters in more complex cases) it becomes possible to develop a theory of the phase transition to any level of accuracy one wishes to aim at, and within any theoretical framework involving the order parameter it may also be possible to encompass a wide range of phenomena. Moreover, aside from the symmetry of the deformation, the exact pattern of atomic displacements may often be of little relevance in any subsequent theory. Continuing the analogy with the normal mode coordinate in the theory of lattice dynamics, the mode eigenvectors do not enter into the thermodynamics of the lattice vibrations. The only quantity that is of importance is the mode frequency, through which we obtain the normal mode coordinate and thence the thermodynamic functions. The point is that it should be possible to obtain an expansion of the thermodynamic free energy in terms of the order parameter. From the free energy it is possible to calculate any other function, and hence it ought to be possible to construct a theoretical framework that has a simple dependence on order parameter and which encompasses all experimental data. This would give a broad unifying description, and could be a predictive tool.

Before we consider some specific examples, it is worth developing a graphical illustration of the central point. In Fig. 12.8 we show three atoms in each unit cell that are involved in the symmetry-breaking distortion associated with a displacive phase transition. One of the atoms lies on a mirror plane, and as a

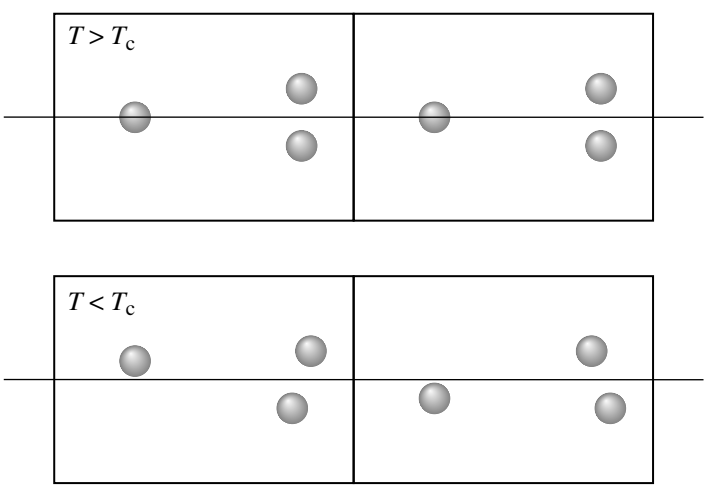


Fig. 12.8 Illustration of symmetry-breaking displacements. The mirror plane is converted to a glide plane as the lattice repeat doubles.

result of the phase transition the atoms in alternate unit cells move in opposite directions normal to the mirror plane. The other atoms are also displaced in order to break the mirror symmetry. There are two effects on the symmetry as a result of the phase transition. The first is the loss of the mirror plane, and the second is the doubling of the repeat distance along the direction of the mirror plane. These two symmetries are replaced by the glide symmetry. The order parameter describes how far the atoms move to break the symmetry – in effect, the order parameter gives a measure of how much the symmetry has been broken by. Of course, a change in symmetry is step-wise: the symmetry is either present or is not present. However, the changes that accompany the loss of symmetry will be gradual to some extent, and the order parameter gives a measure of these gradual changes. This is very similar to the discussion of symmetry-breaking displacements associated with normal modes of vibrations given in Sections 3.4 and 8.5.4.

We noted earlier that some phase transitions may actually require the use of more than one order parameter. One example is the ferroelectric phase transition in the perovskites, in which the polarization could point along any of the three axes. We denote these three components as P_x, P_y, and P_z. For the phase transition in $PbTiO_3$, $P_y = P_z = 0$, and we only need to consider the variation of P_z. For $BaTiO_3$, the first phase transition on cooling is similar to that in $PbTiO_3$. The second phase transition then involves forming $P_y \neq 0$. Finally, all three components become active with $P_x = P_y = P_z \neq 0$. Similarly, the rotational phase transitions in the perovskites can be defined by three rotation angles R_x, R_y, and R_z. For $SrTiO_3$, $R_y = R_z = 0$. However, there are some rotational phase transitions in which the behaviour is more complicated. In $LaAlO_3$, there are equal rotations about all three axes, i.e. $R_x = R_y = R_z$. But in other perovskites, such as $CaTiO_3$, there are unequal rotations about the three axes, $R_x \neq R_y \neq R_z \neq 0$. Complete descriptions of these phase transitions will require the use of all three order parameters.

12.2.2 Examples of order parameters for specific phase transitions

The concept of the order parameter is illustrated by several of the examples of phase transitions we have already encountered. In the octahedral tilting phase transition in $SrTiO_3$ the order parameter is simply the angle through which the TiO_6 octahedra rotate. These rotations involve displacements of oxygen atoms, and there is only one type of atomic displacement. The phase transitions in quartz and cristobalite involve rotations of SiO_4 tetrahedra as analogues to the octahedral-tilting phase transition in $SrTiO_3$. However, these rotations are accompanied by displacements of the tetrahedra, and there is a more complex pattern of atomic displacements. Nevertheless, the pattern of displacements can be described by a single order parameter in each case.

In none of these examples is there a macroscopic analogue of the order parameter. On the other hand, the ferroelectric phase transitions do have a clear macroscopic equivalent of the order parameter. We recall that in $PbTiO_3$ there are displacements of cations and anions. The anions lying along the z direction from the Ti cation are allowed to have different displacements to the oxygens in the same x–y plane as the cations. The electrical polarization that is generated

by these displacements is given as

$$P = N\left[q(\mathrm{Ti})\delta(\mathrm{Ti}) + q(\mathrm{Pb})\delta(\mathrm{Pb}) + \sum_{j=1}^{3} q(\mathrm{O}_j)\delta(\mathrm{O}_j)\right] \qquad (12.2)$$

12.2.3 Order parameters in other phase transitions

We have illustrated the concept of the order parameter through a number of different displacive phase transitions. However, the concept of the order parameter actually transcends specific details of the type of phase transition, and can be used to describe any phase transition. The common example is a magnetic phase transition. A magnetic system will contain ions with magnetic moments. What changes as a result of the phase transition is not the displacements of the magnetic moments, but their orientations. In the high-temperature paramagnetic phase the orientations of the individual magnetic moments will be random (although they may be constrained to a few particular orientations). Thus the average moment of the crystal obtained by summing over all moments is zero. Below the phase transition, the orientations of all atomic moments become progressively ordered. These two situations are illustrated in Fig. 12.9. If we write the individual moments as $\mathbf{m}_j$, the order parameter can be written as

$$\langle \mathbf{m} \rangle = \frac{1}{N} \sum_j \mathbf{m}_j \qquad (12.3)$$

In this case, the phase transition involves a real ordering of the orientations of the magnetic moment, and it is in this context that the name **order parameter** has its origin.

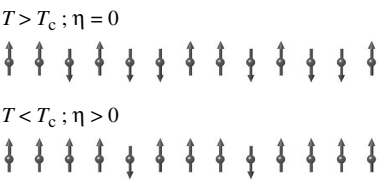


Fig. 12.9 Illustration of the calculation of the order parameter for the ordering of an array of spins.

Similar ordering processes can be seen in phase transitions involving orientations of molecules. Examples from earlier in the book include the ordering of the orientations of N_2 and C_{60} molecules. The same idea will apply to the ordering of orientations of long-chain molecules in some liquid crystals.

Another example of an ordering process occurs in alloys such as Cu_3Au. In the high-temperature phase, the Cu and Au atoms are distributed randomly across all sites in a ccp structure, but on cooling below a particular temperature the Cu and Au atoms begin to occupy certain preferred sites, until at low temperature there is complete ordering of the positions of the different atoms in a tetragonal unit cell. We use x_{Cu} to represent the occupancy of the low-temperature Cu sites by Cu atoms. At low temperature $x_{\mathrm{Cu}} = 1$, and in the high-temperature disordered phase $x_{\mathrm{Cu}} = 3/4$. We can therefore quantify the degree of order by the order parameter

$$4\left(x_{\mathrm{Cu}} - \frac{3}{4}\right)$$

This type of site-ordering phase transition is found in many system where there are cations that are able to occupy identical sites without too high a cost in energy.

12.2.4 Experimental measurements of order parameter

For ferroelectric phase transitions, the dielectric polarization is a macroscopic representation of the order parameter, and can be measured using macroscopic

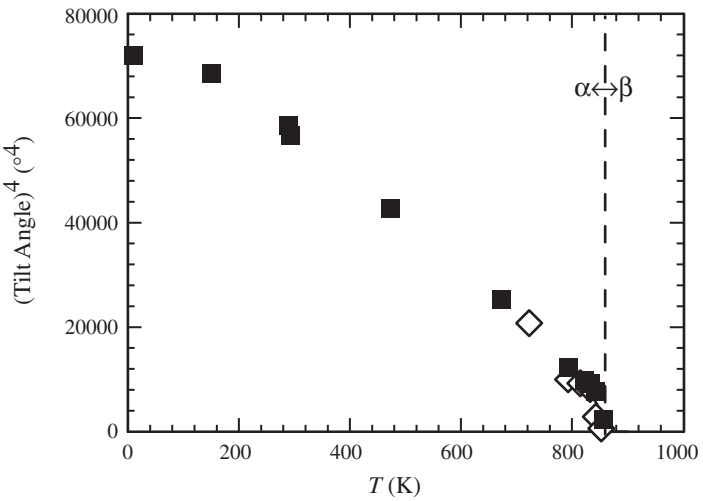


Fig. 12.10 Temperature dependence of the fourth power of the angle of rotation of SiO_4 tetrahedra associated with a displacive phase transition in quartz – the fourth power is plotted because the temperature dependence is almost linear except for the first-order discontinuity at the phase transition temperature.

techniques. However, in most cases there will be no macroscopic representation of the order parameter, and then it will be necessary to perform a direct measurement of the atomic displacements through measurements of the crystal structure, or else obtain information about the order parameter through indirect measurements.

Figure 12.10 shows the temperature dependence of the angle through which the SiO_4 tetrahedra rotate following the displacive phase transition in quartz (the rotation is shown in Fig. 12.1). The figure combines both X-ray and neutron diffraction data (Chapter 6).

Indirect measurements of the order parameter require that some measurable property of the material will change as a result of the phase transition, and the change can be related to the value of the order parameter. The value of an indirect measurement of the order parameter is that it can often be obtained with higher precision than through direct diffraction measurements. Examples of indirect measurements are intensities of Bragg peaks that vanish on heating into the high-symmetry phase (see Problem 12.12), changes in the lattice parameters that accompany the phase transitions, or changes in phonon frequencies measured by infrared or Raman spectroscopy (Chapter 10; Section 12.5.1). Often, as in measurements of changes in vibrational frequencies, the change will be independent of the sign of the order parameter, and hence will be proportional to the square of the order parameter.

In Fig. 12.11 we show measurements of the piezoelectric coefficients of quartz. One of the coefficients becomes zero in the high-temperature phase, and by symmetry its variation with temperature below the phase transition mirrors the temperature dependence of the order parameter.

12.2.5 First- and second-order phase transitions

We have met examples of the temperature dependence of the order parameter and quantities that depend on the order parameter. In some cases the order parameter falls continuously to zero at the transition temperature T_c, and in

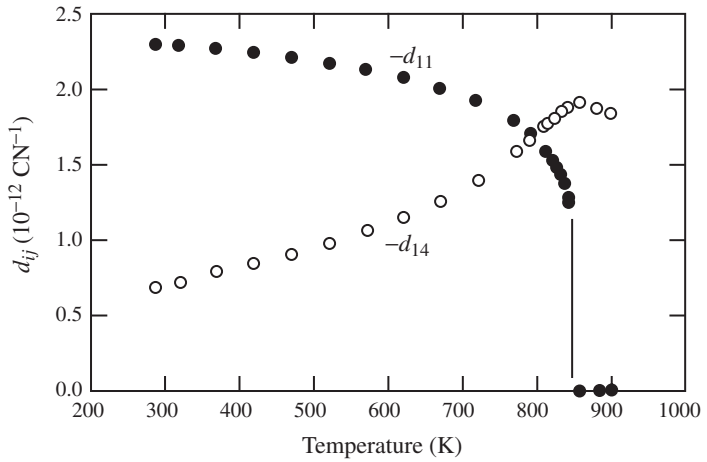


Fig. 12.11 Temperature dependence of the piezoelectric coefficients of quartz.

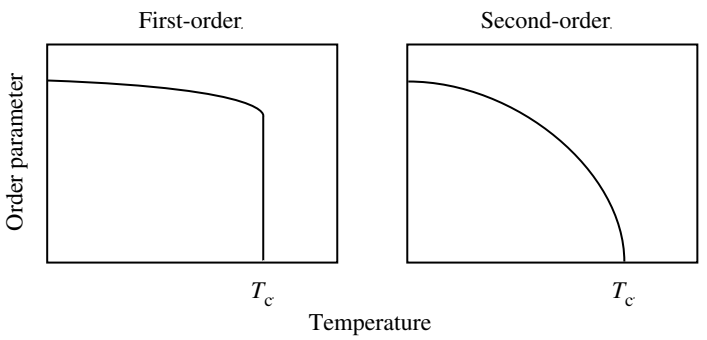


Fig. 12.12 Schematic representation of temperature dependence of order parameters at first and second order phase transitions.

other cases there is a discontinuous change at the transition temperature. These differences define two types of phase transitions: the case where there is a discontinuous change is called a **first-order** phase transition, and the case where the change is continuous is called a **second-order** phase transition. The names come from an old characterization, where the order of a phase transition was associated with the lowest-order derivative of the free energy with temperature that changes discontinuously at the phase transition. Thus at a first-order phase transition, the entropy $S = -\partial F/\partial T$ changes discontinuously at the transition temperature, and at a second-order phase transition the heat capacity $c = T\partial S/\partial T = -T\partial^2 F/\partial T^2$ changes discontinuously. In principle one might imagine the possibility of a third-order phase transition, but this concept is not useful since it represents a special case that can be understood within the concept of second-order phase transitions. The ideas can be generalized for pressure-induced displacive phase transitions. The two types of phase transitions are illustrated in Fig. 12.12.

In many second-order phase transitions, the order parameter, which we label as η (this is often given the symbol Q, but this is liable to get confused with the symbol for normal mode coordinate), varies as

$$\eta \propto |T_c - T|^{1/2} \tag{12.4}$$

In this case, any property that varies as η^2 will fall linearly to zero at the phase transition. In first-order phase transitions, the analytical form of η is more complex (to be discussed below; Section 12.3.4). Some phase transitions seem to be almost exactly part-way between second and first order, and in these cases the order parameter varies as

$$\eta \propto |T_c - T|^{1/4} \tag{12.5}$$

Such cases are called **tricritical phase transitions**, because in some cases this behaviour is found close to a tricritical point in a phase diagram. Quartz is an example of a phase transition that is strictly first order, but the discontinuity is so small that for a wide range of temperatures the order parameter varies in a way that is close to that of a tricritical phase transition. The temperature dependence of the SiO_4 rotation angle in quartz in Fig. 12.10 is plotted in a way that highlights this point.

12.3 Landau theory of displacive phase transitions

12.3.1 Qualitative behaviour of the free energy

Displacive phase transitions lend themselves to a reasonably accurate general free energy analysis. In this section we will develop an empirical but quantitative thermodynamic description of phase transitions within the context of what is known as **Landau theory**. Later in this chapter (Section 12.5) we will show that the basic ideas can be justified from a lattice dynamics perspective.

As we have constantly stressed, the starting point has to be the free energy, since this contains information about the equilibrium structure. In this chapter we will use the Helmholtz free energy. Consider a material at a temperature above that of a second-order phase transition. Its free energy as a function of η will have a single minimum at $\eta = 0$, as shown in Fig. 12.13.

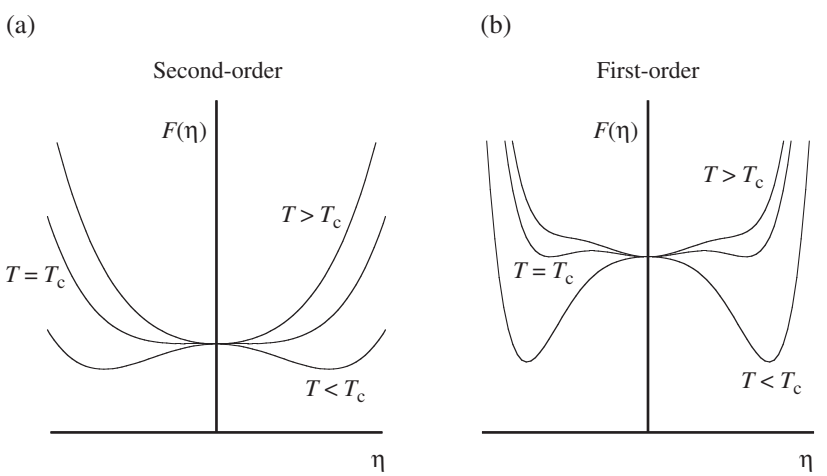


Fig. 12.13 Variation of free energy curves with order parameter for second-order and first-order phase transitions and temperatures below, at and above the transition temperature.

At temperatures far below the transition temperature, the free energy curve, $F(\eta)$, will have a global minimum at non-zero values of η. In a second-order phase transition, the position of this minimum gradually moves closer to $\eta = 0$ on heating towards T_c, as shown in Fig. 12.13. For temperatures above T_c there is only a single minimum at $\eta = 0$.

At a first-order phase transition, the picture is more complicated. At the transition temperature, there must be minima at both $\eta = 0$ and at non-zero values of η. The overall variation of the free energy curves for a first-order phase transition is shown in Fig. 12.13.

12.3.2 Expansion of the free energy function for a second-order phase transition

At high temperatures, the free energy curves have a single minimum, and we can express the free energy function as

$$F(\eta) = F_0 + \frac{1}{2}A\eta^2 + \cdots \tag{12.6}$$

where A is a positive coefficient. F_0 is the part of the free energy that does not change through the phase transition, and we will not always include this in the subsequent equations. At the transition temperature, the free energy function at $\eta = 0$ switches from being a minimum to a maximum. This corresponds to the coefficient A changing sign at $T = T_c$. Since the process is gradual, one simple representation of the temperature dependence of A is

$$A(T) = a(T - T_c) \tag{12.7}$$

Since the free energy function curves downwards around $\eta = 0$ where $T < T_c$, the function must contain an additional positive term with a higher-order power of η in order that to $F(\eta)$ will curve upwards again at larger values of η. The lowest-order term that we can use is the quartic term, so we can now expand the free energy function as

$$F(\eta) = F_0 + \frac{1}{2}a(T - T_c)\eta^2 + \frac{1}{4}b\eta^4 \tag{12.8}$$

The development of this form of the free energy function has been intuitive, but it is worth rehearsing the principles behind the intuition. We have used the constraints on the form of the free energy to ensure that it gives the correct minima across a range of temperatures. We have also used symmetry constraints, namely that $F(\eta) = F(-\eta)$, to demonstrate that the free energy has no odd powers of η. We have also used the property that F varies continuously through the phase transition, so that all parameters must be continuous functions of temperature.

12.3.3 Calculation of properties for a second-order phase transition

The principle that the equilibrium state is that for which the free energy is a minimum gives us a method for calculating the temperature dependence of the

properties. The equilibrium value of the order parameter for any temperature is that given by the solution of the equations

$$\frac{\partial F}{\partial \eta} = 0 \qquad \frac{\partial^2 F}{\partial \eta^2} > 0 \tag{12.9}$$

The first differential gives

$$a(T - T_c)\eta + b\eta^3 = 0 \tag{12.10}$$

There are three solutions in principle. One is $\eta = 0$. When $T > T_c$, this solution corresponds to the minimum of the free energy, and there are no other real solutions. When $T < T_c$, the solution $\eta = 0$ corresponds to a maximum point of the free energy, and the other two solutions now correspond to the minima in the free energy function:

$$\eta = \pm\sqrt{\frac{a}{b}}\,|T_c - T|^{1/2} \tag{12.11}$$

This describes the temperature dependence of the order parameter found in many phase transitions.

We now unpack the free energy to obtain its separate contributions:

$$F - F_0 = \Delta U - T\Delta S$$

$$\Rightarrow \quad \Delta U = -\frac{1}{2}aT_c\eta^2 + \frac{1}{4}\eta^4 \tag{12.12}$$

$$\Delta S = -\frac{1}{2}a\eta^2 \tag{12.13}$$

Thus the internal energy, which is assumed to be independent of temperature, is a double-well function. The entropy decreases as a quadratic function of the order parameter, which we will understand as arising from the increase in phonon frequencies as the structure is distorted.

The contribution to the heat capacity from the phase transition is obtained as

$$\Delta c(T < T_c) = T\frac{\partial S}{\partial T} = -\frac{1}{2}aT\frac{\partial \eta^2}{\partial T} = \frac{a^2 T}{2b} \tag{12.14}$$

$$\Delta c(T > T_c) = 0 \tag{12.15}$$

Thus the total heat capacity shows a small step change at the second-order phase transition temperature.

The susceptibility can be obtained from the second-order differential of the free energy with respect to the order parameter (Problem 12.1):

$$\chi^{-1} = \frac{\partial^2 F}{\partial \eta^2} = a(T - T_c) + b\eta^2 \tag{12.16}$$

When $T > T_c$, $\eta = 0$ and

$$\chi^{-1} = a(T - T_c) \tag{12.17}$$

On the other hand, when $T < T_c$, substitution of the equation for $\eta(T)$ (eqn 12.11) yields

$$\chi^{-1} = 2a(T_c - T) \tag{12.18}$$

Clearly $\chi^{-1} \to 0$ as $T \to T_c$, which corresponds to a divergence of the susceptibility at the phase transition. In a ferroelectric phase transition, this accounts for the divergence of the dielectric constant as seen in Figs 1.8 and 2.31.

12.3.4 First-order phase transitions

At a second-order displacive phase transition, the free energy function at $\eta = 0$ changes from being a minimum to a maximum at the phase transition. However, at a first-order phase transition, a minimum at $\eta = 0$ will coexist with a second minimum at a non-zero value of η. We will restrict our analysis to the case where $F(\eta) = F(-\eta)$, so there will be three coexisting minima at the phase transition, but there are cases where this symmetry is not required. The symmetric case is shown in Fig. 12.13. This is achieved by a small modification of the Landau free energy function for the second-order phase transition:

$$F(\eta) = \frac{1}{2}a(T - T_0)\eta^2 - \frac{1}{4}b\eta^4 + \frac{1}{6}c\eta^6 \tag{12.19}$$

where we use T_0 rather than T_c because T_0 is not to be identified with the transition temperature. The main difference from the free energy function for the second-order phase transition is that the fourth-order term is now negative (b is a positive number), and in order to ensure that the free energy function rises to large positive values for large values of $|\eta|$ it is necessary to have the positive sixth-order term.

The temperature dependence of the order parameter can be obtained by minimization of the free energy, yielding (see Problem 12.2)

$$\eta^2 = \frac{b + \sqrt{b^2 - 4ca(T - T_0)}}{2c} \tag{12.20}$$

The transition temperature is given as the temperature at which values of the free energy at each minimum are zero (Problem 12.2):

$$T_c = T_0 + \frac{3b^2}{16ac} \tag{12.21}$$

At this temperature the order parameter jumps from zero to $\pm(3b/4c)^{1/2}$. The change in internal energy at this temperature is

$$\Delta U = -\frac{3abT_c}{8c} = \frac{aT_c\eta^2}{2} \tag{12.22}$$

This is the latent heat for the phase transition.

There may be several reasons why a phase transition is first order. One is associated with the coupling to the spontaneous strain. We could expand the free energy as

$$F(\eta) = \frac{1}{2}a(T - T_0)\eta^2 + \frac{1}{4}b\eta^4 + \frac{1}{2}\lambda\epsilon_s\eta^2 + \frac{1}{2}C_{el}\epsilon_s^2 + \cdots \tag{12.23}$$

where ϵ_s is a spontaneous strain produced by the phase transition, C_{el} is the corresponding elastic constant, and λ is another constant. The term of the form $\epsilon_s\eta^2$ has the appropriate symmetry for a phase transition that is not ferroelastic, which is seen by solving for the equilibrium condition:

$$\frac{\partial F}{\partial \epsilon_s} = 0 \quad \Rightarrow \quad \epsilon_s = -\frac{\lambda\eta^2}{2C_{el}} \tag{12.24}$$

(see Section 12.1.2.) Substituting into eqn 12.23 gives

$$F(\eta) = \frac{1}{2}a(T - T_0)\eta^2 + \frac{1}{4}\left(b - \frac{\lambda^2}{2C_{\text{el}}}\right)\eta^4 + \cdots \tag{12.25}$$

The important point is that the coupling to strain has reduced the coefficient of the fourth-order term, and if the coupling were sufficiently large this coefficient could become negative. In this case, the phase transition would become first-order (and we would then need to incorporate the sixth-order term in the expansion of F).

12.3.5 The range of validity of Landau theory

Landau theory provides a very powerful description of the thermodynamics of phase transitions. This power is derived from the fact that the free energy can be written in terms of the order parameter, which in turn allows the free energy to written in terms that encompass other quantities (such as elastic strain).

Although Landau theory is rooted in thermodynamics, the Landau free energy function $F(\eta)$ is not itself a true thermodynamic function. This is best indicated by the low-temperature behaviour. For a second-order phase transition the temperature dependence of η has the same form for all temperatures. More precisely, the derivative of the order parameter with respect to temperature is obtained from

$$2\eta\frac{\partial\eta}{\partial T} = -\frac{2a}{b} \tag{12.26}$$

In the limit $T \to 0$ this result contrasts sharply with the usual thermodynamic limiting behaviour $\partial\eta/\partial T|_{T\to 0} = 0$. This is obtained by starting from the basic definition of the internal energy

$$\mathrm{d}U = T\,\mathrm{d}S + H\,\mathrm{d}\eta \tag{12.27}$$

where H is a field conjugate to η, which need not be possible to produce in the laboratory (not to be confused here with the enthalpy). The differential form of the internal energy gives

$$T = \left(\frac{\mathrm{d}U}{\mathrm{d}S}\right)_\eta; \qquad H = \left(\frac{\mathrm{d}U}{\mathrm{d}\eta}\right)_S \tag{12.28}$$

These then yield

$$\frac{\mathrm{d}T}{\mathrm{d}\eta} = \frac{\mathrm{d}^2U}{\mathrm{d}S\,\mathrm{d}\eta} = \frac{\mathrm{d}H}{\mathrm{d}S} \tag{12.29}$$

This pair of equations implies that

$$\frac{\mathrm{d}\eta}{\mathrm{d}T} = \frac{\mathrm{d}S}{\mathrm{d}H} \tag{12.30}$$

Since $S = 0$ for all values of the conjugate field H at $T = 0$, it follows that, at $T = 0$,

$$\left.\frac{\mathrm{d}\eta}{\mathrm{d}T}\right|_{T=0} = 0 \tag{12.31}$$

This contrasts with eqn (12.26), which gives a non-zero gradient for $\eta(T)$ at $T = 0$. The point is that the formulation of Landau theory is clearly not appropriate for low temperature. We will understand this point better in Section 12.5.

12.4 Soft mode theory of displacive phase transitions

12.4.1 Basic idea of the soft mode

If we perform a lattice dynamics calculation on a high-temperature phase of a material that undergoes a displacive phase transition, at least one of the harmonic values of ω^2 will be calculated as negative (such a calculation corresponds to $T = 0$). This is because the crystal is at a potential energy maximum with respect to the distortions that accompany the phase transition. These distortions will correspond to the atomic displacements associated with this particular phonon. The negative value of ω^2 is a signature of a displacive instability. Such a calculation would correspond to a temperature of zero, leading to the conclusion that the stable phase at low temperatures is the lower-symmetry distorted phase. However, at high temperatures the high-symmetry phase becomes stable, and we therefore expect that this phonon will now have a positive value of ω^2. This line of reasoning suggests that ω^2 has an intrinsic temperature dependence of the form sketched in Fig. 12.14, increasing in value (i.e. becoming less negative) on heating. Eventually the value of ω^2 will reach zero, and for temperatures above this point the high-symmetry phase will be stable. The temperature at which ω^2 reaches zero is the transition temperature for a second-order phase transition.

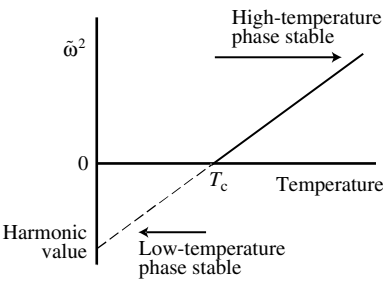


Fig. 12.14 Schematic representation of the temperature dependence of the square of the frequency of a soft mode. Below the temperature T_c the frequency is imaginary and hence unstable. The frequency at $T = 0\,\mathrm{K}$ is the harmonic value.

It is instructive to turn the picture round. We then have a high-temperature phase that has one phonon whose frequency falls on cooling, eventually reaching zero, at which point the material undergoes a displacive phase transition to a lower-symmetry structure. The phonon frequency is said to **soften**, because its value is determined by restoring forces that appear to soften on cooling. This phonon is called the **soft mode**. The atomic motions associated with the soft mode correspond closely to the way the structure distorts through the phase transition – one can think of the lower-symmetry distorted structure in terms of the symmetric structure modulated by a frozen-in phonon distortion (recall the discussion in Section 12.1.2). For example, in the ferroelectric perovskites the soft mode involves the cations moving in opposite directions to the anions.

From the theory of the anharmonic phonon developed in Chapter 11 we can gain an understanding of the origin of the temperature dependence of the soft mode. Specifically we recall that the intrinsic temperature dependence of a phonon frequency due to direct anharmonic interactions (eqn 11.30) is given by

$$\omega^2 = \omega_0^2 + \frac{k_B T}{2} \sum_{\mathbf{k}',\nu'} \alpha^{(4)}_{\mathbf{k}',\nu'} / \omega^2(\mathbf{k}', \nu') \qquad (12.32)$$

ω_0^2 is the negative harmonic value that we would calculate in the lattice dynamics calculation, and $\alpha^{(4)}_{\mathbf{k}',\nu'}$ is the coefficient of the fourth-order anharmonic

interaction that couples the soft mode to a phonon of wave vector $\mathbf{k}'$, mode label ν', and angular frequency $\omega(\mathbf{k}', \nu')$. Clearly, if sufficient values of $\alpha^{(4)}_{\mathbf{k}',\nu'}$ are positive, the effect of increasing temperature is to increase the value of ω^2 until it becomes positive. The phase transition occurs at the temperature at which $\omega^2 = 0$:

$$k_B T_c = \frac{-2\omega_0^2}{\sum_{\mathbf{k}',\nu'} \alpha^{(4)}_{\mathbf{k}',\nu'}/\omega^2(\mathbf{k}', \nu')} \tag{12.33}$$

If we now make use of the Einstein model, we can define an average angular frequency $\langle\omega\rangle$ and an average coupling coefficient $\langle\alpha^{(4)}\rangle$ (the averages can be weighted in order to ensure consistency), and therefore write

$$RT_c = \frac{-2\omega_0^2\langle\omega\rangle^2}{3\langle\alpha^{(4)}\rangle} \tag{12.34}$$

The temperature dependence of the soft mode is then given as

$$\omega^2 = \frac{|\omega_0^2|}{T_c}(T - T_c) \tag{12.35}$$

12.4.2 Ferroelectric soft modes

The soft mode theory of phase transitions was originally developed to explain the origins and mechanisms of ferroelectric phase transitions. In fact, the idea of the soft mode for ferroelectric phase transitions is implicit in the Lyddane–Sachs–Teller relation introduced in Section 8.5.6 to explain why the longitudinal (LO) and transverse (TO) optic modes at $\mathbf{k} = 0$ in ionic materials are not equivalent. We recall the main result (eqn 8.61):

$$\frac{\epsilon(\omega = 0)}{\epsilon(\omega = \infty)} = \frac{\omega_{LO}^2}{\omega_{TO}^2} \tag{12.36}$$

The point is that since the static dielectric constant, $\epsilon(\omega = 0)$, diverges at the ferroelectric phase transition, the LST relation implies that $\omega_{TO}^2 \to 0$ at the

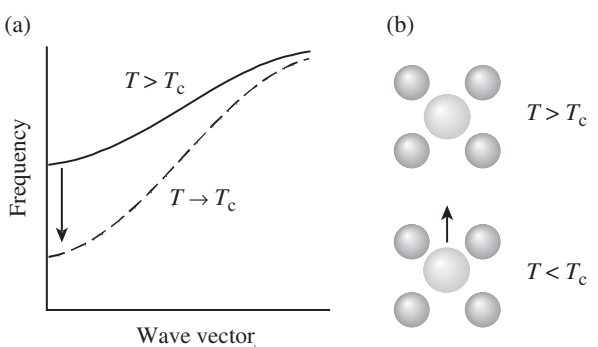


Fig. 12.15 Schematic representations of ferroelectric soft mode behaviour: (a) behaviour of the phonon dispersion curves; (b) atomic displacements.

same time. The softening of the transverse optic phonon branch is illustrated in Fig. 12.15, together with a schematic illustration of the atomic displacements associated with the soft mode.

The temperature dependence of the frequency of the soft mode associated with the ferroelectric phase transition in $PbTiO_3$ is shown in Fig. 12.16. The frequencies were measured using both inelastic neutron scattering (high-temperature phase) and Raman spectroscopy (low-temperature phase). At temperatures below the transition temperature, the soft mode hardens again. The two frequency values are for oscillations of the atoms in directions parallel to and normal to the direction of the dielectric polarization. The phase transition in $PbTiO_3$ is first order, so the soft mode frequency does not actually reach zero.

There are also phase transitions that have soft modes with zero wave vector but which are not ferroelectric. One example is quartz, which undergoes a hexagonal–trigonal phase transition at 846 K. Although both the high and low temperature phases are non-centrosymmetric, there is no accompanying change in the dielectric polarization or a divergence of the dielectric constant. The soft phonon mode involves the rotations of the SiO_4 tetrahedra shown in Fig. 12.1.

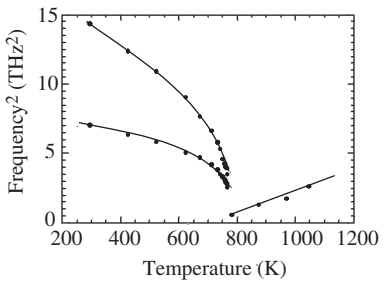


Fig. 12.16 Temperature dependence of the ferroelectric soft mode in $PbTiO_3$. Data in the low-temperature phase are taken from the Raman scattering study of Burns and Scott (*Phys. Rev. Lett.* **25**, 167, 1970), and data in the high-temperature phase are taken from the inelastic neutron scattering study of Shirane *et al.* (*Phys. Rev.* B **2**, 155, 1970).

12.4.3 Zone boundary (antiferroelectric) phase transitions

A number of crystals undergo phase transitions which involve soft modes with wave vectors at Brillouin zone boundaries. In these cases the soft phonons can be either acoustic or optic modes. Because of mixing of eigenvectors, it is often the case that the distinction between the two types of mode at zone boundary wave vectors is not clear-cut (as pointed out in Chapter 8). The different types of soft mode, and the atomic displacements, are shown schematically in Fig. 12.17. One of the results of a zone boundary soft mode phase transition is that the unit cell of the low-temperature phase is doubled in one or more directions. In some cases neighbouring unit cells of the high-temperature form develop dipole moments, but as these are in opposite directions the unit cell at low temperature has no net moment.

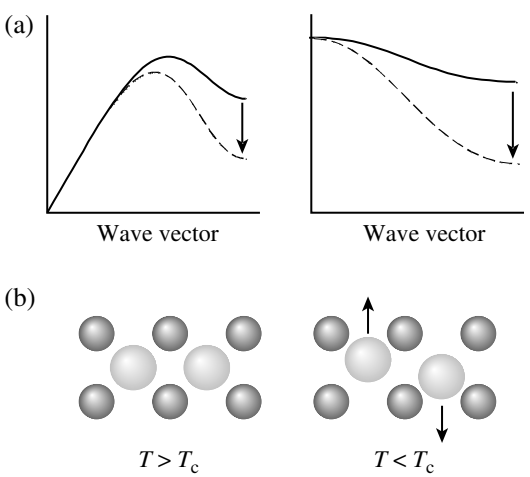


Fig. 12.17 Schematic behaviour of (a) zone boundary acoustic and optic soft modes; (b) atomic displacements showing doubling of the unit cell and cancelling induced dipole moments.

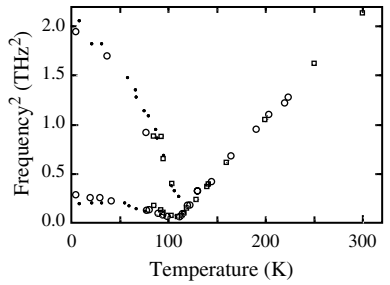


Fig. 12.18 The temperature dependence of the soft mode in $SrTiO_3$, as measured by a combination of inelastic neutron scattering (open circles and squares, data from Sirane and Yamada, *Phys. Rev.* **177**, 858, 1969, and Cowley *et al.*, *Sol. State Comm.* **7**, 181, 1969, respectively) and Raman spectroscopy (closed circles, data from Fleury *et al.*, *Phys. Rev. Lett.*, **21**, 16, 1968).

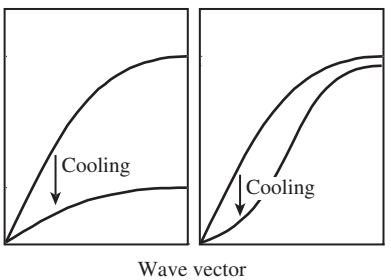


Fig. 12.19 Schematic representation of the softening of the acoustic mode. The left-hand figure shows the case where the acoustic mode softens across the whole branch. The right-hand figure shows the case where the softening is restricted to small wave vectors. In both cases, the softening leads to a reduction in the slope of the acoustic mode dispersion curve in the limit $k \rightarrow 0$.

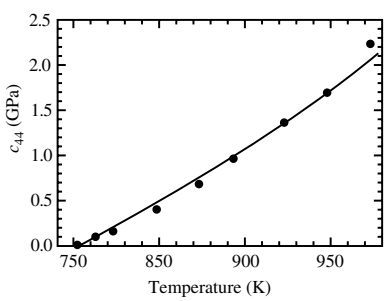


Fig. 12.20 Temperature dependence of the soft c_{44} elastic constant associated with the ferroelastic phase transition in Na_2CO_3 measured by inelastic neutron scattering. The data have been adapted from Harris *et al.*, *Phys. Rev. Lett.*, **79**, 4846, 1997.

Undoubtedly the best example of a zone boundary phase transition is the cubic–tetragonal transition in the perovskite $SrTiO_3$ (transition temperature 110 K). The soft mode involves the rotations of the interconnected TiO_6 octahedra about [001], as shown in Fig. 12.6, with neighbouring octahedra in the (001) plane rotating in opposite directions. This distortion leads to a doubling of the repeat distance in all three directions, although this gives a face-centred tetragonal unit cell that can be reduced to a body-centred tetragonal unit cell by a 45° rotation of the axes about [001]. The soft mode has wave vector $[1/2, 1/2, 1/2]$. The soft mode in $SrTiO_3$ has been measured by both neutron scattering and Raman spectroscopy, and its temperature dependence is shown in Fig. 12.18.

12.4.4 Ferroelastic phase transitions

The soft modes considered above have been optic modes, or of optic character. Acoustic modes at $k = 0$ can also soften, but here the basic picture will be somewhat different since at $k = 0$ the frequency is already zero. In this case, the softening involves the slope of the acoustic mode falling to zero, and this corresponds to a softening of the elastic constant. Different ways in which this can happen are shown in Fig. 12.19. The phase transitions caused by a softening of an acoustic mode are the ferroelastic phase transitions discussed earlier in this chapter. The gradient of the soft acoustic mode as $k \rightarrow 0$ gives the velocity of sound, the square of which gives one of the elastic constants. The temperature dependence of the soft c_{44} elastic constant in ferroelastic Na_2CO_3 (see Section 12.1.2) has been determined from inelastic neutron scattering measurements of the acoustic modes, and the results are shown in Fig. 12.20.

12.4.5 Incommensurate phase transitions

Most soft-mode phase transitions involve modes with wave vectors at $k = 0$ or at one of the Brillouin zone boundaries. However, it is possible for the instability to occur at a wave vector that is at a more general position in the Brillouin zone. A schematic representation of the modulation of a crystal due to a frozen-in normal mode with a general wave vector is given in Fig. 12.21.

This type of phase transition is the **incommensurate phase transition** introduced in Chapter 2, and is so called because the wavelength of the modulation associated with the soft mode is incommensurate with the underlying crystal repeat distance (in metals the wave vector of the incommensurate soft mode will be associated with wave vector of the Fermi surface, Section 5.4.7). One example of an incommensurate phase transition is quartz: the famous α–β phase transition actually has an incommensurate phase existing over a small temperature range between the two better-known phases. The mechanism for this involves a softening of an optic mode and an anti-crossing interaction with an acoustic mode (anti-crossing was described in Section 8.5.7). The effects of the anti-crossing increase with k, since the optic and acoustic modes have different symmetry at $k = 0$, so the result is that the frequency of the acoustic mode is driven to zero at a non-zero value of k. This is illustrated in Fig. 12.22.

12.5 Lattice dynamical theory of the low-temperature phase

12.5.1 Lattice dynamical theories

The theory of the soft mode developed so far is strictly applicable to the dynamics of the high-temperature phase. From the theory it is possible to understand how the anharmonic interactions stabilize the soft mode, and the theory has given an equation for the transition temperature. Because the theory is only applicable to the high-temperature phase, it does not provide information about the temperature dependence of the order parameter. This is best obtained using a lattice dynamic model of the free energy, which we describe in this section. In this case, the soft mode becomes a static distortion of the crystal – we note that the order parameter η is now firmly identified with the static equivalent of the normal mode coordinate of the soft mode. In this case the potential energy part of the Hamiltonian has the form

$$\begin{aligned}\mathcal{H} &= \frac{1}{2}\sum_k \omega_k^2 Q_k Q_{-k} + \frac{1}{4}\alpha_k Q_k Q_{-k}\eta^2 \\ &= \frac{1}{2}\sum_k \left(\omega_k^2 + \frac{1}{2}\alpha_k \eta^2\right) Q_k Q_{-k} \end{aligned} \tag{12.37}$$

In this equation k denotes all phonon modes excluding the soft mode, and the coefficient α_k denotes the anharmonic coupling of the order parameter with any phonon. This is equivalent to the coefficient $\alpha^{(4)}_{\mathbf{k}_0,-\mathbf{k}_0,\mathbf{k},-\mathbf{k}}$, where we denote $\mathbf{k}$, ν by k, and $\mathbf{k}_0$, ν_0 are the wave vector and branch label of the soft mode. The new Hamiltonian indicates that when there is a static distortion, the phonon frequencies are modified to give

$$\tilde{\omega}_k^2 = \omega_k^2 + \frac{1}{2}\alpha_k \eta^2 \tag{12.38}$$

If α_k is positive, the frequency is increased on cooling into the low-temperature phase, and this will have the effect of decreasing the phonon entropy as required. We also remark that the dependence of the phonon frequency on the order parameter given in this relationship can be measured in a spectroscopy experiment, giving an indirect method to determine the temperature dependence of η (see Section 12.2.4).

12.5.2 Potential energy of the crystal

Before we consider the phonons further, we also need to account for the change in potential energy of the crystal due to the phase transition. We write this as

$$E(\eta) = -\frac{1}{2}\kappa_2 \eta^2 + \frac{1}{4}\kappa_4 \eta^4 \tag{12.39}$$

This form has been anticipated by eqn 12.12, which gives the internal energy contribution to the Landau free energy function. Since we can identify the order parameter η as the static component of the frozen-in normal mode that is the

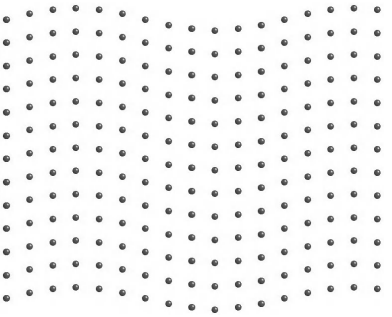


Fig. 12.21 Schematic representation of the modulation of a lattice caused by a soft mode with a general wave vector.

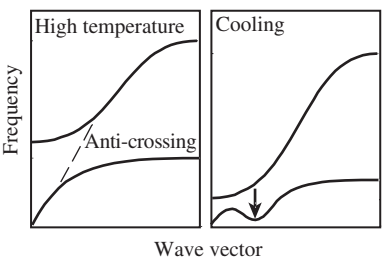


Fig. 12.22 Schematic representation of the mechanism of the incommensurate softening of an acoustic mode in quartz as a result of the softening of an optic branch.

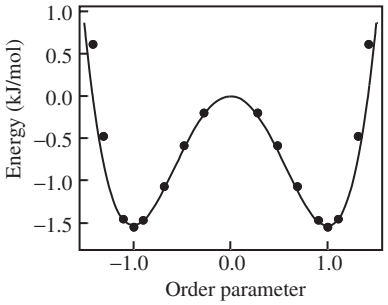


Fig. 12.23 Variation of the potential energy of quartz with changing order parameter. The points are energies calculated using a pair potential model as described in Chapter 5, and the curve is described by eqn 12.39 with fitted values of κ_2 and κ_4. (The calculation was originally reported by Dove *et al.*, *Phys. Chem. Min.*, **26**, 344, 1999.)

soft mode for the phase transition, κ_2 can be identified with the soft mode ω^2. Of course, η can be scaled by an arbitrary factor, in which case κ_4 will be inversely scaled by the same factor, but in the way we develop the theory this arbitrary scaling will be cancelled. The equilibrium value of the order parameter at zero temperature can be obtained from minimization of the potential energy:

$$\frac{\partial E}{\partial \eta} = -\kappa_2 \eta + \kappa_4 \eta^3 = 0 \quad \Rightarrow \quad \eta_0^2 = \kappa_2/\kappa_4 \tag{12.40}$$

An example of this energy function is shown in Fig. 12.23, where calculations of $E(\eta)$ for quartz are fitted by a function of the form of eqn 12.39.

12.5.3 Phonon free energy

We now proceed by considering the phonon free energy (Chapter 9)

$$F_{\text{ph}} = k_{\text{B}} T \sum_k \ln[2 \sinh(\hbar \omega_k / 2 k_{\text{B}} T)] \tag{12.41}$$

For the moment we will make the high-temperature approximation appropriate for $k_{\text{B}} T > \omega_k$, to give

$$\begin{aligned} F_{\text{ph}} &= k_{\text{B}} T \sum_k \ln(\hbar \omega_k / k_{\text{B}} T) \\ &= 3RT \ln(\hbar \langle \tilde{\omega} \rangle / k_{\text{B}} T) \end{aligned} \tag{12.42}$$

We have now replaced the sum over all modes by average values, so that we have from eqn 12.38

$$\langle \tilde{\omega} \rangle^2 = \langle \omega \rangle^2 + \frac{1}{2} \langle \alpha \rangle \eta^2 \tag{12.43}$$

The phonon free energy can be written as a Taylor expansion in the order parameter about the point $\eta = 0$:

$$F_{\text{ph}}(\eta) = F_{\text{ph}}(\eta = 0) + \frac{1}{2} \eta^2 \left(\frac{\partial^2 F_{\text{ph}}}{\partial \eta^2} \right)_{\eta = 0} + \cdots \tag{12.44}$$

There is no linear term because $\partial F / \partial \eta = 0$ at equilibrium, and we will not consider the higher-order terms. We have the following differentials:

$$\frac{\partial F_{\text{ph}}}{\partial \eta} = \frac{3RT}{\langle \tilde{\omega} \rangle} \frac{\partial \langle \tilde{\omega} \rangle}{\partial \eta} \tag{12.45}$$

$$\frac{\partial^2 F_{\text{ph}}}{\partial \eta^2} = \frac{3RT}{\langle \tilde{\omega} \rangle} \left[\frac{\partial^2 \langle \tilde{\omega} \rangle}{\partial \eta^2} - \frac{1}{\langle \tilde{\omega} \rangle} \left(\frac{\partial \langle \tilde{\omega} \rangle}{\partial \eta} \right)^2 \right] \tag{12.46}$$

From eqn 12.43 we can obtain the following differentials of the frequencies:

$$\frac{\partial \langle \tilde{\omega} \rangle}{\partial \eta} = \frac{\langle \omega \rangle \eta}{2 \langle \tilde{\omega} \rangle}; \quad \left(\frac{\partial \langle \tilde{\omega} \rangle}{\partial \eta} \right)_{\eta = 0} = 0 \tag{12.47}$$

$$\frac{\partial^2 \langle \tilde{\omega} \rangle}{\partial \eta^2} = \frac{\langle \omega \rangle}{2 \langle \tilde{\omega} \rangle} - \frac{\langle \omega \rangle \eta}{2 \langle \tilde{\omega} \rangle^2} \frac{\partial \langle \tilde{\omega} \rangle}{\partial \eta}; \quad \left(\frac{\partial^2 \langle \tilde{\omega} \rangle}{\partial \eta^2} \right)_{\eta = 0} = \frac{\langle \omega \rangle}{2 \langle \tilde{\omega} \rangle} \tag{12.48}$$

As a result we can write

$$\left(\frac{\partial^2 F_{\rm ph}}{\partial \eta^2}\right)_{\eta=0} = \frac{3\langle\alpha\rangle RT}{2\langle\omega\rangle^2} \tag{12.49}$$

and

$$F_{\rm ph}(\eta) = F_{\rm ph}(\eta=0) + \frac{3\langle\alpha\rangle RT}{4\langle\omega\rangle^2}\eta^2 \tag{12.50}$$

12.5.4 Full free energy and the Landau free energy function

We now combine the phonon free energy with the potential energy to obtain

$$\begin{aligned} F(\eta) &= F_{\rm ph}(\eta) + E(\eta) \\ &= F_{\rm ph}(\eta=0) + \frac{3\langle\alpha\rangle RT}{4\langle\omega\rangle^2}\eta^2 - \frac{1}{2}\kappa_2\eta^2 + \frac{1}{4}\kappa_4\eta^4 \\ &= F_{\rm ph}(\eta=0) + \frac{1}{2}\left(\frac{3\langle\alpha\rangle RT}{2\langle\omega\rangle^2} - \kappa_2\right)\eta^2 + \frac{1}{4}\kappa_4\eta^4 \\ &= F_{\rm ph}(\eta=0) + \frac{1}{2}\left(\frac{3\langle\alpha\rangle R}{2\langle\omega\rangle^2}\right)\left(T - \frac{2\kappa_2\langle\omega\rangle^2}{3\langle\alpha\rangle R}\right)\eta^2 + \frac{1}{4}\kappa_4\eta^4 \end{aligned} \tag{12.51}$$

This looks exactly like the Landau free energy function (eqn 12.8)

$$F(\eta) = F(\eta=0) + \frac{1}{2}a(T - T_{\rm c})\eta^2 + \frac{1}{4}\eta^4 \tag{12.52}$$

with

$$T_{\rm c} = \frac{2\kappa_2\langle\omega\rangle^2}{3\langle\alpha\rangle R} \tag{12.53}$$

Since κ_2 is the negative of the square of the harmonic frequency of the soft mode, this equation for $T_{\rm c}$ is exactly the same as eqn 12.34 obtained earlier from the model of the soft mode. We have, however, achieved more than simply replicating the earlier result. We have shown how Landau theory falls out of the lattice dynamics, which gives some justification for the use of Landau theory in the study of displacive phase transitions. We have also obtained microscopic interpretations of the coefficients in the Landau free-energy function.

12.5.5 Low-temperature behaviour

The theory above was developed assuming that we could make the high-temperature approximation. Clearly this is not appropriate at low temperatures. In principle we can take the complete free energy function and expand this in terms of η^2 following the same procedure. We would end up with something similar, albeit with more complicated differentials, and the final equation would not end up looking exactly like Landau theory. In this sense we can deduce that the standard Landau theory is itself a high-temperature approximation, and we already know that at low temperatures it is not appropriate.

Rather than do a complete analysis of the low-temperature behaviour, we will restrict ourselves to $T = 0\,\mathrm{K}$. Here the free energy is the potential energy plus the zero point phonon motion, giving

$$F(\eta, T = 0) = \frac{3}{2}N_\mathrm{A}\hbar\langle\tilde{\omega}\rangle - \frac{1}{2}\kappa_2\eta^2 + \frac{1}{4}\kappa_4\eta^4 \tag{12.54}$$

To obtain the equilibrium value of η at $T = 0\,\mathrm{K}$ we calculate the differential of F, using the results for the differential of $\langle\tilde{\omega}\rangle$ given above:

$$\begin{aligned}\left.\frac{\partial F}{\partial \eta}\right|_{T=0} &= \frac{3}{2}N_\mathrm{A}\hbar\frac{\partial\langle\tilde{\omega}\rangle}{\partial\eta} - \kappa_2\eta + \kappa_4\eta^3 \\ &= \left(3N_\mathrm{A}\hbar\frac{\langle\alpha\rangle}{4\langle\tilde{\omega}\rangle} - \kappa_2\right)\eta + \kappa_4\eta^3\end{aligned} \tag{12.55}$$

Setting the differential to zero allows us to obtain the equilibrium zero-temperature value of the order parameter, η_0:

$$\begin{aligned}\left(3N_\mathrm{A}\hbar\frac{\langle\alpha\rangle}{4\langle\tilde{\omega}\rangle} - \kappa_2\right)\eta + \kappa_4\eta^3 = 0 \\ \Rightarrow \quad \eta_0^2 = \frac{\kappa_2}{\kappa_4} - \frac{3N_\mathrm{A}\hbar\overline{\alpha}}{4\kappa_4\langle\tilde{\omega}\rangle}\end{aligned} \tag{12.56}$$

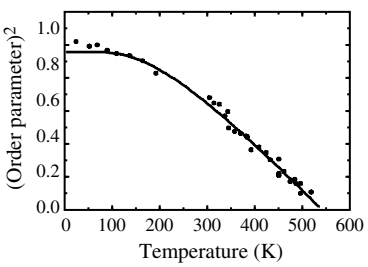


Fig. 12.24 Temperature dependence of the square of the order parameter associated with the displacive phase transition in the aluminosilicate $CaAlSi_3O_8$. The curve represents a fit of the theory of this chapter to the data, taking account of the quantum terms at low temperature.

Note that because we are not now expanding around $\eta = 0$, the equation for η_0 contains $\langle\tilde{\omega}\rangle$ rather than $\langle\omega\rangle$; the former is the frequency modified by the non-zero value of η, and the latter is the frequency extrapolated from the high-temperature phase. This result shows that the effect of zero point motion is to lower the zero-temperature value of η below that given by minimization of $E(\eta)$ alone.

An example of the complete quantum mechanical version of this theory fitted to experimental data for the order parameter is shown in Fig. 12.24. This clearly shows a linear dependence of η^2 close to T_c, and the expected curvature away from this linear relation at low temperatures.

Summary of chapter

- **Displacive phase transitions** occur on changing temperature or pressure, and involve small symmetry-breaking displacements of atoms that can be reversed on changing temperature or pressure.
- The symmetry-breaking atomic displacements associated with **ferroelectric** phase transitions lose the centre of symmetry and give rise to the formation of an electric dipole moment in the unit cell. The resultant dielectric polarization of the crystal can be reversed by application of an electric field. The most common examples of ferroelectric phase transitions are found in the perovskite family of structures, e.g. $PbTiO_3$ and $BaTiO_3$. Ferroelectric phase transitions are accompanied by a divergence of the dielectric constant.
- The atomic displacements associated with some displacive phase transitions are different in neighbouring unit cells, leading to a new lattice

whose unit cells have volumes that are multiples of the initial unit cell volumes. One example is the displacive phase transition in $SrTiO_3$, which involves rotations of the TiO_6 octahedra.

- **Ferroelastic** phase transitions are those for which the primary instability is a shear strain of the lattice (that is, the change in symmetry of the crystal must be equivalent to the change in symmetry associated with the shear strain). Ferroelastic phase transitions are accompanied by the softening of an elastic constant.
- The symmetry-breaking distortions associated with a displacive phase transition can be described by an overall **order parameter** (or a small number of independent order parameters). Examples of order parameters are the size of the dielectric polarization associated with a ferroelectric phase transition (in this case it is possible to define separate order parameters for the three possible vector components of the polarization), or the angle of polyhedral rotation.
- Phase transitions in which the order parameter falls continuously to zero at the transition temperature are defined as **second-order phase transitions**. Those for which there is a discontinuous change in the order parameter at the transition temperature are defined as **first-order phase transitions**. First-order phase transitions also have a discontinuous change in the entropy. For many second-order displacive phase transition the square of the order parameter falls linearly to zero at the transition temperature.
- The thermodynamics associated with the phase transition can be described by an expansion of the free energy in what is known as **Landau theory**. This can describe both first and second-order phase transitions, and is able to incorporate variables such as strain. Although Landau theory is a phenomenological model, it properly describes symmetry in the way that it incorporates different variables, and as a result has some degree of predictive power. Landau theory's main failing is to describe the behaviour of the system at low temperatures.
- An important component of the mechanism of a displacive phase transition is the **soft mode**, namely a phonon mode whose frequency falls to zero at the phase transition. ω^2 represents the restoring force against a set of atomic displacements, which vanishes to allow the phase transition to take place. The temperature dependence of the soft mode arises from **anharmonic phonon interactions**.
- The thermodynamics of the phase transition can be represented by incorporating the dependence of the phonon frequencies on the order parameter, giving results for the transition temperature that are in accord with the transition temperature calculated by anharmonic phonon theory. The Landau free energy function can be obtained from an expansion of the phonon free energy and lattice energy as a power series in the order parameter in the high-temperature limit.
- The low-temperature behaviour of the order parameter is determined by the quantum mechanics of phonons. At absolute zero, the dependence of the zero point motions on the order parameter will cause a reduction of the order parameter, and in extreme cases can suppress the phase transition and stabilize the high-temperature phase.

Further reading

The topic of phase transitions is not usually covered in textbooks on solid state physics other than in discussions of dielectric properties and magnetism; both Kittel (1996) and Ashcroft and Mermin (1976) give introductions to the subject. Dove (1993, 1997), Rao and Rao (1978), and Salje (1990) give complementary comprehensive treatments at an introductory level. A more advanced treatment is given by Bruce and Cowley (1981). The statistical mechanics of phase transitions are discussed by Yeomans (1992). Landau theory is discussed to an advanced level by Tolédano and Tolédano (1987). Some wider issues are discussed in a review article written by the author (M. T. Dove, 1997, *American Mineralogist*, **82**, 213–44.)

Exercises

(12.1) A **tricritical** phase transition is on the border between first-order and second-order behaviour. Such a phase transition might be described by a Landau free energy function of the form of eqn 12.19 with $b = 0$. Obtain expressions for the temperature dependence of the order parameter, susceptibility, and heat capacity for $T < T_c$, and contrast the results with those for second-order and first-order phase transitions (Sections 12.3.3 and 12.3.4 respectively).

(12.2) Consider a first-order phase transition described by Landau theory, eqn 12.19. At the transition temperature, there are minima of $F(\eta)$ at a non-zero value of η^2 with $F(\eta) = 0$. By combining this condition with the condition that $\partial F/\partial\eta = 0$ at the minima, show that the minimum occurs for $\eta^2 = 3b/4c$. Use this result to show that the transition temperature is given by eqn 12.21. Show that eqn 12.20 is the stable solution of $\partial F/\partial\eta = 0$ for $T < T_c$.

(12.3) Show that at the temperature $T_0 + b^2/4ac$ the Landau free energy function for a first-order phase transition, eqn 12.19, develops a new minimum at $\eta^2 = b^2/4c$.

(12.4) Obtain the parameters T_c, b/a, and c/a from the data for the temperature dependence of the ferroelectric polarization in $PbTiO_3$ given in Fig. 2.30 ($\eta = P$). This will require the use of the minimization function in a spreadsheet program together with judicious use of the data. Use the results to plot the form of $F(P)$ for various temperatures.

(12.5) For some phase transitions, symmetry allows the Landau free energy to be written as

$$F(\eta) = \frac{1}{2}a(T - T_c)\eta^2 + \frac{1}{4}\eta^4 + \frac{1}{2}\lambda\epsilon_s\eta + \frac{1}{2}C_{el}\epsilon_s^2$$

where ϵ_s is the spontaneous strain. Show (a) that the equilibrium condition $\partial F/\partial\epsilon_s = 0$ gives $\epsilon_s = -\lambda\eta/2C_{el}$; (b) that this coupling leads to a new transition temperature at $T_c + \lambda^2/4aC_{el}$; (c) that the corresponding elastic constant falls to zero at the phase transition temperature.

(12.6) It is often found that a doubly-degenerate phonon will split at a phase transition to give two new frequencies as $\omega^2 = \omega_0^2 \pm \gamma\eta$. Show that

$$\left.\frac{\partial\omega}{\partial\eta}\right|_{\eta=0} = \frac{\gamma}{2\omega_0}$$

$$\left.\frac{\partial^2\omega}{\partial\eta^2}\right|_{\eta=0} = \pm\frac{\gamma^2}{4\omega_0^3}$$

Substitute these contributions into eqns 12.44–12.46 to show that each split mode gives a contribution to the free energy in the high-temperature limit of

$$f = -\frac{\gamma^2 k_B T}{4\omega_0^3}\eta^2$$

Demonstrate that this has the effect of raising the transition temperature.

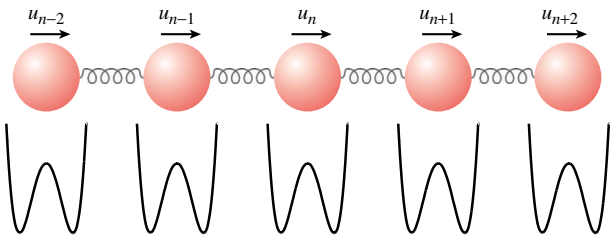
(12.7) Equation 12.53 gives T_c obtained in the high-temperature approximation, namely for $k_B T > \hbar\langle\omega\rangle$. By comparing eqns 12.53 and 12.56, show that if the calculated value of $k_B T_c < \hbar\langle\omega\rangle/2$ the phase transition is suppressed through quantum fluctuations.

(12.8) Show that at zero temperature, eqn 12.32 for the soft mode frequency is given by

$$\omega^2 = \omega_0^2 + \frac{3N_A\overline{\alpha}}{4\langle\omega\rangle}$$

Equation 12.34 gives an alternative formulation of T_c in terms of the squares of the harmonic frequency of the soft mode at zero temperature, ω_0^2, again calculated within the high-temperature limit. Show that if the calculated value of $k_B T_c < \hbar\langle\omega\rangle/2$, the soft mode frequency is prevented from becoming unstable at zero temperature.

(12.9) The figure below shows a simple model for a phase transition.



Each atom interacts with its neighbours through harmonic forces, characterized by the force constant J, but also is subject to forces through a local potential energy function of the form

$$E(u) = -\frac{\kappa_2}{2}u^2 + \frac{\kappa_4}{4}u^4$$

Show that the energy of atom j in a one-dimensional chain can be written as

$$E_j = \frac{2J - \kappa_2}{2}u_j^2 - Ju_j(u_{j+1} + u_{j-1}) + \frac{\kappa_4}{4}u^4$$

Obtain the dispersion curve for the harmonic component of this energy.

The anharmonic component can be approximated as $u^4 \simeq 3u^2\langle u^2\rangle$, and we can approximate $\langle u^2\rangle = k_B T/\omega_0^2$. Modify the dispersion curve to incorporate this quasi-harmonic term, and show that the model gives a phase transition at $k_B T_c = \kappa_2\omega_0^2/3\kappa_4$.

(12.10) Insights into the rotational phase transition in the perovskite structure can be gained by treating the octahedra as nearly rigid objects. If each octahedron is assigned a single variable to describe its orientation, θ, two neighbouring octahedra have lowest energy if they rotate by exactly the same amount in opposite directions. Show that the harmonic energy between two neighbouring octahedra can be written as

$$E(i, j) = \frac{1}{2}K(\theta_i + \theta_k)^2$$

For a phase transition to occur, it can be presumed that each octahedron also experiences a local double-well potential of the form

$$E(\theta) = -\frac{\kappa_2}{2}\theta^2 + \frac{\kappa_4}{4}\theta^4$$

Show that when the harmonic term is combined with the interactions between neighbours, a soft mode is obtained at $\mathbf{k} = (\frac{1}{2}, \frac{1}{2}, 0)$.

(12.11) The Landau theory for the ferroelectric phase transition in the perovskite structure can be written in the following form to account for the components of the polarization along the three axes:

$$F = \frac{1}{2}a(T - T_c)\left(P_x^2 + P_y^2 + P_z^2\right) + \frac{1}{4}b\left(P_x^4 + P_y^4 + P_z^4\right) + \frac{1}{4}\gamma\left(P_x^2 + P_y^2 + P_z^2\right)^2$$

Describe the conditions under which the stable solutions are

(a) $P_x \neq 0$ and $P_y = P_z = 0$
(b) $P_x = P_y \neq 0$ and $P_z = 0$
(c) $P_x = P_y = P_z \neq 0$.

(12.12) A given material undergoes a displacive phase transition at $T_c = 950$ K. The vibrations have an average frequency of 10 THz, and the phase transition causes the vibrational frequencies to increase by an average of 0.02 THz on cooling to low temperature. The phase transition is accompanied by a step change in the heat capacity of $0.124\,\mathrm{J\,mol^{-1}\,K^{-1}}$. Calculate (*a*) the coefficients of the Landau free energy for the phase transition on the assumption that the data can be interpreted in the high-temperature limit; (*b*) the change in the value of the order parameter at low temperature due to the effects of zero point motions. (*Note*: you will need to determine values of κ_2 and κ_4; their exact values will depend on how you chose to normalize the value of the order parameter at low temperature.)

Use a spreadsheet to minimize the free energy with respect to the order parameter for a range of temperatures in order to obtain the temperature dependence of the order parameter, using the values of the parameters obtained in the first part of the question. From your plot, ascertain the validity of the high-temperature approximation for the analysis of the free energy around T_c. At what temperatures does a plot of the square of the order parameter against temperature deviate from a nearly linear relationship. Is this temperature in accord with the average vibrational frequency?

A Real crystals!

A.1 Reality against ideality

Throughout the chapters of this book we have tacitly assumed that we always have perfect lattices, and that the atoms remain tied to their sites. We have allowed atoms to vibrate about their equilibrium positions, and have developed a detailed theory of thermal motions, but thermal motions are not imperfections. The ideas in the main chapters are not significantly modified if one admits that crystals are not perfect, but there are some phenomena that cannot be explained without invoking the presence of defects. These include:

- Understanding the forces required to fracture a crystal. If one simply calculated the force required to separate two layers of atoms in a crystal, the result would come out to be much larger that the actual force required to fracture a solid. It therefore appears as if a crystal breaks in stages, with each stage requiring a rather smaller force.
- The ability of a crystal to deform irreversibly under stress. The primary example is called **slip**, where application of a stress acts as a shear stress on particular crystal planes which appear to slide over each other. This is described in Section 1.1.3, and illustrated in Fig. 1.6.
- The origin of colour in many materials that would otherwise be transparent. A well-known example is the red colour of ruby, Al_2O_3, which is caused by Cr^{3+} impurities on the Al sites.
- The presence of a small but non-zero electrical conductivity in insulators. These are caused by the motion of charged ions as in a salt solution. Ionic conductivity is much higher in materials that contain vacant sites as defects.
- The ability of atoms to diffuse within a solid, such as the diffusion of carbon atoms in iron. The processes are related to those that allow electrical conductivity.
- Crystal growth, which often appears to proceed through a process that involves the formation of a spiral defect.

The range of phenomena that can only be explained by the presence of imperfections, and the fact that there is a wide range of different type of defects, means that there has been a lot of interest in this side of solid state physics for many years. In some cases the presence of defects can be exploited in applications, and in other cases the presence of defects causes technological problems, and both mean that it is important to understand the behaviour of defects. In fact, we show below that real crystals will contain some types of imperfections as an equilibrium state.

In this appendix we will briefly review a number of ways in which real materials depart from the perfect ideal structures we have focused on the main part of this book. There is a mixture of equilibrium imperfections, imperfections that arise from the way that a crystal is grown, and imperfections that can arrive from processes that occur within the crystal itself. We will separate two types of imperfections, namely those that are associated with individual sites in the crystal, and those that span space. Since the principal focus of the book is concerned with phenomena that are not significantly affected by crystal imperfections, our review can be little more than cursory, but is included because it is important to be aware of them.

A.2 Point defects

A.2.1 Vacancies: Schottky defects

Consider the simplest crystal defect, namely that of an atomic site becoming vacant, with the missing atom migrating to the crystal surface. This process will be energetically unfavourable, costing the system a change in energy that we denote as E_V. If we have a crystal with N crystal sites containing atoms and n vacant sites, the number of arrangements of the vacant sites is

$$W = \frac{(N+n)!}{N!n!} \tag{A.1}$$

The corresponding entropy is

$$S = k_B \ln W = k_B(\ln(N+n)! - \ln N! - \ln n!) \tag{A.2}$$

We can use Stirling's approximation for large N,

$$\ln N! = N \ln N - N \tag{A.3}$$

to express the entropy as

$$S = k_B((N+n)\ln(N+n) - N \ln N - n \ln n) \tag{A.4}$$

The free energy is then written as

$$F = nE_V - k_B T((N+n)\ln(N+n) - N \ln N - n \ln n) \tag{A.5}$$

The equilibrium number of vacant sites can be obtained by minimization of the free energy with respect to n:

$$\frac{\partial F}{\partial n} = E_V - k_B T(\ln(N+n) - \ln n) = 0 \tag{A.6}$$

By rearrangement of this equation, and assuming that $n/N \ll 1$, we obtain

$$n = N \exp(-E_V/k_B T) \tag{A.7}$$

This analysis shows that there will be a finite number of vacant sites as an equilibrium property, and this number will increase on heating.

The simple case of a vacant site is called a **Schottky defect**. In order to maintain a neutral charge distribution across a local length scale, it is common for both positive and negative vacant sites to be produced in thermal equilibrium and to be evenly distributed throughout the sample. This is represented in Fig. A.1.

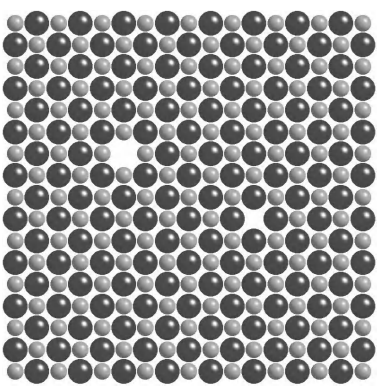


Fig. A.1 Cation and anion charge-balanced Shottky defects in NaCl.

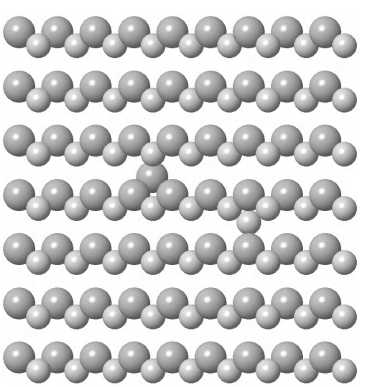


Fig. A.2 Pair of charge-balanced Frenkel defects in AgI.

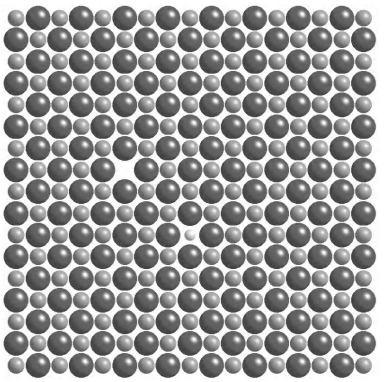


Fig. A.3 Substitution of a Ca^{2+} cation for a Na^{+} cation in NaCl, accompanied by the formation of a vacant cation site in order to maintain charge neutrality.

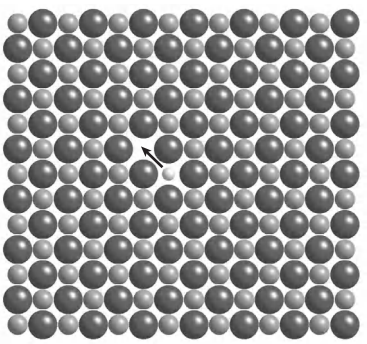


Fig. A.4 Diffusion of a cation in NaCl assisted through the presence of Shottky defects.

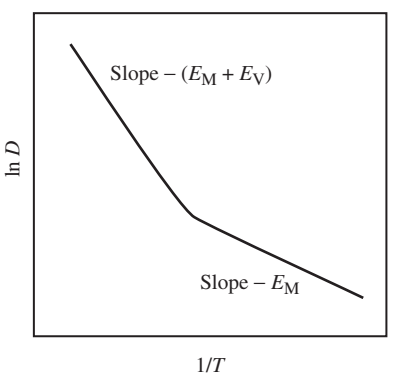


Fig. A.5 Diffusion of a cation in NaCl assisted through the presence of Shottky defects.

A.2.2 Interstitial defects: Frenkel defects

In crystals that do not pack with high efficiency, it is possible for atoms to occupy sites that are normally vacant, called **interstitial sites**. A **Frenkel defect** occurs when at atom leaves its normal site to create a vacancy, and is then displaced into one of the interstitial sites. This process is illustrated in Fig. A.2. This type of defect is found in materials such as silver halides.

A.2.3 Coupled charge substitutions and vacancies

The vacancies we have described occur as equilibrium processes. Other point defects can be produced as non-equilibrium structures through the process of crystal growth. A common defect is the substitution of a different type of atom, usually one of similar charge. For example, NaCl can contain K^{+} defects substituting for the cation sites, or F^{-} substituting on the anion sites. A substitution of a different charged cation will require production of a compensating charge defect. For example, substituting a Ca^{2+} cation in NaCl will require the formation of a charge-compensating cation vacancy. This structure is illustrated in Fig. A.3.

A.2.4 Colour centres

The charge substitutions we have considered have been restricted to ions. It is also possible for electrons to occupy vacant anion sites in order to maintain charge neutrality. The electron forms its own energy bands. The colour centre is able to absorb electromagnetic radiation in the visible spectrum, and this gives colour to what would otherwise be a transparent crystal. As a result, the presence of the colour centre can be detected in an optical absorption experiment.

A.2.5 Diffusion and atomic mobility

The existence of point defects allow for the possibility of diffusion of atoms. Consider a Ca^{2+} cation within a crystal of NaCl. Any motion of the Ca^{2+} cation is considerably assisted if its neighbouring site is vacant, as illustrated in Fig. A.4.

The steady-state diffusion of atoms in one-dimension is described by **Fick's law**:

$$J = -D\frac{\mathrm{d}c}{\mathrm{d}x} \tag{A.8}$$

where J is the flux of atoms (number crossing unit area in unit time), c is the concentration of atoms, and D is the **diffusion constant**. The form of this equation means that there is a net flow of atoms against the gradient of concentration. In princple the flux and concentration gradients are both first-rank tensors, and the diffusion constant is a second-rank tensor (Chapter 7), but we limit our discussion here to isotropic materials. The diffusion constant typically depends on temperature according to an **Arrhenius relationship**:

$$D = D_0 \exp(-E_A/k_B T) \tag{A.9}$$

E_A is called the **activation energy**, and this represents an energy barrier that must be crossed to allow an atom to move. We can identify two factors. The first, which is clearly seen in the picture of the diffusion pathway in the NaCl structure (Fig. A.4), is that energy is required to allow the diffusing atom to move past the atoms that are blocking the pathway into the vacant site. Writing E_M as the energy barrier against mobility, we simply have $E_A = E_M$. The second factor concerns the probability of a diffusing atom to have a vacant neighbouring site to move into. At low temperatures, the number of vacant sites will be determined by the level of impurities (e.g., the number of vacant sites in NaCl will be determined by the number of Ca^{2+} cations present in the sample). However, at high temperatures the number of vacant sites will be dominated by the number that is created thermally, eqn A.7. The diffusion constant is proportional to the concentration of vacant sites, so we can write $E_A = E_M + E_V$.

There are several experimental methods available to measure the diffusion constant, some of which are based on macroscopic measurements (e.g., use of radioactive tracer elements), and others which are microscopic (ionic conductivity, magnetic resonance). It is common to plot data for eqn A.9 in the form of ln D *vs.* $1/T$, as shown in Fig. A.5. The slope of the plot is $-E_A/k_B$. The plot can be divided into two regions: the low-temperature region where $E_A = E_M$, and the high-temperature region where $E_A = E_M + E_V$.

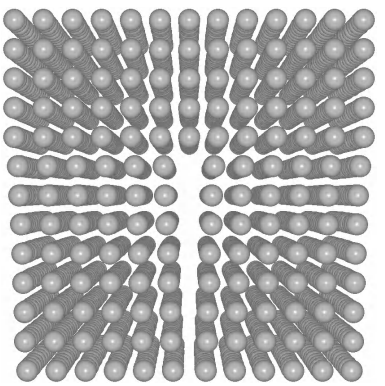


Fig. A.6 Schematic representation of an edge dislocation.

A.3 Large-scale imperfections

A.3.1 Dislocations

Dislocations are defects that are extended over a line. In particular, they involve missing partial planes of atoms, and the dislocation is the line that defines the edge of the missing partial plane. The simplest type of dislocation to visualize is the edge dislocation, which is illustrated schematically in Fig. A.6. The missing part of the plane of atoms can clearly be seen, and the dislocation marks the edge of this plane. The pure edge dislocation is the easiest to visualize. It is obvious that the normal to the additional plane of atoms is orthogonal to the edge of the extra plane.

In general there can be any orientational relationship between the line of the dislocation and the Burgers vector. If the two are parallel rather than orthogonal, the dislocation is called a **screw dislocation** because the mismatch screws around the dislocation.

A complex line defect is imaged in Fig. A.7, and an interesting zig-zag shaped line defect is imaged in Fig. A.8.

Dislocations were first investigated in order to understand a significant discrepancy between the calculated and experimental shear strengths of metals. Dislocations give the mechanism by which large-scale plastic deformation of a crystal, such as slip, can occur.

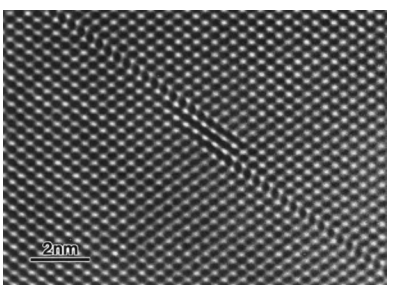


Fig. A.7 Line defect in CdTe observed by electron microscopy. (Photograph courtesy of Jeol Ltd.)

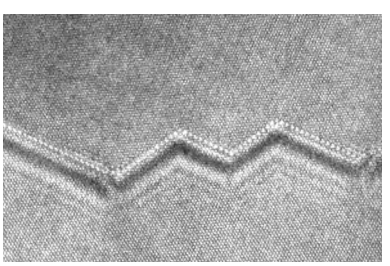


Fig. A.8 Unusual zig-zag shaped line defect in gold observed by electron microscopy. (Picture courtesy of A. Kirkland and P. Midgely, University of Cambridge.)

A.3.2 Grain boundaries

Grain boundaries are the boundaries along which two crystallites of different orientations are joined. There need be no orientational relationship between

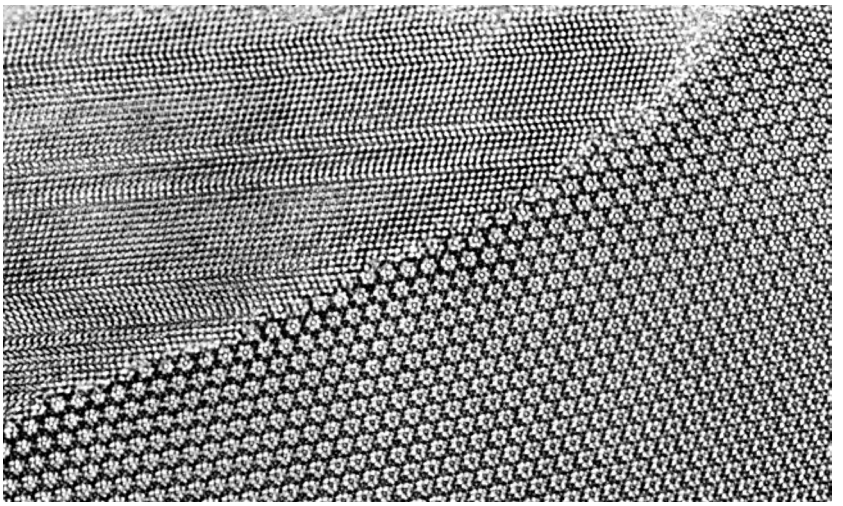


Fig. A.9 Two grains of gold observed in a transmission electron diffraction experiment. The photograph shows the structure of the grain boundary clearly. The figure also shows stacking faults in one of the grains. (Picture courtesy of A. Kirkland and P. Midyley, University of Cambridge.)

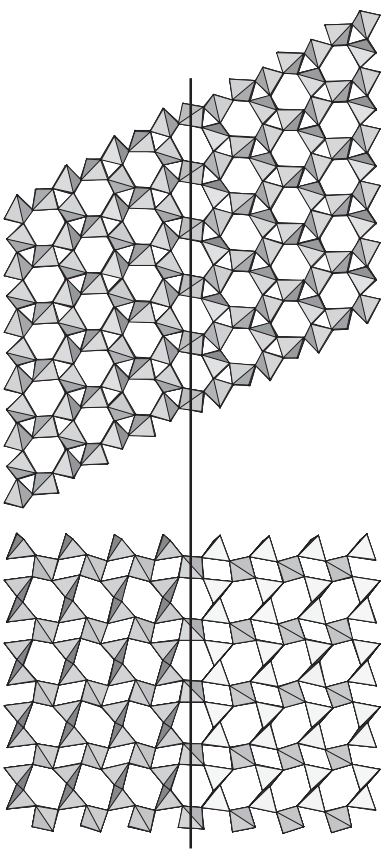


Fig. A.10 Structure of the domain wall in quartz viewed along two orthogonal crystal axes. The position of the domain wall is denoted by the vertical line. (Data for the diagram kindly provided by Mark Calleja, University of Cambridge.)

different grains, but often the surfaces of two crystallites will mesh together more easily if there is some matching of their orientations. The grain boundaries will usually have a lower density than the bulk, and for this reason grain boundaries provide low-energy pathways for diffusion of atoms. Figure A.9 shows a grain boundary in gold. The same figure also shows stacking faults in one of the grains.

A.3.3 Domains and domain walls

When a material undergoes a phase transition, it is unlikely that the whole crystal will transform into the new phase in exactly the same way. Instead, it is more probable that different regions will transform independently. Take, for example, the ferroelectric phase transition in $PbTiO_3$. The ferroelectric moment can lie along each of the three cubic axes in either a positive or negative sense. As a result, there are six ways in which the phase transition can proceed, and it is likely that different regions of a crystal of $PbTiO_3$ will transform to give each of the six possibilities on cooling below the phase transition. Once a small region of a crystal has started to transform into one of the ferroelectric states, the crystal in the immediate environment will follow and the transformation process will grow. The small region of the sample that transforms in one particular way is called either a **domain** or a **twin**.

Because the orientations of neighbouring domains are fixed by the structure of the parent unit cell, the orientations of the domain walls may also be constrained. Fig. A.10 shows a domain wall in quartz (the phase transition in quartz is illustrated in Chapter 12), and Fig. A.11 shows a domain wall in the perovskite $CaTiO_3$. The two structures either side of the domain all have an opposite sign of the order parameter. The two domains fit together with the aid of a local shear. No tetrahedra need to be distorted through the formation of the domain wall.

A.3.4 Surfaces and surface reconstructions

The study of surfaces is one of the major activities of modern solid state physics, and we can do little more here than allude to this work. The point we wish to

make here is that the forces on the atoms at a surface will necessarily be different from those in the bulk, and the atomic structure at the surface will reflect these differences. A clean-cut surface may lead to some atoms losing bonds. In some cases, the missing bonds will be satisfied by grafting on a layer of bonded hydrogen atoms. For example, a surface of MgO (NaCl structure) will contain $(OH)^-$ molecular ions. In other cases, the atoms on the surface will rearrange in order to bond to each other, a process called **surface reconstruction**. This process occurs on surfaces of silicon, where each atom seeks to have four bonds. It is also found that the spacing between atomic layers at the surface will be slightly modified. There is often a pattern of displacements of the layers that decays exponentially with the distance from the surface.

It should be stressed that different surface cuts will give very different energies, and only a small subset of all possible surfaces are actually energetically favourable. One issue is whether the surface has charge neutrality. The $(10\bar{4})$ plane of calcite forms an ideal surface because it achieves charge neutrality – the atomic structure of this surface is shown in Fig. 4.2, and this surface is the predominant surface that gives crystals of calcite their characteristic shape, Fig. 1.1.

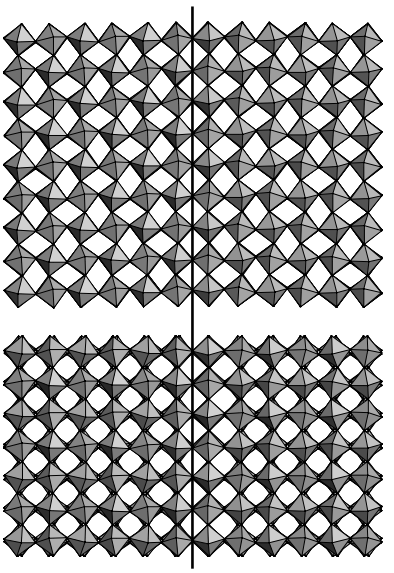


Fig. A.11 Structure of a domain wall in $CaTiO_3$ perovskite viewed along two orthogonal crystal axes. The position of the domain wall is denoted by the vertical line. (Data for the diagram kindly provided by Mark Calleja, University of Cambridge.)

Summary of appendix

- Atom vacancies are produced in a crystal as an equilibrium feature. These are known as **Schottky defects**.
- **Frenkel defects** involve formation of a vacancy with the ion moving into an interstitial site.
- Ionic mobility is governed by an **activation energy** that describes an energy barrier an atom must cross, which is coupled with the energy required to form a vacant site.
- Large-scale defects include **dislocations**, **grain boundaries**, and **domain walls**.

Further reading

Both Kittel (1996) and Ashcroft and Mermin (1976) give excellent reviews of the wide range of defects in crystalline materials. More detailed treatments, particularly for dislocations, are given by Hirth and Lothe (1968), Hull and Bacon (1984), Kosevich (1999), Nabarro (1967), Tilley (1987), and Whelan (1990).

B Fourier analysis

B.1 Fourier transforms as the extension of Fourier series

The equations developed to describe the diffraction of radiation from matter are examples of the mathematical technique of Fourier transformation. The Fourier transform is an extension of the Fourier series. In a one-dimensional Fourier series, a repeating function $y(x)$ that has period a is analysed in terms of a superposition of waves:

$$y(x) = \sum_n A_n \cos(2\pi nx/a) + B_n \sin(2\pi nx/a) \tag{B.1}$$

where A_n and B_n are the amplitudes of the cosine and sine components, each of which has wavelength a/n. Equivalently, each wave has wave vector $k = n2\pi/a$.

The amplitudes are given by

$$A_n = \frac{1}{a}\int_0^a y(x)\cos(2\pi nx/a)\,\mathrm{d}x$$
$$B_n = \frac{1}{a}\int_0^a y(x)\sin(2\pi nx/a)\,\mathrm{d}x$$

The Fourier transform is an extension of the Fourier series, in which any function is represented in terms of a superposition of waves, but in the general case there is no repetition of the function, and therefore waves of all wavelength will contribute. The Fourier transform can be viewed as a Fourier analysis of a repeating function where the period $a \to \infty$. This means that there will be an infinitesimally small difference between two wave vectors $k = n2\pi/a$ and $k + \Delta k = (n+1)2\pi/a$. Thus the sum in eqn B.1 over all values of n, and hence over k, can be replaced by an integral over all values of k. The final result is a distribution function for the components of k. The equations of the Fourier transform given in this appendix can readily be generalized to three dimensions.

B.2 One-dimensional Fourier transform

The Fourier transform of the one-dimensional function $f(x)$ is defined as

$$g(k) = \int_{-\infty}^{\infty} f(x)\exp(2\pi ikx)\,\mathrm{d}x \tag{B.2}$$

The function $g(k)$ gives the distribution function for the decomposition of $f(x)$. The function $f(x)$ can be reconstructed by the reverse Fourier transform:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(k)\exp(-2\pi ikx)\,\mathrm{d}k \tag{B.3}$$

B.3 Some one-dimensional Fourier transforms

B.3.1 Dirac δ function

The Dirac δ function is defined in Chapter 3, eqns 3.5 and 3.6. In one dimension we write

$$\delta(x-x') = \begin{cases}\infty & \text{if } x = x' \\ 0 & \text{if } x \neq x'\end{cases} \tag{B.4}$$

with $\int \delta(x-x')\,\mathrm{d}x = 1$. The Fourier transform is written as

$$\begin{aligned}\int_{-\infty}^{\infty}\delta(x-x')\exp(ikx)\,\mathrm{d}x &= \lim_{\delta x\to 0}\int_{x'-\delta x}^{x'+\delta x}\delta(x-x')\exp(ikx)\,\mathrm{d}x \\ &= \exp(ikx')\int_{x'-\delta x}^{x'+\delta x}\delta(x-x')\,\mathrm{d}x \\ &= \exp(ikx')\end{aligned} \tag{B.5}$$

Thus the one-dimensional Fourier transform of a Dirac δ function is simply a wave with period $2\pi/x'$.

B.3.2 Slit function

The **slit function** (so called because it represents a slit in optical diffraction, and also known as the **box function**) is defined as

$$f(x) = \begin{cases}1/2a, & |x| < a \\ 0, & |x| > a\end{cases} \tag{B.6}$$

The function is normalized to give $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. The Fourier transform is given as

$$\begin{aligned}g(k) &= \int_{-\infty}^{\infty} f(x)\exp(ikx)\,\mathrm{d}x \\ &= \frac{1}{2a}\int_{-a}^{a}\cos(kx)\,\mathrm{d}x \\ &= \sin(ka)/ka\end{aligned} \tag{B.7}$$

This function is called the **sinc function**, and is illustrated in Fig. B.1.

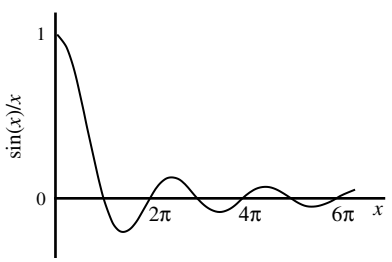


Fig. B.1 Sinc function as the Fourier transform of the slit function.

B.3.3 Symmetric exponential function

The **symmetric exponential function** is defined as

$$f(x) = \frac{1}{a}\exp(-|x|/a) \tag{B.8}$$

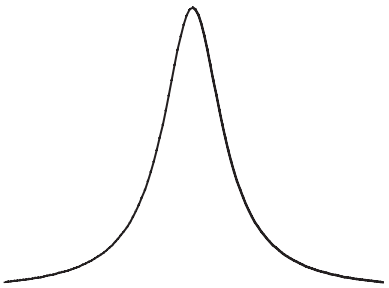


Fig. B.2 Representation of the Lorentzian function.

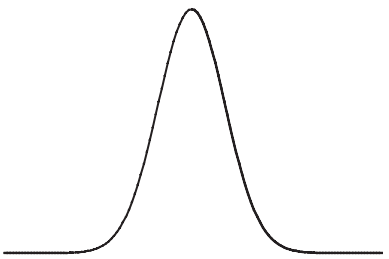


Fig. B.3 Representation of the Gaussian function.

The Fourier transform is

$$g(k) = \frac{1}{a}\int_{-\infty}^{0} \exp(x/a)\exp(ikx)\,\mathrm{d}x + \frac{1}{a}\int_{0}^{\infty} \exp(-x/a)\exp(ikx)\,\mathrm{d}x$$
$$= \frac{2}{1+(ak)^2} \tag{B.9}$$

This is the **Lorentzian function**, and is shown in Fig. B.2.

B.3.4 Gaussian function

The one-dimensional Gaussian function is defined as

$$f(x) = \frac{1}{\sqrt{\pi}a}\exp(-x^2/a^2) \tag{B.10}$$

This function is normalized to give an integral of 1, and is shown in Fig. B.3. The Fourier transform is

$$g(k) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}a}\exp(-x^2/a^2)\exp(ikx)\,\mathrm{d}x = \exp(-k^2a^2/4) \tag{B.11}$$

which is also a Gaussian function.

B.4 Convolution theorem

Central to the discussion of diffraction in Chapter 6 is the theorem that the Fourier transform of the function obtained by the convolution of two other functions is the product of the Fourier transforms of the two convolved functions. For two-dimensional functions $g(x)$ and $h(x)$, the convolution is written as

$$f(x) = g(x) \otimes h(x) = \int g(x)h(x-x')\,\mathrm{d}x' \tag{B.12}$$

If the Fourier transforms of $g(x)$ and $h(x)$ are denoted as $G(k)$ and $H(x)$, the convolution theorem holds that the Fourier transform of $f(x)$ is given as

$$F(k) = G(k) \times H(k) \tag{B.13}$$

The convolution theorem is easily demonstrated by multiplying $G(k)$ and $H(k)$:

$$G(k) \times H(k) = \int g(x)\exp(ikx)\,\mathrm{d}x \times \int h(x')\exp(ikx')\,\mathrm{d}x'$$
$$= \int g(x)h(x')\exp(ik(x+x'))\,\mathrm{d}x\,\mathrm{d}x'$$
$$= \int g(x)h(x''-x)\exp(ikx'')\,\mathrm{d}x\,\mathrm{d}x'' \tag{B.14}$$

This is simply equal to the Fourier transform of $g(x) \otimes h(x)$ (eqn B.12), as required.

Summary of appendix

- The mathematics of one-dimensional Fourier transforms is reviewed.
- The Fourier transforms of the Dirac delta function, slit function, exponential and Gaussian functions are derived.
- The convolution theorem by which the Fourier transform of $g(x) \otimes h(x)$ is equal to the product of the separate Fourier transforms of $g(x)$ and $h(x)$ is proven.

Further reading

This appendix has only attempted to cover the Fourier transforms of relevance to this book. More general treatments will be given in almost all undergraduate textbooks on mathematics for the physical sciences.

C Schoenflies representation of the point groups

C.1 The Schoenflies and International systems

There are two main systems in general use for labelling point groups, namely the International system that is used throughout this book, and the Schoenflies system. The two are used together in Chapter 3, and the purpose of this appendix is to explain the rationale of the Schoenflies system in order to understand how the two systems compare. If it is possible to generalize, it is often the case that the International system is used by crystallographers, and the Schoenflies system by spectroscopists. In Chapter 3 we follow the International system in developing the formalism for the space groups. However, a similar system of labelling the space groups based on the Schoenflies system is in use, and both are given in listings of the space groups.

C.2 Schoenflies labelling of non-cubic point groups

The starting point is the n-fold rotation axis, which is given the label C_n (the international symbol being n). The mirror planes are then added to the rotation axes in either of two ways. The symbol C_{nh} represents the case where the mirror plane is normal to the mirror (the equivalent of n/m), and C_{nv} is the case where the rotation axis lies within the mirror planes (as in $mm2$).

The special symbol D_n is given for the case where an n-fold rotation axis is combined with a set of 2-fold rotation axes normal to the main rotation axis (e.g. 222). This can then be combined with a mirror plane normal to the main rotation axis to give D_{nh}, or with mirror planes that include the 2-fold rotation axes to give D_{nd}.

The main difference between the Schoenflies system and the International system is in the way that the Schoenflies system treats the rotoinversion axis. In the Schoenflies system the rotoinversion axis is not defined. Instead it uses what is called the rotation-reflection axis. This combines an n-fold rotation with a reflection through the plane normal to the rotation. Thus S_6 and S_4 are the Schoenflies equivalent of $\overline{3}$ and $\overline{4}$ respectively. The centre of symmetry, $\overline{1}$, is obtained by the S_2 rotation-reflection in this description. On the other hand, the Schoenflies equivalent of $\overline{6}$ is C_{3h}, which is the parallel of the equivalent symbol $3/m$.

Three other symbols are also used. First, C_i can be used instead of S_2 for $\overline{1}$. This can be combined with an n-fold rotation axis to give C_{ni}. The one example that is commonly used is C_{3i}, which is equivalent to S_6 (i.e. $\overline{3}$ in the International system). Other examples are represented by the main alternative symbols (e.g. C_{2h} is used instead of C_{2i}). The C_{ni} combination should not be confused with the rotoinversion axis.

The second symbol in common use is V, to represent D_2 (222 in the International system). This can be combined with mirror planes normal to the 2-fold axes to give V_h (mmm in the International system), which is equivalent to D_{2h}. V_d can also be defined (recall that in this case the mirror planes will include the rotation axes); this is equivalent to D_{2d} ($\overline{4}2m$).

The third symbol that is sometimes used is C_s to represent a single mirror plane as an alternative to C_{1h}.

C.3 Schoenflies labelling of the cubic point groups

The basic cubic point groups are based on either tetrahedral or octahedral symmetry, which are given the labels T and O respectively; these are the equivalent of 23 and 432 in the International system. The first can be combined with a mirror plane to give T_d, the equivalent of $\overline{4}3m$. Both cubic groups can be combined with mirror planes normal to the 2-fold axes to give T_h and O_h, the equivalent of $m3$ and $m3m$ respectively in the International system.

Summary of appendix

- The **Schoenflies** and **International systems** of classification of point groups are compared.
- One important difference between the two systems is that the Schoenflies system uses the **rotation-reflection axis** instead of the **rotoinversion axis** of the international system.

Further reading

Ashcroft and Mermin (1976) give a description of symmetry that covers both formalisms; similarly Volume I of the **International Tables of Crystallography** present both formalisms together with their adaptation for the symbols for space groups.

D

Rhombohedral, trigonal, and hexagonal unit cells

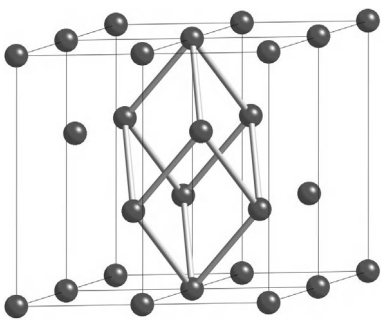


Fig. D.1 One mapping of a primitive rhombohedral unit cell onto non-primitive trigonal axes.

As noted in Chapter 3, the trigonal and hexagonal lattice parameters have the same relationships between them, namely $a = b \neq c$ and $\alpha = \beta = 90°$, $\gamma = 120°$. On the other hand, the rhombohedral lattice parameters are given as $a = b = c$, $\alpha = \beta = \gamma$. The defining symmetry for the hexagonal crystal is the presence of a 6-fold or $\bar{6}$-fold symmetry, whereas the defining symmetry for the trigonal or rhombohedral crystals is the presence of a 3-fold or $\bar{3}$-fold symmetry. It is this defining symmetry, rather than the relationships between lattice parameters, that puts the trigonal and rhombohedral lattices together rather than the trigonal and hexagonal lattices.

The important point is that the rhombohedral lattice can be recast in the form of a non-primitive trigonal lattice. This is shown in Fig. D.1. We can write this in terms of the transformation matrices given in Chapter 4:

$$\begin{pmatrix} \mathbf{a}_{\text{rhom}} \\ \mathbf{b}_{\text{rhom}} \\ \mathbf{c}_{\text{rhom}} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \mathbf{a}_{\text{trig}} \\ \mathbf{b}_{\text{trig}} \\ \mathbf{c}_{\text{trig}} \end{pmatrix} \tag{D.1}$$

and

$$\begin{pmatrix} \mathbf{a}_{\text{trig}} \\ \mathbf{b}_{\text{trig}} \\ \mathbf{c}_{\text{trig}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{a}_{\text{rhom}} \\ \mathbf{b}_{\text{rhom}} \\ \mathbf{c}_{\text{rhom}} \end{pmatrix} \tag{D.2}$$

Summary of appendix

- Two settings of the rhombohedral lattice are compared.

Further reading

The rhombohedral lattice is discussed in some detail in Volume I of the International Tables of Crystallography.

Space groups

E

E.1 Space group symbols

Although one may feel that the existence of as many as 230 space groups makes the use of symmetry seem rather daunting, it is the case that their use can be extremely helpful. We are aided by the fact that there is a logic to the development of any space group, and we need not contemplate trying to learn the details beyond the basic principles.

The convention used for labelling the different space groups is similar to that used for the labelling of point groups. The space-group symbol is prefixed with the following lattice symbols:

P	primitive
I	body centred
F	face centred (all faces)
R	rhombohedral
A, B, or C	centred on one of the A, B, or C faces

The lattice symbol is followed by symbols to denote the symmetry along each of the three crystallographic axes:

m	mirror place
a, b or c	glide plane
n	n glide plane
d	diamond glide place
2, 3, 4, or 6	rotation axis
2_1, 3_1, 3_2, 4_1 etc., 6_1 etc.	screw axis

In composing the labels for the space groups, if there is no special symmetry along two axes only the symmetry of the third axis is given, following the convention for point groups. Also, as for point groups, the label of the space group will only include the minimum information required to generate the complete symmetry, and will not necessarily include all the symmetry operations that are generated by the given symmetry operations.

The ideas are illustrated by considering a straightforward example. Consider an orthorhombic crystal that has an n-glide plane with reflection in (100), a mirror plane with reflection in (010), and an a-glide with reflection in (100). This combination will have the space group symbol $Pnma$. Just as with point groups, such as mmm, in this example there will be rotational symmetry along

each axis that is implied by the three planes of reflection see below. These are not included in the symbol for the space group.

E.2 Defining symmetry

Each space group has a defining set of symmetry operations. We met this idea when discussing point groups. Let us take the example of the point group $2/m$. The symmetry means that the following four points, defined with respect to the origin of the symmetry and with the unique axis lying along [010], are equivalent:

$$\begin{array}{ll} x, y, z & \bar{x}, \bar{y}, \bar{z} \\ x, \bar{y}, z & \bar{x}, y, \bar{z} \end{array}$$

In this case, the existence of any two symmetry operations, whether the 2-fold rotation axis, the mirror plane, or the centre of symmetry, automatically generates the other symmetry operation.

We can illustrate this by consideration of the space group $Pnma$. We can write down the four symmetry operations defined by the symbol, assuming that all three planes of reflection intersect at the origin:

$$\begin{array}{rl} \text{Identity:} & x, y, z \\ n\text{-glide on (100):} & \bar{x}, \frac{1}{2}+y, \frac{1}{2}+z \\ \text{mirror plane on (010):} & x, \bar{y}, z \\ a\text{-glide on (001):} & \frac{1}{2}+x, y, \bar{z} \end{array}$$

Now we can combine the different operations:

$$\begin{array}{rcl} n\,(100) \otimes m\,(010) & \Rightarrow & \bar{x}, \frac{1}{2}-y, \frac{1}{2}+z = 2_1\,[001] \,@\, 0, \frac{1}{4}, z \\ n\,(100) \otimes a\,(001) & \Rightarrow & \frac{1}{2}-x, \frac{1}{2}+y, \frac{1}{2}-z = 2_1\,[010] \,@\, \frac{1}{4}, y, \frac{1}{4} \\ m\,(010) \otimes a\,(001) & \Rightarrow & \frac{1}{2}+x, \bar{y}, \bar{z} = 2_1\,[100] \,@\, x, 0, 0 \end{array}$$

Finally, we can combine *any* of the planes of reflection with the rotation axis normal to that plane to obtain, for example,

$$m\,(010) \otimes 2_1\,[010] \,@\, \tfrac{1}{4}, y, \tfrac{1}{4} \;\Rightarrow\; \tfrac{1}{2}-x, \tfrac{1}{2}-y, \tfrac{1}{2}-z = \bar{1} \,@\, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}$$

Because the rotation axis normal to one of the planes of reflection is generated by the other two planes of reflection, this result is effectively the combination of all three planes of reflection (which is why it does not matter which combination of reflection plane and normal rotation axis is used).

The centre of symmetry that is generated by the other symmetry operations is located at position 1/4, 1/4, 1/4. It is conventional to use a set of coordinates in which the centre of symmetry is located at the origin, 0, 0, 0. This can be achieved by shifting the positions of the rotation axes and planes of reflection by $-1/4, -1/4, -1/4$, which gives a new set of positions:

$$\begin{array}{llll} x, y, z; & \frac{1}{2}+x, \frac{1}{2}-y, \frac{1}{2}-z; & \bar{x}, \frac{1}{2}+y, \bar{z}; & \frac{1}{2}-x, \bar{y}, \frac{1}{2}+z; \\ \bar{x}, \bar{y}, \bar{z}; & \frac{1}{2}-x, \frac{1}{2}+y, \frac{1}{2}+z; & x, \frac{1}{2}-y, z; & \frac{1}{2}+x, y, \frac{1}{2}-z \end{array}$$

These are called the **general equivalent positions**, and define the complete set of symmetry operators. These operators are now located at

n (100): $x = \frac{1}{4}, x = \frac{3}{4}$

m (010): $y = \frac{1}{4}, y = \frac{3}{4}$

a (001): $z = \frac{1}{4}, z = \frac{3}{4}$

$\bar{1}$: $0,0,0\ \ \frac{1}{2},0,0\ \ 0,\frac{1}{2},0\ \ 0,0,\frac{1}{2}\ \ \frac{1}{2},\frac{1}{2},0\ \ \frac{1}{2},0,\frac{1}{2}\ \ 0,\frac{1}{2},\frac{1}{2}\ \ \frac{1}{2},\frac{1}{2},\frac{1}{2}$

2_1 [100]: $x, \pm\frac{1}{4}, \pm\frac{1}{4}$

2_1 [010]: $0, y, 0\ \ \frac{1}{2}, y, 0\ \ 0, y, \frac{1}{2}\ \ \frac{1}{2}, y, \frac{1}{2}$

2_1 [001]: $\pm\frac{1}{4}, 0, z\ \ \pm\frac{1}{4}, \frac{1}{2}, z$

We are aided in our analysis by noting that we can assign a point group to each space group. In so doing, screw axes play the role of rotation axes, and glide planes play the role of mirror planes. Thus the space group $Pnma$ belongs to the point group mmm. This point group implies the existence of three 2-fold rotation or 2_1 screw axes, as we found above. The point group also implies the existence of the centre of symmetry. The number of symmetrically related sites in the point group will be equal to the number of general equivalent positions in the primitive unit cell of the space group.

To pick up on the final point in the previous section, the symbol $Pnma$ is used rather than the full symbol $P\frac{2_1}{n}\frac{2_1}{m}\frac{2_1}{a}$, just as mmm is used instead of $\frac{2}{m}\frac{2}{m}\frac{2}{m}$ in the point group label. We have seen above that the 2_1 axes and the centre of symmetry are implied by the label $Pmna$.

E.3 General and special positions

The general equivalent positions of the space group $Pnma$ have been given above. The point group symmetry of these positions is simply 1. We can also define a set of special positions in the unit cell of space group $Pnma$ that have higher symmetry. In this space group we can define the following **special equivalent positions**

Symmetry	Special equivalent positions			
m	$x, \frac{1}{4}, z$	$\bar{x}, \frac{1}{4}, \bar{z}$	$\frac{1}{2}-x, \frac{3}{4}, \bar{z}$	$\frac{1}{2}+x, \frac{1}{4}, \frac{1}{2}-z$
$\bar{1}$	$0,0,\frac{1}{2}$	$0,\frac{1}{2},\frac{1}{2}$	$\frac{1}{2},0,0$	$\frac{1}{2},\frac{1}{2},0$
$\bar{1}$	$0,0,0$	$0,\frac{1}{2},0$	$\frac{1}{2},0,\frac{1}{2}$	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$

Although there are eight positions with a centre of symmetry, an atom at the origin will only be replicated by the symmetry at three other positions: four of the symmetry operations will simply replicate the set of four positions. Thus the other group of centres of symmetry constitute another set of special equivalent positions.

E.4 *The International Tables of Crystallography*

The complete sets of general and special equivalent positions for each of the 230 space groups are given in the International Tables of Crystallography, Volume 1. This large volume is an indispensible comprehensive source of information. Although it is not difficult to work out the general equivalent positions for an orthorhombic space group, it is rather harder for some of the higher symmetry lattices. The entries in the International Tables also give diagramatic representations of the symmetry, which can help visualize the symmetry. The entries also give the systematic absences seen in a diffraction experiment due to the symmetry (see Chapter 6).

E.5 Relating general equivalent positions to actual atomic positions

The set of general equivalent positions allow you to generate the entire crystal structure given the fractional coordinates of a small set of atoms, some of which will be on general positions, and some on special positions. A crystal structure of space group *Pnma* is shown in Fig. E.1. Some of the atoms lie on the centres of symmetry, some lie on the mirror planes (in the plane of the diagram), and others lie on general positions. The actions of the various symmetry operations can be seen by close inspection. In order to generate the whole structure, it is necessary to define only the coordinates of a few atoms, together with the lattice parameters ($a = 10.225$ Å, $b = 5.994$ Å, $c = 4.762$ Å):

Mg(1)	0, 0, 0
Mg(2)	0.278, 1/4, 0.990

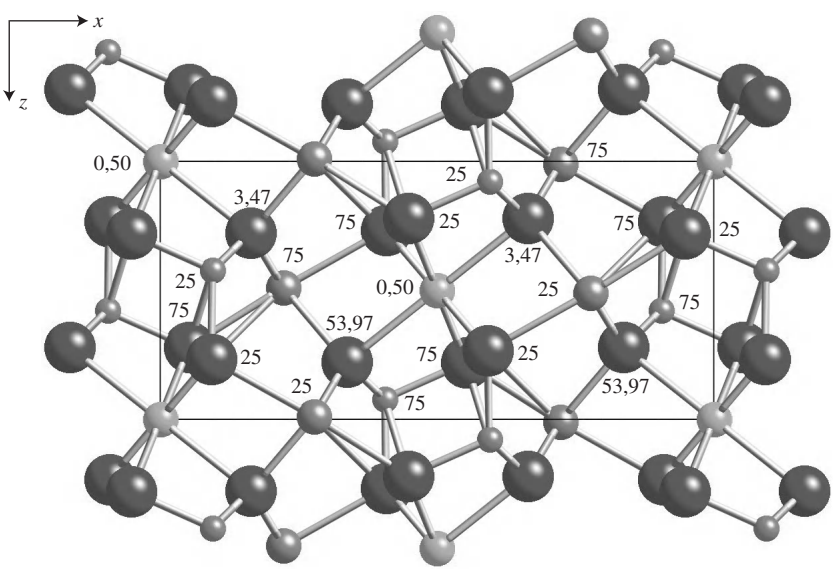


Fig. E.1 Crystal structure of the orthorhombic phase of Mg_2SiO_4 showing the positions of the atoms. The heights of the atoms are given in fractional coordinates multiplied by 100.

Si	0.095, 1/4, 0.423
O(1)	0.092, 1/4, 0.767
O(2)	0.448, 1/4, 0.220
O(3)	0.163, 0.034, 0.278

These coordinates and the operations of the space group are sufficient to generate the positions of all 28 atoms in the unit cell.

Summary of appendix

- Space groups are defined through a combination of lattice type and the rotation or reflection symmetry about each of the crystal axes.
- The combination of symmetry leads to the existence of sets of symmetrically related positions (called the general equivalent positions), from which the overall crystal structure can be constructed.

Further reading

The essential source book is Volume I of the International Tables of Crystallography. Space groups are discussed in many textbooks on crystallography to a greater or lesser extent. Examples are Giacovazzo *et al*. (1992) and Woolfson (1997). A more detailed treatment can be found in Burns and Glazer (1990).

F Lattice energy minimization

A number of methods for minimization of the lattice energy are used in practical computational schemes, most of which are based around the following expansion of the energy:

$$E(\mathbf{p}+\Delta\mathbf{p}) = E(\mathbf{p}) + \mathbf{g}^{\mathrm{T}} \cdot \Delta\mathbf{p} + \frac{1}{2}\Delta\mathbf{p}^{\mathrm{T}} \cdot \mathbf{h} \cdot \Delta\mathbf{p} + \cdots \tag{F.1}$$

where $\mathbf{p}$ is a column matrix of the set of structural parameters $[p_l]$, and $\Delta\mathbf{p}$ is a column matrix of changes in $[p_l]$. $\mathbf{g}$ is a column matrix of the first derivatives of the energy with respect to the structural parameters:

$$g_l = \frac{\partial E}{\partial p_l} \tag{F.2}$$

$\mathbf{h}$ is a square-symmetric matrix of the second differentials of the energy:

$$h_{k,l} = \frac{\partial^2 E}{\partial p_k \partial p_l} \tag{F.3}$$

The energy expansion of eqn F.1 is not usually continued beyond the quadratic term, as is appropriate if the energy is close to its minimum. Now we take the case when $\Delta\mathbf{p}$ gives all the changes in the structure required to reach the equilibrium structure. Differentiation with respect to the set of $\Delta\mathbf{p}$ gives

$$0 = \mathbf{g}^{\mathrm{T}} + \mathbf{h} \cdot \Delta\mathbf{p} \tag{F.4}$$

Rearranging gives

$$\Delta\mathbf{p} = -\mathbf{h}^{-1} \cdot \mathbf{g} \tag{F.5}$$

This equation provides the basis of a numerical algorithm to find the equilibrium structure starting from a given structure and a model for the interatomic potential. Of course, it is not exact because we have neglected terms beyond the second derivative of the energy, but the process can be repeated for a few times because successive applications will bring the structure closer to equilibrium.

Summary of appendix

- An algorithm for the numerical minimization of the lattice energy is described.

Some notes on the variational theorem

G

We consider the general form of the Schrödinger equation for any wave function Ψ:

$$\hat{\mathcal{H}}\Psi = E\Psi \tag{G.1}$$

It will prove helpful to use the following standard shorthand:

$$\langle \phi^* | \hat{\mathcal{H}} | \psi \rangle = \int \phi^* \hat{\mathcal{H}} \psi \, \mathrm{d}\mathbf{r} \tag{G.2}$$

$$\langle \phi^* | \psi \rangle = \int \phi^* \psi \, \mathrm{d}\mathbf{r} \tag{G.3}$$

With this formalism, the energy can be written as

$$E = \frac{\langle \Psi^* | \hat{\mathcal{H}} | \Psi \rangle}{\langle \Psi^* | \Psi \rangle} \tag{G.4}$$

We now consider a wave function that is close to the ground state wave function:

$$\Psi = \Psi_0 + \delta\Psi \tag{G.5}$$

The energy is given as

$$\begin{aligned} E &= \frac{\langle \Psi_0^* + \delta\Psi^* | \hat{\mathcal{H}} | \Psi_0 + \delta\Psi \rangle}{\langle \Psi_0^* + \delta\Psi^* | \Psi_0 + \delta\Psi \rangle} \\ &= \frac{\langle \Psi_0^* | \hat{\mathcal{H}} | \Psi_0 \rangle + 2\langle \delta\Psi^* | \hat{\mathcal{H}} | \Psi_0 \rangle + \langle \delta\Psi^* | \hat{\mathcal{H}} | \delta\Psi \rangle}{\langle \Psi_0^* | \Psi_0 \rangle + 2\langle \delta\Psi^* | \Psi_0 \rangle + \langle \delta\Psi^* | \delta\Psi \rangle} \\ &= \frac{E_0 \langle \Psi_0^* | \Psi_0 \rangle + 2E_0 \langle \delta\Psi^* | \Psi_0 \rangle + \langle \delta\Psi^* | \hat{\mathcal{H}} | \delta\Psi \rangle}{\langle \Psi_0^* | \Psi_0 \rangle + 2\langle \delta\Psi^* | \Psi_0 \rangle + \langle \delta\Psi^* | \delta\Psi \rangle} \\ &= E_0 + O((\delta\Psi)^2) \end{aligned} \tag{G.6}$$

This demonstrates that a change in wave function will lead to a second-order change in the energy. This also shows that the best wave function is that for which the energy is a minimum.

Now we establish how the variational theory also applies when the wave function is expanded as an expansion of basis functions. We define a trial wave function as

$$\Psi = \sum_i a_i \phi_i \tag{G.7}$$

where the functions ϕ_i are true eigenfunctions of $\hat{\mathcal{H}}$ (but their exact forms are not known), and the coefficients a_i are constrained to give overall normalization of Ψ. The energy is therefore given as

$$\begin{aligned} E &= \sum_{ij} a_i a_j \langle \phi_i^* | \hat{\mathcal{H}} | \phi_j \rangle \\ &= \sum_{ij} a_i^2 E_i \end{aligned} \tag{G.8}$$

where we have used the fact that the functions ϕ_i are orthogonal so that terms of the form $\langle \phi_i^* | \hat{\mathcal{H}} | \phi_j \rangle$, where $i \neq j$, are equal to zero. The true ground state energy, E_0, is, by definition, lower than all values of E_i, so we can write

$$\langle \Psi^* | \hat{\mathcal{H}} | \Psi \rangle > \langle \Psi_0^* | \hat{\mathcal{H}} | \Psi_0 \rangle = E_0 \tag{G.9}$$

Thus the trial wave function that best approximates the ground state wave function will be that for which the energy is a minimum.

Summary of appendix

- The **variational principle** provides a method to optimize a trial wave function to give the best approximation of the true wave function by adjusting parameters to give the lowest energy.

Further reading

The variational theory is discussed in some undergraduate-level textbooks, including Rae (1992). Ashcroft and Mermin (1976) have a useful appendix on this topic.

Ewald sphere

H

Diffraction involves an incoming and scattered beam of radiation with wave vectors having the same length, $2\pi/\lambda$, but different directions. Diffraction only occurs when the vector difference between the two wave vectors is equal to a reciprocal lattice vector. The complete diffraction of a crystal involves all relative orientations of the wave vectors of the incident and scattered beam that differ by a reciprocal lattice vector. The locus of all relative orientations of the two wave vectors is the surface of a sphere of radius $2\pi/\lambda$, as shown in Fig. H.1. The sphere is pinned to the origin of reciprocal space, and the relative orientations of the sphere and reciprocal space are allowed to change as the orientation between the crystal and the incident beam is changed. This sphere is called the **Ewald sphere**. Diffraction occurs when the orientation of the Ewald sphere has a point on the surface that passes through the reciprocal lattice point, as shown in Fig. H.1.

The concept of the Ewald sphere can be useful to interpret diffraction processes in a wide range of experimental situations. We consider first the case where a crystal in an X-ray beam is rotated about a single axis. This situation can be intepreted as the rotation of the Ewald sphere in reciprocal space, pinned to the origin and rotating about the same axis about which the crystal rotates. The rotation of the Ewald sphere will sweep across a spherical volume of twice the radius of the Ewald sphere, as shown in Fig. H.2. As the surface of the sphere

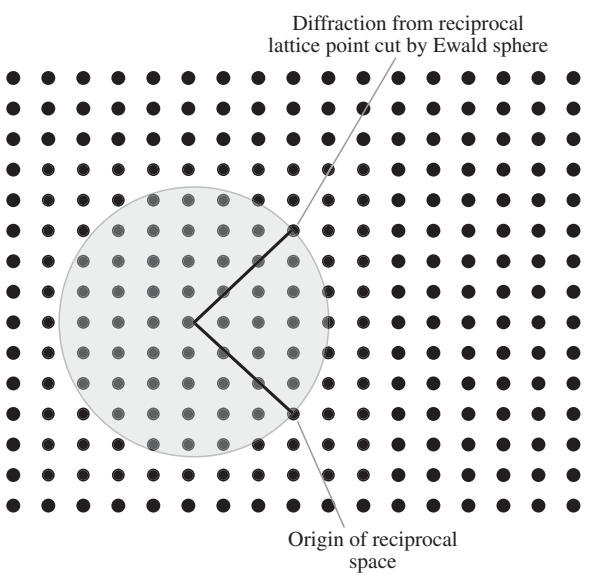


Fig. H.1 Two-dimensional section of reciprocal lattice, showing the circular cut of the Ewald sphere with an orientation for which the surface passes through a reciprocal lattice point. The radius of the Ewald sphere is equal to $2\pi/\lambda$.

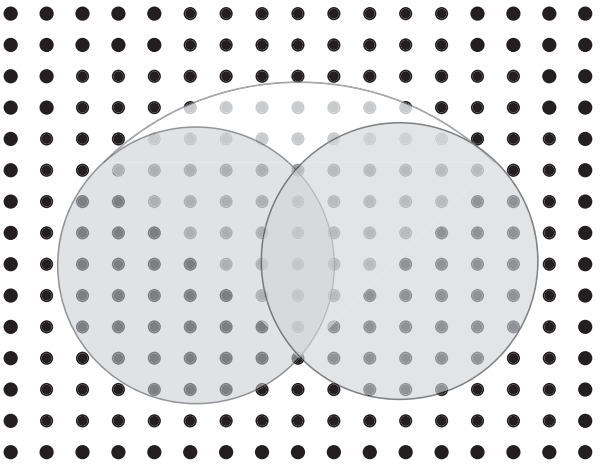


Fig. H.2 Two-dimensional section of reciprocal lattice, showing the sweep of the Ewald sphere as it rotates about the axis normal to the plane of the diagram. The shaded region represents the volume of reciprocal space that has given diffraction during the rotation of the Ewald sphere.

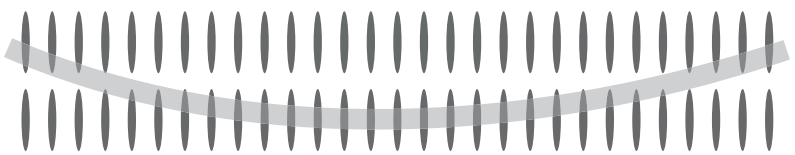


Fig. H.3 Schematic representation of two planes of reciprocal lattice points that are broadened due to the sample being thin, together with the trace of the Ewald sphere for an electron diffraction experiment. The large radius of the Ewald sphere coupled with the broadening of the reciprocal lattice points enables a large area of the diffraction pattern to be measured in a single experiment.

crosses each reciprocal lattice point, diffraction of the corresponding reflection will occur. The effect of rotating the crystal will ensure that the diffraction from a wide range of reflections can be measured.

In a powder diffraction experiment, the sample contains crystallites with (ideally) a uniform distribution of orientations. This is equivalent to the situation with a continuous distribution of orientations of the Ewald sphere in reciprocal space. As a result, all reflections within a sphere of radius that is twice the radius of the Ewald sphere will have a subset of the distribution of Ewald spheres passing through their reciprocal lattice points, and simultaneous diffraction from all reflections will occur.

One of the most important uses of the Ewald sphere is in understanding the origin of diffraction in the transmission electron microscope, where a thin sample is held at a fixed orientation. Because the sample is a thin flake, the reciprocal lattice point is broadened through convolution with the Fourier transform of the function that describes the shape of the crystal flake (Section 6.7.2). The wavelength of the electron beam is much shorter than the dimensions of the unit cell, which means that the radius of the Ewald sphere is much larger that the spacings between reciprocal lattice points. The situation is shown schematically in Fig. H.3. Because the curvature of the Ewald sphere is so slight compared to the widths of the broadened reciprocal lattice points, the surface of the Ewald sphere sweeps through several reciprocal lattice points simultaneously, with no need to rotate the crystal. Furthermore, the sphere also

cuts through the next layers of reciprocal lattice points further away from the origin of reciprocal space.

Summary of appendix

- For a given orientation of a crystal with respect to an incident beam of radiation, the locus of all possible scattering vectors forms a spherical surface in reciprocal space which passes through the origin of reciprocal space. This is the **Ewald sphere**.
- Diffraction occurs when the crystal is oriented so that the surface of the Ewald sphere cuts through a reciprocal lattice point.
- In a powder diffraction experiment, reflections are associated with all reciprocal lattice points that lie within a sphere of twice the radius of the Ewald sphere centred on the origin of reciprocal space.
- In **electron diffraction**, a section of reciprocal space can be measured in a single setting of the crystal orientation because the curvature of the Ewald sphere is shallow enough for its surface to pass through the reciprocal lattice points that are broadened by the use of a thin sample.

Further reading

The Ewald sphere is treated in a number of books on diffraction, including Cullity (2001), Cowley (1992), Giacovazzo *et al*. (1992), and Woolfson (1997).

I The Wilson plot

It is useful in the analysis of diffraction data to try to place the reflection intensities on a normalized scale. We measure a relative structure factor that is related to the real structure factor by the scale factor s:

$$F_{\text{rel}}(hkl) = s|F(hkl)| \tag{I.1}$$

If we assume that the thermal motions are isotropic, we can write

$$F_{\text{rel}}(hkl) = s\exp(-B\sin^2\theta/\lambda^2)|\tilde{F}(hkl)| \tag{I.2}$$

where $\tilde{F}(hkl)$ is the structure factor with the temperature factor taken out. Taking averages of the squares of the structure factors over small ranges of values of $\sin\theta/\lambda$ gives

$$s^2\exp(-2B\sin^2\theta/\lambda^2) = \frac{\langle F_{\text{rel}}^2(hkl)\rangle}{\langle|\tilde{F}(hkl)|^2\rangle} \tag{I.3}$$

We can write

$$\begin{aligned}|\tilde{F}(hkl)|^2 &= \left(\sum_j f_j\exp(i\mathbf{Q}\cdot\mathbf{r}_j)\right)\times\left(\sum_j f_j\exp(-i\mathbf{Q}\cdot\mathbf{r}_j)\right)\\ &= \sum_j f_j^2 + \sum_{j\neq j'} f_j f_{j'}\exp(i\mathbf{Q}\cdot[\mathbf{r}_j-\mathbf{r}_{j'}])\end{aligned} \tag{I.4}$$

Taking the average over all $\mathbf{Q}$ in a small range of $\sin^2\theta/\lambda^2$ causes the second term to average to zero, so we can write

$$\langle|\tilde{F}(hkl)|^2\rangle = \sum_j f_j^2 \tag{I.5}$$

Thus we have

$$s^2\exp(-2B\sin^2\theta/\lambda^2) = \frac{\langle F_{\text{rel}}^2(hkl)\rangle}{\sum_j f_j^2} \tag{I.6}$$

Taking logarithms of both sides gives

$$2\ln s - 2B\sin^2\theta/\lambda^2 = \ln K \tag{I.7}$$

where for shorthand we have written

$$K = \frac{\langle F_{\text{rel}}^2(hkl)\rangle}{\sum_j f_j^2} \tag{I.8}$$

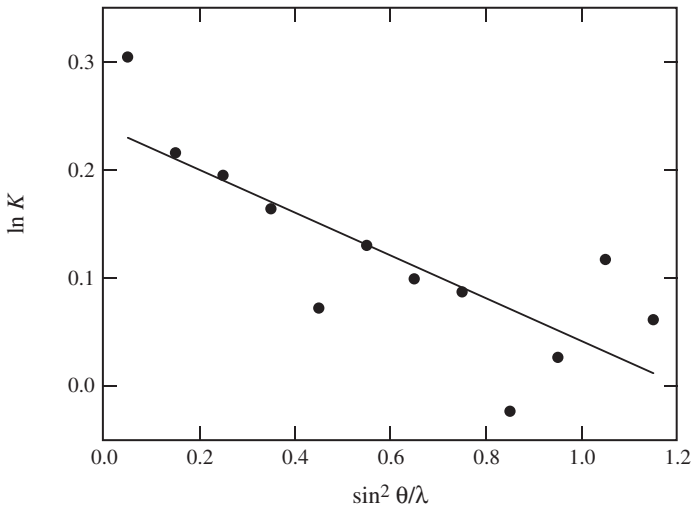


Fig. I.1 Wilson plot for the mineral symesite, $Pb_{10}(SO_4)O_7Cl_4 \cdot H_2O$. (Plot constructed from structure factor data kindly provided by Mark Cooper, University of Manitoba, and Mark Welch, Natural History Museum, London, with the final structure analysis reported in *Am. Min.* **85**, 1526, 2000.)

If the diffraction data are analysed to produce a set of averages K over a number of ranges of values of $\sin^2\theta/\lambda^2$, a plot of $\ln K$ against $\sin^2\theta/\lambda^2$ will give a straight line and will yield approximate values for s and B. This plot is called a **Wilson plot**.

An example of a Wilson plot for the mineral symesite, formula $Pb_{10}(SO_4)O_7Cl_4 \cdot H_2O$, is shown in Fig. J.1. This plot was constructed from over 4000 reflections, and shows the level of accuracy that is obtained from this type of analysis. The value of $2B$ extracted from the plot is 0.20 ± 0.05 Å^2.

Summary of appendix

- The **Wilson plot** is a method to estimate values of the scale and temperature factors from a set of measurements of structure factors.

Further reading

The Wilson plot is discussed in several text books on X-ray crystallography, including Cullity (2001), Giacovazzo *et al.* (1992), Hammond (2001), Ladd and Palmer (1977), and Woolfson (1997).

J

Diffraction from isotropic materials

J.1 Basic diffraction equations

The basic diffraction equation for a process in which a beam of radiation is scattered from a collection of N atoms at positions $\mathbf{r}_j$ with a change in wave vector $\mathbf{Q}$ is

$$S(\mathbf{Q}) = \frac{1}{N} \sum_{j,k} b_j b_k \exp[i\mathbf{Q} \cdot (\mathbf{r}_j - \mathbf{r}_k)] \tag{J.1}$$

where b_j is the scattering factor for atom j. The essential point of this equation is that it shows how diffraction gives direct information about the separation of atoms. In fact, as this equation stands it provides no direct information about the positions of the atoms, only about their separation. In Chapter 6 on diffraction, we were only able to recover information about the positions of atoms by supposing that the atomic separations would be consistent with a periodic arrangement of atoms that were moving as if they were independent oscillators (such motions are consistent with harmonic interactions between atoms).

For a homogeneous isotropic material, such as a glass or liquid, or even a polycrystalline material, there is neither a special origin nor a special orientation of the sample. This means that knowledge of the absolute positions of atoms has little relevance. Instead, the important atomic distributions are the shells of neighbours around each atom. For example, in amorphous silica, SiO_2, the silicon atoms will be surrounded by four oxygen atoms in a tetrahedral arrangement. The next neighbours for silicon atoms will be other silicon atoms that are linked through a Si–O–Si connection. There will then be shells of neighbouring oxygen atoms that are bonded to the linked silicon atoms. For the oxygen atoms, the nearest neighbours are the bonded silicon atoms. The second shell of atoms arise from the other oxygen atoms within the SiO_4 tetrahedra. There are then subsequent shells of silicon and oxygen atoms in neighbouring SiO_4 tetrahedra. The important information concerns the distribution of distances between atoms, not the actual separations of specific atom pairs. Such a distribution will have three components, first the mean separation, second the width of the distribution (that is, the spread of separations), and third the integral of the distribution that gives the number of atom pairs within the distribution.

J.2 Isotropic orientational averages

Because the sample is isotropic, any distribution of atomic bonds will involve all orientations. This means that the terms in the scattering law, namely the

$\exp(i\mathbf{Q}\cdot(\mathbf{r}_j-\mathbf{r}_k))$ components, need to be averaged over all relative orientations of $\mathbf{Q}$ and $\mathbf{r}_j - \mathbf{r}_k$. We write $r_{jk} = |\mathbf{r}_j - \mathbf{r}_k|$, and we denote the angle between $\mathbf{Q}$ and $\mathbf{r}_j - \mathbf{r}_k$ by θ. The average over all orientations follows as

$$\langle \exp(i\mathbf{Q}\cdot(\mathbf{r}_j - \mathbf{r}_k))\rangle = \frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \exp(iQr_{jk}\cos\theta)\sin\theta\,\mathrm{d}\theta \quad \text{(J.2)}$$

$$= \frac{1}{2}\int_0^{\pi} \exp(iQr_{jk}\cos\theta)\sin\theta\,\mathrm{d}\theta \quad \text{(J.3)}$$

To solve the integral, we substitute $x = \cos\theta$, with $\mathrm{d}x = -\sin\theta\,\mathrm{d}\theta$. Thus

$$\int_0^{\pi} \exp(iQr_{jk}\cos\theta)\sin\theta\,\mathrm{d}\theta = -\int_{+1}^{-1} \exp(iQr_{jk}x)\,\mathrm{d}x$$

$$= 2\frac{\sin(Qr_{jk})}{Qr_{jk}} \quad \text{(J.4)}$$

Thus we can write

$$\langle \exp(i\mathbf{Q}\cdot(\mathbf{r}_j - \mathbf{r}_k))\rangle = \frac{\sin(Qr_{jk})}{Qr_{jk}} \quad \text{(J.5)}$$

J.3 Pair distribution functions

In performing the isotropic average, we (correctly) remove the dependence on the specific bonds. However, the scattering law involves a sum over all atom pairs, which now has to be replaced by the appropriate distribution functions. We define the number density ρ_j to give the number of atoms of type j per unit volume, and c_j as the proportion of type j. We take one atom of type j, and now represent the number of atoms of type k lying within a spherical shell of radius r and thickness $\mathrm{d}r$ about atom j to be equal to $4\pi r^2\rho_k g_{jk}(r)\,\mathrm{d}r$. $g(r)$ is the pair distribution function. By this definition, we expect $g_{jk}(r)$ to be equal to 0 for values of r_{jk} lower than the smallest possible interatomic contact distance. Moreover, for large values of r we expect there to be no correlations in the positions of atoms j and k, so that the number of atoms of type j within the shell is equal to $4\pi r^2\rho_k\,\mathrm{d}r$. Thus $g_{jk}(r) \to 1$ in the limit $r \to \infty$. We also expect $g_{jk}(r)$ to have a peak around values of r corresponding to a specific short-range contact (or bond) distance. If we have a bond with a spread of distances between $R \pm \delta R$, the coordination number will be equal to

$$n_{jk} = 4\pi\rho_k \int_{R-\delta R}^{R+\delta R} r^2 g_{jk}(r)\,\mathrm{d}r \quad \text{(J.6)}$$

For example, if atom j is silicon and atom k is oxygen, there will be a peak in $g_{\text{Si–O}}(r)$ at around $R = 1.61$ Å, with $n_{\text{Si–O}} = 4$.

Now we can consider the form of the scattering law. We separate the contribution from atoms with zero separation, and write

$$S(Q) = F(Q) + \sum_i c_i b_i^2 \quad \text{(J.7)}$$

All information about pairs of atoms is contained in the function $F(Q)$. Using the formalism of the pair distribution functions, we can write

$$F(Q) = \rho \sum_{j,k} c_j c_k b_j b_k 4\pi \int_0^\infty r^2 g_{jk}(r) \frac{\sin(Qr)}{Qr} \, \mathrm{d}r \tag{J.8}$$

where we have replaced ρ_j by $c_j \rho$, with ρ representing the total number density (i.e. the number of atoms of all types per unit volume). It becomes convenient to then separate $F(Q)$ into two parts:

$$\begin{aligned} F(Q) = &\sum_{j,k} c_j c_k b_j b_k 4\pi \int_0^\infty r^2 [g_{jk}(r) - 1] \frac{\sin(Qr)}{Qr} \, \mathrm{d}r \\ &+ \sum_{j,k} c_j c_k b_j b_k \int_0^\infty 4\pi r^2 \frac{\sin(Qr)}{Qr} \, \mathrm{d}r \end{aligned} \tag{J.9}$$

The second integral is only non-zero if $Q = 0$, which is experimentally inaccessible. If we write

$$G(r) = \sum_{j,k} c_j c_k b_j b_k [g_{jk}(r) - 1] \tag{J.10}$$

the scattering law becomes

$$S(Q) = \rho_0 \int_0^\infty 4\pi r^2 G(r) \frac{\sin(Qr)}{Qr} \, \mathrm{d}r + \sum_j c_j b_j^2 \tag{J.11}$$

J.4 Reverse Fourier transform

We can define the reverse Fourier transform:

$$G(r) = \frac{1}{(2\pi)^3 \rho} \int_0^\infty 4\pi Q^2 F(Q) \frac{\sin(Qr)}{Qr} \, \mathrm{d}Q \tag{J.12}$$

When comparing eqns 3.11 and 3.12, it can be seen that the quantities related by the Fourier transforms are $QF(Q)$ and $rG(r)$. For this reason, the pair distribution functions are often presented in the form

$$D(r) = 4\pi\rho r G(r) \tag{J.13}$$

In this spirit, we can define partial pair distribution functions

$$d_{jk}(r) = r[g_{jk}(r) - 1] \tag{J.14}$$

$$\Rightarrow D(r) = 4\pi\rho \sum_{jk} c_j c_k b_j b_k d_{jk}(r) \tag{J.15}$$

From the equations above, $G(r)$ is equal to $-(\sum c_j b_j)^2$ as $r \to 0$. Thus $D(r)$ has a negative linear form at small r. As a result, it is common to also use the function

$$T(r) = D(r) + 4\pi\rho r \left(\sum c_j b_j \right)^2 \tag{J.16}$$

This has the property that $T(r) = 0$ for values of r lower than the shortest interatomic contact distances, and tends towards a linear function at large r.

The forms of $T(r)$, $D(r)$ and $G(r)$ are compared in Fig. J.1.

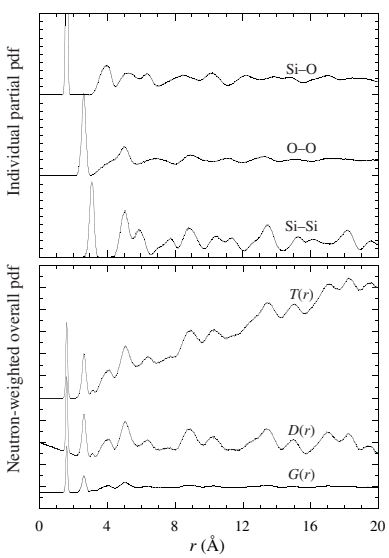


Fig. J.1 Individual pair distribution functions $g(r)$ for Si–O, O–O and Si–Si interatomic distances in the high-temperature disordered phase of cristobalite (Section 2.3). These functions are assembled with the neutron scattering lengths to give the functions $T(r)$ (eqn J.16), $D(r)$ (eqn J.15) and $G(r)$ (eqn J.12) for easy comparison of the forms of these three functions.

J.5 General approach to analysis of diffraction data

From the above equations, the important quantity to extract from experimental data is $QF(Q)$. The quantity measured is $S(Q)$, and the difference between $F(Q)$ and $S(Q)$ is $\sum c_j b_j^2$. This difference is an absolute number, and for the conversion to work it is necessary to be able to remove all background contributions to the experimental measurement of $S(Q)$ and then normalize $S(Q)$ onto an absolute scale. $F(Q)$ should tend towards zero at large Q, which provides a check on the corrections applied to $S(Q)$. If the corrections are not perfect, the form of $QF(Q)$ will veer away from oscillations about zero as Q increases.

Summary of appendix

- In diffraction from isotropic materials, the scattering function needs to be averaged over all possible orientations of scattering vector and interatomic vector. This average leads to the dependence of scattering on $\sin(Qr)/Qr$.
- Diffraction from isotropic materials can be analysed in terms of **pair distribution functions**. The pair distribution functions can be obtained as the Fourier transform of the scattering function.

Further reading

The basic approach of this appendix is described in only a few textbooks, including Bacon (1973), Elliott (1990), Nield and Keen (2001), and Squires (1978). The reader should be aware that the nomenclature varies between different authors, often using the same symbols for different quantities, and this can be very confusing. Keen (*Journal of Applied Crystallography*, **34**, 172–177, 2001) has compared the different presentations of the formalism, and this paper is a useful companion to further reading.

K Calculation of physical properties

K.1 Expansion of the crystal energy

Macroscopic physical properties, such as the piezoelectric and elastic constants discussed in Chapter 7, are related to interatomic interactions. In turn, the interatomic potentials can be represented by analytical parametrized functions. The link between the macroscopic physical properties and the microscopic interatomic forces is obtained by writing the lattice energy in terms of small strains and small displacements of atoms, taking account of correlations between these different parameters. Formally, we define a 1×6 column matrix $\boldsymbol{\epsilon}$ that contains the six strain components, and a $1 \times 3N$ matrix $\mathbf{u}$ containing the atomic displacements for a crystal containing N atoms in the unit cell (this can be generalized when taking account of ionic polarizability by a shell model, Chapter 5). The atomic displacement matrix contains the x, y, z components, the transpose being given as

$$\mathbf{u}^{\mathrm{T}} = (u_x(1), u_y(1), u_z(1), u_x(2), \ldots, u_z(N)) \tag{K.1}$$

The energy of the crystal can be written as an expansion about the equilibrium structure:

$$E = E_0 + \frac{1}{2}\mathbf{u}^{\mathrm{T}} \cdot \mathbf{W}_{\mathbf{uu}} \cdot \mathbf{u} + \mathbf{u}^{\mathrm{T}} \cdot \mathbf{W}_{\mathbf{u}\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}} \cdot \mathbf{W}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} \tag{K.2}$$

where E_0 is the equilibrium energy of the crystal. The three $\mathbf{W}$ matrices represent the second derivatives of the energy, containing terms of the form:

$$W_{\mathbf{uu}} = \frac{\partial^2 E}{\partial u_\alpha(j) u_{\alpha'}(j')} \tag{K.3}$$

$$W_{\mathbf{u}\boldsymbol{\epsilon}} = \frac{\partial^2 E}{\partial u_\alpha(j) \partial \epsilon_{\alpha'}} \tag{K.4}$$

$$W_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} = \frac{\partial^2 E}{\partial \epsilon_\alpha \partial \epsilon_{\alpha'}} \tag{K.5}$$

K.2 Equilibrium condition and the elastic constant tensor

To provide the link to physical properties, we use the equilibrium condition

$$\frac{\partial E}{\partial \mathbf{u}} = 0 \tag{K.6}$$

Thus we can write

$$\mathbf{W_{uu}} \cdot \mathbf{u} + \mathbf{W_{u\epsilon}} \cdot \boldsymbol{\epsilon} = 0 \tag{K.7}$$

so that

$$\mathbf{u} = -\mathbf{W_{uu}^{-1}} \cdot \mathbf{W_{u\epsilon}} \cdot \boldsymbol{\epsilon} \tag{K.8}$$

This is substituted back into the equation for the energy to give

$$\begin{aligned} E &= E_0 - \frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}} \cdot \mathbf{W_{\epsilon u}} \cdot \mathbf{W_{uu}^{-1}} \cdot \mathbf{W_{u\epsilon}} \cdot \boldsymbol{\epsilon} + \frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}} \cdot \mathbf{W_{\epsilon\epsilon}} \cdot \boldsymbol{\epsilon} \\ &= E_0 + \frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}} \cdot \left[\mathbf{W_{\epsilon\epsilon}} - \mathbf{W_{\epsilon u}} \cdot \mathbf{W_{uu}^{-1}} \cdot \mathbf{W_{u\epsilon}}\right] \cdot \boldsymbol{\epsilon} \end{aligned} \tag{K.9}$$

By comparison with the equation for the elastic energy of a crystal:

$$\frac{E - E_0}{V} = \frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}} \cdot \mathbf{c} \cdot \boldsymbol{\epsilon} \tag{K.10}$$

we obtain an equation for the elastic constant tensor $\mathbf{c}$:

$$\mathbf{c} = \frac{1}{V_{\mathrm{c}}}\left[\mathbf{W_{\epsilon\epsilon}} - \mathbf{W_{\epsilon u}} \cdot \mathbf{W_{uu}^{-1}} \cdot \mathbf{W_{u\epsilon}}\right] \tag{K.11}$$

where V_{c} is the volume of the unit cell (provided that, as is usual, the energy and its derivatives are calculated as per unit cell).

The bulk modulus can be obtained directly from the elastic constant tensor. The **elastic compliance** tensor is the inverse of the elastic constant tensor, $\mathbf{s} = \mathbf{c}^{-1}$. The bulk modulus is obtained from $\mathbf{s}$ by

$$K = \frac{1}{s_{11} + s_{22} + s_{33} + 2(s_{12} + s_{23} + s_{31})} \tag{K.12}$$

K.3 Piezoelectric and dielectric tensors

To extend the calculations to properties including the electric field (dielectric constants, piezoelectric coefficients) we extend the energy to include the interaction with the electric field. We first need to define a charge matrix

$$\mathbf{Q} = \begin{pmatrix} Q_1 & \cdot & \cdot \\ \cdot & Q_1 & \cdot \\ \cdot & \cdot & Q_1 \\ Q_2 & \cdot & \cdot \\ & \vdots & \\ \cdot & \cdot & Q_N \end{pmatrix} \tag{K.13}$$

The energy of the crystal can be expanded to

$$E = E_0 + \frac{1}{2}\mathbf{u}^{\mathrm{T}} \cdot \mathbf{W_{uu}} \cdot \mathbf{u} + \mathbf{u}^{\mathrm{T}} \cdot \mathbf{W_{u\epsilon}} \cdot \boldsymbol{\epsilon} + \frac{1}{2}\boldsymbol{\epsilon}^{\mathrm{T}} \cdot \mathbf{W_{\epsilon\epsilon}} \cdot \boldsymbol{\epsilon} - \mathbf{u}^{\mathrm{T}} \cdot \mathbf{Q} \cdot \mathbf{E} \tag{K.14}$$

where $\mathbf{E}$ is the 1×3 column matrix containing the electric field vector. The new equilibrium condition gives

$$\mathbf{W_{uu}} \cdot \mathbf{u} + \mathbf{W_{u\epsilon}} \cdot \boldsymbol{\epsilon} - \mathbf{Q} \cdot \mathbf{E} = 0 \tag{K.15}$$

and hence

$$\mathbf{u} = -\mathbf{W_{uu}^{-1}} \cdot \mathbf{W_{u\epsilon}} \cdot \boldsymbol{\epsilon} + \mathbf{W_{uu}^{-1}} \cdot \mathbf{Q} \cdot \mathbf{E} \tag{K.16}$$

We note that the dielectric polarization field $\mathbf{P}$ can now be written in the form

$$\begin{aligned} \mathbf{P} &= \frac{4\pi}{V_c} \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{u} \\ &= \frac{4\pi}{V_c} \mathbf{Q}^{\mathrm{T}} \cdot \Big[-\mathbf{W_{uu}^{-1}} \cdot \mathbf{W_{u\epsilon}} \cdot \boldsymbol{\epsilon} + \mathbf{W_{uu}^{-1}} \cdot \mathbf{Q} \cdot \mathbf{E} \Big] \end{aligned} \tag{K.17}$$

We can identify this equation with the general form

$$\mathbf{P} = \mathbf{d} \cdot \boldsymbol{\epsilon} + \boldsymbol{\chi} \cdot \mathbf{E} \tag{K.18}$$

where we now have equations for the piezoelectric coefficent tensor $\mathbf{d}$:

$$\mathbf{d} = -\frac{4\pi}{V_c} \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{W_{uu}^{-1}} \cdot \mathbf{W_{u\epsilon}} \tag{K.19}$$

and the susceptibility tensor $\boldsymbol{\chi}$:

$$\boldsymbol{\chi} = \frac{4\pi}{V_c} \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{W_{uu}^{-1}} \cdot \mathbf{Q} \tag{K.20}$$

To obtain the dielectric constant tensor, it is necessary to add the unit matrix to the susceptibility tensor. To obtain the high-frequency dielectric constant, the matrix of second derivatives, $\mathbf{W_{uu}^{-1}}$, must be calculated using only the atomic shells. Without the shell model, the high-frequency dielectric constant will be unity.

Summary of appendix

- Expressions for the tensors of **dielectric constants**, **piezoelectric coefficients**, and **elastic constants** are derived from the derivatives of the energy with respect to strain and atomic displacements. The expressions include atomic polarizability through the shell model.

Further reading

The main source for this appendix has been *Catlow and MacKrodt* (1982). The programming of these formulae is described by *Gale* (1997) *Journal of the Chemical Society – Faraday Transactions*, **93**, 629–639.

Partition function: some key results

L

L.1 The definition and use of the partition function

Consider a system with different energy levels E_j. The probability of the j-th state existing is P_j:

$$P_j = \frac{\exp(-\beta E_j)}{\sum_i \exp(-\beta E_i)} \tag{L.1}$$

where $\beta = 1/k_B T$. The denominator is called the **Partition function**:

$$Z = \sum_i \exp(-\beta E_i) \tag{L.2}$$

An example of the use of the partition function is given by the calculation of the mean energy of a system, which is equivalent to the **internal energy**. If there are possible states with different energies, we calculate the mean energy by summing over all states and weighting by their relative probabilities:

$$\langle E \rangle = U = \frac{\sum_i E_i \exp(-\beta E_i)}{\sum_i \exp(-\beta E_i)} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial \ln Z}{\partial \beta} \tag{L.3}$$

L.2 The free energy

The free energy can be obtained from the partition function through

$$F = -\frac{1}{\beta} \ln Z \tag{L.4}$$

We justify this result by a self-consistency argument. The free energy is related to the internal energy and entropy via

$$F = U - TS = U - \frac{1}{k_B \beta} S \tag{L.5}$$

The entropy is obtained from the free energy by

$$\begin{aligned} S &= -\frac{\partial F}{\partial T} = k_B \beta^2 \frac{\partial F}{\partial \beta} \\ &= k_B \ln Z - \frac{k_B \beta}{Z}\frac{\partial Z}{\partial \beta} \end{aligned} \tag{L.6}$$

When combining the equations for the internal energy and entropy we obtain

$$U - \frac{1}{k_B\beta}S = \left(-\frac{1}{Z}\frac{\partial Z}{\partial \beta}\right) - \frac{1}{k_B\beta}\left(k_B \ln Z - \frac{k_B\beta}{Z}\frac{\partial Z}{\partial \beta}\right) = -\frac{1}{\beta}\ln Z \quad \text{(L.7)}$$

This is identical to the original proposition for the free energy, eqn L.4.

L.3 Some results

L.3.1 Heat capacity

The heat capacity is given by

$$\begin{aligned} C_V &= \frac{\partial U}{\partial T} = -k_B\beta^2\frac{\partial U}{\partial \beta} \\ &= -\frac{k_B\beta^2}{Z^2}\left(\frac{\partial Z}{\partial \beta}\right)^2 + \frac{k_B\beta^2}{Z}\left(\frac{\partial^2 Z}{\partial \beta^2}\right) \end{aligned} \quad \text{(L.8)}$$

The first term is clearly related to $\langle E\rangle^2$ from eqn L.3. The second term can be related to $\langle E^2\rangle$, which is obtained from the partition function by

$$\langle E^2\rangle = \frac{1}{Z}\sum_i E_i^2 \exp(-\beta E_i) = \frac{1}{Z}\frac{\partial^2 Z}{\partial \beta^2} \quad \text{(L.9)}$$

Equation L.8 therefore reduces to

$$C_V = k_B\beta^2\left(\langle E^2\rangle - \langle E\rangle^2\right) \quad \text{(L.10)}$$

L.3.2 Susceptibility

If a crystal is subject to a field H, there will be a response Q. For example, the applied field can be a magnetic field, and the response will be the magnetization. Or, with reference to Chapter 7, if the applied field is an electric field, the response will be the dielectric polarization. The susceptibility is defined as

$$\chi = \frac{\partial Q}{\partial H} \quad \text{(L.11)}$$

For a given state of the system, j, the applied field will generate a response Q_j. The energy associated with the response is $-HQ_j$. The partition function is therefore given as

$$Z = \sum_j \exp(\beta H Q_j) \quad \text{(L.12)}$$

It follows that

$$\langle Q\rangle = \frac{1}{\beta Z}\frac{\partial Z}{\partial H} \quad \text{(L.13)}$$

$$\langle Q^2\rangle = \frac{1}{\beta^2 Z}\frac{\partial^2 Z}{\partial H^2} \quad \text{(L.14)}$$

The susceptibility is given from eqn L.11 as

$$\chi = \frac{\partial \langle Q \rangle}{\partial H} = \frac{1}{\beta Z} \frac{\partial^2 Z}{\partial H^2} - \frac{1}{\beta^2 Z} \left(\frac{\partial Z}{\partial H} \right)^2$$

Hence it follows from eqns L.13 and L.14 that the susceptibility can be written as

$$\chi = \beta \left(\langle Q^2 \rangle - \langle Q \rangle^2 \right) \tag{L.15}$$

Summary of appendix

- The **partition function** is defined, and the **free energy** is defined in terms of the partition function.
- The formalism of the partition function is used to obtain expressions for the **heat capacity** and **susceptibility** in terms of **fluctuations** of energy and order parameter respectively.

Further reading

The partition function is discussed in many textbooks on thermodynamics and statistical mechanics, because of its central role in the subject. Yeomans (1992) gives a good account in relationship to the subject of phase transitions.

M Lattice sums

A number of results in the study of crystalline materials rely on sums over a lattice, whether the crystal lattice or reciprocal lattice. The lattice component of the sum can simplify the results. The two fundamental results are

$$\frac{1}{N}\sum_{\mathbf{R}} \exp(i\mathbf{k}\cdot\mathbf{R}) = \delta_{\mathbf{k},0} \tag{M.1}$$

where $\mathbf{k}$ is a wave vector and $\mathbf{R}$ is a lattice vector, and

$$\frac{1}{N}\sum_{\mathbf{G}} \exp(i\mathbf{G}\cdot\mathbf{r}) = \delta_{\mathbf{r},0} \tag{M.2}$$

where $\mathbf{r}$ is a vector in real space, and $\mathbf{G}$ is a reciprocal lattice vector.

These two results are plausible from the following reasoning. Clearly, when $\mathbf{k} = 0$ in eqn M.1 and $\mathbf{r} = 0$ in eqn M.2, the sum is over N values of 1, and the final result when normalized by $1/N$ is 1. However, when $\mathbf{k} \neq 0$ or $\mathbf{r} \neq 0$, the complex exponentials can take on all values spanned by $\pm 1 \pm i$, positive and negative, and the sum will converge to zero.

Ashcroft and Mermin (1978) point out that these results are consistent with the fact that in a periodic system the summations should be independent of the definition of the origin. That is, when a small arbitrary vector $\delta\mathbf{R}$ is added to each lattice point $\mathbf{R}$ in eqn M.1, or $\delta\mathbf{G}$ to each reciprocal lattice point $\mathbf{G}$ in eqn M.2, we expect the value of the summation to be unchanged:

$$\exp(i\mathbf{k}\cdot\delta\mathbf{R})\sum_{\mathbf{R}} \exp(i\mathbf{k}\cdot\mathbf{R}) = \sum_{\mathbf{R}} \exp(i\mathbf{k}\cdot\mathbf{R}) \tag{M.3}$$

$$\exp(i\delta\mathbf{G}\cdot\mathbf{r})\sum_{\mathbf{G}} \exp(i\mathbf{G}\cdot\mathbf{r}) = \sum_{\mathbf{G}} \exp(i\mathbf{G}\cdot\mathbf{r}) \tag{M.4}$$

The only solutions to these equations are eqns M.1 and M.2 respectively.

Summary of appendix

- General results for summations over the real or reciprocal lattices are obtained.

Further reading

Ashcroft and Mermin (1976) have a useful appendix on these mathematical results.

Mean-square atomic displacement and temperature factors

N

Consider a single wave (wave vector $\mathbf{k}$, angular frequency ω) acting on an atom of mass m, causing a displacement $\mathbf{u}$:

$$\mathbf{u} = \mathbf{u}_0 \exp(i[\mathbf{k} \cdot \mathbf{r} - \omega t]) \tag{N.1}$$

The kinetic energy of the atom is given by the time derivative:

$$\frac{1}{2} m \dot{u}^2 = \frac{1}{2} m \omega^2 u^2 \tag{N.2}$$

For a harmonic oscillator, the average kinetic energy of a wave is equal to half the total energy (Section 9.3), so that

$$m\langle \dot{u}^2 \rangle = m\omega^2 \langle u^2 \rangle = \hbar\omega \left(\frac{1}{2} + n(\omega, T) \right) \tag{N.3}$$

For simplicity, we make the approximation that each wave has the same average frequency ω. For a crystal containing N atoms, the energy of a wave will be shared equally between each atom. But there are $3N$ vibrations, so the mean-square displacement of a single atom is

$$\langle u^2 \rangle = \frac{3\hbar}{m\omega} \left(\frac{1}{2} + n(\omega, T) \right) \simeq \frac{3k_\mathrm{B}T}{m\omega^2} \tag{N.4}$$

in the high-temperature limit. The mean-square displacement for an monatomic crystal with atomic mass 40 amu and average frequency 8 THz is shown in Fig. N.1. Note that the mean-square displacement is proportional to temperature above 200 K.

A more sophisticated analysis would take account of all phonons explicitly, giving rise to a sum over all normal modes (we have effectively worked within the Einstein approximation). However, the central result represented by the form of the temperature dependence of $\langle u^2 \rangle$ shown in Fig. N.1 is recovered in a more exact treatment. In a full treatment, the mean-square atomic displacement is allowed to be different along different directions – the complete representation of the three-dimensional atomic displacement is called the anisotropic temperature factor.

As noted in Chapter 6, the mean-squared atomic displacement of each atom can be determined from analysis of diffraction data. In the refinement of a crystal structure (Section 6.9.4), the values of the anisotropic temperature factors can be varied with the atomic coordinates to give the best agreement of the calculated structure factors with the measured structure factors. This analysis is mathematically straightforward, but in practice there are complicating factors.

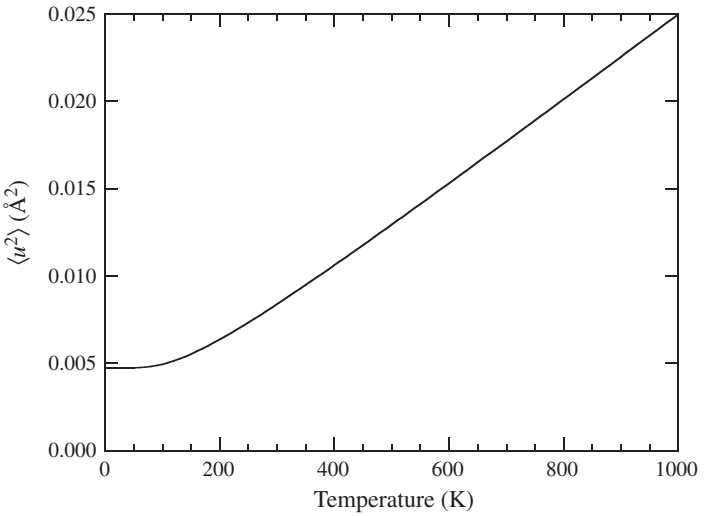


Fig. N.1 Mean-squared displacement for the crystal described in the text.

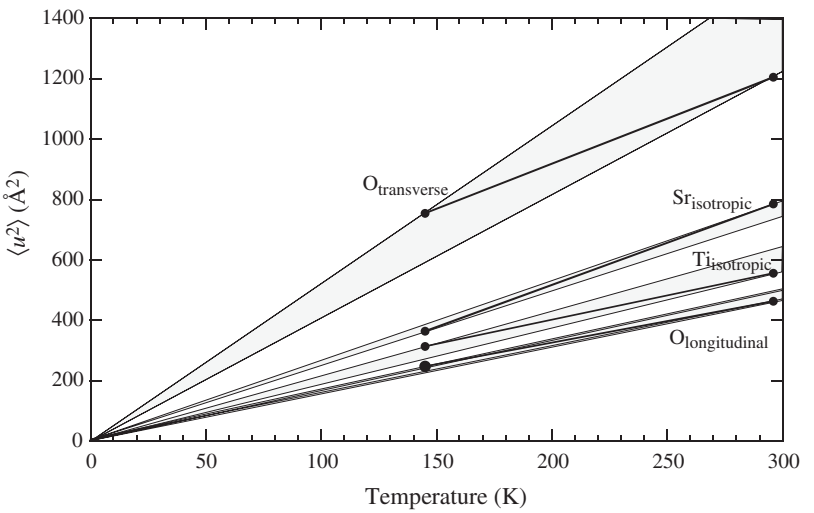


Fig. N.2 Mean-squared displacements, $\langle u^2 \rangle$, for the atoms in $SrTiO_3$ measured at two temperatures. The connecting lines are given as guides to the eye, and the shaded regions encompass lines that pass through the data and the origin, indicating the level of consistency with $\langle u^2 \rangle \propto T$. (Data taken from Abramov *et al.*, *Acta Cryst.*, B **51**, 947, 1995.)

The refined values of the temperature factors are very susceptible to effects of systematic errors in the data, such as the failure to properly account for absorption of the X-ray beam by the sample. Moreover, the precision of the values of temperature factors is frequently very low.

An example of a good determination of the temperature factors is a recent study of the thermal motion of the three types of atoms in $SrTiO_3$. These have been determined at two temperatures from X-ray diffraction, and the results are shown in Fig. N.2. The following points should be noted:

- The individual values of $\langle u^2 \rangle$ are close to varying linearly with temperature as in eqn N.4.
- The limits of the precision of the values of $\langle u^2 \rangle$ are highlighted in Fig. N.2.
- For the oxygen atoms, the value of $\langle u^2 \rangle$ for motions perpendicular to the Ti–O–Ti vector are larger than the motions along this vector. This

implies that the predominant overall motions determining the motions of individual oxygen atoms are rotations of the TiO_6 octahedra, which are the motions associated with the phase transition (Chapter 12).

Summary of appendix

- The mean-square amplitude of the atomic motion is proportional to temperature and inversely proportional to the mean square phonon frequency in the high-temperature limit.
- The mean-square atomic displacement is an anisotropic quantity.

Further reading

The standard text on the treatment of the temperature factor is Willis and Pryor (1967). Dove (1993) provides an alternative discussion of the role of lattice dynamics in determining the origin of the temperature factor.

Solutions to exercises

Chapter 2

(2.1) (a) $\sqrt{3/8}l$; (b) $l/\sqrt{3}$; (c) $\sqrt{2/3}l$.

(2.2) c is equal to twice the distance from a centre of a face of a tetrahedron to the opposite apex. Using the nomenclature in the previous question, l is equivalent to twice the atomic radius. Hence we have $c = 2 \times \sqrt{2/3}l = 4\sqrt{2/3}r$.

(2.4) For the simple cubic structure with one atom per unit cell, the volume within the atom is $4\pi r^3/3$, and the volume of the unit cell is $8r^3$ since $a = 2r$. The packing efficiency is the ratio of the two volumes, namely $4\pi r^3/3 \times 1/8r^3 = \pi/6$. For the diamond structure, the volume contained within the eight atoms in the unit cell is $32\pi r^3/3$. The unit cell parameter is $a = 8r/\sqrt{3}$, giving the volume of the unit cell as $512r^3/3\sqrt{3}$. The packing efficiency is therefore equal to $32\pi r^3/3 \times 3\sqrt{3}/512r^3 = \sqrt{3}\pi/16 = 0.24$.

(2.5) P lattice with centre of symmetry at the origin. Mirror planes normal to cube axes passing through the origin. 2-fold rotation axes parallel to the cube axes passing through the origin. 3-fold axes along the cube diagonals. 4_2 axes parallel to the cube axes interleaved with the 2-fold axes. W(1) has 12 neighbouring atoms separated by 2.84 Å.

(2.6) For the NaCl structure, the volume contained within the ions is equal to $4 \times 4\pi(r_+^3 + r_-^3)/3$, and the volume of the unit cell is equal to $8(r_+ + r_-)^3$. The packing efficiency is therefore $2\pi(r_+^3 + r_-^3)/3(r_+ + r_-)^3$. For the CsCl structure, the volume contained within the ions is equal to $4\pi(r_+^3 + r_-^3)/3$, and the volume of the unit cell is equal to $8(r_+ + r_-)^3/3\sqrt{3}$. The packing efficiency is therefore $\sqrt{3}\pi(r_+^3 + r_-^3)/2(r_+ + r_-)^3$. The ratio of the two packing efficiencies is $4/3\sqrt{3}$. In the limit where the sizes of cation and anion are equal, we expect that the ratio of the packing efficiencies should be equal to the ratio of the packing efficiencies for the simple cubic and bcc structures, which is equal to $\pi/6 \times 8/\sqrt{3}\pi = 4/3\sqrt{3}$.

(2.7) For the cubic ZnS structure, $\sqrt{3}a = 4(r_+ + r_-)$. For the stated ionic radii, $a = 2.31$ Å.

(2.8) SiO_4 and AlO_4 tetrahedra, and AlO_6 octahedra.

(2.9) Writing the distance between the centre and corners of the octahedron as r, for displacements δr towards one of the anions, the first-order change in the electrostatic energy from the interaction with a pair of anions on

opposite corners is proportional to

$$-\frac{1}{r(1+\delta)} - \frac{1}{r(1-\delta)} \simeq -\frac{2}{r(1-\delta^2)}$$

The point $\delta = 0$ is a maximum.

(2.10) Following the discussion in Section 2.4.1, we can form the following table

A cation	r_A (Å)	$\sqrt{2}(r_A + r_O)$	B cation	r_B (Å)	$2(r_B + r_O)$
Li^+	1.06	3.27	Vr^{5+}	0.68	3.86
Na^+	1.32	3.63	Nb^{5+}	0.78	4.06
K^+	1.65	4.10	Ta^{5+}	0.78	4.06
Rb^+	1.75	4.24			
Cs^+	1.88	4.43			

If the stability requires close packing of the BO_6 octahedra with the A cation more loosely bound, the niobates and tantalates will be stable with the Li, Na and K alkali cations, and the vanadates will be stable for only the La and Na alkali cations.

(2.11) From a plan, it can be seen that $a = (2+\sqrt{2})l$ and $c = \sqrt{2}l$, giving $a/c = 1 + 1/\sqrt{2} = 1.707$. The actual lattice parameters give $a/c = 1.55$. The ratio a/c can be reduced when the closest oxygen atoms within planes parallel to (001) move together to reduce the O–Ti–O angle (the experimental angle is 81° rather than 90°).

(2.12) The packing efficiencies are

$$\text{Fluorite} \qquad \frac{\sqrt{3}\pi}{4}\frac{r_+^3 + 2r_-^3}{(r_+ + r_-)^3}$$

$$\text{Rutile} \qquad \frac{8\pi}{3(4+3\sqrt{2})}\frac{r_+^3 + 2r_-^3}{(r_+ + r_-)^3}$$

$$\text{Cristobalite} \qquad \frac{\sqrt{3}\pi}{16}\frac{r_+^3 + 2r_-^3}{(r_+ + r_-)^3}$$

The coordination number of the cation in the fluorite, rutile and cristobalite structure is 8, 6 and 4 respectively, and the anions have coordination numbers 4, 3 and 2 respectively. The density is higher for higher coordination numbers.

Chapter 3

(3.7) n-glides parallel to (100) and (010) at positions $x = \pm\frac{1}{4}$ and $y = \pm\frac{1}{4}$ respectively. Mirror plane parallel to (001) at $z = 0, \frac{1}{2}$. Centre of symmetry at origin. $2_1 \parallel x$ passing through $0, \pm\frac{1}{4}, \pm\frac{1}{4}$; $2_1 \parallel y$ passing through $\pm\frac{1}{4}, 0, \pm\frac{1}{4}$; $2 \parallel z$ passing through origin. $Pccm$.

(3.8) (a) $x, y, z; \overline{y}, x-y, z; \overline{x}+y, \overline{x}, z$.

(b) $x, y, z; \overline{y}, x-y, z; \overline{x}+y, \overline{x}, z; \overline{x}, \overline{y}, z; y, \overline{x}+y, z; x-y, x, z$.

(c) $x,\ y,\ z;\overline{y},\ x-y,\ z;\overline{x}+y,\ \overline{x},\ z;\overline{x},\ \overline{y},\ \overline{z};y,\ \overline{x}+y,\ \overline{z};x-y,\ \overline{x},\ \overline{z}.$

(d) $x,\ y,\ z;\overline{y},\ x-y,\ z;\overline{x}+y,\ \overline{x},\ z;x,\ y,\ \overline{z};\overline{y},\ x-y,\ \overline{z};\overline{x}+y,\ \overline{x},\ \overline{z}.$

(3.9) (a) $x,\ y,\ z;\overline{y},\ x-y,\ \frac{1}{3}+z;\overline{x}+y,\ \overline{x},\ \frac{2}{3}+z.$

(b) $x,\ y,\ z;x-y,\ x,\ \frac{1}{6}+z;\overline{y},\ x-y,\ \frac{1}{3}+z;\overline{x},\ \overline{y},\ \frac{1}{2}+z;\overline{x}+y,\ \overline{x},\ \frac{2}{3}+z;\ y,\ \overline{x}+y,\ \frac{5}{6}+z.$

(c) $x,\ y,\ z;\overline{x},\ \overline{y},\ z;x-y,\ x,\ \frac{1}{3}+z;\overline{x}+y,\ \overline{x},\ \frac{1}{3}+z;\overline{y},\ x-y,\ \frac{2}{3}+z;\ y,\ \overline{x}+y,\ \frac{2}{3}+z.$

(3.10) Na at origin has point symmetry $\overline{3}m$; other Na and C have point symmetry $\overline{6}m2$; O has point symmetry $mm2$.

(3.11) The non-intersecting rotation axes generate a 2_1 axis passing through the origin. In the first case, the set of coordinates is $x,\ y,\ z;\ x,\ \overline{y},\ \overline{z};\ \overline{x},\ y,\ \overline{z}$ and $\overline{x},\ \overline{y},\ z$. The set of coordinates for the second case is $x,\ y,\ z;\ x,\ \overline{y},\ \overline{z};\overline{x},\ y,\ \frac{1}{2}-z$ and $\overline{x},\ \overline{y},\ \frac{1}{2}+z$.

(3.12) (a) $x,\ y,\ x;\frac{1}{2}-x,\ y,\ z;x,\ \overline{y},\ z;\frac{1}{2}+x,\ y,\ \overline{z};\overline{x},\ \overline{y},\ \overline{z};\frac{1}{2}+x,\ \overline{y},\ \overline{z};\ \overline{x},\ y,\ \overline{z};\frac{1}{2}-x,\ \overline{y},\ z.\ 2_1 \parallel x;2 \parallel y;2 \parallel z.$

(b) $x,\ y,\ x;\overline{x},\ y,\ \frac{1}{2}+z;x,\ \frac{1}{2}-y,\ \frac{1}{2}+z;x,\ \frac{1}{2}+y,\overline{z};\overline{x},\ \overline{y},\ \overline{z};x,\ \overline{y},\ \frac{1}{2}-z;\overline{x},\ \frac{1}{2}+y,\ \frac{1}{2}-z;\overline{x},\ \frac{1}{2}-y,\ z.\ 2 \parallel x;2_1 \parallel y;2 \parallel z.$

(c) $x,\ y,\ x;\frac{1}{2}+x,\ \frac{1}{2}-y,\ \overline{z};x,\ \overline{y},\ z;\frac{1}{2}+x,\ \frac{1}{2}+y,\ \overline{z}.$

(d) $x,\ y,\ x;\overline{x},\ \frac{1}{2}+y,\ z;x,\ \overline{y},\ z;x,\ \frac{1}{2}+y,\ \overline{z};\overline{x},\ \overline{y},\ \overline{z};\overline{x},\ \frac{1}{2}-y,\ z;\ x,\ \frac{1}{2}-y,\ \overline{z};\overline{x},\ y,\ \overline{z}.\ 2 \parallel x;2 \parallel y;2 \parallel z.$

(3.13) Both space groups have a 2_1 parallel to z. However, $Pmna$ has 2-fold rotation axes parallel to both x and y, whereas $Pnma$ has 2_1 axes parallel to both x and y.

Chapter 4

(4.1) For the first part, $[U_1V_1W_1]\cdot(hkl)=0$ and $[U_2V_2W_2]\cdot(hkl)=0$. These define two equations for two unknowns – there is an arbitrary scale factor on (hkl) – and the result follows as the solution. A similar approach gives the second part.

(4.2) $(4\overline{1}0)$, $(1\overline{2}0)$, (210), $(3\overline{4}0)$.

(4.3) The neat approach is to write the lattice vectors as the matrix $\mathbf{A}$, with $\mathbf{A}^{-1}$ containing the components of the reciprocal lattice vectors. $\mathbf{A}^{-1}\times\mathbf{A}=2\pi\mathbf{1}$, where $\mathbf{1}$ is the unit matrix. The volumes of the real-space and reciprocal-space unit cells are $\det[\mathbf{A}]$ and $\det[\mathbf{A}^{-1}]$ respectively. Since $\det[2\pi\mathbf{1}]=(2\pi)^3$, it follows that $\det[\mathbf{A}^{-1}]=(2\pi)^3/\det[\mathbf{A}]$.

(4.5) Write $\mathbf{A}'=\det[\mathbf{M}]\times\det[\mathbf{A}]$. The volumes of the two unit cells are $\det[\mathbf{A}']$ and $\det[\mathbf{A}]$ respectively. Hence $\det[\mathbf{M}]=\det[\mathbf{A}']/\det[\mathbf{A}]$.

(4.6) The (100) point in the reciprocal space of the fcc lattice has two neighbouring reciprocal lattice points (000) and (200) at distance a^*, and four neighbouring reciprocal lattice points at (111), $(11\overline{1})$, $(1\overline{1}1)$, and $(1\overline{1}\overline{1})$ at distance $\sqrt{2}a^*$. The (110) point has two neighbouring reciprocal lattice points (111) and $(11\overline{1})$ at distance a^*, and four neighbouring reciprocal lattice points at (000), (200), (020), and (220) at distance $\sqrt{2}a^*$. Both arrangements of points have identical geometry but different orientation. The analysis can easily be extended to arbitrary distance

because we simply add more reciprocal lattice vectors. The issue of orientation is not relevant when considering symmetry.

The (100) point in the reciprocal space of the bcc lattice has six neighbouring reciprocal lattice points at distance a^* in perfect octahedral arrangement: (000), (200), (110), $(1\bar{1}0)$, (101), and $(10\bar{1})$. The (111) point has a similar octahedral arrangement of neighbours: (110), (112), (011), (211), (101), and (121). The argument then follows as for the fcc case above.

(4.7) The matrix describing the lattice vectors of the primitive lattice $\mathbf{A}_P$ is related to the matrix of the C-centred lattice, $\mathbf{A}_C$, through

$$\mathbf{A}_P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{A}_C$$

Accordingly the Miller indices of the two lattices are related through

$$\begin{pmatrix} h \\ k \\ l \end{pmatrix}_P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h \\ k \\ l \end{pmatrix}_C$$

Each of the hkl indices of the primitive lattice need to be integers. This is only possible if the Miller indices of the reciprocal lattice of the C-centred lattice are subject to the constraint that $h + k$ is an even number. For A and B-centred lattices, the constant conditions apply to $k + l$ and $h + l$ respectively.

(4.8) The result follows from the relationship between the Miller indices for the two settings, which is obtained from Appendix D:

$$\begin{pmatrix} h_{\text{rhom}} \\ k_{\text{rhom}} \\ l_{\text{rhom}} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} h_{\text{trig}} \\ k_{\text{trig}} \\ l_{\text{trig}} \end{pmatrix} \tag{4.5}$$

It is essential that the Miller indices in the primitive rhombohedral setting are integers.

Chapter 5

(5.1) $\sigma = 2^{-2/3}a = 2.813$ Å, $\epsilon = 1.88/6 = 0.313$ kJ mol^{-1}.

(5.3) We obtain

$$\frac{\partial^2 U}{\partial V^2} = \left(\frac{\partial V}{\partial R}\right)^{-2} \frac{\partial^2 U}{\partial R^2}$$

Additional terms involving $\partial U/\partial R$ are equal to zero by the condition of equilibrium. $V = N_{\text{A}}\gamma R^3$.

(5.4) Use the same method as for Problem 5.3.

(5.9) There are two electrons per $\mathbf{k}$. The number of wave vectors in a spherical shell in reciprocal space (radius k, thickness $\mathrm{d}k$) is

$$\frac{4\pi k^2\,\mathrm{d}k}{(2\pi)^3/V}$$

The energy is obtained by integrating the energy of a shell over all shells:

$$\begin{aligned} E &= 2\int_0^{k_\mathrm{F}} \frac{4\pi k^2}{(2\pi)^3/V}\frac{\hbar^2k^2}{2m}\,\mathrm{d}k \\ &= \frac{V}{\pi^2}\frac{\hbar^2}{2m}\frac{k_\mathrm{F}^2}{5} \end{aligned}$$

Using $k_\mathrm{F}^3 = 3\pi^2 n/V$ and substituting for v/π^2 gives the final stage of the proof.

(5.10) At zero pressure, kyanite is stable to 450 K, andalusite is stable from 450 to 1050 K, and sillimanite is stable for temperatures above 1050 K. The slope of the kyanite–andalusite phase boundary is positive, but the slope of the andalusite–sillimanite phase boundary is negative. The two phase boundaries meet at a triple point at 810 K and 0.5 GPa. For higher pressures only kyanite and sillimanite are stable.

Chapter 6

(6.1) From the momentum relationship, $\lambda = h/mv = ht/ml = 2d\sin\theta$. (a) 20 m; (b) 100 m.

(6.2) (a) $Q = 2.7876\,\text{Å}^{-1}$, $\mathbf{Q} = (0.9534, -2.6195, 0)$. $\sin\theta/\lambda = 0.2218\,\text{Å}^{-1}$, and from the graph $f = 2.0$. From the atomic coordinates, $F(\text{positions}) = 2\cos(\mathbf{Q}\cdot\mathbf{r})90 = 1.777$. Final structure factor $= F(\text{positions}) \times f = 3.55$.
(b) $Q = 5.2390\,\text{Å}^{-1}$, $\mathbf{Q} = (3.3675, -4.0133, 0)$. $\sin\theta/\lambda = 0.417\,\text{Å}^{-1}$, and from the graph $f = 1.65$. From the atomic coordinates, $F(\text{positions}) = 2\cos(\mathbf{Q}\cdot\mathbf{r}) = -0.225$. Final structure factor $= F(\text{positions}) \times f = -0.372$.

(6.3) The area is a sphere with radius twice the radius of the Ewald sphere, i.e. $4\pi/\lambda$.

(6.4) $(h00)$, $|h| \leq 8$; $(h02)$, $|h| \leq 7$.

(6.5) Pyrophyllite: $F(001) \simeq 10$; $F(002) \simeq -15$; $F(003) \simeq 60$.
Muscovite: $F(001) \simeq 25 - 18 = 7$; $F(002) \simeq -30 + 18 = -12$; $F(003) \simeq 66 - 18 = 47$.

(6.6) Lowest order reflection is (111, $I(1273)/I(293) = 0.507$.
Highest order reflection is 420, $I(1273)/I(293) = 0.011$.

(6.7) (a) $Pba2$ or $Pbam$; (b) $Pccn$; (c) $P2_1n2_1$ or $Pmnm$; (d) $P2_1am$ or $Pmam$.

(6.8) Systematic absences are

	(a)	(b)	(c)	(d)
$0kl$:	l odd		$k+l$ odd	
$h0l$:	l odd	—	$h+l$ odd	$h+l$ odd
$hk0$:	h odd	$h+k$ odd	—	—
$h00$:	h odd	h odd	h odd	h odd
$0k0$:	—	k odd	k odd	—
$00l$:	l odd	—	l odd	l odd

Chapter 7

(7.1) Demonstrate that $\mathbf{a}^{\mathrm{T}} \cdot \mathbf{a}$ gives the unit matrix.

(7.2) 68.4° from [001]. 1.19×10^{-3}.

(7.3) $\sigma_{ij} = \sigma_3'/9$ for all i, j.

(7.5) $P_1 = P_2 = P_3 = 2d_{14}\sigma_3'/9$. 2.58×10^{-9} C.

(7.6) Reading from the graph $d_{11} = -2.3 \times 10^{-12}\ \mathrm{C\,N^{-1}}$ and $d_{14} = -0.7 \times 10^{-12}\ \mathrm{C\,N^{-1}}$. $\epsilon_2 = -\epsilon_1 = 2.3 \times 10^{-6}$, $\epsilon_4 = -0.7 \times 10^{-6}$.

(7.7) If the 2-fold axis is along z, the non-zero components are d_{31}, d_{32}, d_{33}, d_{24}, and d_{15}.

(7.8) If the $\overline{4}$ axis is along z, the non-zero components are $d_{31} = -d_{32}$, $d_{24} = -d_{15}$, $d_{14} = d_{25}$ and d_{36}.

(7.10) The condition for c_{44} follows to give stability against a non-zero e_4. The other two conditions follow from the condition that the eigenvalues of $\mathbf{c} > 0$ for stability.

Chapter 8

(8.2) $\partial\phi/\partial x = (x/r)(\partial\phi/\partial r)$. The second derivative follows as $\partial^2\phi/\partial x^2 = (\partial(x/r)/\partial x)(\partial\phi/\partial r) + (x/r)^2(\partial^2\phi/\partial r^2)$. The first term vanishes since for the special case where we only have nearest-neighbour interactions, $\partial\phi/\partial r = 0$.

(8.3) (a) For $\mathbf{k} \parallel [100]$, the coordinate system is the same as for the crystal. The four vectors to the nearest neighbours in the next (100) plane are $a(1, \pm 1, 0)/2$ and $a(1, 0, \pm 1)/2$. The result for K is obtained by summing the four values of $(x/r)^2$ for the LA mode force constant and the four values of either $(y/r)^2$ or $(z/r)^2$ for the TA mode. (b) For $\mathbf{k} \parallel [111]$, the coordinate system is $x \parallel \mathbf{k}$, $y \parallel [1\bar{1}0]$, and $z \parallel [11\bar{2}]$. The three vectors to the nearest-neighbour atoms in the next (111) plane are $a(\sqrt{2}, \pm\sqrt{3}/2, -1/2)/\sqrt{6}$ and $a(\sqrt{2}, 0, 1)/\sqrt{6}$. The result for K for the LA and TA modes follows from the same procedure. (c) For $\mathbf{k} \parallel [110]$, the coordinate system is $x \parallel \mathbf{k}$, $y \parallel [1\bar{1}0]$, and $z \parallel [001]$. The four nearest-neighbour atoms in the first (110) plane are at $a(\pm 1, \pm 1, \sqrt{2})/2\sqrt{2}$, and the nearest-neighbour atom in the second-neighbour plane is at $a(1, 0, 0)/\sqrt{2}$. The LA and two TA values of K follow the same procedure, but now we can also define a non-zero value of K for the second-neighbour plane.

(8.4) The best value of **k** to fit is at the zone boundary point, where $m\omega^2 = 4J$. Averaging over the zone boundary mode frequencies gives an average value of K of give the following estimates of K: $K_{\mathrm{L}}(100) = 2.58\,\mathrm{J\,mol^{-1}\,\AA^{-2}}$. (Note that the value obtained from the TA(111) mode is somewhat lower than the other three values.) The value for K calculated from values of ϵ and σ obtained in Chapter 5 is $2.26\,\mathrm{J\,mol^{-1}\,\AA^{-2}}$, which is a little lower than indicated from the phonon dispersion curves.

(8.5) This result is useful because the idea will be used in Chapter 12.

(8.6) Considering nearest-neighbour interactions for $\mathbf{k} \parallel [100]$, the next plane has atomic neighbours at the four positions $a(1, \pm 1, \pm 1)/2$. Following the analysis for Ne, the sums of $(x/r)^2$, $(y/r)^2$, and $(z/r)^2$ are equal. Hence the TA and LA modes have identical contributions from the nearest-neighbour force constants. The fact that the experimental values are so similar points to the dominance of nearest-neighbour forces.

(8.7) It is convenient to define the distance between the (111) planes as d, so that the zone boundary along the plane normals can be written as π/d. One approach to the analysis of the dispersion curves is to analyse the solutions of the one-dimensional model with interactions up to third neighbour at $k = \pi/3d$, $k = 2\pi/3d$, and $k = \pi/d$. We have the following three equations for each of these wave vectors:

$$\begin{aligned} m\omega^2(k = \pi/3) &= 4J_1 \sin^2(\pi/6) + 4J_2 \sin^2(\pi/3) + 4J_3 \sin^2(\pi/2) \\ &= J_1 + 3J_2 + 4J_3 \\ m\omega^2(k = 2pi/3) &= 4J_1 \sin^2(\pi/3) + 4J_2 \sin^2(2\pi/3) \\ &= 3J_1 + 3J_2 \\ m\omega^2(k = \pi/2) &= 4J_1 + 4J_3 \end{aligned}$$

The three frequency values are roughly 2.25, 1.0, and 2.25 respectively, giving $m\omega^2$ values of 7.815×10^4, 1.544×10^4, and $7.815 \times 10^4\,\mathrm{J\,mol^{-1}\,\AA^{-2}}$ respectively. The solutions to the three equations give $J_1 = J_2 = 0.26 \times 10^4\,\mathrm{J\,mol^{-1}\,\AA^{-2}}$ and $J_3 = 1.70 \times 10^4\,\mathrm{J\,mol^{-1}\,\AA^{-2}}$. J_3 is much larger because it involves a direct compression of the nearest-neighbour bond.

(8.8) (a) The explanation is the same as for the (100) modes in potassium in the preceeding question. (b) Analysis of nearest-neighbour force constants gives this result. (c) For the (110) direction, we define $x \parallel (100)$. The molecules in the nearest-neighbour plane have $x = a/\sqrt{2}$ and $y = 0$. The LA force constant is determined by $(x/r)^2$, and the TA force constant by $(y/r)^2$. This latter value is zero within the limits of this model.

(8.10) For the limit $k \to 0$, the acoustic mode solution is $\omega^2 = Jk^2a^2/2(M + m)$, and the optic mode solution is $\omega^2 = 2J(M + m)/Mm$. For $k = \pi/a$, the two solutions are $\omega^2 = 2J/M$ and $\omega^2 = 2J/m$.

(8.11) The full equations of motion can be written in matrix form:

$$\begin{pmatrix} m\omega^2 - (J_1 + J_2) & J_1 + J_2\exp(-ika) \\ J_1 + J_2\exp(ika) & m\omega^2 - (J_1 + J_2) \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0$$

The frequencies are given as

$$\omega^2 = \frac{J_1 + J_2}{m} \pm \frac{1}{m}\sqrt{(J_1 + J_2)^2 - 4J_1J_2\sin^2(ka/2)}$$

For $k = \pi/a$ the two solutions are $\omega^2 = 2J_1/m$ and $\omega^2 = 2J_2/m$. Substituting these solutions into the matrix equation gives $u_1 = -u_2$ and $u_1 = u_2$ for the two modes respectively. The reason that this behaviour is different from the case of two different masses but equal force constants is that the off-diagonal matrix elements vanish in the latter case.

(8.12) The solutions for optic and acoustic modes of L or T character at the zone boundary are $\omega^2 = 2J/M$ and $\omega^2 = 2J/m$, where M and m are the masses of the two ions. The ratio of the masses for NaCl is 1.54. From the data we have $\omega^2(\mathrm{LO})/\omega^2(\mathrm{LA}) = 1.59$ and $\omega^2(\mathrm{TO})/\omega^2(\mathrm{TA}) = 1.29$. For $k = 0$, $\omega^2 = 2J(M+m)/Mm$, which is equal to the sum of the zone boundary ω^2. From the experimental data, $\omega^2(k = 0)$ is 15% and 18% lower for the T and L modes respectively than the corresponding sum of ω^2 values for the zone boundary. The mapping of the model into the diatomic chain could be said to be optimistic due to the fact that Coulomb interactions are long range, so in the light of this we could say that the mapping is really quite good. The (111) modes are the most suitable for this mapping because neighbouring planes of atoms contain only cations or only anions.

Chapter 9

(9.1) For $x = 1$, $n(x) + 1/2 = 1.082$, i.e., the high-temperature approximation gives an error of only 8% on the internal energy.

(9.3) Generalizing the argument that led to the derivation of $g(\omega)$ for three dimensions, we have $g(k)\,\mathrm{d}k \propto k^{d-1}$. Hence $g(\omega)\,\mathrm{d}\omega = g(k)\,\mathrm{d}k \propto (\omega/c)^{d-1}\,\mathrm{d}\omega/c$, giving the required result.

(9.4) Use the lattice sums of Appendix M to show that the sum over pairs of wave vectors reduces to a single sum, and use the orthonormal condition of phonon eigenvectors to show that the sum over pairs of normal modes reduces to a single sum.

(9.5) Use the same approach as problem 5. For the high-temperature limit, use eqn 9.45.

(9.6) The equation is a sum over $\langle Q(k)Q(-k)\rangle$, which in the high-temperature limit is proportional to T/ω^2. The sum can be replaced by the integral weighted by the density of states.

In three dimensions, $g(\omega)/\omega^2$ is a constant, and the integral is well behaved. In two dimensions, $g(\omega)/\omega^2 \propto \omega^{-1}$, and the integral has a logarithmic divergence. In one dimension, $g(\omega)/\omega^2 \propto \omega^{-2}$, and the integral clearly diverges. Thus for both $d = 1$ and $d = 2$ the mean-square atomic displacement is infinite, and the system is effectively a liquid.

Chapter 10

(10.1) In the equation for $S(\mathbf{Q}, E = 0)$, the Fourier transform of the average $\langle \exp(i\mathbf{Q}\cdot\mathbf{r}_j(t)) \exp(-i\mathbf{Q}\cdot\mathbf{r}_k(0))\rangle$ becomes an integral over all time.

Over most time the two exponential functions have no correlation, so the integral becomes equal to a delta function multiplied by $\langle \exp(i\mathbf{Q} \cdot \mathbf{r}_j)\rangle \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}_k)\rangle$. When added over all pairs of atoms, we have $|\sum_j b_j \langle \exp(i\mathbf{Q} \cdot \mathbf{r}_j)\rangle|^2$, which is the intensity for Bragg scattering as described in Chapter 6.

(10.2) The $m = 0$ term in eqn 10.16 is simply equal to 1. When we put back all the other parts into eqn 10.17 we are left with scattering from the average positions of atoms together with the temperature factors, which is equivalent to the squared modulus of the structure factor for diffraction as developed in Chapter 6.

(10.3) Equation M.1 in Appendix M should be applied, so the sum over lattice vectors for the terms $\exp(i(\mathbf{Q} + \mathbf{k}) \cdot \mathbf{r}_i)$ will be given as $N\delta(\mathbf{Q} + \mathbf{k})$, so that $\mathbf{k} = -\mathbf{Q}$. The minus sign can be reversed by noting that the sum over $\mathbf{k}$ includes both positive and negative values. The same argument is applied to the sum over j, which involves $\mathbf{k}'$. The remaining summations over atoms is only over atoms in the unit cell. The condition $\nu = \nu'$ follows from mode orthogonality.

(10.4) The integral of $S(\mathbf{Q}, E)$ over all E is equivalent to the Fourier transform of $S(\mathbf{Q}, E)$ evaluated at $t = 0$. Thus we have the instantaneous quantity $\sum_{j,k} b_j b_k \langle \exp(i\mathbf{Q}\cdot(\mathbf{r}_j - \mathbf{r}_k))\rangle$, which is the sum scattering arising from the interference scattering from pairs of atoms, giving information about the separations of atoms.

(10.5)
(a) LA mode for $\mathbf{k} = (0.2, 0, 0)$.
(b) TA mode, polarization [001], for $\mathbf{k} = (0.2, 0, 0)$.
(c) TA mode, polarization [112], for $\mathbf{k} = (0.3, 0.3, -0.3)$.
(d) LA mode and TA mode, polarization $[1\bar{1}0]$, for $\mathbf{k} = (1, 1, 0)$, with relative weightings from $(\mathbf{Q}\cdot\mathbf{e})^2$ factor 16 and 4 respectively (although the TA mode will have additional weighting through the $1/\omega^2$ factor).
(e) LA mode and TA mode, polarization [001], for $\mathbf{k} = (0.2, 0.2, 0)$, with relative weightings from $(\mathbf{Q}\cdot\mathbf{e})^2$ factor 10.04 and 16 respectively (although the TA mode will have additional weighting through the $1/\omega^2$ factor).
(f) LA mode (very weak) and TA modes, polarizations [110] and [001], for $\mathbf{k} = (0.6, -0.6, 0)$, with relative weightings from $(\mathbf{Q} \cdot \mathbf{e})^2$ factor 0.64, 36, and 4 respectively.

(10.6) Only the B_{1u} mode gives rise to an electrical dipole moment, so this is the only mode that could be IR active, although if these four atoms are the only atoms in the unit cell the mode will be an acoustic mode and would not be observed in an IR absorption measurement.

(10.7) The second equation follows from the first because we do not expect the spin of the nucleus to be correlated with its position in the crystal (the magnetic dipole interactions between nuclei are so weak that they do not have effects until very low temperatures). Similarly $\overline{b_j b_k} = \bar{b}_j \bar{b}_k$ unless $j = k$, so $\overline{b_j b_k} - \bar{b}_j \bar{b}_k$ is zero when $j \neq k$. $S_{\text{incoherent}}(\mathbf{Q}, E)$ reduces to a sum over individual atoms directly as a result. The incoherent scattering factor gives spatial and dynamic information about the motions of individual atoms. In practice this can be used

for measurements of vibrational densities of states, since the motions of individual atoms are determined by all lattice vibrations, or about the motions in fast diffusion.

(10.8) $E/h = 5.25\,\text{THz}$.

(10.9) *Neutron scattering advantages*:

- Can measure phonons for all wave vectors, hence enabling measurements of full phonon dispersion curves
- Weak selection rules can be varied by performing measurements in several Brillouin zones, thus enabling measurements of **all** phonons

Neutron scattering disadvantages:

- Weak scattering, leading to need for lower resolution and larger samples, and difficulties in measurement of high-frequency modes
- Need to use central-site facilities, with associated need for long-term planning

Raman and IR spectroscopy advantages:

- High intensity allows high resolution and measurement of high-frequency modes, and ability to perform accurate measurements in short times enables measurements to be performed over wide range of sample conditions
- Strict selection rules can be exploited to provide information about symmetry, particularly when both types of experiment are performed on the same sample
- Experiments can be performed in scientists' own laboratories

Raman and IR spectroscopy disadvantages:

- Restriction to $\mathbf{k} \simeq 0$
- Strict selection rules mean that some phonons at $\mathbf{k} \simeq 0$ cannot be measured

(10.10) The intensity for energy loss is proportional to $n + \frac{1}{2}$, whereas the intensity for energy gain is proportional to n. The ratio $(n + \frac{1}{2})/n$ will give a measurement of temperature.

Chapter 12

(12.1)
$$\eta = \left(\frac{a}{c}\right)^{\frac{1}{4}} (T_\text{c} - T)^{1/4}$$
$$\chi^{-1} = 4a(T_\text{c} - T)$$
$$\Delta C = \frac{1}{4}\left(\frac{a^3}{c}\right)^{\frac{1}{2}} T(T_\text{c} - T)^{-1/2}$$

The exponent in the expression for η is half that for a second-order phase transition, the prefactor for χ^{-1} is twice that for a second-order phase transition, and the heat capacity diverges at the tricritical phase transitions whereas there is a small step associated with a second-order phase transition.

(12.3) The result follows as the solution to $\partial^2 F/\partial\eta^2 = 0$.

(12.4) Exact values will depend on the procedures followed. The transition temperature can be fixed from experimental data.

(12.5) (a) This follows as the solution of $\partial F/\partial\epsilon_s = 0$. (b) Substitution of ϵ_s into F gives a modified term for η^2, which can be interpreted in terms of a modified transition temperature. (b) Substitution of $\epsilon_s = -2C_{el}/\lambda$ into F gives a Landau free energy function in terms of ϵ_s with the same modified transition temperature. The elastic constant is given by $\partial^2 F/\partial\epsilon_s^2$, giving a function that falls to zero at the modified transition temperature following χ^{-1}.

(12.7) The proof is obtained by setting

$$\frac{3N_A\hbar\overline{\alpha}}{4\kappa_4\langle\tilde{\omega}\rangle} > \frac{\kappa_2}{\kappa_4}$$

and substituting the value of T_c from eqn 12.53.

(12.8) At zero temperature, $\langle Q(k)Q(-k)\rangle = \hbar/2\omega(k)$. The value of ω^2 for the soft mode at zero frequency is obtained by using this average rather than the high-temperature limit in eqn 12.32.

(12.9) The harmonic solution is

$$m\omega^2 = 2J(1 - \cos ka) - \kappa_2$$

This is modified by addition of the term $3\kappa_4/4k_BT$ when we account for the anharmonic coefficient. The phase transition occurs when $\kappa_2 = 3\kappa_4/4k_BT$.

(12.10) The model can be mapped onto the monatomic chain for $\mathbf{k}$ parallel to (110), but with harmonic terms of the form $(\theta_j + \theta_k)^2$ rather than $(u_j - u_k)^2$ – the change of sign within the brackets is significant. The harmonic terms give

$$\omega^2 = -\kappa_2 + \frac{4K}{I}\sin^2(kd/2)$$

where I is the moment of inertia, and d is the spacing of (110) planes. The key point is that ω^2 has its minimum value not at $k = 0$ but at the zone boundary.

(12.11) The three cases give, with the non-zero components written as P

(a)
$$P^2 = \frac{a}{b+\gamma}(T_c - T)$$

$$F = -\frac{a^2}{4(b+\gamma)}$$

(b)
$$P^2 = \frac{2a}{2b+4\gamma}(T_c - T)$$

$$F = -\frac{4a^2}{4(2b+4\gamma)}$$

(c)
$$P^2 = \frac{3a}{3b + 9\gamma}(T_c - T)$$

$$F = -\frac{9a^2}{4(3b + 9\gamma)}$$

The stability conditions depend on the relative sizes of $b + \gamma$, $(b + 2\gamma)/2$ and $(b + 3\gamma)/3$ respectively, the smaller giving the most stable structure. c is the stable solution.

(12.12) At low temperature, and for small changes of frequency, we have $\alpha = 4\omega\Delta\omega = 31.65 \times 10^{24}\,\mathrm{s}^{-2}$ if we set $\eta = 1$ at low temperature. We have $\kappa_2 = 3\alpha R T_c/\omega_0^2 = 63.2\,\mathrm{J\,mol}^{-1}$. In the high-temperature limit, the step in the heat capacity is given by

$$\Delta C = \frac{3\alpha R}{2\omega_0^2}\frac{\kappa_2}{\kappa_4}$$

from which we obtain $\kappa_2/\kappa_4 = 1.24$, and hence $\kappa_4 = b = 51\,\mathrm{J\,mol}^{-1}$. The reduction in η^2 at low temperature due to the zero point motions is equal to $\kappa_2/\kappa_4 - 1$.

References

Alcock, N.W. (1990). Bonding and structure: structural principles in inorganic and organic chemistry. Ellis Horwood, London, New York.

Ashcroft, N.W. and Mermin, N.D. (1976). Solid state physics. Holt, Rinehart & Winston, New York.

Aslanov, L.A., Fetisov, G.V., and Howard, J.A.K. (1998). Crystallographic instrumentation. Oxford University Press, Oxford.

Azároff, L.V. (1974). X-ray diffraction. McGraw-Hill, New York.

Bacon, G.E. (1973). Neutron diffraction. Clarendon Press, Oxford.

Bacon, G.E. (1987). Fifty years of neutron diffraction: the advent of neutron scattering. Adam Hilger, Bristol.

Barron, T.H.K. and White, G.K. (1999). Heat capacity and thermal expansion at low temperatures. Kluwer Academic/Plenum, New York, London.

Bée, M. (1988). Quasielastic neutron scattering: principles and applications in solid state chemistry, biology and materials science. Adam Hilger, Bristol.

Borg, R.J. and Dienes, G.J. (1992). The physical chemistry of solids. Academic Press, Boston, London.

Böttger, H. (1983). Principles of the theory of lattice dynamics. Physik-Verlag, Weinheim.

Bruce, A.D. and Cowley, R.A. (1981). Structural phase transitions. Taylor & Francis, London.

Burdett, J.K. (1995). Chemical bonding in solids. Oxford University Press, Oxford.

Burns, G. and Glazer, A.M. (1990). Space groups for solid state scientists. Academic Press, London.

Clegg, W. (1998). Crystal structure determination. Oxford University Press, Oxford.

Collings, P.J. (1997). Introduction to liquid crystals chemistry and physics. Taylor & Francis, London.

Collings, P.J. and Patel, J.S. (1997). Handbook of liquid crystal research. Oxford University Press, Oxford, New York.

Coppens, P. (1992). Synchrotron radiation crystallography. Academic Press, London.

Coppens, P. (1997). X-ray charge densities and chemical bonding. Oxford University Press, Oxford.

Cowley, J.M. (1992). Electron diffraction techniques. Oxford University Press, Oxford.

Cullity, B.D. (2001). Elements of X-ray diffraction (3rd edn). Prentice-Hall International, London.

Di Bartolo, B. and Powell, R.C. (1976). Phonons and resonances in solids. Wiley, New York.

Diem, M. (1993). Introduction to modern vibrational spectroscopy. Wiley, Chichester, New York.

Dobrzynski, L. and Blinowski, K. (1994). Neutrons and solid state physics. Ellis Horwood, New York, London.

Dorner, B. (1982). Coherent inelastic neutron scattering in lattice dynamics. Springer-Verlag, Berlin, New York.

Dove, M.T. (1993). Introduction to lattice dynamics. Cambridge University Press, Cambridge.

Durig, J.R. (1990). Applications of FT-IR spectroscopy. Elsevier, Amsterdam.

Elliott, S.R. (1990). Physics of amorphous materials (2nd edn). Longman Scientific & Technical, Harlow.

Elliott, S.R. (1998). The physics and chemistry of solids. Wiley, Chichester.

Giacovazzo, C. *et al*. (1992). Fundamentals of crystallography. Oxford University Press, Oxford.

Glusker, J.P. and Trueblood, K.N. (1985). Crystal structure analysis: a primer (2nd edn). Oxford University Press, Oxford.

Hammond, C. (2001). The basics of crystallography and diffraction. Oxford University Press, New York.

Hardy, J.R. and Karo, A.M. (1979). The lattice dynamics and statistics of alkali halide crystals. Plenum Press, New York, London.

Hazen, R.M. and Finger, L.W. (1982). Comparative crystal chemistry: temperature, pressure, composition and the variation of crystal structure. Wiley, Chichester.

Hirth, J.P. and Lothe, J. (1968). Theory of dislocations. Wiley, Chichester, New York.

Hull, D. and Bacon, D.J. (1984). Introduction to dislocations. Pergamon, Oxford.

Iqbal, Z. and Owens, F.J. (1984). Vibrational spectroscopy of phase transitions. Academic Press, Orlando.

Janot, C. (1997) Quasicrystals: a primer (2nd edn). Clarendon Press, Oxford.

Keen (2001). Journal of Applied Crystallography, **34**, 172–7.

Kelly, A., Groves, G.W., and Kidd, P. (2000). Crystallography and crystal defects. Wiley, Chichester.

Kittel, C. (1996). Introduction to solid state physics (7th edn). Wiley, Chichester, New York.

Kosevich, A.M. (1999). The crystal lattice: phonons, solitons, dislocations. Wiley-VCH, Chichester.

Krishnan, R.S., Srinivasan, R., and Devanarayanan, S. (1979). Thermal expansion of crystals. Pergamon, Oxford.

Ladd, M.F.C. (1994). Chemical bonding in solids and fluids. Ellis Horwood, London.

Ladd, M.F.C. (1999). Crystal structures: lattices and solids in stereoview. Ellis Horwood, Chichester.

Ladd, M.F.C. and Palmer, R.A. (1977). Structure determination by X-ray crystallography. Plenum Press, New York.

Lines, M.E. and Glass, A.M. (1977). Principles and applications of ferroelectrics and related materials. Clarendon Press, Oxford.

Lovesey, S.W. (1984). Theory of neutron scattering from condensed matter. Clarendon Press, Oxford.

Lovett, D.R. (1989). Tensor properties of crystals. Adam Hilger, Bristol.

Mak, T.C.W. (1997). Crystallography in modern chemistry: a resource book of crystal structures. Wiley, Chichester.

McKie, D. and McKie, C. (1986). Essentials of crystallography. Blackwell Scientific, Oxford.

Mitsui, T. and Tatsuzaki, I. (1976). An introduction to the physics of ferroelectrics. Gordon and Breach, New York.

Nabarro, F.R.N. (1967). Theory of crystal dislocations. Clarendon Press, Oxford.

Newport, R.J., Rainford, B.D., and Cywinski, R. (1988). Neutron scattering at a pulsed source. Adam Hilger, Bristol.

Nichols, C.S. (1995). Structure and bonding in condensed matter. Cambridge University Press, Cambridge.

Nield, V.M. and Keen, D.A. (2001). Diffuse neutron scattering from crystalline materials. Clarendon Press, Oxford.

Nye, J.F. (1985). Physical properties of crystals: their representation by tensors and matrices. Clarendon Press, Oxford.

Pettifor, D.G. (1995). Bonding and structure of molecules and solids. Clarendon Press, Oxford.

Rae, A.I.M. (1992). Quantum mechanics (3rd edn). Institute of Physics, Bristol.

Rao, C.N.R. and Gopalakrishnan, J. (1986). New directions in solid state chemistry: structure, synthesis, properties, reactivity and materials design. Cambridge University Press, Cambridge.

Rao, C.N.R. and Rao, K.J. (1978). Phase transitions in solids: an approach to the study of the chemistry and physics of solids. McGraw-Hill, New York, London.

Salje, E.K.H. (1990). Phase transitions in ferroelastic and co-elastic crystals: an introduction for mineralogists, material scientists and physicists. Cambridge University Press, Cambridge.

Sands, D.E. (1982, 1995). Vectors and tensors in crystallography. Addison-Wesley, Reading, Mass. (1982); Dover, New York (1995).

Schenk, H. (1984). An introduction to direct methods: the most important phase relationships and their application in solving the phase problem. University College Cardiff Press, Cardiff.

Senechal, M. (1995). Quasicrystals and geometry. Cambridge University Press, Cambridge.

Squires, G.L. (1978). Introduction to the theory of thermal neutron scattering. Cambridge University Press, Cambridge.

Sköld, K. and Price, D.L. (1986). Neutron scattering. Academic Press, Orlando.

Srivastava, G.P. (1990). The physics of phonons. Adam Hilger, Bristol.

Stout, G.H. and Jensen, L.H. (1989). X-ray structure determination: a practical guide. Wiley, Chichester.

Sutton, A.P. (1993). Electronic structure of materials. Clarendon Press, Oxford.

Tilley, R.J.D. (1987). Defect crystal chemistry and its applications. Blackie, Glasgow.

Tolédano, J.-C. and Tolédano, P. (1987). The Landau theory of phase transitions: application to structural, incommensurate, magnetic, and liquid crystal systems. World Scientific, Singapore.

Whelan, M.J. (1990). Worked examples in dislocations. Institute of Metals, London.

Windle, A.H. (1977). A first course in crystallography. G. Bell, London.

Woolfson, M.M. (1997). An introduction to X-ray crystallography (2nd edn). Cambridge University Press, Cambridge.

Yeomans, J.M. (1992). Statistical mechanics of phase transitions. Clarendon Press, Oxford.

Young, R.A. (1993). The Rietveld method. Oxford University Press, Oxford.

Ziman, J. (2000). Electrons and phonons: the theory of transport phenomena in solids. Oxford University Press, Oxford.

Index