IUCr Texts on Crystallography · 9

Theories and Techniques of Crystal Structure Determination

URI SHMUELI

INTERNATIONAL UNION OF CRYSTALLOGRAPHY
OXFORD SCIENCE PUBLICATIONS

INTERNATIONAL UNION OF CRYSTALLOGRAPHY
BOOK SERIES

IUCr BOOK SERIES COMMITTEE

E. N. Baker, *New Zealand*
J. Bernstein, *Israel*
G. R. Desiraju, *India*
A. M. Glazer, *UK*
J. R. Helliwell, *UK*
P. Paufler, *Germany*
H. Schenk *(Chairman), The Netherlands*

IUCr Monographs on Crystallography

1 *Accurate molecular structures*
A. Domenicano, I. Hargittai, editors
2 *P. P. Ewald and his dynamical theory of X-ray diffraction*
D. W. J. Cruickshank, H.J. Juretschke, N. Kato, editors
3 *Electron diffraction techniques, Vol. 1*
J. M. Cowley, editor
4 *Electron diffraction techniques, Vol. 2*
J. M. Cowley, editor
5 *The Rietveld method*
R. A. Young, editor
6 *Introduction to crystallographic statistics*
U. Shmueli, G. H. Weiss
7 *Crystallographic instrumentation*
L. A. Aslanov, G. V. Fetisov, G. A. K. Howard
8 *Direct phasing in crystallography*
C. Giacovazzo
9 *The weak hydrogen bond*
G. R. Desiraju, T. Steiner
10 *Defect and microstructure analysis by diffraction*
R. L. Snyder, J. Fiala and H. J. Bunge
11 *Dynamical theory of X-ray diffraction*
A. Authier
12 *The chemical bond in inorganic chemistry*
I. D. Brown

13 *Structure determination from powder diffraction data*
W. I. F. David, K. Shankland, L. B. McCusker, Ch. Baerlocher, editors
14 *Polymorphism in molecular crystals*
J. Bernstein
15 *Crystallography of modular materials*
G. Ferraris, E. Makovicky, S. Merlino
16 *Diffuse x-ray scattering and models of disorder*
T. R. Welberry
17 *Crystallography of the polymethylene chain: an inquiry into the structure of waxes*
D. L. Dorset
18 *Crystalline molecular complexes and compounds: structure and principles*
F. H. Herbstein
19 *Molecular aggregation: structure analysis and molecular simulation of crystals and liquids*
A. Gavezzotti
20 *Aperiodic crystals: from modulated phases to quasicrystals*
T. Janssen, G. Chapuis, M. de Boissieu
21 *Incommensurate crystallography*
S. van Smaalen

IUCr Texts on Crystallography

1 *The solid state*
A. Guinier, R. Julien
4 *X-ray charge densities and chemical bonding*
P. Coppens
5 *The basics of crystallography and diffraction, second edition*
C. Hammond
6 *Crystal structure analysis: principles and practice*
W. Clegg, editor
7 *Fundamentals of crystallography, second edition*
C. Giacovazzo, editor
8 *Crystal structure refinement: a crystallographer's guide to SHELXL*
P. Müller, editor
9 *Theories and techniques of crystal structure determination*
U. Shmueli
10 *Advanced structural inorganic chemistry*
Wai-Kee Li, Gong-Du Zhou, Thomas Mak

Theories and Techniques of Crystal Structure Determination

Uri Shmueli

School of Chemistry, Tel Aviv University

INTERNATIONAL UNION OF CRYSTALLOGRAPHY

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 Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

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

© Uri Shmueli 2007

The moral rights of the authors have been asserted
Database right Oxford University Press (maker)

First published 2007

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

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

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Printed in Great Britain
on acid-free paper by Biddles Ltd. hhbiddles.co.uk

ISBN 978–0–19–921350–4 (Hbk) 978–0–19–921966–7 (Pbk)

1 3 5 7 9 10 8 6 4 2

Preface

This book arose mainly from lectures on crystallography delivered by the author to graduate students of chemistry at Tel Aviv University. Although it is not intended to be a "primer", efforts have been made to present the material in a self-contained manner. Previous exposure to elementary crystallography may be helpful but is not a prerequisite. The reader is expected to be a graduate in chemistry, physics, or materials science with a mathematical background corresponding to the first-year curricula of these fields of study. Specifically, a working knowledge of elementary vector algebra–in its applications to geometry– is expected to be helpful. Other mathematical techniques employed in this book are basic linear algebra, some calculus, and the notion of Fourier series. I have sought to provide detailed derivations and material which may disrupt the line of reasoning is collected into four Appendices.

The first chapter introduces lattice geometry, relationships between direct and reciprocal lattices, permissible rotational symmetries of a 3-D lattice and allowed combinations of rotational symmetries. A discussion of the seven crystal systems concludes the chapter. The second and third chapters employ the building blocks secured in the first chapter to present the theories of the point-group and space-group symmetries of a crystal, which is conceived as a triply periodic array of material units. The first three chapters thus contain a self-consistent introduction to the geometrical and symmetry aspects of classical crystallography.

The fourth chapter deals with the crystallographic experiment. We introduce the conventional diffraction conditions and their representation, introduce the reader to the production and detection of X-rays by traditional and modern methods, and present a concise survey of the apparatus used for the collection of diffracted intensity data. The modern methods are exemplified by the principles and applications of synchrotron radiation and imaging plates.

The fifth chapter presents a discussion of X-ray scattering by charged particles and proceeds to the scattering from a periodic array of material units, followed by a derivation of the structure factor and its relation to the electron density within the repeating unit of the crystal. Next, effects of space-group symmetry on the diffraction pattern are considered, and hence the first steps in the determination of the crystal symmetry.

Given some basic ideas about the geometry and symmetry of a crystal, a survey of experimental techniques for the collection of intensity data, and the structure-dependent model in terms of which these data are interpreted, we are ready for actual methods of crystal structure determination. The sixth chapter deals with one such method, based on the Patterson function. Common to all

its applications is the use of interatomic vectors, usually aided by symmetry considerations. After the introduction of this function, some classical techniques based on it are discussed as well as some of its advanced applications.

The seventh chapter deals with structure-factor statistics, as applied to scaling the intensities and completing the space-group determination. Classical as well as modern methods are discussed. This chapter was chosen to precede the next one, on direct methods, since the latter methods are, in part, based on probabilistic considerations and the seventh chapter may serve as an introduction.

The eighth chapter deals with the basics of direct determination of the phases of the structure factors and hence determination of the structure with no previous assumptions. Space-group symmetry and probabilistic considerations are of importance in most variants of these methods.

The ninth chapter deals in some detail with small displacements of the atoms from their mean positions in the crystal, including small departures from perfect periodicity and how these affect the structure factors and hence the intensities. Actually, this chapter could have followed the fifth one but the details presented here are not required until the next chapter.

The tenth and last chapter deals with refinement of structural parameters. This process is essentially a confrontation of the experimental data with the model used for their interpretation. In this confrontation, the parameters on which the model is based are systematically varied so that the agreement of the model with experiment improves. Among the existing refinement techniques, we shall discuss only the least-squares technique. The discussion is based on conventional atomic parameters, rigid-body parameters in constrained refinement calculations, and geometrical parameters included in the section on restraints.

It is expected that this concise presentation of the subject will serve as a good introduction to its principles, and also as a reminder of some topics which call for clarification.

Tel Aviv, 29 June, 2006. Uri Shmueli

Acknowledgments

I wish to remember the late Arthur Wilson and Ken Trueblood, with whom I had fruitful and enjoyable collaborations. Thanks are due to George Weiss for instructive and very friendly cooperation and, last but not least, to my students at Tel Aviv University who made me think about the principles of crystallography. The School of Chemistry and the Computation Center at Tel Aviv University put at my disposal facilities which are most gratefully acknowledged. I am very grateful to Michael Dacombe of the International Union of Crystallography and to Mary Hale of Argonne National Laboratory for the kind and prompt granting of copyright permissions. Finally, I wish to thank Sönke Adlung and his team at the Oxford University Press for their interest and support during the preparation of the final version of this book.

Several illustration and tables, appearing in this book, were reproduced, in whole or in part, from the literature, and I am grateful to their owners for the permissions to do so. Figures 1.1 and 3.1 were reproduced from *Symmetry Aspects of M. C. Escher's Periodic Drawings* by Caroline MacGillavry, with copyright permission of the International Union of Crystallography (IUCr). Figures 3.2, 3.3, 3.4, 3,5, 3.6, 3.7 and 3.8 were reproduced, in part, from Volume A of the *International Tables for Crystallography*, with copyright permission of the IUCr. Figures 4.4, 4.5. 4.7 and 4.12 were reproduced from *An Introduction to the Scope, Potential and Applications of X-ray Analysis* by Michael Laing, with copyright permission of the IUCr. Figure 4.8 illustrating the Advanced Photon Source (APS) synchrotron was kindly provided by Argonne National Laboratory, and copyright permission has been granted. Figures 4.6 and 4.9 were reproduced from the Web sites of Drs. Norman and Garrett, respectively, with their kind permissions. Tables 3.2 and 3.4 were reproduced, in part, from Volume A and Volume B, respectively, of the *International Tables for Crystallography*, with copyright permission of the IUCr.

Uri Shmueli

This page intentionally left blank

Contents

1. Symmetry in crystals: fundamentals **1**
- 1.1 Introduction 1
- 1.2 The lattice plane 5
- 1.3 The reciprocal lattice 10
- 1.4 Permissible rotational symmetries of a three-dimensional lattice 13
- 1.5 Every lattice is centrosymmetric 16
- 1.6 Orientation of symmetry axes with respect to the lattice 17
- 1.7 Permissible combinations of axes of rotation 18
- 1.8 Crystal systems 22
- 1.9 Exercises for Chapter 1 26

2. Point groups and lattice types **29**
- 2.1 Introduction 29
- 2.2 Fundamentals of group theory 29
  - 2.2.1 Some important definitions 33
  - 2.2.2 Matrix representation of symmetry operations 33
  - 2.2.3 Isomorphism, coset decomposition, and the Lagrange theorem 34
- 2.3 Crystallographic point groups 37
  - 2.3.1 The axial point groups 38
  - 2.3.2 Completion of the presentation of the point groups 39
- 2.4 Bravais Lattices 45
- 2.5 Exercises for Chapter 2 47

3. Space-group symmetry **49**
- 3.1 Introduction 49
- 3.2 Space-group symmetry operators 49
- 3.3 Nonsymmorphic space-group operators 52
- 3.4 Some general considerations 59
- 3.5 Classification and tabulation of crystallographic space groups 65
- 3.6 Location of symmetry elements 65
- 3.7 Graphical representation of space groups 68
- 3.8 Some computational aspects 76
  - 3.8.1 Description of Table 3.4 - explicit space-group symbols 82
- 3.9 Exercises for Chapter 3 83

4. X-ray diffraction techniques **85**
- 4.1 Introduction 85

4.2 Diffraction conditions 86
4.2.1 The Laue and Bragg equations 86
4.2.2 Ewald's sphere of reflection 90
4.3 Production of X-rays 91
4.3.1 The X-ray tube 91
4.3.2 Synchrotron radiation 96
4.4 Detectors of X-rays 100
4.4.1 X-ray film 100
4.4.2 Imaging plate 101
4.4.3 Charge-coupled device (CCD) detector 102
4.5 The rotating-crystal method 102
4.6 Moving-crystal–moving-film methods 105
4.6.1 The Weissenberg method 106
4.6.2 The de Jong–Bouman method 106
4.6.3 The Buerger precession method 106
4.7 The four-circle diffractometer 107
4.7.1 Geometrical considerations 107
4.7.2 The orientation matrix 110
4.7.3 Coordinates and angles 111
4.7.4 Comments on the experiment 112
4.8 The Laue method 113
4.8.1 Principle of the method 113
4.8.2 Calculation of the Laue pattern 114
4.9 Exercises for Chapter 4 117

5. The structure factor and the electron density 119
5.1 Introduction 119
5.2 Scattering by a free charged particle 119
5.3 Scattering by a material unit 123
5.4 Scattering by a periodic array of material units 125
5.5 The atomic scattering factor 128
5.6 Space-group symmetry and the structure factor 130
5.7 The Fourier synthesis of the electron density 131
5.7.1 Computational considerations 133
5.8 The phase problem 134
5.9 Effects of space-group symmetry 135
5.9.1 Effect of the lattice type 138
5.9.2 Effect of screw axes and glide planes 139
5.10 Exercises for Chapter 5 142

6. The Patterson function 143
6.1 Introduction 143
6.2 The Patterson function 143
6.2.1 Periodicity of the Patterson function 145
6.2.2 The heavy atom method 146

6.2.3 Overlap of interatomic vectors 147
6.3 Classification of Patterson vectors 148
6.4 Symmetry of the Patterson function 149
6.5 Harker peaks 151
6.6 Overview of Advanced Applications 152
6.6.1 Superposition methods 152
6.6.2 Rotation functions 153
6.6.3 Translation functions 155
6.6.4 Molecular replacement 157
6.7 Exercises for Chapter 6 159

7. Structure-factor statistics 161
7.1 Introduction 161
7.2 Wilson's method of scaling the intensities 162
7.3 Basic Wilson statistics 164
7.3.1 Cumulative distributions 167
7.3.2 Moments of $|E|$ 168
7.4 Non-ideal structure-factor statistics 170
7.5 Exercises for Chapter 7 173

8. Direct methods 177
8.1 Introduction 177
8.2 Phase invariants and seminvariants 179
8.2.1 Phase invariants 179
8.2.2 Structure seminvariants 181
8.3 Specification of the origin 183
8.4 Inequalities 186
8.5 The Sayre equation and the tangent formula 190
8.6 Conditional probability density of a three-phase invariant 192
8.7 Other probabilistic considerations 198
8.8 Some practical comments on the solution of the phase problem 201

9. Atomic displacement parameters 207
9.1 Introduction 207
9.2 Atomic displacement parameters in the Gaussian approximation 210
9.2.1 Non-Cartesian representations 211
9.2.2 Cartesian representations 213
9.3 A rigid-body model 215
9.3.1 Introduction 215
9.3.2 Rotational displacement 215
9.3.3 Rigid-body model parameters 217

10. Refinement of structural parameters 219
10.1 Introduction 219
10.2 The linear least-squares method 219

10.2.1 An example of a linear least-squares problem 223
10.3 The non-linear least-squares method 224
10.4 Conventional refinement 226
10.4.1 Rigid-group constraints 228
10.4.2 Introduction of restraints 231

Appendix

A. Some geometrical considerations 235
A.1 Transformation to a Cartesian system 235
A.2 Change of basis and trace invariance 236
A.3 The finite rotation operator 237

B. Fundamentals of tensor notation 241
B.1 Introduction 241
B.2 Transformations 242
B.3 The scalar product and metric tensors 242
B.4 Examples 243

C. Basic notions from theory of probability 247
C.1 Introduction 247
C.2 Random variables 247
C.3 The probability density function and related quantities 248
C.4 Joint and conditional pdfs 249
C.5 Characteristic function and cumulants 250

D. The discrete Fourier transform 253
D.1 Some properties of the discrete Fourier transform 253
D.2 The Fast Fourier Transform 257

References 261

Index 267

1 Symmetry in crystals: fundamentals

1.1 Introduction

Crystalline solids possess a property which is absent from the amorphous, liquid and gaseous states, namely that their fundamental building blocks are arranged to a good approximation in a triply periodic array. Strict periodicity is, of course, an idealization, and the various (minor) deviations therefrom are subject to study in the advanced stages of structure determination. We shall, however, assume such periodicity while introducing the geometrical aspects of crystallography. The above-mentioned periodic arrangement of the building blocks underlies all of the geometry and symmetry of a crystal structure, and we shall try to describe it in some detail.

Consider a microscopic material unit; this can be an atom, a molecule or a group of molecules. Let us generate a row of identical units by shifting the above material unit as a whole, parallel to itself, through the vectors $\pm\mathbf{a}$, $\pm 2\mathbf{a}$, $\ldots$, $\pm n\mathbf{a}$, where n can be an arbitrarily large integer ($n \to \infty$). We obtain an infinite array of identical units in the same orientation, and such that any unit can be brought into coincidence with another unit in the row if it is translated as a whole and parallel to itself through some integer multiple of $\mathbf{a}$. Our row is a one-dimensional periodic arrangement of material units, the period being the vector $\mathbf{a}$, and the fact that any unit can be brought into coincidence with another one in the row by parallel translation through an integer multiple of $\mathbf{a}$ shows that the row possesses *translational symmetry*. The set of vectors $\{n\mathbf{a}\}$, where n is any integer, is called a one-dimensional vector lattice. It fully defines the periodicity of the one-dimensional arrangement, but its zero vector can be associated with any point within the starting or reference material unit.

Let us now choose a well-defined origin of the starting unit, for example its center of mass. It is clear that the position vectors of the centers of mass of all the units are elements of the infinite set $\{n\mathbf{a}\}$. The corresponding set of the endpoints of these position vectors is called a *one-dimensional point lattice*, a vector of this lattice being defined as $\mathbf{r}_u = u\mathbf{a}$, where u can be any integer (including zero). A more general element of a one-dimensional point lattice has the form $\mathbf{r} + u\mathbf{a}$, where $\mathbf{r}$ is the position vector of any point within the starting unit.

Let us now expand the row of material units into a sheet of such units by shifting the row as a whole, parallel to itself, through the vectors $\pm\mathbf{b}$, $\pm 2\mathbf{b}$, $\ldots$, $\pm n\mathbf{b}$, where n can be an arbitrarily large integer ($n \to \infty$) and the vector $\mathbf{b}$ is not collinear with $\mathbf{a}$, may differ from $\mathbf{a}$ in its magnitude, and does not have to be

perpendicular to it. We have now obtained a *two-dimensional periodic arrangement* which possesses a *two-dimensional translational symmetry*. The position vectors of the centers of mass of all the units are elements of the infinite set $\{p\mathbf{a}+q\mathbf{b}\}$, where each of p and q ranges over all the integers. The corresponding set of the endpoints of these position vectors is called a *two-dimensional point lattice*, a vector of this lattice now being defined as $\mathbf{r}_{uv} = u\mathbf{a} + v\mathbf{b}$, where u and v can be any integers (including zero).

We proceed, finally, to a triply periodic array of material units, which is the conventional representation of a three-dimensional ideal single crystal. This is obtained on expanding the sheet of material units by shifting the infinite sheet as a whole, parallel to itself through the vectors $\pm\mathbf{c}$, $\pm 2\mathbf{c}$, $\ldots$, $\pm n\mathbf{c}$, where n can be an arbitrarily large integer ($n \to \infty$) and the vector $\mathbf{c}$ is not collinear with either $\mathbf{a}$ or $\mathbf{b}$, may differ from them in its magnitude and does not have to form a right angle with either. We have now obtained a *three-dimensional periodic arrangement* which possesses a *three-dimensional translational symmetry*. The position vectors of the centers of mass of all the units are elements of the infinite set $\{p\mathbf{a} + q\mathbf{b} + r\mathbf{c}\}$, where each of p, q and r ranges over all the integers. The corresponding set of the endpoints of these position vectors is called a *three-dimensional point lattice*, a vector of this lattice now being defined as

$$\mathbf{r}_{uvw} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}, \tag{1.1}$$

where u, v, and w can be any integers (including zero). The complete structure of the three-dimensional arrangement is thus defined by the structure of a single material unit and the linear combinations of the three basis vectors of the vector lattice which represents the periodicity of that arrangement. It is customary in the crystallographic literature to denote the lattice vector given in eqn (1.1) by the symbol $[uvw]$. We shall use this notation occasionally, but the explicit notation will be preferred.

Although most of our discussions will involve three-dimensional lattices, it is often useful to simplify them by dealing with their two-dimensional projections. A portion of a two-dimensional periodic arrangement is illustrated in Fig. 1.1 by a drawing by the well known Dutch graphic artist M. C. Escher. The unit of the pattern (the equivalent of a material unit) can be chosen in several ways, for example as two horizontally adjacent white and gray horses. Once a choice has been made, it is easy to see how the whole pattern can be reconstructed by parallel translations of the unit chosen, by linear combinations of two basis vectors with integer coefficients. Since, however, all the horses have identical shapes we can also think of a translation followed by a change of color as a symmetry operation relating,for example, adjacent white and gray horses. We shall not deal in this book with the advanced subject of "black and white" or, in general, color symmetry, so well illustrated by some of Escher's drawings, but merely point out its existence. It has little to do with structure determination but may be of considerable importance in the study of the physical properties of crystals. Another question which suggests itself is: how shall we choose the

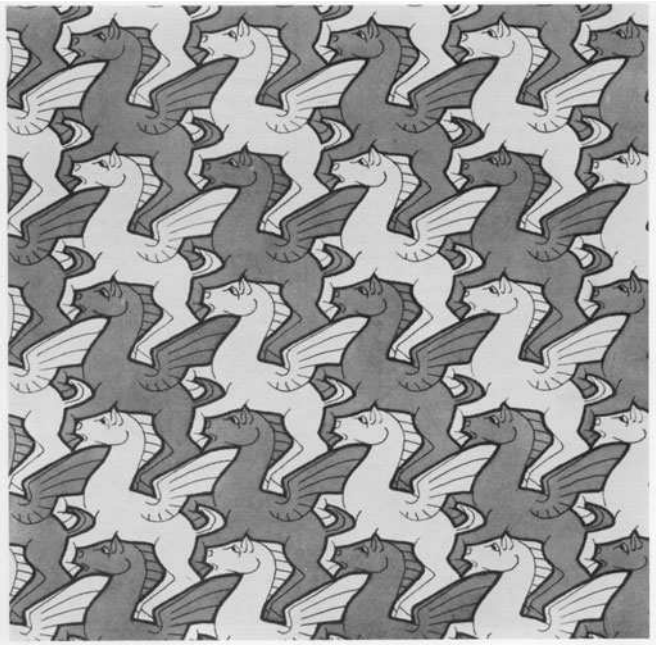


Fig. 1.1 A two-dimensional periodic pattern. Reproduced from MacGillavry (1965), with copyright permission of the International Union of Crystallography (IUCr).

origin of the point lattice? The answer is: it may be *any* point within a unit of the pattern. The above choice of the center of mass was merely a convenience. These questions are discussed further in the Exercises and in the following chapters. It is evident that in order to reconstruct the whole pattern in Fig. 1.1 all we need to know is the internal structure of the unit of the pattern, its origin, and the two basis vectors of its two-dimensional lattice, properly oriented with respect to the unit. Similarly, the structure and structure-dependent properties of three-dimensional crystals can be reconstructed if the internal structure of the fundamental material unit is known, and so are the basis vectors of the underlying three-dimensional lattice and its origin.

While basically correct, and qualitatively consistent with the above generation of the periodic arrangement, the above picture is not, however, suitable for a quantitative implementation. This is so, because we ultimately desire to relate the atomic positions–and more generally, the density function of scattering matter–to a convenient frame of reference. In order to achieve this, we shall first define the smallest volume which, when translated by all the lattice vectors, generates the three-dimensional space in which the lattice exists. This can be taken as a parallelepiped with edges coinciding with the basis vectors **a**, **b**, and **c**, and one of its corners taken as the origin of the lattice. This parallelepiped is called *the unit cell* (Fig. 1.2). Of course, all the eight corners of the unit cell are

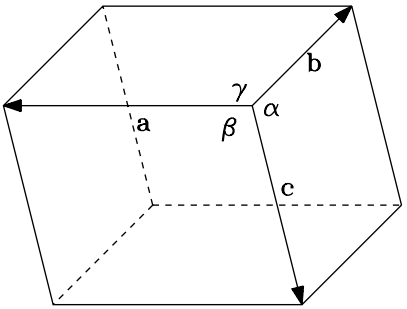


Fig. 1.2 The unit cell

lattice points, and each lattice point corresponds to a material unit. However, each such unit is shared by eight unit cells having in common a single lattice point. It follows that the unit cell contains just one material unit. This point is illustrated further in the Exercises. We note here that a unit cell containing lattice points only at its corners is called a *primitive unit cell*. This will be discussed further below. The lengths of the unit cell edges are simply designated by a, b, and c and the angles between the basis vectors **abc** are defined as follows: α is the angle between **b** and **c**, β is the angle between **c** and **a**, and γ is the angle between **a** and **b**. Thus in the general case the unit cell is described by the six quantities a, b, c, α, β, and γ which are known as the unit-cell parameters or, sometimes, unit-cell constants. It is important to remember that the vectors **a**, **b**, and **c** are conventionally taken as a right-handed triad.

The next step is the definition of a coordinate system (see Fig. 1.3). Let us choose the X axis along **a**, the Y axis along **b**, and the Z axis along **c**, and try to express the position vector of point P in Fig. 1.3 in terms of its coordinates $X_P = OA, Y_P = OB, Z_P = DP$, and the basis vectors **abc**. We have from the figure

$$\mathbf{r}_P = \mathbf{OA} + \mathbf{OB} + \mathbf{DP}. \tag{1.2}$$

But the vector **OA** is just X_P multiplied by a unit vector along **a**, etc. Hence

$$\mathbf{r}_P = X_P \frac{\mathbf{a}}{a} + Y_P \frac{\mathbf{b}}{b} + Z_P \frac{\mathbf{c}}{c}. \tag{1.3}$$

Now, eqn (1.3) can be readily rewitten as

$$\mathbf{r}_P = \frac{X_P}{a}\mathbf{a} + \frac{Y_P}{b}\mathbf{b} + \frac{Z_P}{c}\mathbf{c} \equiv x_P\mathbf{a} + y_P\mathbf{b} + z_P\mathbf{c} \tag{1.4}$$

with the following consequences: in eqn (1.3), the basis vectors are unit vectors, that is they are dimensionless, and the coordinates X_P, Y_P, Z_P have the dimension of length. These coordinates of a point within the unit cell may have values not exceeding the lengths of the corresponding unit-cell edges. In eqn (1.4), on

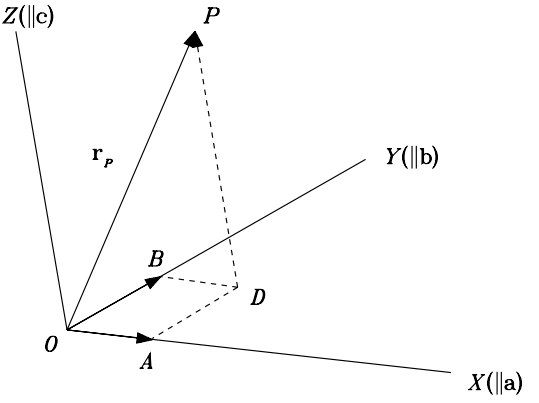


Fig. 1.3 A crystallographic system of coordinates.

the other hand, the basis vectors are just the basis vectors of the lattice and have dimensions of length, and their coefficients–the coordinates–are dimensionless. These coordinates of a point *within the unit cell* may have values between 0 and 1 only, and are called *fractional coordinates*. Of course, any point in the space spanned by the basis vectors **abc** can be described by its fractional coordinates. Only if each of the three fractional coordinates of a point is an integer does the point represent one of the lattice points. That is, the integers uvw in eqn (1.1) can be regarded as fractional coordinates of a lattice point. The great majority of the existing reports of crystal structures employ the fractional coordinates xyz in presenting atomic positions.

1.2 The lattice plane

We shall now deal with some fundamental properties of lattice planes, as these are indispensable for the understanding of the geometrical basis of some important diffraction phenomena. Only elementary vector algebra will be employed for this purpose and the discussion will be presented in the conventional notation.

A lattice plane is defined by three noncollinear lattice points. These three points define two lattice vectors, and the shortest lattice vectors in their directions can be taken as the basis vectors of a two-dimensional lattice. We shall proceed to show that a three-dimensional lattice can be represented as a family of parallel and equidistant lattice planes, find the general form of the equation of such a lattice plane, and derive an expression for the distance between any two successive lattice planes within the family. A schematic illustration of such

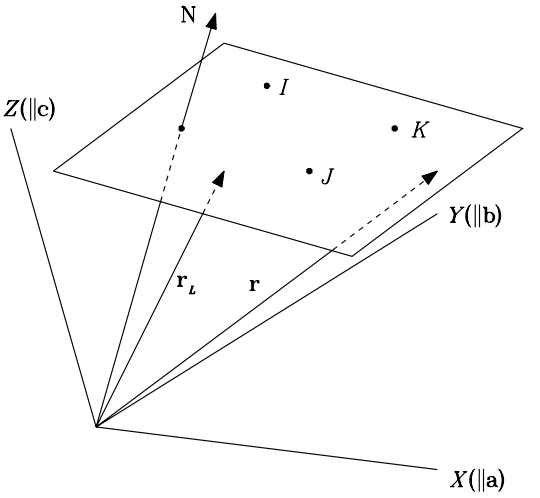


Fig. 1.4 A schematic illustration of a lattice plane.

a lattice plane is given in Fig. 1.4. Let us choose in the three-dimensional lattice, a set of primitive basis vectors, **a**, **b**, and **c**, such that the unit cell constructed from these three vectors is primitive, and fix the origin at a lattice point. The coordinate axes – to be used later on – will be denoted by X, Y, and Z, where X is parallel to **a**, Y is parallel to **b** and Z is parallel to **c**. The position vectors of the three lattice points determining the plane can be written as

$$\begin{aligned}\mathbf{r}_I &= u_{(I)}\mathbf{a} + v_{(I)}\mathbf{b} + w_{(I)}\mathbf{c}\\ \mathbf{r}_J &= u_{(J)}\mathbf{a} + v_{(J)}\mathbf{b} + w_{(J)}\mathbf{c}\\ \mathbf{r}_K &= u_{(K)}\mathbf{a} + v_{(K)}\mathbf{b} + w_{(K)}\mathbf{c},\end{aligned}$$

where, for example, $u_{(I)}v_{(I)}w_{(I)}$ are the fractional coordinates of the Ith lattice point. Since the basis vectors form a primitive unit cell the fractional coordinates of all the lattice points may be any integers. The above three lattice points define two in-plane lattice vectors,

$$\begin{aligned}\mathbf{Q} &= \mathbf{r}_I - \mathbf{r}_J\\ &= (u_{(I)} - u_{(J)})\mathbf{a} + (v_{(I)} - v_{(J)})\mathbf{b} + (w_{(I)} - w_{(J)})\mathbf{c}\\ &\equiv Q_x\mathbf{a} + Q_y\mathbf{b} + Q_z\mathbf{c}\end{aligned}$$

and

$$\begin{aligned}\mathbf{R} &= \mathbf{r}_K - \mathbf{r}_J\\ &= (u_{(K)} - u_{(J)})\mathbf{a} + (v_{(K)} - v_{(J)})\mathbf{b} + (w_{(K)} - w_{(J)})\mathbf{c}\\ &\equiv R_x\mathbf{a} + R_y\mathbf{b} + R_z\mathbf{c},\end{aligned}$$

where the components $Q_xQ_yQ_z$ and $R_xR_yR_z$ may be any integers. We can now proceed to formulate the equation of the lattice plane. A vector normal to the plane is given, for example, by the vector product of the two in-plane lattice vectors defined above. This plane normal is given by

$$\begin{aligned}\mathbf{N} &= \mathbf{R} \times \mathbf{Q} \\ &= (R_yQ_z - R_zQ_y)(\mathbf{b} \times \mathbf{c}) + (R_zQ_x - R_xQ_z)(\mathbf{c} \times \mathbf{a}) \\ &+ (R_xQ_y - R_yQ_x)(\mathbf{a} \times \mathbf{b}) \end{aligned} \tag{1.5}$$

and this can be rewritten briefly as

$$\mathbf{N} = h(\mathbf{b} \times \mathbf{c}) + k(\mathbf{c} \times \mathbf{a}) + l(\mathbf{a} \times \mathbf{b}), \tag{1.6}$$

where $h \equiv (R_yQ_z - R_zQ_y)$, $k \equiv (R_zQ_x - R_xQ_z)$, and $l \equiv (R_xQ_y - R_yQ_x)$, the coefficients of the vector products in eqn (1.5), are arbitrary integers. If, however, we divide h, k, and l in eqn (1.6) by their greatest common factor–other than $+1$ or -1–the resulting coefficients of the vector products in eqn (1.6) will be *relatively prime* integers and the direction of $\mathbf{N}$, the plane normal, will of course remain unchanged. This imparts a uniqueness to the plane normal, which will be essential to our discussion of lattice planes. We shall therefore assume, in this discussion, that h, k, and l in eqn (1.6) are relatively prime integers.

The orientation of the lattice plane is now determined, and we proceed to define its location. Let

$$\mathbf{r}_L = u\mathbf{a} + v\mathbf{b} + w\mathbf{c} \tag{1.7}$$

be the position vector of a lattice point lying in the plane and let

$$\mathbf{r} = \frac{X}{a}\mathbf{a} + \frac{Y}{b}\mathbf{b} + \frac{Z}{c}\mathbf{c} \tag{1.8}$$

be the position vector of a point (not necessarily a lattice point) lying in the plane. Its coordinates XYZ have dimensions of length and $\mathbf{r}$ is a lattice point if each of the X, Y, and Z is an integer multiple of the corresponding a, b and c respectively. The ratios X/a, Y/b and Z/c are the fractional coordinates of the point defined by the vector $\mathbf{r}$. Let us assume further that $\mathbf{r}$ and $\mathbf{r}_L$ do not coincide. It is obvious that the vector $\mathbf{r} - \mathbf{r}_L$ is parallel to the lattice plane, and therefore its scalar product with the plane normal must vanish:

$$\mathbf{N} \cdot (\mathbf{r} - \mathbf{r}_L) = 0 \tag{1.9}$$

or

$$\mathbf{N} \cdot \mathbf{r} = \mathbf{N} \cdot \mathbf{r}_L. \tag{1.10}$$

A straightforward substitution of eqn (1.6), eqn (1.7) and eqn (1.8) into eqn (1.10) leads to mixed products, such as $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ and $\mathbf{a} \cdot (\mathbf{c} \times \mathbf{a})$. These mixed products vanish if any two vectors are the same, and equal the volume of the unit cell, V,

if the vectors are a cyclic permutation of **a**, **b** and **c**. Only cyclic permutations are obtained in the triplets of unequal basis vectors. We therefore obtain

$$\left(h\frac{X}{a}+k\frac{Y}{b}+l\frac{Z}{c}\right)V=(hu+kv+lw)V. \tag{1.11}$$

Since V cancels out, and the expression in parentheses in the right-hand side of eqn (1.11) can be an arbitrary integer, the equation of the above lattice plane becomes

$$h\frac{X}{a}+k\frac{Y}{b}+l\frac{Z}{c}=n. \tag{1.12}$$

The relatively prime integers hkl are known as *Miller indices* of the lattice plane, XYZ are the coordinates of the point **r** (with dimensions of length) (see eqn (1.8)), and n may assume any integer value. We thus have a family of lattice planes, with the same Miller indices hkl but differing in the value of n. All these planes belonging to the (hkl) family are obviously parallel (they have the same plane normal), but it remains to be shown that they are also equidistant.

Assume that n is positive, and locate the intercepts of the plane in eqn (1.12) on the coordinate axes X, Y, and Z. We have

$$\begin{aligned}
Y=Z=0 &\rightarrow X_n=\frac{na}{h},\\
Z=X=0 &\rightarrow Y_n=\frac{nb}{k},\\
X=Y=0 &\rightarrow Z_n=\frac{nc}{l}.
\end{aligned}$$

The intercepts of the $(n-1)$th plane will be $(n-1)a/h$, $(n-1)b/k$ and $(n-1)c/l$, and so on for other intercepts. The distance between two successive intercepts will be a/h on the X axis, b/k on the Y axis and l/c on the Z axis, from which it readily follows that the adjacent parallel planes in the (hkl) family are equidistant.

Let us now find an expression for this interplanar distance. Rather than considering distances between successive intercepts, let us consider the vector leading from one intercept to the next along one of the axes; on the X axis this will be $\mathbf{a}/h$ etc. The distance between two successive lattice planes in the (hkl) family is just the projection of the vector $\mathbf{a}/h$ on the direction of the plane normal, and this is simply found as a scalar product of the vector to be projected onto the direction of **N** and a unit vector along this direction. If we denote the interplanar distance by d_{hkl}, we have

$$d_{hkl}=\frac{\mathbf{a}}{h}\cdot\frac{\mathbf{N}}{|\mathbf{N}|}. \tag{1.13}$$

However, we have from eqn (1.6)

$$\begin{aligned}\frac{\mathbf{a}}{h}\cdot\mathbf{N} &= \frac{1}{h}\{h[\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})]+k[\mathbf{a}\cdot(\mathbf{c}\times\mathbf{a})]+[\mathbf{a}\cdot(\mathbf{a}\times\mathbf{b})]\}\\ &= \frac{1}{h}h[\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})]\\ &= V.\end{aligned}$$

It follows that

$$d_{hkl} = \frac{V}{|\mathbf{N}|} = \frac{1}{|\mathbf{N}/V|} \equiv \frac{1}{|\mathbf{h}|}, \tag{1.14}$$

where

$$\mathbf{h} = h\frac{\mathbf{b}\times\mathbf{c}}{V} + k\frac{\mathbf{c}\times\mathbf{a}}{V} + l\frac{\mathbf{a}\times\mathbf{b}}{V}. \tag{1.15}$$

The vector defined in eqn (1.15) is perpendicular to the family of (hkl) planes and its magnitude is the reciprocal of the interplanar distance of this family. We shall generalize the geometrical properties of $\mathbf{h}$ in the next section, and encounter its physical and mathematical interpretations in the following chapters of this book. In the meantime, let us apply it to the calculation of the interplanar distance for the simplest example of a cubic lattice. We have here: $a = b = c \equiv a_0$, $\alpha = \beta = \gamma = 90^\circ$ and $\mathbf{a} = a_0\mathbf{i}, \mathbf{b} = a_0\mathbf{j}, \mathbf{c} = a_0\mathbf{k}$, where $\mathbf{ijk}$ are Cartesian unit vectors in the directions of $\mathbf{abc}$, respectively. The expression for $\mathbf{h}$ becomes

$$\mathbf{h} = \frac{h}{a_0}\mathbf{i} + \frac{k}{a_0}\mathbf{j} + \frac{l}{a_0}\mathbf{k},$$

and we obtain for the interplanar distance

$$d_{hkl} = \frac{1}{|\mathbf{h}|} = \frac{1}{(\mathbf{h}\cdot\mathbf{h})^{1/2}} = \frac{a_0}{\sqrt{h^2+k^2+l^2}}. \tag{1.16}$$

Expressions for interplanar distances in other crystal systems can be readily obtained if we make use of the relationships between direct and reciprocal bases (see below).

We conclude this section by indicating a convenient method for graphical representation of lattice planes. Because of the periodicity of the lattice, this can be done for the planes with $n = 1$ and $n = 0$. The intercepts of the lattice plane with $n = 1$ on the X, Y and Z axes are, according to the above discussion, a/h, b/k and c/l respectively, and can therefore be found within one of the eight unit cells about the origin. However, one unit cell is sufficient if we are free to shift its origin to each of its eight corners (for the purpose of graphical representation only). Several examples of such lattice planes are shown in Fig. 1.5. According to the above, a concise definition of the Miller indices is the following: *A lattice plane with Miller indices hkl, conventionally denoted by (hkl), intercepts the X axis at a/h, the Y axis at b/k and the Z axis at c/l.* For example, the (312) plane intercepts the X axis at a/3, the Y axis at b and the Z axis at c/2. However, the above definition could in principle be applicable to any hkl, but would they always describe a lattice plane? Let us consider, as an illustration, the (100)

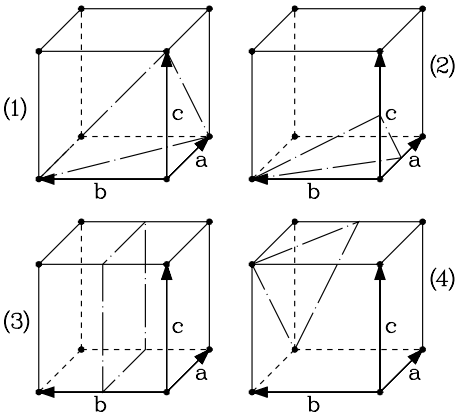


Fig. 1.5 Lattice planes with n=1. The planes are (1)(111), (2)(212), (3)(020), (4)($\overline{121}$). Note that (020) does not pass through any lattice point and is therefore not a lattice plane.

and (200) planes, for $n = 1$. The (100) plane intercepts the X axis at a, and is parallel to the Y and Z axes. This plane and all the remaining ones in the (100) family pass through lattice points in planes parallel to the (**b**, **c**) plane. The (200) plane intercepts the X axis at $a/2$, and is also parallel to the Y and Z axes. However, *it does not pass through any lattice point*. We would find the same result for the (222) plane and, in general, for any plane (hkl) (with $n = 1$) for which the indices h, k, and l are not relatively prime. The above argument serves the purpose of an illustration, but does not constitute a rigorous proof that the indices h, k, and l represent a family of lattice planes *if and only if* h, k, and l are relatively prime integers. Such a proof, employing number-theoretical arguments, can be found in the article of Deas and Hamill (1957) but is outside the scope of our treatment.

1.3 The reciprocal lattice

The above discussion of families of lattice planes and of interplanar distances within those families led to the appearance of the vector **h**, given by eqn (1.15). It is interesting to consider this vector and its components in some detail, in view of its numerous applications in crystallography.

Let us first rewrite **h** as

$$\mathbf{h} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*, \tag{1.17}$$

where

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{V}, \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{V}, \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{V}, \tag{1.18}$$

and V is the volume of the unit cell.

Let us now allow hkl to take on any integer values, that is we shall not restrict them to relatively prime integers. The vector $\mathbf{h}$ in eqn (1.17) now becomes a linear combination of the basis vectors $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$, the coefficients of this combination being any integers. Mathematically, this means that $\mathbf{h}$ in eqn (1.17) is a lattice vector, and is in this sense analogous to $\mathbf{r}_{uvw}$ in eqn (1.1). However, these two lattices are different: $\mathbf{r}_{uvw}$ and its basis vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ have dimensions of *length*, while $\mathbf{h}$ and its basis vectors $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$ have dimensions $length^{-1}$. For this reason, as well as others which will become clear later, we call the set of all the points with position vectors

$$u\mathbf{a} + v\mathbf{b} + w\mathbf{c},$$

where uvw are any integers, the *direct lattice*, and the set of all the points with position vectors

$$h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*,$$

where hkl are any integers, the *reciprocal lattice*.

We shall now proceed to inspect the relationships between the bases of the direct and reciprocal lattices. It is evident from eqn (1.18) that $\mathbf{a}^*$ is perpendicular to $\mathbf{b}$ and to $\mathbf{c}$, $\mathbf{b}^*$ is perpendicular to $\mathbf{c}$ and to $\mathbf{a}$, and $\mathbf{c}^*$ is perpendicular to $\mathbf{a}$ and to $\mathbf{b}$. We therefore have

$$\mathbf{a}^* \cdot \mathbf{b} = \mathbf{a}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{b} = 0. \tag{1.19}$$

We further have

$$\mathbf{a} \cdot \mathbf{a}^* = \mathbf{a} \cdot \frac{\mathbf{b} \times \mathbf{c}}{V} = \frac{V}{V} = 1 \tag{1.20}$$

and similarly for the scalar products involving $\mathbf{b}$ and $\mathbf{c}$. That is

$$\mathbf{a} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{b}^* = \mathbf{c} \cdot \mathbf{c}^* = 1. \tag{1.21}$$

Equation (1.19) indicates further that $\mathbf{a} \propto \mathbf{b}^* \times \mathbf{c}^*$, $\mathbf{b} \propto \mathbf{c}^* \times \mathbf{a}^*$ and $\mathbf{c} \propto \mathbf{a}^* \times \mathbf{b}^*$. If we combine this result with eqn (1.21), we obtain

$$\mathbf{a} = \frac{\mathbf{b}^* \times \mathbf{c}^*}{V^*}, \quad \mathbf{b} = \frac{\mathbf{c}^* \times \mathbf{a}^*}{V^*}, \quad \mathbf{c} = \frac{\mathbf{a}^* \times \mathbf{b}^*}{V^*}, \tag{1.22}$$

where

$$V^* = \mathbf{a}^* \cdot (\mathbf{b}^* \times \mathbf{c}^*) = \mathbf{b}^* \cdot (\mathbf{c}^* \times \mathbf{a}^*) = \mathbf{c}^* \cdot (\mathbf{a}^* \times \mathbf{b}^*)$$

is the volume of the unit cell of the reciprocal lattice, based on the vectors $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$.

Note further that

$$\begin{aligned} V^* &= \mathbf{a}^* \cdot (\mathbf{b}^* \times \mathbf{c}^*) \\ &= \frac{(\mathbf{b} \times \mathbf{c}) \cdot (\mathbf{b}^* \times \mathbf{c}^*)}{V} \\ &= (\mathbf{b} \cdot \mathbf{b}^*)(\mathbf{c} \cdot \mathbf{c}^*) - (\mathbf{b} \cdot \mathbf{c}^*)(\mathbf{c} \cdot \mathbf{b}^*)V \\ &= \frac{1}{V}. \end{aligned} \tag{1.23}$$

That is, the volumes of the direct-lattice and reciprocal-lattice unit cells also turn out to be mutually reciprocal. Here use has been made of the standard vector-algebraic expression for a scalar product of two vector products and of eqs (1.18), (1.19), and (1.21).

Let us now find an expression for V in terms of the unit-cell parameters $a, b, c, \alpha, \beta, \gamma$. A straightforward approach to the calculation of the volume of a parallelepiped in terms of its edges and angles is rather cumbersome, and we shall follow the method of Buerger (1941) which is far simpler. If we refer the basis vectors **a**, **b**, and **c** to a Cartesian system, say

$$\begin{aligned} \mathbf{a} &= \mathbf{i}a_x + \mathbf{j}a_y + \mathbf{k}a_z, \\ \mathbf{b} &= \mathbf{i}b_x + \mathbf{j}b_y + \mathbf{k}b_z, \\ \mathbf{c} &= \mathbf{i}c_x + \mathbf{j}c_y + \mathbf{k}c_z, \end{aligned} \tag{1.24}$$

then the volume of the parallelepiped constructed from these vectors can be readily shown to equal

$$V = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}. \tag{1.25}$$

If we now make use of the fact that a transposition of the rows and columns of a determinant does not affect its value, the squared volume of the parallelepiped can be computed as

$$V^2 = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix} \begin{vmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{vmatrix} = \begin{vmatrix} \mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} & \mathbf{a}\cdot\mathbf{c} \\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{c} \\ \mathbf{c}\cdot\mathbf{a} & \mathbf{c}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c} \end{vmatrix} = \begin{vmatrix} a^2 & ab\cos\gamma & ac\cos\beta \\ ba\cos\gamma & b^2 & bc\cos\alpha \\ ca\cos\beta & cb\cos\alpha & c^2 \end{vmatrix}, \tag{1.26}$$

and upon expanding the rightmost determinant in eqn (1.26) and taking the square root we obtain

$$V = abc(1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma)^{1/2}. \tag{1.27}$$

An analogous expression for the volume of the reciprocal unit cell can be obtained as

$$V^* = a^* b^* c^* (1 - \cos^2 \alpha^* - \cos^2 \beta^* - \cos^2 \gamma^* + 2 \cos \alpha^* \cos \beta^* \cos \gamma^*)^{1/2}. \quad (1.28)$$

We now, finally, find the relationship between the interaxial angles of the reciprocal unit cell and of the real unit cell. We can do this, for example, by comparing different expressions for the same scalar products of two basis vectors. We have

$$\mathbf{b}^* \cdot \mathbf{c}^* = \frac{|\mathbf{c} \times \mathbf{a}|}{V} \frac{|\mathbf{a} \times \mathbf{b}|}{V} \cos \alpha^* = \frac{a^2 bc}{V^2} \sin \beta \sin \gamma \cos \alpha^* \quad (1.29)$$

and, on the other hand

$$\begin{aligned} \mathbf{b}^* \cdot \mathbf{c}^* &= \frac{\mathbf{c} \times \mathbf{a}}{V} \cdot \frac{\mathbf{a} \times \mathbf{b}}{V} \\ &= \frac{(\mathbf{c} \cdot \mathbf{a})(\mathbf{a} \cdot \mathbf{b}) - (\mathbf{c} \cdot \mathbf{b})(\mathbf{a} \cdot \mathbf{a})}{V^2} \\ &= \frac{a^2 bc}{V^2} (\cos \beta \cos \gamma - \cos \alpha). \end{aligned} \quad (1.30)$$

If we equate eqn (1.29) and eqn (1.30), we obtain

$$\cos \alpha^* = \frac{\cos \beta \cos \gamma - \cos \alpha}{\sin \beta \sin \gamma}. \quad (1.31)$$

Analogous calculations lead to

$$\cos \beta^* = \frac{\cos \gamma \cos \alpha - \cos \beta}{\sin \gamma \sin \alpha}, \quad (1.32)$$

$$\cos \gamma^* = \frac{\cos \alpha \cos \beta - \cos \gamma}{\sin \alpha \sin \beta}, \quad (1.33)$$

and to the inverse relationships

$$\cos \alpha = \frac{\cos \beta^* \cos \gamma^* - \cos \alpha^*}{\sin \beta^* \sin \gamma^*}, \quad (1.34)$$

$$\cos \beta = \frac{\cos \gamma^* \cos \alpha^* - \cos \beta^*}{\sin \gamma^* \sin \alpha^*}, \quad (1.35)$$

$$\cos \gamma = \frac{\cos \alpha^* \cos \beta^* - \cos \gamma^*}{\sin \alpha^* \sin \beta^*}. \quad (1.36)$$

The relations in eqns (1.31)–eqn (1.36) can also be obtained by methods of spherical trigonometry (see, for example, Buerger 1956).

1.4 Permissible rotational symmetries of a three-dimensional lattice

The universally present symmetry of a lattice is, by definition, a translation by a lattice vector. However, a lattice may also be symmetric with respect to a rotation about a certain direction, through certain angular intervals. For example,

any two-dimensional lattice is symmetric with respect to a rotation through 180° about an axis perpendicular to the lattice and passing through a lattice point, a square two-dimensional lattice is symmetric with respect to rotation through 90° about a perpendicular axis and possesses additional symmetry elements, and so on. It is well known that the permissible angles of rotation about axes of symmetry of a three-dimensional lattice are multiples of 360°, 180°, 120°, 90°, and 60° *only*, and many illustrations of this result have been presented in the crystallographic literature. Rigorous proofs of the above are rather scarce, however, and we shall try to present one, along the lines of that given by Zachariasen (1945). Most of the mathematical details involved in this proof will, however, be collected in Appendix A, which will also be referred to from other chapters of the book.

Let a three-dimensional lattice be symmetric with respect to a rotation through an angle α about some, as yet unspecified, direction. This means that the lattice is brought by this rotation into self-coincidence and that each vector of the lattice before the rotation took place is carried into a vector of the rotated lattice. Let us assume that the basis vectors have been chosen so that they define a primitive unit cell, and their coefficients in any lattice vector $\mathbf{r}_L = u\mathbf{a}+v\mathbf{b}+w\mathbf{c}$ are therefore necessarily integers. The operation of rotation can be written as

$$\begin{pmatrix} u_f \\ v_f \\ w_f \end{pmatrix} = \begin{pmatrix} P_{11}(\alpha)\ P_{12}(\alpha)\ P_{13}(\alpha) \\ P_{21}(\alpha)\ P_{22}(\alpha)\ P_{23}(\alpha) \\ P_{31}(\alpha)\ P_{32}(\alpha)\ P_{33}(\alpha) \end{pmatrix} \begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix}, \tag{1.37}$$

where the subscript i denotes the lattice vector before the rotation, $P_{kl}(\alpha)$ is an element of the matrix of the symmetric rotation acting on the components of $\mathbf{r}_{Li} = u_i\mathbf{a}+v_i\mathbf{b}+w_i\mathbf{c}$ and the subscript f denotes the lattice vector resulting from this rotation. Since the whole lattice is symmetric with respect to the rotation $\boldsymbol{P}(\alpha)$, the rotation in eqn (1.37) must carry an infinite number of triples of integers into triples of integers. This can be so only if all the elements of the rotation matrix $\boldsymbol{P}(\alpha)$ are also integers. In particular, we must have

$$\mathrm{Tr}(\boldsymbol{P}) = P_{11}(\alpha) + P_{22}(\alpha) + P_{33}(\alpha) = n, \tag{1.38}$$

where n is an integer. We show in Appendix A that the trace of a square matrix is invariant under a transformation of the basis to which the matrix is referred, provided the transformation matrix is nonsingular. We also show there that such a transformation from the crystal to a Cartesian system can be readily found. We then find the general form of an operator that performs a rotation through an angle α about an axis coinciding with a given unit vector. When this operator is referred to a Cartesian system, it is seen that its trace is given by $1 + 2\cos\alpha$, and thus depends on the angle of rotation alone. Since, in our case, the trace of the matrix of symmetric rotation must be an integer, we have

$$\mathrm{Tr}(\boldsymbol{P}) = 1 + 2\cos\alpha = m, \tag{1.39}$$

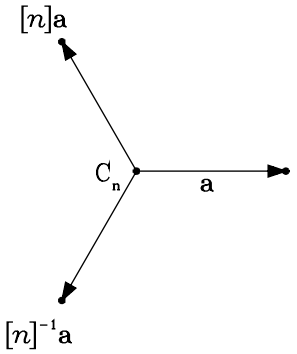


Fig. 1.6 Permissible symmetric rotations of a two-dimensional lattice.

where m is an integer. This integer is of course restricted, since

$$-1 \leq \cos\alpha = \frac{m-1}{2} \leq 1. \tag{1.40}$$

The possible values of m are: -1, 0, 1, 2 and 3 and these values correspond to $\alpha = 180°$, $120°$, $90°$, $60°$ and $360°$, respectively, which was to be shown.

The above result is obtained in the crystallographic literature by a variety of methods, which vary from the rigorous method outlined above to qualitative justifications. It may be instructive to present also another geometrical derivation, which, however, deals with the permissible symmetric rotations of a two-dimensional lattice about a normal to its plane.

Consider the vector diagram presented in Fig. 1.6. We assume that:

1. The lattice plane in the figure is symmetric with respect to a rotation by $2\pi/n$ about the normal to the plane.
2. The origin is chosen on the axis of rotation.
3. The vector $\mathbf{a}$ is the shortest in-plane lattice vector.

If we denote the operator of rotation through $2\pi/n$ about the symmetry axis by $[n]$, the vectors $[n]\mathbf{a}$ and $[n]^{-1}\mathbf{a}$ must also be in-plane lattice vectors, their magnitudes are equal to $|\mathbf{a}|$, and their sum

$$[n]\mathbf{a} + [n]^{-1}\mathbf{a} \equiv \mathbf{r}_L \tag{1.41}$$

is also an in-plane lattice vector, which must be parallel or antiparallel to $\mathbf{a}$. Also, since $\mathbf{a}$ is the shortest in-plane lattice vector, $\mathbf{r}_L$ in eqn (1.41) must be an integer multiple of $\mathbf{a}$, that is. $\mathbf{r}_L = k\mathbf{a}$, where k is an integer. Of course, we

also have $|\mathbf{r}_L| = k|\mathbf{a}|$. Because of the orientation of $\mathbf{r}_L$, the magnitude $|\mathbf{r}_L|$ is the projection of $\mathbf{r}_L$ on the direction of $\mathbf{a}$,

$$\begin{aligned} |\mathbf{r}_L| &= \mathbf{r}_L \cdot \frac{\mathbf{a}}{|\mathbf{a}|} \\ &= ([n]\mathbf{a} + [n]^{-1}\mathbf{a}) \cdot \frac{\mathbf{a}}{|\mathbf{a}|} \\ &= 2|\mathbf{a}| \cos\frac{2\pi}{n}, \end{aligned} \tag{1.42}$$

since the angles which are formed by $[n]\mathbf{a}$ and $\mathbf{a}$ and by $[n]^{-1}\mathbf{a}$ and $\mathbf{a}$ are $2\pi/n$ and $-2\pi/n$, respectively. It follows from the above that $2\cos(2\pi/n) = k$, where k is an integer. Finally,

$$-1 \leq \cos\frac{2\pi}{n} = \frac{k}{2} \leq +1.$$

The possible values of k are -2, -1, 0, 1, and 2, and to these correspond the values of n 2, 3, 4, 6, and 1, respectively.

We have arrived at the same result as that obtained by the first method; however, only a two-dimensional lattice was considered here. Of course, no other values of n, that is other orders of rotation, are possible in a three-dimensional lattice, since they would have to be obeyed by its two-dimensional projections (which are two-dimensional lattices).

1.5 Every lattice is centrosymmetric

A three-dimensional lattice possesses, by definition, translational symmetry and can also have a restricted number of rotational symmetries, as shown in the previous section. There is, however, another kind of symmetry which must be obeyed by all possible lattices, independently of their (integral) dimensionality.

Let $\{\mathbf{a}_i\}$ be a set of linearly independent basis vectors. A typical lattice vector, constructed from this basis, is a linear combination of the basis vectors, with integer coefficients. That is, it has the form

$$\mathbf{r}_L = \sum_{i=1}^{N} u_i \mathbf{a}_i \tag{1.43}$$

where $u_i, i = 1, \ldots, N$ are any integers. We can therefore find for each set of $\{u_i\}$ a set of $\{-u_i\}$ or, in other words, to each lattice point with position vector $\mathbf{r}_L$ there corresponds in the lattice a point with position vector $-\mathbf{r}_L$. It follows that if the lattice is inverted through the origin, chosen to be at a lattice point, it is brought into self-coincidence. The point through which the lattice is symmetrically inverted is called a center of symmetry of the lattice. Of course, each lattice point is a center of symmetry of the lattice and, as will be seen later, there are more such points.

Returning to the case of the three-dimensional lattice and the conventional notation, to each lattice point with coordinates uvw there corresponds a lattice point with coordinates $\overline{uvw}$, where uvw are all integers. The possible locations of centers of symmetry will be discussed further in the Exercises.

1.6 Orientation of symmetry axes with respect to the lattice

The above fundamental features of the possible lattice symmetries are the building blocks of the theory of crystal symmetry, which will be dealt with in the next two chapters. There exist, however, restrictions on the possible orientations of symmetry axes and planes with respect to the lattice, which are indispensable for convenient classification of crystal symmetries. We shall deal here with axes only, and defer the analogous treatment of planes to the Exercises. Consider a line in a lattice which coincides with an axis of rotational symmetry (see Fig. 1.7). Let the corresponding symmetry operation be a rotation through $2\pi/n$, where n may equal 2, 3, 4, or 6, as shown above (the trivial value $n = 1$ being omitted), and let the corresponding operator be denoted by $\mathbf{P}_n$. Let $\mathbf{r}_L$ be a lattice vector, referred to an origin on the axis of symmetry, such that $\mathbf{r}_L$ is neither perpendicular nor parallel to the axis. By definition, $\mathbf{P}_n\mathbf{r}_L$,$\mathbf{P}_n^2\mathbf{r}_L$, ..., $\mathbf{P}_n^{n-1}\mathbf{r}_L$ are also lattice vectors, and so is their sum. Let us now decompose $\mathbf{r}_L$ into components parallel and perpendicular to the axis, say $\mathbf{r}_L = \mathbf{r}_\parallel + \mathbf{r}_\perp$. It can be shown that the perpendicular components

$$(\mathbf{I} + \mathbf{P}_n + \cdots + \mathbf{P}_n^{n-1})\mathbf{r}_\perp,$$

where $\mathbf{I}$ is a unit operator, must equal zero for $n = 2$, 3, 4 or 6. Therefore, the lattice vector

$$(\mathbf{I} + \mathbf{P}_n + \cdots + \mathbf{P}_n^{n-1})\mathbf{r}_L = (\mathbf{I} + \mathbf{P}_n + \cdots + \mathbf{P}_n^{n-1})\mathbf{r}_\parallel = n\mathbf{r}_\parallel$$

must be parallel to the axis of symmetry. Summing up, if a lattice is symmetric with respect to a (permissible) rotation, there must exist a lattice line parallel to that axis.

Let us now choose two lattice vectors, $\mathbf{r}'_L$ and $\mathbf{r}''_L$, each of which is neither parallel nor perpendicular to the axis of symmetry, and which are not coplanar with that axis, and decompose them as before into components that are parallel and perpendicular to the axis. The vectors

$$\mathbf{r}'_\parallel - \mathbf{P}_n\mathbf{r}'_\parallel \quad \text{and} \quad \mathbf{r}''_\parallel - \mathbf{P}_n\mathbf{r}''_\parallel$$

obviously vanish, and

$$\mathbf{r}'_\perp - \mathbf{P}_n\mathbf{r}'_\perp \quad \text{and} \quad \mathbf{r}''_\perp - \mathbf{P}_n\mathbf{r}''_\perp$$

are lattice vectors perpendicular to the axis of symmetry. If we take the shortest lattice vectors in the directions of those vectors, and all their linear combinations

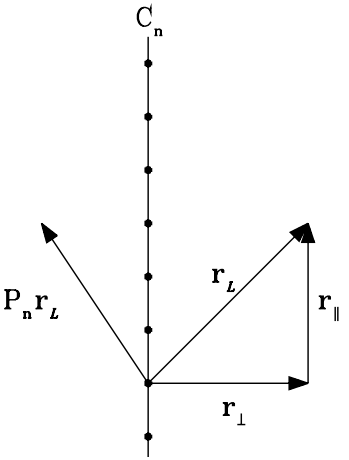


Fig. 1.7 Orientation of symmetry axes.

with integer coefficients, we generate a lattice plane which is perpendicular to the axis of symmetry. It is interesting to note that if a unit cell is constructed from these shortest vectors it may contain a lattice point at its center (see Fig. 1.8). Such a unit cell is known as *nonprimitive* and will be discussed repeatedly later. It follows from the above considerations that *if a lattice is symmetric with respect to a rotation about an axis, the lattice must contain a lattice line parallel to that axis of symmetry and a lattice plane perpendicular to it.* This theorem, and a similar one related to symmetry planes to be dealt with in the Exercises, are the basis of the conventional classification of crystal systems which will conclude this chapter.

1.7 Permissible combinations of axes of rotation

In order to characterize the full symmetry of the direct lattice, we have to solve two additional problems. First, what are the admissible pairs of differently oriented allowed axes of rotation, such that their successive application still leaves the lattice unchanged? Second, what additional symmetry operations arise as a consequence of the ever-present center of symmetry of a lattice (see above)? We shall solve the first of these problems in this section, and defer the solution of the second one to the next chapter, which will be largely devoted to the classification of crystallographic point groups.

If a lattice is symmetric with respect to two rotations about differently oriented axes, the product of these two operations must also be a symmetry op-

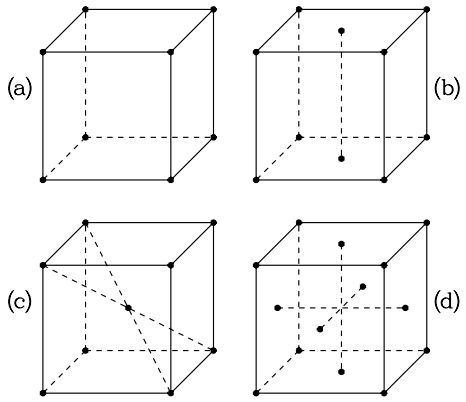


Fig. 1.8 Primitive and nonprimitive unit cells.

eration of the lattice, and this operation must be a rotation about a certain direction. Let us denote an operator corresponding to a rotation through an angle $2\pi/n$ by $[n]$. Suppose $[2']$ and $[2'']$ correspond to two twofold axes of symmetric rotation, passing through the origin of the lattice and forming an angle ϕ (see Fig. 1.9). An application of the operator $[2']$ to the vector $\mathbf{r}_L$ yields the vector $[2']\mathbf{r}_L$ (see Fig. 1.9), and a subsequent application of the operator $[2'']$ to the vector $[2']\mathbf{r}_L$ yields the vector $[2'']\,[2']\mathbf{r}_L$. It is readily seen that the endpoints of the vectors $\mathbf{r}_L$ and $[2'']\,[2']\mathbf{r}_L$ are at the same height above the plane containing the two twofold axes. These points are therefore simply related by a rotation about an axis which is perpendicular to the plane containing the two twofold axes. Figure 1.9 shows further that the angle of that rotation is exactly twice as large as the angle enlosed between the two two-fold axes, that is, it equals 2ϕ. We know, however, that the angle of rotation (apart from the trivial 360°) can only be 180°, 120°, 90°, or 60°, and therefore the values of ϕ are restricted to 90°, 60°, 45° or 30°, respectively. Summing up, we have

$$\mathbf{r}''_L = [2'']\,[2']\mathbf{r}_L = [n]\mathbf{r}_L, \tag{1.44}$$

where $n = 2$, 3, 4, or 6 for $\phi = \pi/2$, $\pi/3$, $\pi/4$, or $\pi/6$, respectively. The axis of $[n]$ is perpendicular to the plane containing the axes of $[2']$ and $[2'']$, and the following four important allowable combinations are obtained: $[2][2'][2'']$, $[3][2'][2'']$, $[4][2'][2'']$, and $[6][2'][2'']$.

The consideration of allowable pairs of axes where at least one is of order higher than 2 is more complicated and is best carried out in a general manner. Let $[p]$ and $[q]$ be rotation operators corresponding to symmetric rotations through

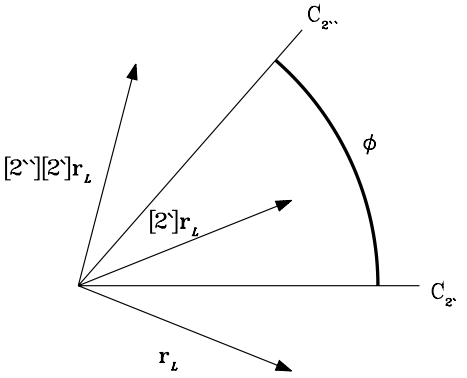


Fig. 1.9 *Permissible combinations of two-fold axes of rotation*

$2\pi/p$ and $2\pi/q$, respectively, and let $\mathbf{r}_L$ be a lattice vector not coinciding with either of the axes of rotation. We then have

$$\mathbf{r}'_L = [p]\mathbf{r}_L$$

$$\mathbf{r}''_L = [q]\mathbf{r}'_L$$

or

$$\mathbf{r}''_L = [q][p]\mathbf{r}_L \equiv [n]\mathbf{r}_L, \tag{1.45}$$

where $[n]$ is a symmetry operator, equivalent to the product $[q][p]$ and corresponding to a rotation through $2\pi/n$, with $n = 2$, 3, 4, or 6. We further have

$$\mathrm{Tr}([n]) = \mathrm{Tr}([q][p]) = 1 + 2\cos(2\pi/n), \tag{1.46}$$

as shown in Appendix A. The trace of $[n]$ thus depends on the product of the representations of the operators $[p]$ and $[q]$ and this, in turn, depends on the angle between the corresponding axes of rotation and on their orders. We shall therefore derive an explicit expression for the trace of the matrix representation of $[q][p]$ and try to form an equation for the angle between the axes of $[p]$ and $[q]$, given fixed orders of $[p]$, $[q]$, and $[n]$.

In view of the trace invariance discussed in Appendix A, we can refer the calculation to a Cartesian system. Let the axes of rotation corresponding to the operators $[p]$ and $[q]$ be located in the XY plane of a Cartesian system, and be denoted by C_p and C_q, respectively. The axes are assumed to form an angle φ and to intersect at the origin of the Cartesian system. Suppose that C_p coincides

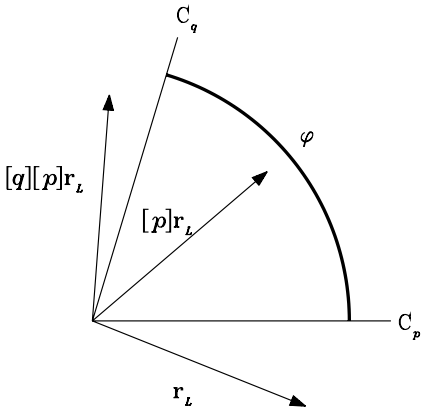


Fig. 1.10 *Permissible combinations of axes of rotation*

with the X axis (see Fig. 1.10). The matrix representation of the operator $[p]$ is then

$$[p] : \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_p & -s_p \\ 0 & s_p & c_p \end{pmatrix}, \tag{1.47}$$

where $c_p \equiv \cos(2\pi/p)$ and $s_p \equiv \sin(2\pi/p)$. The matrix representation of the operator $[q]$ can be obtained by making direct use of the matrix in eqn (A.23), the components of the unit vector along C_q being $k_1 = \cos\varphi$, $k_2 = \sin\varphi$ and $k_3 = 0$. An alternative, easily visualized three-stage approach leading to the same result, is the following: (i) rotate the Cartesian system about the Z axis through $\pi/2 - \varphi$, so that the Y axis coincides with C_q; (ii) apply, in the rotated system a rotation about C_q through $2\pi/q$; and (iii) restore the Cartesian system to its original position. The matrix representation of $[q]$ is thus given by the three-matrix product

$$[q] : \begin{pmatrix} s_\varphi & c_\varphi & 0 \\ -c_\varphi & s_\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_q & 0 & s_q \\ 0 & 1 & 0 \\ -s_q & 0 & c_q \end{pmatrix} \begin{pmatrix} s_\varphi & -c_\varphi & 0 \\ c_\varphi & s_\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.48}$$

where $c_\varphi \equiv \cos\varphi$, $s_\varphi \equiv \sin\varphi$, $c_q \equiv \cos(2\pi/q)$ and $s_q \equiv \sin(2\pi/q)$. The result, whether computed from eqn (1.48) or from eqn (A.23), is

$$[q] : \begin{pmatrix} c_\varphi^2(1-c_q)+c_q & s_\varphi c_\varphi(1-c_q) & s_\varphi s_q \\ s_\varphi c_\varphi(1-c_q) & s_\varphi^2(1-c_q)+c_q & -c_\varphi s_q \\ -s_\varphi s_q & c_\varphi s_q & c_q \end{pmatrix}. \tag{1.49}$$

The required matrix representation of the product $[q][p]$ is obtained as

$$[q][p] : \begin{pmatrix} c_\varphi^2(1-c_q)+c_q & s_\varphi c_\varphi(1-c_q)c_p + s_\varphi s_p s_q & -s_\varphi c_\varphi(1-c_q)s_p + s_\varphi c_p s_q \\ s_\varphi c_\varphi(1-c_q) & c_\varphi^2(c_p c_q - c_p) + c_p - c_\varphi s_p s_q & -c_\varphi^2(s_p c_q - s_p) - s_p - c_\varphi c_p s_q \\ -s_\varphi s_q & c_\varphi c_p s_q + s_p c_q & -c_\varphi s_p s_q + c_p c_q \end{pmatrix}.$$

and the trace of this matrix representation is obtained as

$$\mathrm{Tr}([q][p]) = A\cos^2(\varphi) + B\cos(\varphi) + C = 1 + 2\cos\frac{2\pi}{n}, \tag{1.50}$$

where $n = 2, 3, 4$ or 6 and

$$\begin{aligned} A = &\ 1 - \cos(2\pi/p) - \cos(2\pi/q) + \cos(2\pi/p)\cos(2\pi/q) \\ = &\ 4\sin^2(\pi/q)\sin^2(\pi/p), \\ B = &-2\sin(2\pi/p)\sin(2\pi/q) \\ = &-8\sin(\pi/q)\sin(\pi/p)\cos(\pi/q)\cos(\pi/p), \\ C = &\ \cos(2\pi/p) + \cos(2\pi/q) + \cos(2\pi/p)\cos(2\pi/q) \\ = &\ 4\cos^2(\pi/q)\cos^2(\pi/p) - 1. \end{aligned}$$

The solution of the quadratic equation, formed from eqn (1.50), with the above simplified forms of the trigonometric expressions, is

$$\cos(\varphi) = \frac{\cos(\pi/q)\cos(\pi/p) \pm \cos(\pi/n)}{\sin(\pi/q)\sin(\pi/p)}. \tag{1.51}$$

It is now possible to undertake a systematic determination of all the allowable combinations of axes of rotational symmetry. The method outlined above is based on the matrix-algebraic derivation of crystallographic groups of symmetry by Seitz (1934). An alternative approach, presented in considerable detail by Buerger (1956), is based on spherical trigonometry. There are altogether six different nontrivial allowable combinations of three axes of rotational symmetry: four of these have been found above for the special case of $p = q = 2$, and the remaining two are [2][3][3] and [2][3][4]. Some angular relationships involved will be discussed in the Exercises.

1.8 Crystal systems

We conclude this chapter with a discussion of the seven crystal systems which define unit cells that conform to the symmetry that exists. This discussion will be based on the results of Sections 1.6 and 1.7.

1. If there is no rotational symmetry other than [1], there is of course no formal indication of the choice of basis vectors and they satisfy, *in the general case*, the relations

$$a \neq b \neq c, \qquad \alpha \neq \beta \neq \gamma \neq 90^o.$$

The unit cell is thus a general parallelepiped, and the crystallographic system of coordinates (referred to more briefly as the crystal system) based on such a cell is called the *triclinic* system.

Note that in the discussion of crystal systems the character " $\neq$ " means "must not be equal to". An accidental approximate equality, not constrained by symmetry, is of course possible.

2. If the lattice is symmetric with respect to rotation about a single type of twofold axis, there exists a lattice line parallel to the axis and a lattice plane perpendicular to it (see Section 1.6). We can therefore choose one of the basis vectors along the axis of symmetry and the other two in the perpendicular plane. Such a choice constrains two angles to the value 90°, and the parameters of the resulting unit cell can be

$$a \neq b \neq c, \qquad \alpha = \beta = 90^\circ, \gamma \neq 90^\circ$$

if the basis vector along the twofold axis is called **c** (this is referred to as the *first setting*), or

$$a \neq b \neq c, \qquad \alpha = \gamma = 90^\circ, \beta \neq 90^\circ$$

if the basis vector along the twofold axis is called **b** (this is referred to as the *second setting*). These are the most frequently encountered unit-cell choices based on this lattice symmetry. The corresponding crystal system is called the *monoclinic* system.

3. If the lattice is symmetric with respect to rotation about either of two mutually perpendicular twofold axes, there must also be a third twofold axis of symmetry, perpendicular to the other two, and by the same argument as that given above there exist three mutually perpendicular lattice lines along which the basis vectors can be chosen. The unit cell based on such a choice is given by

$$a \neq b \neq c, \qquad \alpha = \beta = \gamma = 90^\circ,$$

and the crystal system based on this lattice symmetry is called the *orthorhombic*. system. This system is orthogonal, but the lengths of the basis vectors are still unconstrained by symmetry.

4. We now consider a lattice which is symmetric with respect to rotation about a single type of fourfold axis (rotation through $2\pi/4$ or 90°). As in the monoclinic system, one of the basis vectors can be chosen along the fourfold axis and the other two in the lattice plane perpendicular to it. When, however, we choose a vector in the perpendicular lattice plane as one of the basis vectors, a symmetric rotation through 90° must carry it into another in-plane vector which is perpendicular to the first, and has *the same length*. The latter vector is chosen to be the second basis vector in the lattice plane perpendicular to the fourfold axis. The unit cell based on such a choice of basis vectors is given by

$$a = b \neq c, \qquad \alpha = \beta = \gamma = 90^\circ$$

and the corresponding crystal system is called the *tetragonal* system(the basis vector along the fourfold axis is always called **c**).

5. We next consider a lattice which is symmetric with respect to rotation about a single type of threefold axis (rotation through $2\pi/3$ or 120°). In the more frequently encountered case, the basis vector **c** is chosen along the threefold axis and the other two in the lattice plane perpendicular to it. However, as in the tetragonal case, we choose here the two in-plane lattice vectors so that they are related by a rotation about the threefold axis, that is, they form an angle of $120^\circ o$ and have the same length. The unit cell based on this choice of basis vectors is given by

$$a = b \neq c, \qquad \alpha = \beta = 90^o, \gamma = 120^o$$

and the corresponding axial system is called the *hexagonal* axial system.

The other case, somewhat less readily visualized, is illustrated in Fig. 1.11. Of course, there exists a lattice line parallel to the threefold axis, but none of the basis vectors is chosen along this line. Instead, one of the basis vectors is chosen to be the shortest lattice vector that radiates from an origin on the threefold axis and is neither parallel nor perpendicular to that axis. Repeated application of the operator [3] generates two more such equivalent vectors, and these three vectors are chosen as the basis vectors of the lattice. They must have the same length and, by symmetry, the angles between each pair of them are the same. The threefold axis then coincides with one of the body diagonals of the resulting symmetric parallelepiped, each face of which is a rhombus. The standard definition of the unit cell based on such a choice of basis vectors is therefore

$$a = b = c \equiv a_0, \qquad \alpha = \beta = \gamma \equiv \alpha_0 \neq 90^\circ$$

and the corresponding axial system is called the *rhombohedral* axial system. It is seen that for $\alpha_0 = 90^o$ the rhombohedral parallelepiped reduces to a cube. The crystal system based on a lattice possessing a single type of threefold axis of symmetry is called the *trigonal* system, and its symmetric unit cell can be represented either by hexagonal or rhombohedral axes, or by both types by means of a suitable transformation.

6. We now consider a lattice which is symmetric with respect to rotation about a single type of sixfold axis (rotation through $2\pi/6$ or 60°). One of the basis vectors is chosen along the lattice line coinciding with the sixfold axis, and is called **c**, while the other two are chosen in the lattice plane perpendicular to it. If **a** is taken as the shortest in-plane lattice vector, the vector **b** is chosen so as to be generated by rotating **a** through 120° about the sixfold axis, resulting in the same hexagonal axial system as that seen above in the trigonal case. However, the vector $\mathbf{a} + \mathbf{b}$ has the same length as **a** and forms an angle of 60° with the latter: hence the compatibility of this axial system with the sixfold symmetry. The unit cell based on this choice of basis vectors is given by

$$a = b \neq c, \qquad \alpha = \beta = 90^\circ, \quad \gamma = 120^\circ$$

and the corresponding *crystal* system is called the *hexagonal* system. The apparent confusion between the hexagonal axial systems within the trigonal and

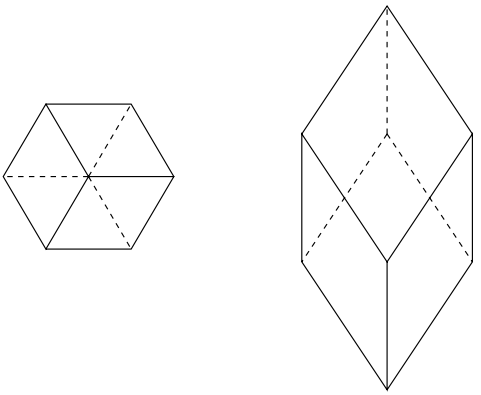


Fig. 1.11 A rhombohedral unit cell and the threefold axis of symmetry.

hexagonal crystal systems has been alleviated by the introduction of the concept of a "crystal family" (Hahn 1983). Thus, the hexagonal family includes both trigonal and hexagonal crystal systems.

The permissible combinations of three operators of rotational symmetry were found in Section 1.7 to be [2][2][2], [3][2][2], [4][2][2], [6][2][2], [2][3][3] and [2][3][4]. The first four combinations are associated with the orthorhombic, trigonal, tetragonal and hexagonal crystal systems respectively, as follows from the above discussion. Each of the last two combinations gives rise, by composition, to four threefold axes of symmetry where each pair of adjacent axes forms an angle of $70°32'$ (or $109°28'$). This can be found with the aid of equation (1.51). These angles are characteristic of the threefold symmetries of a regular tetrahedron, a cube, and a regular octahedron. The basis vectors are chosen to attain the highest symmetry, so that the unit cell is a cube and the threefold axes are accommodated along the four body diagonals of the cube, the axes being parallel to the lattice directions [111], $[\bar{1}11]$, $[1\bar{1}1]$ and $[11\bar{1}]$. The unit cell is a special case of the rhombohedral cell,

$$a = b = c \equiv a_0, \qquad \alpha = \beta = \gamma = 90°$$

and the crystal system is called the *cubic* system.

Note that in none of the seven cases discussed above have we assumed that the unit cell conforming to the symmetry requirements was primitive. We shall take up this question in our discussion of lattice types in Section 2.4. The crystal systems are summarized in Table 1.1.

Table 1.1 The crystal syatems

Name of the system	Symmetric unit cell	Comments
Triclinic	$a \neq b \neq c, \quad \alpha \neq \beta \neq \gamma \neq 90^\circ$	
Monoclinic	$a \neq b \neq c, \quad \alpha = \beta = 90^\circ, \quad \gamma \neq 90^\circ$	first setting
	$a \neq b \neq c, \quad \alpha = \gamma = 90^\circ, \quad \beta \neq 90^\circ$	second setting
Orthorhombic	$a \neq b \neq c, \quad \alpha = \beta = \gamma = 90^\circ$	
Tetragonal	$a = b \neq c, \quad \alpha = \beta = \gamma = 90^\circ$	
Trigonal	$a = b \neq c, \quad \alpha = \beta = 90^\circ, \quad \gamma = 120^\circ$	hexagonal axes
	$a = b = c \equiv a_0, \quad \alpha = \beta = \gamma \neq 90^\circ$	rhombohedral axes
Hexagonal	$a = b \neq c, \quad \alpha = \beta = 90^0, \quad \gamma = 120^\circ$	
Cubic	$a = b = c \equiv a_0, \quad \alpha = \beta = \gamma = 90^\circ$	

1.9 Exercises for Chapter 1

1. Describe the pattern in Fig. 1.1 in terms of three different pairs of basis vectors, chosen so that each pair defines a primitive unit cell. Calculate the areas of these cells. Explain why the areas of all possible primitive unit cells, in a given two-dimensional lattice, should be exactly the same.

2. The compound papaverine hydrochloride, with empirical formula $C_{20}H_{21}NO_4{\cdot}HCl$, crystallizes in a lattice, in which a primitive unit cell has the dimensions:

$$a = 13.059, \quad b = 15.620, \quad c = 9.130\text{Å}, \alpha = 90^\circ, \quad \beta = 92.13^\circ, \quad \gamma = 90^\circ$$

   The density of the crystal is 1.33 g cm^{-1}. Find the number of molecules of papaverine hydrochloride in the primitive unit cell.

   Accurate measurements of unit-cell dimensions and specific gravity are the basis of accurate determination of the Avogadro number.

3. Show, with reference to the discussion of the distance between successive lattice planes (Section 1.2), that for a family of lattice planes with Miller indices h, k and l the following three equations,

$$\begin{aligned} \mathbf{a} \cdot \mathbf{h} &= h, \\ \mathbf{b} \cdot \mathbf{h} &= k, \\ \mathbf{c} \cdot \mathbf{h} &= l, \end{aligned} \tag{1.52}$$

   must be simultaneously satisfied. We shall see in a later chapter that equations closely analogous to those in eqn (1.52) are of central importance in the theory of diffraction by crystals.

4. Sketch the lattice planes (102), $(\bar{1}24)$, and $(1\bar{1}1)$, assuming that the value of n in eqn (1.12) is 1 for each of them. Indicate in your diagrams the directions of the basis vectors **a**, **b** and **c** (remember that they have to form a right-handed triad).

5. Verify eqns (1.34)–eqn (1.36).

6. Show, with the aid of reciprocal-lattice relationships, that if the unit cell dimensions satisfy $a = b \neq c$ and $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$, then the interplanar distance between successive lattice planes in the (hkl) family is given by

$$d_{hkl} = \left[\frac{4(h^2 + k^2 + hk)}{3a^2} + \frac{l^2}{c^2}\right]^{-1/2}.$$

7. Show that if a lattice is symmetric with respect to a plane of reflection, the lattice must contain a lattice line perpendicular to that plane and a lattice plane parallel to it. This completes the considerations of Section 1.6.

8. Show that the allowable combinations of axes of symmetry, where two of them must be twofold, obtained from the geometrical argument at the beginning of Section 1.7 are also obtainable from the more general eqn (1.51). Interpret geometrically the positive and negative values of $\cos(\varphi)$.

9. Convince yourself that the components of the unit vector **k** along the axis of rotation C_q in Fig. 1.10 are $k_1 = \cos\varphi$, $k_2 = \sin\varphi$ and $k_3 = 0$ and show that eqn (1.49) can be obtained directly from eqn (A.23).

10. Examine in detail the interaxial angles associated with the allowable combination of a twofold, a threefold and a fourfold axis of rotational symmetry. Explain, without detailed calculations, how eqns (A.27)–eqn (A.29) could be used in such considerations.

11. It was pointed out in the text that centers of symmetry may also be located at points other than lattice points. Show that a symmetric primitive unit cell of a lattice contains seven centers of symmetry not related by lattice translations, and one center at a lattice point.

12. Show, by considering the derivations in Section 1.2, that the normal to a family of lattice planes in the reciprocal lattice is a vector in the direct lattice with relatively prime components, and that the interplanar distance within this family is the reciprocal of the magnitude of that direct-lattice vector.

13. Show that the cosines of the angles ν_i and ν_i^*, where $i = 1, 2, 3$, ν_i is the angle between the direct basis vector $\mathbf{a}_i$ and the reciprocal-lattice vector $\mathbf{h}(h_1h_2h_3)$ perpendicular to the family of lattice planes $(h_1h_2h_3)$ in the direct lattice, and ν_i^* is the angle between the reciprocal basis vector $\mathbf{a}_i^*$ and the direct-lattice vector perpendicular to the family of planes $(u_1u_2u_3)$ in the reciprocal lattice, are given by

$$\cos\nu_i = h_i \frac{d(h_1h_2h_3)}{|\mathbf{a}_i|} \quad \text{and} \quad \cos\nu_i^* = u_i \frac{d^*(u_1u_2u_3)}{|\mathbf{a}_i^*|}$$

where

$$d^*(u_1u_2u_3) = |u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3|^{-1}$$

(see the previous exercise).

This page intentionally left blank

2 Point groups and lattice types

2.1 Introduction

Crystallographic symmetry can be considered from several points of view. The main ones are (i) the *macroscopic* symmetry of the crystal and its physical properties, (ii) the symmetry of a three-dimensional point lattice, and (iii) the symmetry of a periodic arrangement of material units, or the symmetry of the ideal crystal. The descriptors of macroscopic symmetry are rotations, introduced in the previous chapter, and rotoinversions (the latter include inversion and reflection). Their classification is one of the main topics of this chapter. The crystal and its geometrical abstraction–the three-dimensional lattice–must also conform to these *macroscopic* operations. However, the full symmetry of the periodic arrangement also includes translational symmetry which may be based on the lattice alone (in the case of lattices and some types of crystals) and on additional combinations of rotations and special types of nonlattice translations (in the general case of ideal crystals). These essentially *microscopic* symmetry operations that involve combinations of rotations and translations will be dealt with in the next chapter.

We begin this chapter with a brief recapitulation of the fundamentals of group theory, which is an invaluable aid in the classification of sets of symmetry operators, both finite macroscopic and infinite microscopic ones. We shall then employ the building blocks prepared in the previous chapter to classify the groups of admissible rotations and rotoinversions, the crystallographic point groups. The last part of the chapter will be devoted to the classification of lattice types, due to Bravais, in accordance with the underlying macroscopic symmetry.

2.2 Fundamentals of group theory

The concept of a group and its basic properties are discussed in practically all textbooks of modern algebra, and have been extensively developed in the literature. The following brief summary is no substitute for a regular introduction to group theory, but rather is aimed at collecting together the most relevant properties of a group that will be required in this book. The summary is based on textbooks on group theory (for example Ledermann, 1976) and crystallographic introductions to the subject (for example Hahn and Wondratschek 1994).

An abstract group is a set of elements for which a composition law is defined and which also satisfies the following four axiomatic conditions:

1. Composition of any two elements, say a and b, of a set $\mathcal{G}$ results in a uniquely determined element, say c, of $\mathcal{G}$, which is usually called the *product* of a and

b. A set possessing this property is said to obey the axiom of *closure* with respect to the composition law assumed. In symbolic form, if $a \in \mathcal{G}$ and $b \in \mathcal{G}$, then there always exists $c \in \mathcal{G}$ such that $c = ab$. The symbol $\in$ stands for "belongs to" or "in".

2. If $a \in \mathcal{G}$, $b \in \mathcal{G}$, and $c \in \mathcal{G}$, then $(ab)c = a(bc)$ and the result can be denoted by abc. This axiom is called the *associative law* with respect to the composition assumed.
3. Among the elements of the set, there exists an *identity element* or *unit element* $e \in \mathcal{G}$, such that $ae = ea = a$ for any element $a \in \mathcal{G}$.
4. For each $a \in \mathcal{G}$, there exists an element $b \in \mathcal{G}$such that $ab = ba = e$. The element b is called the *inverse* of a and is denoted by a^{-1}.

Summing up, if for a certain law of composition the set obeys *closure* and *associativity*, has an *identity* element, and each of its elements has an *inverse* element, also belonging to the set, then the set is a group. Let us consider some examples. The set of all the integers (including zero) is a group with respect to addition, chosen as the composition law. This is so because (i) it obeys the axiom of closure (the sum of any two integers is an integer), (ii) it obeys the associative law ($(p+q)+r = p+(q+r)$, where p, q, r are integers), (iii) its identity element is zero ($p + 0 = 0 + p = p$, for any integer p), and (iv) the inverse element of an integer p is simply $-p$ ($p + (-p) = (-p) + p = 0$).

If we reconsider the set of all the integers but choose the subtraction as the composition law, the set is not a group, as can be readily verified. Also, if multiplication is chosen as the composition law, with unity as the identity element, the set of all the integers cannot be a group, because none of its elements, except the identity, possesses an inverse within the set. If, on the other hand, we consider the set of all positive rational numbers, of the form p/q where p and q are natural numbers, and choose multiplication as the composition law, with unity as the identity element, this set can be readily shown to be a group.

Our first example considered above has a most important application to crystallographic symmetry. Since the set of all the integers is a group with respect to addition of integers, the set of all the vectors of a one-dimensional lattice, $u_i\mathbf{a}$, is a group with respect to vector addition. This is a translational group of the one-dimensional lattice. More importantly, we can easily show that a three-dimensional lattice also forms a group with respect to the addition of triplets of coefficients of basis vectors or, equivalently, with respect to addition of lattice vectors.

Consider a lattice expressible in terms of basis vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, which define a primitive unit cell. The coordinates of the lattice points are therefore integers. A typical lattice vector can be written as $\mathbf{r}_{Li} = u_i\mathbf{a} + v_i\mathbf{b} + w_i\mathbf{c}$ and the set of all such vectors can be denoted by $\Gamma = \{\mathbf{r}_{Li}\}$.

1. If $\mathbf{r}_{Li} \in \Gamma$ and $\mathbf{r}_{Lj} \in \Gamma$, then also $\mathbf{r}_{Li} + \mathbf{r}_{Lj} \equiv \mathbf{r}_{Lk} \in \Gamma$, since the coordinates of $\mathbf{r}_{Lk}$ are of necessity integers. The condition of closure is satisfied.

2. For any $\mathbf{r}_{Li} \in \Gamma$, $\mathbf{r}_{Lj} \in \Gamma$, and $\mathbf{r}_{Lk} \in \Gamma$ we have $(\mathbf{r}_{Li} + \mathbf{r}_{Lj}) + \mathbf{r}_{Lk} = \mathbf{r}_{Li} + (\mathbf{r}_{Lj} + \mathbf{r}_{Lk})$, and associativity holds true.
3. The zero vector $\mathbf{0}$, corresponding to $u_i = v_i = w_i = 0$, is the identity element of Γ, since $\mathbf{r}_{Li} + \mathbf{0} = \mathbf{0} + \mathbf{r}_{Li}$ for any $\mathbf{r}_{Li} \in \Gamma$.
4. The inverse element of any $\mathbf{r}_{Li} \in \Gamma$ is $-\mathbf{r}_{Li}$, which also belongs to Γ.

The set Γ is therefore an infinite group of lattice vectors or, equivalently an infinite group of coordinate triplets. This group is of a fundamental importance to crystallographic symmetry.

There are several definitions and concepts we shall require, however, it will be convenient to introduce an example involving a finite set of symmetry operators and use it to illustrate those concepts and definitions.

Let us consider all the symmetries of an equilateral triangle (see Fig. 2.1). We fix the origin at its center, choose the basis vectors $\mathbf{a}$ and $\mathbf{b}$ as indicated in the figure so that they form an angle of 120° and have equal lengths, and take the coordinate axes X and Y to be parallel to $\mathbf{a}$ and $\mathbf{b}$, respectively. We also number the vertices of the triangle, and indicate its mirror lines as explained below. The obvious symmetry operators are rotation of the triangle through 120, 240 and 360 degrees about an axis perpendicular to the triangle and passing through its center, which we denote by $[3], [3]^2$, and $[3]^3 \equiv [1]$; reflection of the triangle in a line passing through vertex 1 so as to interchange vertices 2 and 3, which we denote by $[m_{23}]$; and two additional reflections in lines passing through the vertices 2 and 3 so as to interchange the remaining vertices, denoted by $[m_{13}]$ and $[m_{12}]$, respectively. These appear to be all the operators which when applied to the triangle will bring it into self-coincidence, or the symmetry operators of the triangle. It is logical that any successive application of two such operators will also bring the triangle into self-coincidence, and should be one of the six operators described above, provided that the list is exhaustive. Let us construct a "multiplication table" showing all the possible products (that is, successive applications) of pairs of symmetry operators. For example, let us first rotate the triangle about the threefold axis through 120° and then apply to the so-rotated triangle the reflection $[m_{23}]$. We can do this simply by keeping track of the vertex numbers during the above movements: the rotation [3] carries vertex 1 into vertex 2, 2 into 3, and 3 into 1; the reflection $[m_{23}]$ now leaves 3 (formerly 1) unchanged and interchanges 1 and 2 (formerly 2 and 3). The anticlockwise sequence of vertices was 123 before the rotation [3], and becomes 321 after successive application of the rotation and (then) the reflection: the position of vertex 2 is now the same as before the combined operation, and the vertices 1 and 3 are interchanged; this is just the result of the reflection $[m_{13}]$. We can therefore write $[m_{23}][3] = [m_{13}]$. Note, however, that reversing the order of the operations leads in this case to a different result: it can be readily shown that $[3][m_{23}] = [m_{12}]$. Table 2.1 shows the results of all the products of pairs of the symmetry operators of the equilateral triangle, where the left factor of a product corresponds to an operator in the leftmost vertical column, and the

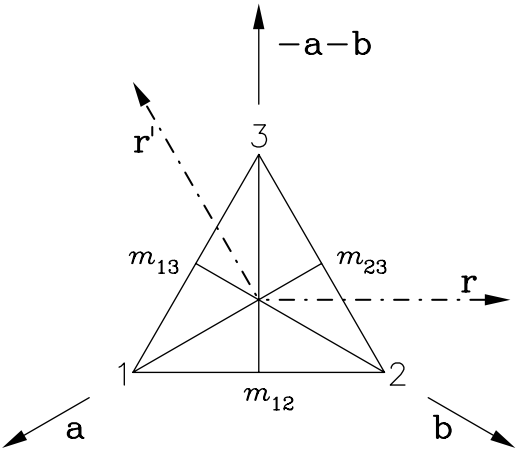


Fig. 2.1 Symmetries of an equilateral triangle.

Table 2.1 Multiplication table of the group of an equilateral triangle

	$[1]$	$[3]$	$[3]^2$	$[m_{23}]$	$[m_{13}]$	$[m_{12}]$
$[1]$	$[1]$	$[3]$	$[3]^2$	$[m_{23}]$	$[m_{13}]$	$[m_{12}]$
$[3]$	$[3]$	$[3]^2$	$[1]$	$[m_{12}]$	$[m_{23}]$	$[m_{13}]$
$[3]^2$	$[3]^2$	$[1]$	$[3]$	$[m_{13}]$	$[m_{12}]$	$[m_{23}]$
$[m_{23}]$	$[m_{23}]$	$[m_{13}]$	$[m_{12}]$	$[1]$	$[3]$	$[3]^2$
$[m_{13}]$	$[m_{13}]$	$[m_{12}]$	$[m_{23}]$	$[3]^2$	$[1]$	$[3]$
$[m_{12}]$	$[m_{12}]$	$[m_{23}]$	$[m_{13}]$	$[3]$	$[3]^2$	$[1]$

right factor corresponds to an operator in the uppermost horizontal row.

It can be seen from Table 2.1 that each row and each column contain all six symmetry operators. Hence the set is closed with respect to successive application of these operators. It can be shown that associativity holds true. The operator $[1]$ is the identity element of the set, and each operator has an inverse in the set. Hence the set of the symmetry operators of the equilateral triangle is a group. We also observe immediately that each reflection is its own inverse, as expected, and that the inverse of $[3]$ is $[3]^2 \equiv [3]^{-1}$, which represents a rotation through 240° or -120°.

2.2.1 Some important definitions

We can now introduce some very relevant definitions and illustrate them in terms of the examples seen above.

- The number of different elements of a group $\mathcal{G}$ is called the *order* of $\mathcal{G}$ and is denoted here by $n(\mathcal{G})$. For example, the order of the translation group of a lattice is infinity, and that of the group of the triangle is 6.
- If for any pair of elements $g_i \in \mathcal{G}$ and $g_j \in \mathcal{G}$ the relation $g_i g_j = g_j g_i$ is valid, the group is called *commutative* or *Abelian*. Clearly, the translation group of a lattice is Abelian, while that of the triangle is not.
- A set of elements, say $a, b, \ldots \in \mathcal{G}$, from which the whole group can be generated by composition, is called a *set of generators* of the group. For example, the group of the triangle can be generated from the operator [3] and any one of the reflections.
- A finite group $\mathcal{G}$ is called *cyclic* if it can be generated by one of its elements, and has the form

$$\mathcal{G} = \{a, a^2, \ldots, a^n \equiv e\}$$

  An example of a cyclic group is $[1], [3], [3]^2$, and its multiplication table is the upper left 3 by 3 block of the multiplication table of the group of the triangle.
- If $\mathcal{G}$ is a group with elements $e = g_1, g_2, \ldots$, then for each element $g_i \in \mathcal{G}$ its cyclic group $\mathcal{G}_i = \{e, g_i, g_i^2, \ldots\}$ can be defined. Its order, defined here as $n(\mathcal{G}_i)$ is called the *order of the element* g_i. For example, in the group of the triangle one can distinguish three orders: the identity [1] is of order 1, each of the three reflections is of order 2 and each of the rotations [3] and $[3]^2 = [3]^{-1}$, is of order 3.
- A subset $\mathcal{H}$ of a group $\mathcal{G}$ is called a *subgroup* of $\mathcal{G}$ if $\mathcal{H}$ obeys the group axioms (it is itself a group). The group $\mathcal{G}$ itself and its identity $\mathcal{I} = \{e\}$ are called *trivial* subgroups of $\mathcal{G}$, while the others are called *proper* subgroups. For example, the proper subgroups of the group of the triangle are $\{[1], [3], [3]^{-1}\}$, $\{[1], [m_{23}]\}$, $\{[1], [m_{13}]\}$, and $\{[1], [m_{12}]\}$.

2.2.2 Matrix representation of symmetry operations

A most convenient representation of symmetry operators is one in terms of matrices, and ultimately in terms of coordinates of symmetry-equivalent positions. A most detailed compilation of such representations may be found in Volume A of the *International Tables for Crystallography* (Arnold 1983), and our purpose here is only to introduce the reader to the meaning and formation of such representations.

We shall now describe briefly a possible method for representing rotation operators by matrices. We base this on the fact that a rotated vector can be expressed either in terms of the old coordinates and a rotated basis or in terms of new coordinates and the old basis. We can therefore do the following:

1. Express a vector in terms of its old coordinates and a temporarily rotated basis.
2. Relate the rotated basis vectors to the old ones.
3. Reexpress the rotated vector in terms of the old basis.

The relationship between the rotated and the old coordinates of the vector, or the matrix representation of the rotation operator, follows readily. For example, consider the rotation of a vector $\mathbf{r}$ through 120°, in the anticlockwise sense about the threefold axis, as shown in Fig. 2.1. The rotated vector $\mathbf{r}'$ can be written in the general form $\mathbf{r}' = [3]\mathbf{r}$. Since $\mathbf{a}$ and $\mathbf{b}$ are basis vectors along the axes X and Y, respectively, forming an angle of 120° and having equal lengths, we have

$$\begin{aligned}[3]\mathbf{r} &= [3](x\mathbf{a} + y\mathbf{b}) \\ &= x[3]\mathbf{a} + y[3]\mathbf{b} \\ &= x\mathbf{b} + y(-\mathbf{a} - \mathbf{b}) \\ &= -y\mathbf{a} + (x - y)\mathbf{b}. \end{aligned} \tag{2.1}$$

If we write $\mathbf{r}' = x'\mathbf{a} + y'\mathbf{b}$, we can relate the coordinates of symmetry related points by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 0 & \overline{1} \\ 1 & \overline{1} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{2.2}$$

and a matrix representation of [3]–in the basis assumed–has been obtained. Proceeding as above, we obtain the matrix representations of the symmetry operators of the group of the triangle as

$$\begin{array}{lll|lll|lll} [1]: & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & & [3]: & \begin{pmatrix} 0 & \overline{1} \\ 1 & \overline{1} \end{pmatrix} & & [3]^{-1}: & \begin{pmatrix} \overline{1} & 1 \\ \overline{1} & 0 \end{pmatrix} \\ [m_{23}]: & \begin{pmatrix} 1 & 0 \\ 1 & \overline{1} \end{pmatrix} & & [m_{13}]: & \begin{pmatrix} 0 & \overline{1} \\ \overline{1} & 0 \end{pmatrix} & & [m_{12}]: & \begin{pmatrix} \overline{1} & 1 \\ 0 & 1 \end{pmatrix}. \end{array}$$

The operators are followed by colons here, rather than by equality or equivalence signs, since the operators have a clear geometrical interpretation, while the matrices depend on the choice of basis. Nevertheless, for a given basis, the above six matrices form a group with respect to matrix multiplication, and the structure of its multiplication table is identical to that given in Table 2.1. This will be discussed further in the Exercises.

2.2.3 Isomorphism, coset decomposition, and the Lagrange theorem

Two groups with multiplication tables of the same structure are said to be *isomorphic*. Take, for example, the subgroups of order 2 of the group of an equilateral triangle. Each of them has a multiplication table of the general (abstract) form

	e	a
e	e	a
a	a	e

and hence any one is isomorphic to any of the other two and, indeed, to any other group of order 2, which is of necessity cyclic (has the form $\{a,\ a^2 = e\}$). It can be shown that groups of order 3 must be cyclic and therefore any pair of such groups must be isomorphic. There are, however, only two different abstract groups of order 4: (i) a cyclic one of the form $\{\ a,\ a^2,\ a^3,\ a^4 = e\ \}$ and (ii) a group of the form $\{\ e,\ a,\ b,\ c\ \}$, where $ab = ba = c, bc = cb = a, ac = ca = b$ and $a^2 = b^2 = c^2 = e$, called the *four group*. Therefore, any group of order 4 must be isomorphic either to the abstract cyclic group of order 4, or to the four group. We shall see later more examples of this property.

A more formal definition of group isomorphism is the following:

DEFINITION 1. *By an isomorphism between two groups $\mathcal{G}$ and $\mathcal{G}'$ is meant a one-to-one correspondence $a \leftrightarrow a'$, between their elements which preserves group multiplication, that is which is such that if $a \leftrightarrow a'$ and $b \leftrightarrow b'$, then $ab \leftrightarrow a'b'$.*

For example, let us compare the group of rotational symmetries of an equilateral triangle, $\{[1], [3], [3]^2\}$, with the group of integers under addition modulo 3 (that is, the remainder after dividing the sum of two integers by 3; this can be only 0, 1, or 2). The multiplication table of the latter group is

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

The above multiplication table has clearly the same structure as the upper 3 by 3 block of the multiplication table shown in Table 2.1, the correspondences being $0 \leftrightarrow [1]$, $1 \leftrightarrow [3]$, and $2 \leftrightarrow [3]^2 \equiv [3]^{-1}$.

A very useful property of a group, which will be of importance in the discussion of space groups in the next chapter, is its decomposability into (right and left) cosets. Its formal definition, to be illustrated below, is:

DEFINITION 2. *If $\mathcal{H}$ is any subgroup of the group $\mathcal{G}$, then by a right coset (or left coset) of the subgroup $\mathcal{H}$ of the group $\mathcal{G}$ is meant any set $\mathcal{H}a$ (or $a\mathcal{H}$) of all the right multiples sa (or left multiples as) of the elements of $\mathcal{H}$ by a fixed element $a \in \mathcal{G}$.*

Let us examine this definition by taking $\mathcal{H}$ as for example, $\{[1], [m_{23}]\}$ and find all the right cosets of $\mathcal{H}$ for the group of the equilateral triangle, with the aid of Table 2.1. We obtain

$$\{[1],[m_{23}]\}[1] = \{[1],[m_{23}]\}, \tag{2.3}$$
$$\{[1],[m_{23}]\}[3] = \{[3],[m_{23}][3]\} = \{[3],[m_{13}]\}, \tag{2.4}$$
$$\{[1],[m_{23}]\}[3]^2 = \{[3]^2,[m_{23}][3]^2\} = \{[3]^2,[m_{12}]\}, \tag{2.5}$$
$$\{[1],[m_{23}]\}[m_{23}] = \{[m_{23}],[m_{23}][m_{23}]\} = \{[m_{23}],[1]\}, \tag{2.6}$$
$$\{[1],[m_{23}]\}[m_{13}] = \{[m_{13}],[m_{23}][m_{13}]\} = \{[m_{13}],[3]\}, \tag{2.7}$$
$$\{[1],[m_{23}]\}[m_{12}] = \{[m_{12}],[m_{23}][m_{12}]\} = \{[m_{12}],[3^2]\}. \tag{2.8}$$

The above example illustrates two important theorems in the theory of groups:

THEOREM 1. *If R and S are any two elements of $\mathcal{G}$ and if $\mathcal{H}$ is a subgroup of $\mathcal{G}$, then the cosets $\mathcal{H}R$ and $\mathcal{H}S$ are identical if and only if $RS^{-1} \in \mathcal{H}$.*

There are indeed three pairs of identical cosets in eqns (2.3)–(2.8), and these conform to the above theorem.

THEOREM 2. *If $\mathcal{H}$ is a subgroup of $\mathcal{G}$, where the orders of these groups are $n(\mathcal{H})$ and $n(\mathcal{G})$, respectively, then $n(\mathcal{H})$ is a divisor of $n(\mathcal{G})$, such that*

$$\frac{n(\mathcal{G})}{n(\mathcal{H})} = m,$$

where m is an integer. The integer m is called the index *of $\mathcal{H}$ in $\mathcal{G}$. This is also the number of distinct (right or left) cosets of $\mathcal{H}$ into which the group $\mathcal{G}$ can be decomposed.*

The division property, contained in Theorem 2 is known in group theory as the Lagrange theorem. Clearly, all the parts of the theorem are illustrated by the formation of all the right cosets of the subgroup $\{[1],[m_{23}]\}$ of the group of the equilateral trangle. This will be discussed further in the Exercises.

A possible decomposition of the group of the triangle into right cosets of the above subgroup is, for example,

$$\{[1],[m_{23}]\} \cup \{[1],[m_{23}]\}[3] \cup \{[1],[m_{23}]\}[3]^2 \tag{2.9}$$

where the symbol $\cup$ indicates a union of two sets.

We now introduce the concept of an *invariant subgroup*, which will be of particular importance in our discussion of space groups in the next chapter. This concept is defined as follows.

DEFINITION 3. *If $\mathcal{H}$ is a subgroup of $\mathcal{G}$ such that, for all $h \in \mathcal{H}$ and all $g \in \mathcal{G}$, $ghg^{-1} \in \mathcal{H}$, then $\mathcal{H}$ is said to be an invariant subgroup (or normal subgroup or self-conjugate subgroup) of $\mathcal{G}$.*

It can also be shown that $\mathcal{H}$ is an invariant subgroup of $\mathcal{G}$ *if and only if* $a\mathcal{H} = \mathcal{H}a$ for all $a \in \mathcal{G}$; that is, *if and only if* all right and left cosets of $\mathcal{H}$ coincide.

We define finally the *outer direct product*, which will be most useful in the classification of the point groups in this chapter. We have the following definition.

DEFINITION 4. *Let $\mathcal{G}$ be a group with subgroups $\mathcal{H}$ and $\mathcal{K}$ such that*

(1) *if $h \in \mathcal{H}$ and $k \in \mathcal{K}$ then $hk = kh$;*
(2) *all $g \in \mathcal{G}$ can be expressed in the form $g = hk$ with $h \in \mathcal{H}$ and $k \in \mathcal{K}$, and*
(3) *the intersection of $\mathcal{H}$ and $\mathcal{K}$, denoted by $\mathcal{H} \cap \mathcal{K} = \{e\}$, is the set consisting of the identity element of $\mathcal{G}$.*

Then $\mathcal{G}$ is called the outer direct product of $\mathcal{H}$ and $\mathcal{K}$, and the factorization in point (2) *above is unique. We write $\mathcal{G} = \mathcal{H} \otimes \mathcal{K} = \mathcal{K} \otimes \mathcal{H}$.*

Examples of such outer direct products will be seen in the next section. It is strongly recommended that the reader reviews the Exercises on the above fundamental aspects of group theory if any reviewing of this subject seems to be desirable.

2.3 Crystallographic point groups

We shall now classify all the crystallographic point groups. These are groups of symmetry operations, the "point" attribute being due to the fact that on the application of such an operation at least one point of the object acted upon remains fixed. In fact, this is strictly so only for the operation of inversion (its fixed point is the center of inversion) and for rotoinversions (see below). In the case of a pure rotation, we have a line of fixed points, while in the case of a reflection in a plane, the whole plane is fixed.

We shall first consider the groups of symmetry operations consisting of pure rotations alone. It was shown in Section 1.4 that the permissible angles of rotation about the axes of symmetry of a three-dimensional lattice are multiples of $2\pi/n$, where $n = 1, 2, 3, 4$ or 6 only. The corresponding rotation operator was denoted by $[n]$. We can therefore have five cyclic groups (see Subsection 2.2.1) generated by the rotation operators [1], [2], [3], [4], and [6], respectively. For example, $\{[4], [4]^2, [4]^3, [4]^4 \equiv [1]\}$ is such a cyclic group. Each of these groups, except $\{[1]\}$, is associated with a single axis of rotation, which is oriented in a definite manner with respect to the periodic arrangement obeying this rotational symmetry. It was also shown in Section 1.7 that there are several permissible combinations of differently oriented axes of symmetry. We have, in fact, six such combinations and a group of symmetry operations can be generated from each of them. Let us consider, for example, the simplest combination of three twofold axes of symmetry. Since these axes must be mutually perpendicular we can choose each of them to be along an axis of a Cartesian system, denote the corresponding rotation operators by $[2_X]$, $[2_Y]$ and $[2_Z]$, and derive easily their matrix representations:

$$[2_X] \leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix} \Bigg| [2_Y] \leftrightarrow \begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix} \Bigg| [2_Z] \leftrightarrow \begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easily shown that the set $\{[1], [2_X], [2_Y], [2_Z]\}$ is a group and is isomorphic to the four group (see Subsection 2.2.3).

2.3.1 The axial point groups

We can now approach the classification of the point groups containing rotations alone, known as the axial point groups. This is done in Table 2.2, where the first column contains the old Schönflies point-group symbol, the second column contains the international or Hermann–Mauguin point-group symbol, and the last column shows all the rotation operators of each point group according to the following convention: The symbol $[n^S_{uvw}]$ denotes an operator of symmetric rotation through an angle $2\pi/n$ about an axis parallel to a lattice vector $u\mathbf{a}+v\mathbf{b}+w\mathbf{c}$ (that is, a lattice direction $[uvw]$), and associated with the crystal or axial system S. The symbols for the system are $M2$, monoclinic (second setting); O, orthorhombic; T, tetragonal; H, hexagonal; and C, cubic. The point groups 3 and 32, belonging to the trigonal system, are considered here in their representation in the hexagonal axial system only, and hence their nontrivial elements carry the superscript H. It is also assumed that the basis vectors have been chosen so as to conform to the crystal/axial system, as discussed above; this suffices for finding the matrix representations of the point-group elements, as will be indicated below.

Some simplifications have been introduced in the listing of the point-group elements and these may, at first sight, disguise the cyclic nature of some groups. For example, the operators $[6^H_{001}]^2$ and $[6^H_{001}]^3$ in the point groups 6 and 622 have been replaced by $[3^H_{001}]$ and $[2^H_{001}]$ respectively, since they have the same meaning. The group–subgroup relations are thereby emphasized.

The representation of the point-group operators by matrices is vital for their further applications. We can do this by the method described in Subsection 2.2.2, as illustrated below.

For example, let us represent the operator $[2^H_{210}]$ by a matrix:

$$\begin{aligned} [2^H_{210}]\mathbf{r} &= [2^H_{210}](x\mathbf{a} + y\mathbf{b} + z\mathbf{c}) \\ &= x[2^H_{210}]\mathbf{a} + y[2^H_{210}]\mathbf{b} + z[2^H_{210}]\mathbf{c} \\ &= x(\mathbf{a} + \mathbf{b}) + y(-\mathbf{b}) + z(-\mathbf{c}) \\ &= x\mathbf{a} + (x - y)\mathbf{b} - z\mathbf{c}. \end{aligned} \tag{2.10}$$

Hence,

$$[2^H_{210}] \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 1 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}.$$

The matrix representations of the rotation operators necessary to generate all the axial point groups listed in Table 2.2 are given in Table 2.3. Some of

Table 2.2 Axial point groups

Schönflies	Hermann-Mauguin	Point group
C_1	1	$\{[1]\}$
C_2	2	$\{[1], [2^{M2}_{010}]\}$
C_3	3	$\{[1], [3^{H}_{001}], [3^{H}_{001}]^{-1}\}$
C_4	4	$\{[1], [4^{T}_{001}], [2^{T}_{001}], [4^{T}_{001}]^{-1}\}$
C_6	6	$\{[1], [6^{H}_{001}], [3^{H}_{001}], [2^{H}_{001}], [3^{H}_{001}]^{-1}, [6^{H}_{001}]^{-1}\}$
D_2	222	$\{[1], [2^{O}_{100}], [2^{O}_{010}], [2^{O}_{001}]\}$
D_3	32	$\{[1], [3^{H}_{001}], [3^{H}_{001}]^{-1}, [2^{H}_{100}], [2^{H}_{110}], [2^{H}_{010}]\}$
D_4	422	$\{[1], [4^{T}_{001}], [2^{T}_{001}], [4^{T}_{001}]^{-1},$ $[2^{T}_{100}], [2^{T}_{010}], [2^{T}_{110}], [2^{T}_{\bar{1}10}]\}$
D_6	622	$\{[1], [6^{H}_{001}], [3^{H}_{001}], [2^{H}_{001}], [3^{H}_{001}]^{-1}, [6^{H}_{001}]^{-1}$ $[2^{H}_{100}], [2^{H}_{210}], [2^{H}_{110}], [2^{H}_{120}], [2^{H}_{010}], [2^{H}_{\bar{1}10}]\}$
T	23	$\{[1], [2^{C}_{100}], [2^{C}_{010}], [2^{C}_{001}],$ $[3^{C}_{\bar{1}11}], [3^{C}_{\bar{1}11}]^{-1}, [3^{C}_{1\bar{1}1}], [3^{C}_{1\bar{1}1}]^{-1},$ $[3^{C}_{11\bar{1}}], [3^{C}_{11\bar{1}}]^{-1}, [3^{C}_{111}], [3^{C}_{111}]^{-1}\}$
O	432	$\{[1], [2^{C}_{011}], [2^{C}_{0\bar{1}1}], [2^{C}_{101}], [2^{C}_{\bar{1}01}], [2^{C}_{110}],$ $[2^{C}_{\bar{1}10}], [3^{C}_{111}], [3^{C}_{111}]^{-1}, [3^{C}_{\bar{1}11}], [3^{C}_{\bar{1}11}]^{-1}, [3^{C}_{1\bar{1}1}],$ $[3^{C}_{1\bar{1}1}]^{-1}, [3^{C}_{11\bar{1}}], [3^{C}_{11\bar{1}}]^{-1}, [4^{C}_{100}], [2^{C}_{100}], [4^{C}_{100}]^{-1},$ $[4^{C}_{010}], [2^{C}_{010}], [4^{C}_{010}]^{-1}, [4^{C}_{001}], [2^{C}_{001}], [4^{C}_{001}]^{-1}\}$

the operators have the same representation in several crystal systems, and this is indicated by a multiple superscript (for example, $[2^{O,T,C}_{100}]$ means that the representation is the same in orthorhombic,tetragonal and cubic systems). A complete listing of the matrix representations of point group operators is given by Arnold (1983, Tables 11.2 and 11.3).

2.3.2 Completion of the presentation of the point groups

Given all the information about the structure of the axial point groups, the derivation of the remaining point groups can be completed in a fairly systematic manner. As shown in Section 1.5, every lattice is centrosymmetric and thus remains invariant under the operations of the group of inversion, $\{[1], [\bar{1}]\}$, denoted

Table 2.3 Representations of axial point-group generators

$[1]$	$\rightarrow$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$[2_{100}^{O,T,C}]$	$\rightarrow$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}$	$[2_{010}^{M2,O,T,C}]$	$\rightarrow$	$\begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}$
$[2_{001}^{all}]$	$\rightarrow$	$\begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$[2_{110}^{T,C}]$	$\rightarrow$	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}$	$[2_{\overline{1}10}^{T,C}]$	$\rightarrow$	$\begin{pmatrix} 0 & \overline{1} & 0 \\ \overline{1} & 1 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}$
$[2_{100}^{H}]$	$\rightarrow$	$\begin{pmatrix} 1 & \overline{1} & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}$	$[2_{210}^{H}]$	$\rightarrow$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}$	$[3_{111}^{C}]$	$\rightarrow$	$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$
$[3_{001}^{H}]$	$\rightarrow$	$\begin{pmatrix} 0 & \overline{1} & 0 \\ 1 & \overline{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$[4_{001}^{T,C}]$	$\rightarrow$	$\begin{pmatrix} 0 & \overline{1} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$[6_{001}^{H}]$	$\rightarrow$	$\begin{pmatrix} 1 & \overline{1} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

by $\overline{1}$ and C_i in the Hermann–Mauguin and Schönflies notations, respectively. This leads to 11 additional point groups, which are obtained by forming the outer direct products of each of the axial point groups with the group of inversion. Such direct product groups exist, since the inversion operator commutes with any point-group operator. When a centrosymmetric point group is formed, we try to find all its subgroups other than those already in the list. When this process has been completed we have 10 more new subgroups (other than the axial ones) and all the 32 crystallographic point groups have been obtained. The 10 new point groups are noncentrosymmetric and not axial. When the inversion operator is combined with rotation operators we obtain rotoinversions, a special case of which is the mirror reflection operator. As above, we shall try to define the operators in an explicit manner–including the orientation of the corresponding symmetry elements, and the axial systems where appropriate–and the following notation will be used:

- The symbol $[m_{uvw}^{S}]$ denotes an operator of symmetric reflection through a plane *perpendicular* to the lattice vector $u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$ (that is, a lattice direction $[uvw]$), and is associated with the crystal or axial system S.
- The symbol $[\overline{n}_{uvw}^{S}]$ denotes an operator of symmetric rotoinversion, defined here as rotation through an angle $2\pi/n$ about an axis *parallel* to the lattice vector $u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$ (that is, a lattice direction $[uvw]$), combined with inversion through a point on the axis of rotation, and is associated with the crystal or axial system S. The point of inversion is a center of symmetry only when n is odd.

For example, $[2_{010}]$ is a twofold axis of rotation *parallel* to $\mathbf{b}$ (or $[010]$), and $[m_{100}]$ is the operator of reflection in a plane *perpendicular* to $\mathbf{a}$ (or $[100]$). The orientation of the axes of rotoinversion is denoted in the same manner as that of axes of rotation. The above formal definitions will be explained further below.

Let us examine in some detail such a direct product of the point groups 2 and $\overline{1}$, assuming the choice of basis vectors on which Table 2.3 is based. We have

$$\begin{aligned} 2 \otimes \overline{1} &= \{[1], [2^{M2}_{010}]\}\{[1], [\overline{1}]\} \\ &= \{[1], [\overline{1}], [2^{M2}_{010}], [2^{M2}_{010}][\overline{1}]\}. \end{aligned} \tag{2.11}$$

Since the matrix representation of the inversion operator is

$$[\overline{1}] \rightarrow \begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix} \tag{2.12}$$

for any crystal system (all three coordinates of a point simply change their sign under this operation), the fourth element of the direct product in eqn (2.11) is

$$[2^{M2}_{010}][\overline{1}] \rightarrow \begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix} \begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.13}$$

It can be seen that the matrix on the right-hand side of eqn (2.13) transforms a column vector with components x, y, and z so that x and z remain unchanged and y changes its sign. Since **b** is perpendicular to **a** and **c** in the system $M2$ (monoclinic, second setting), this matrix represents a reflection in the xz plane. We shall denote this reflection operator by $[m^{M2}_{010}]$ or $[\overline{2}^{M2}_{010}]$ in line with the formal definition given above. If we omit the superscript $M2$ for simplicity, the direct product in eqn (2.11) becomes

$$2 \otimes \overline{1} = \{[1], [\overline{1}], [2_{010}], [m_{010}]\}. \tag{2.14}$$

It is readily shown that the set on the right-hand side of eqn (2.14) is a group. It can be seen that each of its elements is its own inverse, and that $[\overline{1}][2_{010}] = [m_{010}]$, $[2_{010}][m_{010}] = [\overline{1}]$, and $[m_{010}][\overline{1}] = [2_{010}]$. This group is therefore isomorphic to the four group.

The group in eqn (2.14) has three nontrivial subgroups: $\{[1], [2_{010}]\}$, denoted by 2 or C_2, $\{[1], [\overline{1}]\}$, denoted by $\overline{1}$ or C_i and $\{[1], [m_{010}]\}$, denoted by m or C_s in the Hermann–Mauguin or the Schönflies notation, respectively. The groups 2 and $\overline{1}$ were used in the formation of the outer direct product (they are "old") and the group m is a "new" one. This example illustrates the process of the derivation of all the crystallographic point groups; another, more complicated example, will be given below.

Let us comment further on the rotoinversion. This is a rotation followed immediately by inversion, or vice versa. We can write this in our explicit notation as $[\overline{1}][n^S_{uvw}]$ and denote it by $[\overline{n}^S_{uvw}]$. It should be noted that the concept of rotation is often generalized to *proper* rotation, which we have called pure rotation and which has a matrix representation with a determinant of $+1$, and *improper*

rotation, which we have called rotoinversion and which has a matrix representation with a determinant of -1. That is, both proper and improper rotations are often referred to as rotations and all the 10 point-group rotations can be written as $[1], [\overline{1}], [n^S_{uvw}], [\overline{n}^S_{uvw}]$, with $n = 2, 3, 4, 6$.

We shall now present the complete Hermann–Mauguin notation for point-group symbols. As seen in Table 2.2, the Hermann-Mauguin symbol for an axial point group shows the orders of the rotation operators present, but does not give any information about the orientations of the axes of rotation. As we shall see, it gives some more information about the relative orientations of planes of reflection and axes of rotation, when both are present in a point group. The Hermann–Mauguin symbols, and their important combinations are presented below.

- n: Rotation through $2\pi/n$ or rotation of order n.
- $\overline{n}$: Rotation of order n followed by inversion, or vice versa.
- m: Reflection in a plane. Also called *mirror* reflection.
- $\frac{n}{m}$: Rotation of order n and a mirror reflection plane perpendicular to the axis of rotation.
- nm: Rotation of order n and a mirror reflection plane parallel to the axis of rotation.

For example, the direct-product group in eqn (2.14) is denoted in Hermann–Mauguin notation by the symbol $2/m$: the inversion is implied by the presence of the twofold axis and a perpendicular mirror plane.

Let us now consider in some detail a more complicated example, the formation of the outer direct product $32 \otimes \overline{1} = \overline{3}\frac{2}{m}$. We shall omit for brevity the superscript H, but bear in mind that the hexagonal axial system is being employed. We have from Table 2.2

$$\begin{aligned} 32 \otimes \overline{1} &= \{[1], [3_{001}], [3_{001}]^{-1}, [2_{100}], [2_{110}], [2_{010}]\} \otimes \{[1], [\overline{1}]\} \\ &= \{[1], [3_{001}], [3_{001}]^{-1}, [2_{100}], [2_{110}], [2_{010}], \\ &\quad [1][\overline{1}], [3_{001}][\overline{1}], [3_{001}]^{-1}[\overline{1}], [2_{100}][\overline{1}], [2_{110}][\overline{1}], [2_{010}][\overline{1}]\} \\ &= \{[1], [3_{001}], [3_{001}]^{-1}, [2_{100}], [2_{110}], [2_{010}], \\ &\quad [\overline{1}], [\overline{3}_{001}], [\overline{3}_{001}]^{-1}, [m_{100}], [m_{110}], [m_{010}]\}. \end{aligned} \tag{2.15}$$

Let us first give a geometrical description of the resulting group. We have here an axis of inversion, $\overline{3}$, three twofold axes in the perpendicular plane, each adjacent pair forming an angle of 60°, and a mirror plane perpendicular to each of these axes appears, as expected.

The direct-product group is of order 12 and, by Lagrange's theorem, the possible orders of its proper subgroups are 6, 4, 3, and 2. We find three subgroups of order 6. One of them is the centrosymmetric group $\overline{3}$, or

$$3 \otimes [\overline{1}] = \{[1], [3_{001}], [3_{001}]^{-1}, [\overline{1}], [\overline{3}_{001}], [\overline{3}]^{-1}\},$$

which can be generated by taking the first six powers of the operator $[\overline{3}_{001}]$, and is therefore cyclic. The other two subgroups of order 6,

$$32 = \{[1], [3_{001}], [3_{001}]^{-1}, [2_{100}], [2_{110}], [2_{010}]\} \tag{2.16}$$

and

$$3m = \{[1], [3_{001}], [3_{001}]^{-1}, [m_{100}], [m_{110}], [m_{010}]\} \tag{2.17}$$

are isomorphic to the symmetry group of an equilateral trangle, discussed in Section 2.2. There are three subgroups of order 4, each having the general form: $\{[1], [\overline{1}], [2_{uvw}], [m_{uvw}]\}$, geometrically analogous to $2/m$ but differently oriented and isomorphic to the four group. Finally, there are seven subgroups of order 2, namely three 2 groups with symmetry related-axes, three m groups with mirror planes perpendicular to the twofold axes, and the group of inversion.

The matrix representations of the generating elements are:

$$[3^H_{001}] \quad \rightarrow \quad \begin{pmatrix} 0 & \overline{1} & 0 \\ 1 & \overline{1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \Bigg| \quad [2^H_{100}] \quad \rightarrow \quad \begin{pmatrix} 1 & \overline{1} & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix} \quad \Bigg| \quad [\overline{1}] \quad \rightarrow \quad \begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix},$$

and all 12 matrix representations of its elements, in the hexagonal axial system, can be constructed from these generators.

We now summarize the 32 point groups, in Hermann-Mauguin notation, in Table 2.4, and add some explanatory comments.

Some remarks on Table 2.4 are in order:

1. The upper part of the Table is based on uniaxial groups, and the lower part is derived from combinations of several different axes of symmetry.
2. We see that whenever an axis of rotation, given in the first column, is of even order, a mirror plane perpendicular to this axis appears in the symbol of the corresponding centrosymmetric direct-product point group. This is so because the cyclic subgroup generated by such a rotation always has 2 as its trivial or proper subgroup. As was shown above, this must give rise to the appearance of such a mirror plane of symmetry.
3. An empty entry in the column labeled "New subgroups" means that the centrosymmetric point group to the left of such an entry does not have any subgroups other than the axial one in the same row, and those in the preceding rows (if any).
4. The symbol m is synonymous with $\overline{2}$. However, the latter is very rarely encountered in the literature.
5. Detailed graphical representations of the point groups and their elements can be found in Chapter 10 of Volume A of the *International Tables for Crystallography* (Hahn and Klapper 1983).

We show in Table 2.5 how the 32 point groups are distributed among the seven crystal systems. Only the short Hermann–Mauguin notation is be used here. This

Table 2.4 The 32 crystallographic point groups

Axial groups	Centrosymmetric groups	New subgroups
1	$\overline{1}$	–
2	$\frac{2}{m}$	m or $\overline{2}$
3	$\overline{3}$	–
4	$\frac{4}{m}$	$\overline{4}$
6	$\frac{6}{m}$	$\overline{6}$

Axial groups	Centrosymmetric groups		New subgroups
	Full	Short	
222	$\frac{2}{m}\frac{2}{m}\frac{2}{m}$	mmm	$mm2$
32	$\overline{3}\frac{2}{m}$	$\overline{3}m$	$3m$
422	$\frac{4}{m}\frac{2}{m}\frac{2}{m}$	$4/mmm$	$4mm, \overline{4}2m$
622	$\frac{6}{m}\frac{2}{m}\frac{2}{m}$	$6/mmm$	$6mm, \overline{6}2m$
23	$\frac{2}{m}\overline{3}$	$m\overline{3}$	–
432	$\frac{4}{m}\overline{3}\frac{2}{m}$	$m\overline{3}m$	$\overline{4}3m$

Table 2.5 Subdivision of the point groups among crystal systems

Crystal system	Point groups
Triclinic	1, $\overline{1}$
Monoclinic	2, m, $2/m$
Orthorhombic	222, $mm2$, mmm
Tetragonal	4, $\overline{4}$, $4/m$, 422, $4mm$, $\overline{4}2m$, $4/mmm$
Trigonal	3, $\overline{3}$, 32, $3m$, $\overline{3}m$
Hexagonal	6, $\overline{6}$, $6/m$, 622, $6mm$, $\overline{6}2m$, $6/mmm$
Cubic	23, $m\overline{3}$, 432, $\overline{4}3m$, $m\overline{3}m$

subdivision is interesting for its own sake, but it is also of considerable practical importance. As we shall see later the determination of the crystal system follows in a straightforward manner from an initial examination of the experimental data and, obviously, narrows down the possible point-group symmetries of the crystal examined.

2.4 Bravais Lattices

We conclude this chapter with a discussion of the various lattice types which are consistent with the point-group symmetries. Our discussion of crystal systems in Section 1.8 showed that if the lattice obeys a certain rotational symmetry, it is possible to choose in it a set of basis vectors that define a unit cell which preserves that symmetry. In fact, a higher symmetry is preserved (see Table 2.5), since each lattice is centrosymmetric. For example, it can be readily shown that if a lattice is symmetric with respect to a single type of twofold axis, its full symmetry is $2/m$, if it is symmetric with respect to two perpendicular twofold axes only, its full symmetry is mmm, and so on. In general, the full symmetry of a lattice corresponding to a given crystal system is that of the highest-order centrosymmetric point group in Table 2.5 among the groups corresponding to that system.

Two lattices of the same point-group symmetry are said to be of different type, if the symmetric unit cells constructed in them differ in their primitivity or departure therefrom. For example, if in one lattice of symmetry mmm we can construct a primitive orthorhombic unit cell, while in the other lattice of the same symmetry the orthorhombic unit cell contains a lattice point at its center, the two lattices are said to be orthorhombic lattices of different types; it is not possible to construct in the lattice with the centered orthorhombic cell a primitive orthorhombic cell by means of a linear transformation.

A prominent French crystallographer of the nineteenth century, Auguste Bravais, examined the problem lattices of different symmetry and found that there were altogether 14 different lattice types. These are known as *Bravais lattices*. We shall describe and illustrate them, but before we do so, we wish to examine in some detail the possible lattice types that are consistent with the symmetry $2/m$. This implies a monoclinic unit cell, and we assume that the notation for the cell parameters follows the second setting, and that the **abc** basis vectors are primitive translations.

Suppose that the unit cell *may* be nonprimitive, but assume that the two-dimensional sublattice based on **a** and **c** is primitive; this assumption will be justified below. Let us find the possible locations of any additional lattice point(s) within the cell. Let $\mathbf{r}'$ be such an additional lattice vector. Its general form is

$$\mathbf{r}' = p\mathbf{a} + q\mathbf{b} + r\mathbf{c}.$$

Since the lattice obeys the point-group symmetry $2/m$, the vectors

$$\mathbf{r}' - [2_{010}]\mathbf{r}' = 2p\mathbf{a} + 2r\mathbf{c},$$
$$\mathbf{r}' + [2_{010}]\mathbf{r}' = 2q\mathbf{b}$$

must be lattice vectors. In view of the above assumptions, this can be so only if p, q, and r are integer multiples of $1/2$, or if

$$\mathbf{r}' = \frac{1}{2}l\mathbf{a} + \frac{1}{2}m\mathbf{b} + \frac{1}{2}n\mathbf{c},$$

where l, m, and n are integers. Let us consider their smallest values, that is, 0 and 1. The values $l = m = n = 0$ obviously correspond to a primitive monoclinic unit cell. A monoclinic lattice in which such a cell can be constructed is called a type P monoclinic Bravais lattice. The combinations 100, 010 and 001 for lmn are impossible, since they would place lattice points at $\mathbf{a}/2$, $\mathbf{b}/2$, and $\mathbf{c}/2$, respectively, contrary to the assumption that $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are primitive translations. Each of the remaining four possibilities leads to a permissible value of $\mathbf{r}'$ but, as will be seen below, they all reduce to one type of non-primitive monoclinic cell.

The combination $l = 0, m = 1, n = 1$ means that there is an additional lattice point in the center of the $\mathbf{bc}$ face of the unit cell. This is called A-centering, and if no additional lattice points are present in the cell, the lattice is called type A monoclinic. Further, the combination $l = 1, m = 1, n = 0$ means that there is an additional lattice point in the center of the $\mathbf{ab}$ face of the unit cell. This is called C-centering, and if no additional lattice points are present in the cell, the lattice is called type C monoclinic. It can be shown, however, that type A and type C monoclinic lattices are equivalent, and can be related by a linear transformation of their basis vectors (see the Exercises).

The combination $l = 1, m = 0, n = 1$ means that there is an additional lattice point in the center of the $\mathbf{ac}$ face of the unit cell (we have assumed that this cannot be the case in a monoclinic lattice, in the second setting). This is called B-centering. If there are no additional lattice points in the cell, this should be type B monoclinic lattice. However, it is clearly redundant, since we can define in the same lattice a new monoclinic cell

$$\mathbf{a}' = \frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{c},$$
$$\mathbf{b}' = \mathbf{b},$$
$$\mathbf{c}' = -\frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{c},$$

which is primitive, and our assumption is thus justified.

Let us now consider the combination $l = m = n = 1$. This means that there is an additional lattice point in the center of the cell. If there are no additional lattice points in the cell, this is called I-centering, and the lattice is called type I monoclinic. As in the case of the A- and C- centerings it can be shown that I-centering is equivalent to A- or C- centerings. If, for example, we define in an I-type monoclinic lattice

Table 2.6 Lattice types and crystal systems

Crystal system	Bravais lattices
Triclinic	P
Monoclinic	P, C
Orthorhombic	P, C, I, F
Tetragonal	P, I
Trigonal	R
Hexagonal	P
Cubic	P, I, F

$$\mathbf{a}' = \mathbf{a} + \mathbf{c}$$
$$\mathbf{b}' = \mathbf{b}$$
$$\mathbf{c}' = \mathbf{c}$$

the lattice point which was in the center of the cell **abc** is now located in the center of the face $\mathbf{a}'\mathbf{b}'$, and is nothing but a C-centering.

We consider finally the coexistence of various types of centering in the same lattice. The only formally admissible coexistence of centerings is that of A, B, and C in the same lattice (remember, however, that the B-centering is redundant). That is, there is a lattice point in the center of each face of the unit cell. A symmetric unit cell so centered is called *face centered* and the lattice is called as an F-type lattice. Again, it is possible to show that in a monoclinic lattice F-centering is related to C-centering by means of a linear transformation. All other coexistences of centerings destroy the lattice periodicity, and are irrelevant.

Summing up, there are two mathematically different types of monoclinic lattice: the type P, and one of the types A, C, I and F. Older literature used to reduce each of the latter four to the type C, but in more recent tables of symmetry groups the types A and I are also given, to avoid harmless but unnecessary transformations of experimental settings; see the 1983 edition of the *International Tables for Crystallography* (Hahn 1983), referred to hereafter as ITA83.

Table 2.6 lists the 14 different Bravais lattices, and Table 2.7 defines the coordinates of the equivalent positions of the lattice points for each of the various types listed.

To summarize, for triply periodic arrangements such as ideal crystals we have 32 groups of point symmetry operations, and 14 lattice types consistent with these point groups. Each such lattice type may be regarded as a translation group. These finite point groups and infinite groups of translations form the basis of the space groups, to which the next chapter is devoted.

2.5 Exercises for Chapter 2

1. Show that the set $\{u_i, v_i, w_i, x_i\}$, where each of u_i, v_i, w_i, and x_i may be any integer, including zero, is a group with respect to addition. Explain, without calcula-

Table 2.7 Geometrical definition of lattice types

Type	Coordinates
P	$0,0,0$
A	$0,0,0;\ 0,\frac{1}{2},\frac{1}{2}$
B	$0,0,0;\ \frac{1}{2},0,\frac{1}{2}$
C	$0,0,0;\ \frac{1}{2},\frac{1}{2},0$
I	$0,0,0;\ \frac{1}{2},\frac{1}{2},\frac{1}{2}$
F	$0,0,0;\ 0,\frac{1}{2},\frac{1}{2};\ \frac{1}{2},\frac{1}{2},0;\ \frac{1}{2},\frac{1}{2},0$
R	$0,0,0;\ \frac{2}{3},\frac{1}{3},\frac{1}{3};\ \frac{1}{3},\frac{2}{3},\frac{2}{3}$

tions, how this result is related to the translational symmetry of a four-dimensional lattice.

2. Indicate, with the aid of a drawing, all the symmetries of a square. It can be shown that they form a group of order 8. Is this group commutative? To which crystallographic point groups is the group of the square isomorphic?

3. According to the Lagrange theorem, the group of the square can have only proper subgroups of orders 4 and 2. How many subgroups of these orders are there in the group of the square, and which of them are cyclic?

4. Check all the matrix representations of the symmetry operators of the group of the triangle given in Section 2.2.2.

5. The point group 32 (see Tables 2.2 and 2.3) is isomorphic to the group of the triangle. Find the correspondences between the elements of the two groups.

6. Show in detail how eqns (2.3)–(2.8) illustrate the Lagrange theorem. Equation (2.9) shows a decomposition of the group of the triangle into right cosets of the subgroup $\{[1],[m_{23}]\}$. Find all the other possible decompositions of the group of the triangle with respect to this subgroup.

7. As pointed out in the text, there are only two different types of monoclinic lattice, a P-type lattice and a centered one. Find the transformations of basis vectors relating A-centered and F-centered monoclinic lattices to the C-centered one.

8. Show that in the tetragonal crystal system, the Bravais lattices of types I and F are equivalent.

9. Show that if the direct lattice of a crystal is of type F, the vectors of the face-centered unit cell being $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, a unit cell constructed in that lattice on the vectors:
$$\mathbf{a}' = (\mathbf{b}+\mathbf{c})/2, \quad \mathbf{b}' = (\mathbf{c}+\mathbf{a})/2, \quad \mathbf{c}' = (\mathbf{a}+\mathbf{b})/2$$
is necessarily primitive.

3 Space-group symmetry

3.1 Introduction

The previous two chapters dealt with the geometrical properties of a three-dimensional point lattice, its admissible rotational symmetries and their classification in terms of point groups, crystal systems and Bravais lattices. The symmetry of an ideal, three-dimensional crystal, that is, a triply periodic (infinite) assembly of material units with an underlying point lattice, must be a combination of the translational and rotational symmetry of its lattice with the possible modifications that may result from the interaction between these two types of symmetry. We shall present a brief discussion of the algebraic and geometric properties of such combined symmetry operators, represent them in the framework of a crystal structure and present some examples of such derivations. More detailed discussions of the above topics are available in the literature, notably in Volume A of the *International Tables for Crystallography* (Hahn 1983) (all the editions of this volume will be referred to here as ITA83). Section 3.7 of this chapter deals with the principles of graphical representation of space groups and their symmetry elements, fully described in ITA83, a database which is indispensable in crystallographic research. The chapter concludes with some comments on the computational aspects of space-group symmetry and a set of computer-readable space-group symbols sufficient for the derivation of this information.

3.2 Space-group symmetry operators

Before we start this chapter we shall describe the notation to be used. Operators of proper or improper rotation and their representations will be denoted by:

- upper-case bold upright letters (for example $\mathbf{P}$ and $\mathbf{I}$) if general (proper or improper) rotation operators, that is, operators not referred to a basis, are meant;
- the symbols introduced in Chapter 2, for example $[n_{uvw}]$ for a proper rotation about an axis of order n, *parallel* to the lattice vector $u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$, and $[m_{uvw}]$ for a reflection in a plane *perpendicular* to the above lattice vector;
- explicit 3×3 matrices (as in Chapter 2);
- upper-case bold italic letters (for example $\boldsymbol{P}$, $\boldsymbol{Q}$, and $\boldsymbol{I}$) if symbolic representations of 3×3 matrices are meant.

Translation operators and their representations will be denoted by:

- lower-case bold upright letters, if general translation operators or linear combinations of basis vectors (for example $\mathbf{t} = t_1\mathbf{a} + t_2\mathbf{b} + t_3\mathbf{c}$) are meant;
- explicit 3×1 matrices (columns of the components t_i);
- lower-case bold italic letters (for example $\boldsymbol{t}$, $\boldsymbol{u}$, and $\boldsymbol{v}$) if symbolic representations of explicit 3×1 matrices are meant.

Position vectors of points and their representations will be denoted by:

- the lower-case bold upright letter $\mathbf{r}$ if a physical vector (length + orientation) or a linear combination of basis vectors (for example $\mathbf{r}{=}x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$) is meant;
- the lower-case bold italic letter $\boldsymbol{r}$ or $\boldsymbol{x}$ if a symbolic representation of a column of coordinates x, y, and z is meant.

The easiest to imagine general symmetry operator on a crystal is a combination of a permissible rotation and a translation by a lattice vector. We shall use the corresponding general symbols in the following general considerations. Let a permissible rotation operator $\mathbf{P}$ act on a position vector $\mathbf{r}$, and let the result be translated by the lattice vector $\mathbf{r}_L$. If the position vector resulting from this combined operation,

$$\mathbf{r}' = \mathbf{P}\mathbf{r} + \mathbf{r}_L \tag{3.1}$$

is equivalent to $\mathbf{r}$ with respect to some property, the transformation in eqn (3.1) is a symmetry transformation relative to that property. If, for example, the property of interest is a suitably averaged electron density function $\rho(\mathbf{r})$, then the equality

$$\rho(\mathbf{r}) = \rho(\mathbf{P}\mathbf{r} + \mathbf{r}_L)$$

expresses the symmetry of the electron density under the transformation in eqn (3.1). This is a most relevant example, since we shall be interested here in the density of crystalline matter, and its influence on the pattern of radiation diffracted from a crystal. The densities one encounters in crystallography are the electron density, the nuclear density and the density of electrostatic potential, affecting the diffraction of X-rays, thermal neutrons, and electrons, repectively. The principles of crystal symmetry are the same for each of these experimental techniques, and can be conveniently discussed without reference to applications.

In mathematics, combined rotation-translation operations–such as those in the right-hand side of eqn (3.1)–are known as *affine transformations*, but this designation will be used only seldom in the present treatment. We shall assume that the rotation operators belong to one of the 32 crystallographic point groups and the lattice vectors to one of the groups of lattice translations (Bravais lattices) discussed in the previous chapter, and convince ourselves that the combined rotation–translation operators form a group. We shall do so in general terms, before specifying a particular basis for the rotation operators and vectors. Let the above groups be

$$\mathcal{G} = \{\mathbf{P}_i\} \quad \text{and} \quad \Gamma = \{\mathbf{r}_{L_n}\}, \tag{3.2}$$

where $\mathcal{G}$ is the (finite) point group of the (infinite) lattice of all the translations belonging to Γ. We shall first introduce, following ITA83, a convenient shorthand notation for these operations,

$$\mathbf{r}' = \mathbf{P}\mathbf{r} + \mathbf{r}_L \equiv (\mathbf{P}, \mathbf{r}_L)\mathbf{r}, \tag{3.3}$$

and regard $(\mathbf{P}, \mathbf{r}_L)$ as a *symmetry operator*. Another notation, equivalent to the above and frequently employed in solid-state physics, is

$$\mathbf{P}\mathbf{r} + \mathbf{r}_L \equiv (\mathbf{P}|\mathbf{r}_L)\mathbf{r}.$$

Consider two such operators, say $(\mathbf{P}_j,\mathbf{r}_{L_n})$ and $(\mathbf{P}_k,\mathbf{r}_{L_p})$, and let us try to evaluate their product by applying them successively to the position vector $\mathbf{r}$:

$$\begin{aligned}(\mathbf{P}_j, \mathbf{r}_{L_n})(\mathbf{P}_k, \mathbf{r}_{L_p})\mathbf{r} &= (\mathbf{P}_j, \mathbf{r}_{L_n})(\mathbf{P}_k\mathbf{r} + \mathbf{r}_{L_p}) \\ &= \mathbf{P}_j\mathbf{P}_k\mathbf{r} + \mathbf{P}_j\mathbf{r}_{L_p} + \mathbf{r}_{L_n} \\ &= (\mathbf{P}_j\mathbf{P}_k, \mathbf{P}_j\mathbf{r}_{L_p} + \mathbf{r}_{L_n})\mathbf{r}. \end{aligned} \tag{3.4}$$

The composition law of these symmetry operators is therefore

$$(\mathbf{P}_j, \mathbf{r}_{L_n})(\mathbf{P}_k, \mathbf{r}_{L_p}) = (\mathbf{P}_j\mathbf{P}_k, \mathbf{P}_j\mathbf{r}_{L_p} + \mathbf{r}_{L_n}). \tag{3.5}$$

Since $\mathbf{P}_j \in \mathcal{G}$ and $\mathbf{P}_k \in \mathcal{G}$, then $\mathbf{P}_j\mathbf{P}_k \equiv \mathbf{P}_l \in \mathcal{G}$. Since, further, $\mathcal{G}$ is the point group of the lattice, $\mathbf{P}_j\mathbf{r}_{L_p}$ must also be a lattice vector, say $\mathbf{r}_{L_q}$. Finally, the sum of the two lattice vectors $\mathbf{r}_{L_q}$ and $\mathbf{r}_{L_n}$ is also a lattice vector, say $\mathbf{r}_{L_r}$. Therefore, successive application of two such symmetry operators results in an operator of the same form. In our case,

$$(\mathbf{P}_j, \mathbf{r}_{L_n})(\mathbf{P}_k, \mathbf{r}_{L_p}) = (\mathbf{P}_l, \mathbf{r}_{L_r}) \tag{3.6}$$

The above result shows that the postulate of *closure* holds true for the set of all of the symmetry operators considered above, the composition law being given by eqn (3.5).

The identity elements of $\mathcal{G}$ and Γ are $\mathbf{I}$ and $\mathbf{0}$, respectively. The operator ($\mathbf{I}$, $\mathbf{0}$) is the identity element of the set of all of the symmetry operators considered above. Indeed, it can be readily seen, making use of eqn (3.5), that

$$(\mathbf{P}_j, \mathbf{r}_{L_n})(\mathbf{I}, \mathbf{0}) = (\mathbf{I}, \mathbf{0})(\mathbf{P}_j, \mathbf{r}_{L_n}) = (\mathbf{P}_j, \mathbf{r}_{L_n}),$$

independently of the values of j and n.

If $(\mathbf{P}'_j, \mathbf{r}'_{L_n})$ is the inverse element of $(\mathbf{P}_j, \mathbf{r}_{L_n})$, it must satisfy

$$(\mathbf{P}'_j, \mathbf{r}'_{L_n})(\mathbf{P}_j, \mathbf{r}_{L_n}) = (\mathbf{P}_j, \mathbf{r}_{L_n})(\mathbf{P}'_j, \mathbf{r}'_{L_n}) = (\mathbf{I}, \mathbf{0}). \tag{3.7}$$

Let us evaluate the first product:

$$(\mathbf{P}'_j, \mathbf{r}'_{L_n})(\mathbf{P}_j, \mathbf{r}_{L_n}) = (\mathbf{P}'_j\mathbf{P}_j, \mathbf{P}'_j\mathbf{r}_{L_n} + \mathbf{r}'_{L_n}) = (\mathbf{I}, \mathbf{0}). \tag{3.8}$$

If we equate the rotational parts $\mathbf{P}'_j\mathbf{P}_j$ and $\mathbf{I}$, we obtain $\mathbf{P}'_j = \mathbf{P}_j^{-1}$. Proceeding similarly with the translational parts, $\mathbf{P}'_j\mathbf{r}_{L_n} + \mathbf{r}'_{L_n}$ and $\mathbf{0}$, we arrive at

$$\mathbf{r}'_{L_n} = -\mathbf{P}_j^{-1}\mathbf{r}_{L_n}, \tag{3.9}$$

which is an element of Γ, since $\mathbf{P}_j \in \mathcal{G}$. Hence, the inverse element of $(\mathbf{P}_j, \mathbf{r}_{L_n})$ is given by

$$(\mathbf{P}_j, \mathbf{r}_{L_n})^{-1} = (\mathbf{P}_j^{-1}, -\mathbf{P}_j^{-1}\mathbf{r}_{L_n}). \tag{3.10}$$

The evaluation of the second product leads to the same result (see the Exercises).

We have confirmed the validity of the postulates of closure, identity and inverse, and leave to the reader the confirmation of last postulate of associativity.

The set of symmetry operators $\{(\mathbf{P}_k, \mathbf{r}_{L_j})\}$ therefore forms a group, known as a *space group*. The remaining part of this chapter will be devoted to a discussion of some important geometrical aspects of space groups, and the formulations of them which are most useful in crystallographic research.

3.3 Nonsymmorphic space-group operators

The space-group operators discussed in the previous section were obtained by simply combining rotation operators of the point group with translations of the lattice. Space groups consisting of such operators are called *symmorphic space groups*. We can show, however, that there exist admissible translations which are not exact lattice vectors. These arise from the interaction between the rotational and translational symmetries, as will be seen in this section. Before we embark upon the appropriate derivation, let us examine another drawing by Escher (Fig, 3.1) which hints at this new type of space-group symmetry. Of course, if any symmetry accompanied by a change of color is disregarded, the unit of pattern in Fig. 3.1 can be taken as two adjacent white and gray knights. We then have pure translational symmetry, just as in Fig. 1.1. Let us now consider two adjacent white and gray knights across an imaginary vertical line. If the white knight is reflected in the vertical line, its color changed to gray, and it is translated along this line by *half* a lattice translation, it is brought into coincidence with a gray knight that is located there. This symmetry operation is called a *gliding reflection* combined with a color change, and the imaginary vertical line functions as the *glide line*. The space group of the pattern thus consists of a group of two-dimensional lattice translations and a color-glide line operator. We can simplify the picture by assuming that all the white knights are turned gray. The glide-line operation then becomes a reflection combined with a translation through half a lattice vector. Such a space-group operator is called *non-symmorphic*.

In a more general manner, let us consider an operator of the form $(\mathbf{P}, \mathbf{t})$, where $\mathbf{P}$ is a rotation operator, of order g, belonging to one of the crystallographic point groups, and $\mathbf{t}$ is an arbitrary translation vector. The question we wish to examine is what are the permissible values of $\mathbf{t}$ for $(\mathbf{P}, \mathbf{t})$ to be a space-group operator.

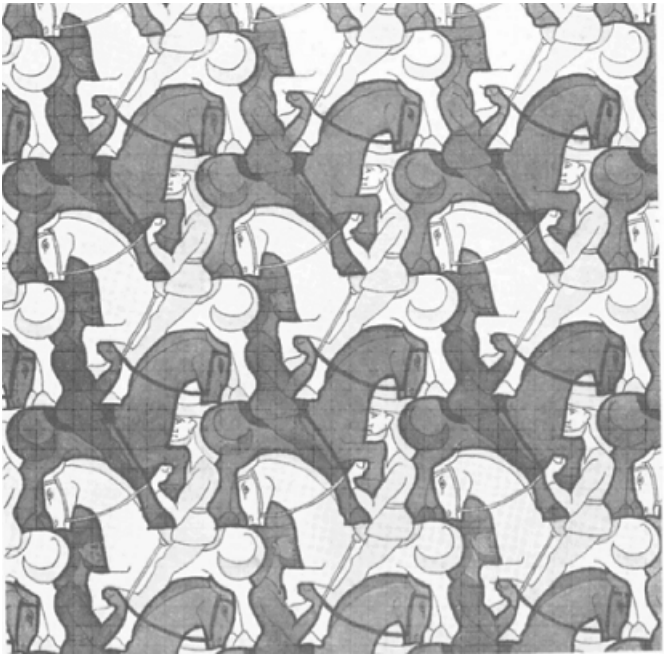


Fig. 3.1 *A non-symmorphic two-dimensional periodic pattern.* Reproduced from MacGillavry (1965), with copyright permission of the International Union of Crystallography (IUCr).

Note that we have not yet defined a basis and fixed its origin, but we shall do this soon.

Let us decompose $\mathbf{t}$ into its components parallel and perpendicular to the axis of rotation,

$$\mathbf{t} = \mathbf{t}_{\parallel} + \mathbf{t}_{\perp} \tag{3.11}$$

where $\mathbf{t}_{\parallel}$ is parallel to the axis and $\mathbf{t}_{\perp}$is perpendicular to it, and calculate the gth power of ($\mathbf{P}$,$\mathbf{t}$),

$$\begin{aligned}(\mathbf{P},\mathbf{t})^g &= (\mathbf{P},\mathbf{t})(\mathbf{P},\mathbf{t})\ldots(\mathbf{P},\mathbf{t}) \\ &= (\mathbf{P}^2,\mathbf{P}\mathbf{t}+\mathbf{t})(\mathbf{P},\mathbf{t})\ldots(\mathbf{P},\mathbf{t}) \\ &= (\mathbf{P}^3,\mathbf{P}^2\mathbf{t}+\mathbf{P}\mathbf{t}+\mathbf{t})\ldots(\mathbf{P},\mathbf{t}) \\ &= [\mathbf{P}^g,(\mathbf{P}^{g-1}+\mathbf{P}^{g-2}+\cdots+\mathbf{I})\mathbf{t}] \\ &= [\mathbf{I},(\mathbf{P}^{g-1}+\mathbf{P}^{g-2}+\cdots+\mathbf{I})\mathbf{t}],\end{aligned} \tag{3.12}$$

where $\mathbf{I}$ is a unit operator. The rotation part of this g-fold product of $(\mathbf{P},\mathbf{t})$ is the identity, $\mathbf{I}$, and in order that it should be a space-group operator, its translation part must of course be a lattice translation. Let us now substitute

the decomposition of $\mathbf{t}$ in eqn (3.11) into eqn (3.12). It is readily shown that the perpendicular component

$$(\mathbf{P}^{g-1} + \mathbf{P}^{g-2} + \cdots + \mathbf{I})\mathbf{t}_\perp \tag{3.13}$$

vanishes identically when $\mathbf{P}$ is an operator of rotation through 180°, 120°, 90°, or 60°. Obviously, only these angles of rotation are associated with nontrivial symmetric rotations of three-dimensional periodic arrays. On the other hand, the parallel component of $\mathbf{t}$ is invariant under a rotation by $\mathbf{P}$ or any of its powers. We then have

$$(\mathbf{P}^{g-1} + \mathbf{P}^{g-2} + \cdots + \mathbf{I})\mathbf{t}_\parallel = g\mathbf{t}_\parallel.$$

Therefore, if $\mathbf{P}$ is a permissible crystallographic symmetric rotation,

$$(\mathbf{P}, \mathbf{t})^g = (\mathbf{I}, g\mathbf{t}_\parallel). \tag{3.14}$$

The vanishing of the perpendicular component in eqn (3.13), for *any* $\mathbf{t}_\perp$, implies that $\mathbf{t}_\perp$ is an arbitrary vector in the plane perpendicular to the axis of rotation. This is in fact the case for a single symmetry operator, since we are free to choose the projection of the origin onto this plane. However, once the three components of the origin have been fixed, the translation parts of any additional operators belonging to the same space group must be consistent with this choice. This will be dealt with below. For the time being, let us assume that $\mathbf{t}_\perp$ is a zero vector, and find the possible (smallest) values of $\mathbf{t}_\parallel$.

Let the rotation operator correspond to a twofold symmetric rotation, in the second setting of the monoclinic system, or briefly, $\mathbf{P} \rightarrow [2^{M2}_{010}]$. The shortest nonzero lattice vector parallel to the axis of rotation is $\mathbf{b}$, and since in this case $g = 2$ we have $\mathbf{t}_\parallel = \mathbf{b}/2$. There are therefore two translationally different symmetry operators based on a twofold axis of rotation in this setting: $([2^{M2}_{010}], \mathbf{0})$, which is a simple rotation about a twofold axis, and $([2^{M2}_{010}], \mathbf{b}/2)$, which is a rotation about a twofold axis, accompanied by a translation through *one-half* of a lattice vector along the axis. The common designation of the symmetry element corresponding to the latter operator is *two-fold screw axis*. Of course, an arbitrary lattice vector can be added to the translation part of each of these two symmetry operators. In general, if $\mathbf{r}_{UVW}$ is the shortest lattice vector parallel to the axis of rotation, the nonlattice translations which are parallel to the axis, are shorter than $\mathbf{r}_{UVW}$ and may appear in the symmetry operator are $\mathbf{r}_{UVW} \times m/g$, where $m = 1, \ldots, g-1$ and g is the order of the rotation operator. The symbolic designation of these various screw axes is simply n_m, where n is the order of the axis and m is a multiple of the basic fractional parallel translation. Thus, if $n = 3$ and the axis is parallel to $\mathbf{c}$, we obtain the operators $([3_{001}], \mathbf{0})$, $([3_{001}], \mathbf{c}/3)$, and $([3_{001}], 2\mathbf{c}/3)$, which are designated as 3, 3_1, and 3_2, respectively, and similarly for fourfold and sixfold axes.

Another kind of nonsymmorphic symmetry element arises from the interaction of reflection and translation. Consider the space-group operator $(\boldsymbol{\sigma}, \mathbf{t})$,

Table 3.1 Glide planes in the monoclinic system

u	v	Reflection operator	Corresponding symmetry element
0	0	$([m_{010}], \mathbf{t}_\perp)$	m (reflection plane)
1	0	$([m_{010}], \frac{1}{2}\mathbf{a} + \mathbf{t}_\perp)$	a (a glide)
0	1	$([m_{010}], \frac{1}{2}\mathbf{c} + \mathbf{t}_\perp)$	c (c glide)
1	1	$([m_{010}], \frac{1}{2}(\mathbf{a} + \mathbf{c}) + \mathbf{t}_\perp)$	n (n glide)

where $\boldsymbol{\sigma}$ is an operator of reflection in a plane, and $\mathbf{t}$ is a translation vector decomposable into components that are parallel and perpendicular to the mirror plane,. that is, $\mathbf{t} = \mathbf{t}_\parallel + \mathbf{t}_\perp$. Two successive applications of this operator are equivalent to

$$(\boldsymbol{\sigma}, \mathbf{t}_\parallel + \mathbf{t}_\perp)(\boldsymbol{\sigma}, \mathbf{t}_\parallel + \mathbf{t}_\perp) = (\mathbf{I}, \mathbf{t}_\parallel - \mathbf{t}_\perp + \mathbf{t}_\parallel + \mathbf{t}_\perp) = (\mathbf{I}, 2\mathbf{t}_\parallel). \tag{3.15}$$

As in the case of a screw axis, the perpendicular component is arbitrary (for a single operator), as it depends on the location of the plane with respect to the origin. On the other hand, twice the parallel component *must* be a lattice vector, for the right-hand side of eqn (3.15) to be a space group operator. Hence, $\mathbf{t}_\parallel$ may be a zero vector, one-half of a lattice vector, or an integer multiple thereof. An operator consisting of a reflection in a plane accompanied by a translation through half a lattice vector parallel to the plane is called a glide-plane operator and the corresponding space-group symmetry element is known as a *glide plane*.

Let us assume in the following example that the reflection operator $([m_{010}], \mathbf{t})$ belongs to a monoclinic space group, in the second setting,that is, with a plane parallel to $\mathbf{a}$ and $\mathbf{c}$ and perpendicular to $\mathbf{b}$. According to the above discussion, the translation part of the right-hand side of eqn (3.15) has the form

$$2\mathbf{t}_\parallel = u\mathbf{a} + w\mathbf{c},$$

where u and w are integers. If we take their values as 0 and 1, we obtain the forms of the reflection operator listed in Table 3.1. The symbolic representations of of mirror and glide planes, as well as of rotation, rotoinversion, and screw axes, as used in space-group tables (for example in ITA83), are given in Table 1.3 of ITA83. We reproduce part of that table in Table 3.2 in a somewhat different notation from that used in the original reference.

The symbols presented in Table 3.2 and letters specifying the lattice type are all that are needed for a representation of a three-dimensional crystallographic space group in terms of the Hermann–Mauguin symbols. The conventions used in these representations are described in detail in ITA83 and in Hahn (1988), and will not be repeated here. The symbols presented in Table 3.2 have graphical representations, which are used in the construction of space-group diagrams. These will be illustrated below. In the meantime we shall present several examples of Hermann–Mauguin symbols and extract from them the (matrix, vector)

representations of the representative elements of the space group, as well as other highly pertinent information.

Let us consider as an example the centrosymmetric orthorhombic space group *Pbcn*. According to the conventions set out in ITA83 (all editions), we have here a b glide perpendicular to $\mathbf{a}$ (parallel to the yz plane, with glide vector $\frac{1}{2}\mathbf{b}$), a c glide perpendicular to $\mathbf{b}$ (parallel to the xz plane, with glide vector $\frac{1}{2}\mathbf{c}$), and an n glide perpendicular to $\mathbf{c}$ (parallel to the xy plane, with glide vector $(\mathbf{a}+\mathbf{b})/2$). These three glide planes must intersect at a point, and let us choose, for the time being, the origin of the space group at this point. This choice constrains the perpendicular components of the translation parts to be equal to zero, and the glide plane operators can therefore be represented as

$$b \rightarrow \left([m_{100}], \frac{1}{2}\mathbf{b}\right), \quad c \rightarrow \left([m_{010}], \frac{1}{2}\mathbf{c}\right), \quad n \rightarrow \left([m_{001}], \frac{1}{2}(\mathbf{a}+\mathbf{b})\right).$$

or, in (matrix,vector) representation,

$$b \rightarrow \left[\begin{pmatrix} \overline{1}\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{2} \\ 0 \end{pmatrix}\right],$$

$$c \rightarrow \left[\begin{pmatrix} 1\,0\,0 \\ 0\,\overline{1}\,0 \\ 0\,0\,1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \end{pmatrix}\right],$$

$$n \rightarrow \left[\begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,\overline{1} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}\right].$$

The superscript O on the reflection operators has been omitted for simplicity. The lattice of *Pbcn* is of type P, that is, there are no lattice translations which may interact with the translation parts of the space-group operators.

We shall now evaluate the products cn, bn, and bc of the glide-plane operators, represented above. We have for cn

$$\begin{aligned}\left([m_{010}], \frac{1}{2}\mathbf{c}\right)\left([m_{001}], \frac{1}{2}(\mathbf{a}+\mathbf{b})\right) &= \left([m_{010}][m_{001}], \frac{1}{2}([m_{010}]\mathbf{a}+[m_{010}]\mathbf{b}+\mathbf{c})\right) \\ &= \left([2_{100}], \frac{1}{2}(\mathbf{a}-\mathbf{b}+\mathbf{c})\right) \\ &\equiv \left([2_{100}], \frac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c})\right). \qquad (3.16)\end{aligned}$$

The product of the reflection operators can be evaluated from geometrical considerations or from their matrix representations. The last line of eqn (3.16) follows from the second by addition of the lattice vector $\mathbf{b}$ to the translation part of the

Table 3.2 Symbols for symmetry elements as given in ITA83

Symbol	Symmetry element and its orientation	Definition of symmetry operation
	Reflection plane	Reflection in plane (3 dimensions)
m	Reflection line	a line (2 dimensions), or a point
	Reflection point	(1 dimension)
	"Axial" glide plane	Glide reflection in a plane with glide vector:
a	$\perp \mathbf{b}$ or $\perp \mathbf{c}$	$\frac{1}{2}\mathbf{a}$
b	$\perp \mathbf{c}$ or $\perp \mathbf{a}$	$\frac{1}{2}\mathbf{b}$
c	$\perp \mathbf{a}$ or $\perp \mathbf{b}$	$\frac{1}{2}\mathbf{c}$
c	$\perp (\mathbf{a}-\mathbf{b})$ or $\perp (\mathbf{a}+\mathbf{b})$	$\frac{1}{2}\mathbf{c}$
c	$\perp \mathbf{a}$ or $\perp \mathbf{b}$ or $\perp (-\mathbf{a}-\mathbf{b})$	$\frac{1}{2}\mathbf{c}$ (hexagonal coordinate system)
c	$\perp (\mathbf{a}-\mathbf{b})$ or $\perp (\mathbf{a}+2\mathbf{b})$ or $\perp (-2\mathbf{a}-\mathbf{b})$	$\frac{1}{2}\mathbf{c}$ (hexagonal coordinate system)
	"Double" glide plane (in centered cells only)	*Two* glide reflections in a plane, with perpendicular glide vectors:
e	$\perp \mathbf{c}$	$\frac{1}{2}\mathbf{a}$ and $\frac{1}{2}\mathbf{b}$
e	$\perp \mathbf{a}$	$\frac{1}{2}\mathbf{b}$ and $\frac{1}{2}\mathbf{c}$
e	$\perp \mathbf{b}$	$\frac{1}{2}\mathbf{a}$ and $\frac{1}{2}\mathbf{c}$
e	$\perp (\mathbf{a}-\mathbf{b})$; $\perp (\mathbf{a}+\mathbf{b})$	$\frac{1}{2}(\mathbf{a}+\mathbf{b})$ and $\frac{1}{2}\mathbf{c}$; $\frac{1}{2}(\mathbf{a}-\mathbf{b})$ and $\frac{1}{2}\mathbf{c}$
e	$\perp (\mathbf{b}-\mathbf{c})$; $\perp (\mathbf{b}+\mathbf{c})$	$\frac{1}{2}(\mathbf{b}+\mathbf{c})$ and $\frac{1}{2}\mathbf{c}$; $\frac{1}{2}(\mathbf{b}-\mathbf{c})$ and $\frac{1}{2}\mathbf{c}$
e	$\perp (-\mathbf{a}+\mathbf{c})$; $\perp (\mathbf{a}+\mathbf{c})$	$\frac{1}{2}(\mathbf{a}+\mathbf{c})$ and $\frac{1}{2}\mathbf{c}$; $\frac{1}{2}(\mathbf{a}-\mathbf{c})$ and $\frac{1}{2}\mathbf{c}$
	"Diagonal" glide plane	Glide reflection in a plane with glide vector:
n	$\perp \mathbf{c}$; $\perp \mathbf{a}$; $\perp \mathbf{b}$	$\frac{1}{2}(\mathbf{a}+\mathbf{b})$; $\frac{1}{2}(\mathbf{b}+\mathbf{c})$; $\frac{1}{2}(\mathbf{a}+\mathbf{c})$
n	$\perp (\mathbf{a}-\mathbf{b})$ or $\perp (\mathbf{b}-\mathbf{c})$ or $\perp (-\mathbf{a}+\mathbf{c})$	$\frac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c})$
n	$\perp (\mathbf{a}+\mathbf{b})$; $\perp (\mathbf{b}+\mathbf{c})$; $\perp (\mathbf{a}+\mathbf{c})$	$\frac{1}{2}(-\mathbf{a}+\mathbf{b}+\mathbf{c})$; $\frac{1}{2}(\mathbf{a}-\mathbf{b}+\mathbf{c})$; $\frac{1}{2}(\mathbf{a}+\mathbf{b}-\mathbf{c})$
	"Diamond" glide plane	Glide reflection in a plane with glide vector:
d	$\perp \mathbf{c}$; $\perp \mathbf{a}$; $\perp \mathbf{b}$	$\frac{1}{4}(\mathbf{a}\pm\mathbf{b})$; $\frac{1}{4}(\mathbf{b}\pm\mathbf{c})$; $\frac{1}{4}(\pm\mathbf{a}+\mathbf{c})$
d	$\perp (\mathbf{a}-\mathbf{b})$ or $\perp (\mathbf{b}-\mathbf{c})$ or $\perp (-\mathbf{a}+\mathbf{c})$	$\frac{1}{4}(\mathbf{a}+\mathbf{b}\pm\mathbf{c})$; $\frac{1}{4}(\pm\mathbf{a}+\mathbf{b}+\mathbf{c})$; $\frac{1}{4}(\mathbf{a}\pm\mathbf{b}+\mathbf{c})$
d	$\perp (\mathbf{a}+\mathbf{b})$; $\perp (\mathbf{b}+\mathbf{c})$; $\perp (\mathbf{a}+\mathbf{c})$	$\frac{1}{4}(-\mathbf{a}+\mathbf{b}\pm\mathbf{c})$; $\frac{1}{4}(\pm\mathbf{a}-\mathbf{b}+\mathbf{c})$; $\frac{1}{4}(\mathbf{a}\pm\mathbf{b}-\mathbf{c})$
	Glide line (2 dimensions)	Glide reflection in a line with glide vector:
g	$\perp \mathbf{b}$; $\perp \mathbf{a}$	$\frac{1}{2}\mathbf{a}$; $\frac{1}{2}\mathbf{b}$

second line. The rotation part of the result contains a twofold axis parallel to **a**; the translation part contains $\mathbf{a}/2$, parallel to the axis, and the perpendicular vector $(\mathbf{b}+\mathbf{c})/2$. Hence the successive application of the n and c glides, defined as above, is equivalent to the application of a twofold screw axis parallel to **a** which, however, does not pass through the origin, since it has a nonzero perpendicular

Table 3.2 Symbols for symmetry elements as given in ITA83 (cont.)

Symbol	Symmetry element and its orientation	Definition of symmetry operation
1	None	Identity
2, 3 ,4 ,6	n-fold rotation axis, n	Counterclockwise rotation by $360/n$ degrees around the axis
$\overline{1}$	Center of symmetry	Inversion through a point.
$\overline{2} = m$, $\overline{3}$, $\overline{4}$, $\overline{6}$	Rotoinversion axis, $\overline{n}$ and inversion point on the axis	Counterclockwise rotation by $360/n$ around the axis, followed by inversion through a point on the axis.
	Screw axes, n_p	
2_1 $3_1, 3_2$ $4_1, 4_2, 4_3$ $6_1, 6_2, 6_3, 6_4, 6_5$		Right-handed screw rotation by $360/n$ degrees around the axis, with screw vector $(p/n)\mathbf{t}$; here $\mathbf{t}$ is the shortest lattice translation vector parallel to the axis in the direction of the screw.

translation component. We have, similarly,

$$bn \equiv \left([2_{010}], \frac{1}{2}\mathbf{a}\right)$$

and

$$bc \equiv \left([2_{001}], \frac{1}{2}(\mathbf{b}+\mathbf{c})\right).$$

We see that bc results in a twofold screw axis parallel to $\mathbf{c}$, while the result of bn is a simple twofold axis parallel to $\mathbf{b}$. Let us now evaluate the result of successive application of a glide and a screw axis perpendicular to it. For example,

$$\begin{aligned}\left([m_{100}], \frac{1}{2}\mathbf{b}\right)\left([2_{100}], \frac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c})\right) &= \left([\overline{1}], -\frac{1}{2}\mathbf{a}+\mathbf{b}+\frac{1}{2}\mathbf{c}\right)\\ &\equiv \left([\overline{1}], \frac{1}{2}(\mathbf{a}+\mathbf{c})\right).\end{aligned}$$

As expected, the result is inversion–but through a point not located at the origin. No new rotational parts will result from other pairwise combinations of the above seven symmetry operators. We must, however, add the identity operator, its translational part being chosen arbitrarily as a zero vector (it can be any lattice vector).

Let us now examine the correspondence of the above operators to their rotational parts:

$$
\begin{array}{lcc}
([1],\mathbf{0}) & \Longrightarrow & [1], \\
([2_{001}],(\mathbf{b}+\mathbf{c})/2) & \Longrightarrow & [2_{001}], \\
([2_{010}],\mathbf{a}/2) & \Longrightarrow & [2_{010}], \\
([2_{100}],(\mathbf{a}+\mathbf{b}+\mathbf{c})/2) & \Longrightarrow & [2_{100}], \\
([\overline{1}],(\mathbf{a}+\mathbf{c})/2) & \Longrightarrow & [\overline{1}], \\
([m_{001}],(\mathbf{a}+\mathbf{b})/2) & \Longrightarrow & [m_{001}], \\
([m_{010}],\mathbf{c}/2) & \Longrightarrow & [m_{010}], \\
([m_{100}],\mathbf{b}/2) & \Longrightarrow & [m_{100}].
\end{array}
$$

The rotation operators on the right-hand sides of the correspondences are just the elements of the orthorhombic point group *mmm*. The symmetry operators on the left-hand sides are representative elements of the space group *Pbcn* (another octet of representative elements could be chosen). The whole space group will be exhausted if we add to each representative, in turn, all the lattice vectors. The space group is split in this way into eight infinite sets, where all the operators of a set have the same rotation part. One of the sets, that containing identity as its rotation operator, is a group, the group of the lattice; it can be shown that the group of the lattice translations is an invariant subgroup of the space group. There is a one-to-one correspondence between each of these sets and a rotation operator of *mmm*. Also, if we apply successively two operators from different sets, we obtain an operator with a rotation part equal to the product of the individual rotation parts. Moreover, the product of all the operators of one set with all the operators of another set will yield one of the existing sets. Similarly, the "inverse" of a set is also one of the existing sets. Summing up, the eight sets (that is, the space group *Pbcn*) are a group which is isomorphic to the point group *mmm*. We may note that the above representations of b, c, and n are a possible set of generating elements of the space group *Pbcn*. It follows from the above that the full space-group symbol of *Pbcn* is $P\frac{2_1}{b}\frac{2}{c}\frac{2_1}{n}$.

3.4 Some general considerations

The above example can be generalized to any crystallographic space group in the following manner. Let $\mathcal{S}=\{(\mathbf{P}_i,\mathbf{t}_k)\}$ be a space group, and let $\Gamma=\{(\mathbf{I},\mathbf{r}_{L_j})\}$ be the group of its lattice. Of course, Γ is a subgroup of $\mathcal{S}$, but we can readily see that it is an invariant subgroup. If $(\mathbf{P}_i,\mathbf{t}_k)$ is any element of $\mathcal{S}$, and $(\mathbf{I},\mathbf{r}_{L_j})$ is any element of Γ, we know from Section 2.2.3 that Γ is an invariant subgroup of $\mathcal{S}$ if $(\mathbf{P}_i,\mathbf{t}_k)(\mathbf{I},\mathbf{r}_{L_j})(\mathbf{P}_i,\mathbf{t}_k)^{-1}$ is also an element of Γ. Indeed,

$$
\begin{aligned}
(\mathbf{P}_i,\mathbf{t}_k)(\mathbf{I},\mathbf{r}_{L_j})(\mathbf{P}_i,\mathbf{t}_k)^{-1} &= (\mathbf{P}_i,\mathbf{t}_k)(\mathbf{I},\mathbf{r}_{L_j})(\mathbf{P}_i^{-1},-\mathbf{P}_i^{-1}\mathbf{t}_k) \\
&= (\mathbf{P}_i,\mathbf{t}_k)(\mathbf{P}_i^{-1},-\mathbf{P}_i^{-1}\mathbf{t}_k+\mathbf{r}_{L_j}) \\
&= (\mathbf{I},-\mathbf{t}_k+\mathbf{P}_i\mathbf{r}_{L_j}+\mathbf{t}_k) \\
&\in \Gamma
\end{aligned}
\tag{3.17}
$$

since $\mathbf{P}_i\mathbf{r}_{L_j}$ is also a vector of the lattice of the space group $\mathcal{S}$. Now let

$$(\mathbf{I},\mathbf{0}),(\mathbf{P}_2,\mathbf{t}_2),\ldots,(\mathbf{P}_g,\mathbf{t}_g)$$

be g representative operators of the space group, such that their translation parts contain all the existing nonsymmorphic translations and specify the origin of the space-group, and the rotation parts are all different (and exhaust the elements of the corresponding point group). The space group $\mathcal{S}$ can now be represented as a union of g right cosets of Γ,

$$\mathcal{S} = \Gamma \cup \Gamma(\mathbf{P}_2, \mathbf{t}_2) \cup \cdots \cup \Gamma(\mathbf{P}_g, \mathbf{t}_g).$$

It can be shown that a representation in terms of a union of left cosets of Γ is also possible.

Another general consideration, of unquestionable practical value, is the effect of a shift of the origin or the change of setting on the translation part of a space-group symmetry operator. For simplicity, let us start with a shift of the origin only. Let the operator be given by $(\mathbf{P}_s, \mathbf{t}_s)$, and let the origin shift operation be given by

$$\mathbf{r}' = \mathbf{r} - \boldsymbol{\rho} \equiv (\mathbf{I}, -\boldsymbol{\rho})\mathbf{r}.$$

The symmetry transformation is

$$\mathbf{r}_s = (\mathbf{P}_s, \mathbf{t}_s)\mathbf{r},$$

and hence

$$\mathbf{r}_s = (\mathbf{P}_s, \mathbf{t}_s)(\mathbf{I}, \boldsymbol{\rho})\mathbf{r}'.$$

The shifted result of the symmetry transformation is therefore

$$\mathbf{r}'_s = (\mathbf{I}, -\boldsymbol{\rho})(\mathbf{P}_s, \mathbf{t}_s)(\mathbf{I}, \boldsymbol{\rho})\mathbf{r}' = (\mathbf{I}, -\boldsymbol{\rho})(\mathbf{P}_s, \mathbf{P}_s\boldsymbol{\rho} + \mathbf{t}_s)\mathbf{r}' = [\mathbf{P}_s, (\mathbf{P}_s - \mathbf{I})\boldsymbol{\rho} + \mathbf{t}_s]\mathbf{r}'. \tag{3.18}$$

Example. Let us find a shift of the origin which will reduce the translation part of the inversion operator in *Pbcn* to a zero vector. That is,

$$([\overline{1}], (\mathbf{a} + \mathbf{c})/2) \longrightarrow ([\overline{1}], \mathbf{0}).$$

In this case, we have $\mathbf{t}_s = (\mathbf{a}+\mathbf{c})/2$, and our requiremement is: $([\overline{1}] - [1])\boldsymbol{\rho} + (\mathbf{a}+\mathbf{c})/2 = \mathbf{0}$. It follows that $\rho = (\mathbf{a}+\mathbf{c})/4$, and the required origin-shift operation is

$$\begin{aligned}([1], -(\mathbf{a}+\mathbf{c})/4)([\overline{1}], (\mathbf{a}+\mathbf{c})/2)([1], (\mathbf{a}+\mathbf{c})/4) &= ([1], -(\mathbf{a}+\mathbf{c})/4)([\overline{1}], (\mathbf{a}+\mathbf{c}/4) \\ &= ([\overline{1}], \mathbf{0})\end{aligned}$$

which was to be shown. The same transformation must of course be applied to all the remaining representative space-group operators. We shall do this in detail for one of them, and leave the others to one of the Exercises. Let us take care of the operator corresponding to the two-fold screw axis parallel to $\mathbf{a}$

$$\begin{aligned}
&\left([1], -\frac{\mathbf{a}+\mathbf{c}}{4}\right)\left([2_{100}], \frac{\mathbf{a}+\mathbf{b}+\mathbf{c}}{2}\right)\left([1], \frac{\mathbf{a}+\mathbf{c}}{4}\right)\\
&= \left([1], -\frac{\mathbf{a}+\mathbf{c}}{4}\right)\left([2_{100}], [2_{100}]\frac{\mathbf{a}+\mathbf{c}}{4} + \frac{\mathbf{a}+\mathbf{b}+\mathbf{c}}{2}\right)\\
&= \left([1], -\frac{\mathbf{a}+\mathbf{c}}{4}\right)\left([2_{100}], \frac{\mathbf{a}-\mathbf{c}}{4} + \frac{\mathbf{a}+\mathbf{b}+\mathbf{c}}{2}\right)\\
&= \left([2_{100}], \frac{\mathbf{a}+\mathbf{b}}{2}\right).
\end{aligned}$$

Proceeding as above, we obtain for all the eight representative operators of *Pbcn*

$$\begin{aligned}
&([1], \mathbf{0}),\\
&([2_{001}], (\mathbf{a}+\mathbf{b}+\mathbf{c})/2),\\
&([2_{010}], \mathbf{c}/2),\\
&([2_{100}], (\mathbf{a}+\mathbf{b})/2),\\
&([\overline{1}], \mathbf{0}),\\
&([m_{001}], (\mathbf{a}+\mathbf{b}+\mathbf{c})/2),\\
&([m_{010}], \mathbf{c}/2),\\
&([m_{100}], (\mathbf{a}+\mathbf{b})/2).
\end{aligned}$$

In this and similar calculations we have omitted integer multiples of lattice vectors and made the translation parts positive by adding lattice vectors (if, for example, the result was -**a**/2, we converted it to **a**/2 by adding **a**).

Let us now consider the effect of a change of setting on the representation of a space-group operator. We shall first deal with the transformation of the coordinates and then proceed to a change of the basis vectors (that is, a change of setting). The symbolic matrix notation will be used.

Let the transformation of the coordinates, with zero origin shift assumed,be given by

$$\boldsymbol{r}' = (\boldsymbol{T}, \boldsymbol{0})\boldsymbol{r},$$

where $\boldsymbol{T}$ is a nonsingular 3×3 matrix. The result of the symmetry transformation in the old system is

$$\begin{aligned}
\boldsymbol{r}_s &= (\boldsymbol{P}_s, \boldsymbol{t}_s)\boldsymbol{r}\\
&= (\boldsymbol{P}_s, \boldsymbol{t}_s)(\boldsymbol{T}^{-1}, \boldsymbol{0})\boldsymbol{r}'
\end{aligned}$$

and in the new system it becomes

$$\begin{aligned}
\boldsymbol{r}'_s &= (\boldsymbol{T}, \boldsymbol{0})(\boldsymbol{P}_s, \boldsymbol{t}_s)(\boldsymbol{T}^{-1}, \boldsymbol{0})\\
&= (\boldsymbol{T}, \boldsymbol{0})(\boldsymbol{P}_s\boldsymbol{T}^{-1}, \boldsymbol{t}_s)\boldsymbol{r}'\\
&= (\boldsymbol{T}\boldsymbol{P}_s\boldsymbol{T}^{-1}, \boldsymbol{T}\boldsymbol{t}_s)\boldsymbol{r}'. \qquad (3.19)
\end{aligned}$$

The matrix $\boldsymbol{T}$premultiplies a column vector of coordinates of a point, denoted here by $\boldsymbol{r}$. Before the transformation, the position vector of this point can be

written as $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$ or, in matrix notation, $\mathbf{r} = \mathbf{A}^{\mathrm{T}}\boldsymbol{r}$, where $\mathbf{A}^{\mathrm{T}}$ is the row of vectors **a**, **b**, and **c**. Since the position vector of a point is independent of the choice of the basis or the system of coordinates, if $\boldsymbol{r}$ has been premultiplied by the matrix $\boldsymbol{T}$, then $\mathbf{A}^{\mathrm{T}}$ must be postmultiplied by the matrix $\boldsymbol{T}^{-1}$ in order to preserve the length and direction of the position vector **r**. We thus have

$$\mathbf{r} = \mathbf{A}^{\mathrm{T}}\boldsymbol{r} = \mathbf{A}^{\mathrm{T}}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{r}. \tag{3.20}$$

Example. Consider the effect of a change of basis on the representation of a symmetry operator. Let, for example, **a** remain unchanged, **b** be replaced by $-\mathbf{c}$, and **c** be replaced by **b**. The transformation can be written as

$$(\mathbf{a}\ \ \overline{\mathbf{c}}\ \ \mathbf{b}) = (\mathbf{a}\ \ \mathbf{b}\ \ \mathbf{c}) \begin{pmatrix} 1\,0\,0 \\ 0\,0\,1 \\ 0\,\overline{1}\,0 \end{pmatrix}. \tag{3.21}$$

The corresponding coordinate transformation matrix must be the inverse of the matrix appearing in eqn (3.21), which happens to be its transpose. Thus

$$\boldsymbol{T} = \begin{pmatrix} 1\,0\,0 \\ 0\,0\,\overline{1} \\ 0\,1\,0 \end{pmatrix}. \tag{3.22}$$

Consider now the *b* glide plane operator of the space group *Pbcn*. The (matrix, vector) representation of this operator is

$$([m_{100}], (\mathbf{a}+\mathbf{b})/2) : \left[\begin{pmatrix} \overline{1}\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} \right]. \tag{3.23}$$

If we now substitute eqns (3.22) and (3.23) into the transformed operator in eqn (3.19) we see that that the matrix corresponding to the mirror plane remains unchanged but the components of the translation vector change from (1/2, 1/2, 0) to (1/2, 0, 1/2). The *b* glide plane thus becomes a *c* glide plane. If we apply this transformation to all the representative elements of *Pbcn* found above, we shall see that the space-group symbol corresponding to this changes from *Pbcn* to *Pcnb*. We shall discuss this further below, in connection with a space-group diagram of an orthorhombic space group.

In practice, we are interested in the structure of a given crystal, possessing a given space-group symmetry. The basis **abc** is fixed in such considerations, and a useful representation of the space-group operator is a $(3\times3\mathrm{matrix}, 3\times1\mathrm{matrix})$ pair (also denoted as a (matrix,vector) pair). The matrix representation of the rotation part is the same as that described in Chapter 2, while the representation of the translation part is simply a column of vector components. So, the representation of $(\mathbf{P}, \mathbf{t})$ is

$$\left[\begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}, \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}\right]$$

where $\mathbf{t} = t_1\mathbf{a} + t_2\mathbf{b} + t_3\mathbf{c}$.

Let us represent in the above manner the representative space-group operators of *Pbcn*:

$$([1], \mathbf{0}) \quad \Longrightarrow \quad \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right],$$

$$([2_{001}], (\mathbf{a}+\mathbf{b}+\mathbf{c})/2) \quad \Longrightarrow \quad \left[\begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}\right],$$

$$([2_{010}], \mathbf{c}/2) \quad \Longrightarrow \quad \left[\begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1/2 \end{pmatrix}\right],$$

$$([2_{100}], (\mathbf{a}+\mathbf{b})/2) \quad \Longrightarrow \quad \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}\right],$$

$$([\overline{1}], \mathbf{0}) \quad \Longrightarrow \quad \left[\begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right],$$

$$([m_{001}], (\mathbf{a}+\mathbf{b}+\mathbf{c})/2) \quad \Longrightarrow \quad \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}\right],$$

$$([m_{010}], \mathbf{c}/2) \quad \Longrightarrow \quad \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & \overline{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1/2 \end{pmatrix}\right],$$

$$([m_{100}], (\mathbf{a}+\mathbf{b})/2) \quad \Longrightarrow \quad \left[\begin{pmatrix} \overline{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}\right].$$

The last stage seems more complicated than the others, however, it leads to a very concise symbolic representation of symmetry transformations, which finds application in all crystallographic formalisms. If $(\boldsymbol{P}, \boldsymbol{t})$ is a matrix representation of a space-group operator, the corresponding representation of a space-group transformation is $\boldsymbol{P}\boldsymbol{r} + \boldsymbol{t}$, and in explicit (matrix, vector component) notation it becomes

$$\begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} P_{11}x_1 + P_{12}x_2 + P_{13}x_3 + t_1 \\ P_{21}x_1 + P_{22}x_2 + P_{23}x_3 + t_1 \\ P_{31}x_1 + P_{32}x_2 + P_{33}x_3 + t_1 \end{pmatrix}.$$

The space-group operations of *Pbcn* can now be represented in terms of the coordinates $x \equiv x_1$, $y \equiv x_2$, and $z \equiv x_3$:

(1) x, y, z; (2) $\overline{x} + \frac{1}{2}, \overline{y} + \frac{1}{2}, z + \frac{1}{2}$; (3) $\overline{x}, y, \overline{z} + \frac{1}{2}$; (4) $x + \frac{1}{2}, \overline{y} + \frac{1}{2}, \overline{z}$;

(5) $\overline{x}, \overline{y}, \overline{z}$; (6) $x + \frac{1}{2}, y + \frac{1}{2}, \overline{z} + \frac{1}{2}$; (7) $x, \overline{y}, z + \frac{1}{2}$; (8) $\overline{x} + \frac{1}{2}, y + \frac{1}{2}, z$.

If the point with coordinates x, y, z in the above list does not coincide with any symmetry element, in which case it will be brought into coincidence with itself, different coordinates are generated for the remaining seven points by the representative operators of the space group *Pbcn*. The positions of these eight points are the *general equivalent positions* of this space group. The elements on which a point can be brought into coincidence with itself are only the twofold axes along [010] and the eight centers of symmetry: this is so because if the point is placed on a screw axis or on a glide plane, the corresponding operators will displace it by the intrinsic translation vector.

Suppose, a point is placed on the simple twofold axis along [010]. It will be shown in the next section that the coordinates of a point on the axis corresponding to the operator $([2_{010}], \mathbf{c}/2)$ are $(0, y, \frac{1}{4})$. If we substitute these coordinates for (x, y, z) in the general equivalent positions we obtain four pairs of identical positions, or only four equivalent positions:

$$0, y, \tfrac{1}{4}; \quad \tfrac{1}{2}, \overline{y} + \tfrac{1}{2}, \tfrac{3}{4}; \quad 0, \overline{y}, \tfrac{3}{4}; \quad \tfrac{1}{2}, \overline{y} + \tfrac{1}{2}, \tfrac{1}{4}.$$

This is an example of *special equivalent positions*–they correspond to points located on the four independent simple twofold axes in the unit cell of the space group *Pbcn*. In general, a special position is brought into coincidence with itself by one or more operators of the space group (for example a point at the intersection of a twofold axis and a mirror plane).

This representation of space-group operations in terms of fractional coordinates lends itself conveniently to a variety of crystallographic applications. It is therefore used very extensively used in standard symmetry tables (Henry and Lonsdale 1952, 1965, and ITA83) and is the representation best known to the crystallographic community. However, as is obvious from the above, this simplicity conceals the group properties of the symmetry operators which are interesting for their own sake but are also most important in various computational approaches. We have therefore dedicated the first few sections of this chapter to an introduction to the general space-group formalism. Of course, once the fractional-coordinates representation is given, it is very simple to recover from it the (3 ×3 matrix, 3 ×1 matrix) representation that is needed for more general considerations (see the Exercises). The above space-group tables present, for

each space group listed the general and special equivalent positions, as well as other information, some of which will be referred to in Chapter 5.

3.5 Classification and tabulation of crystallographic space groups

The first comprehensive classifications of the 230 three-dimensional crystallographic space groups were published by Schönflies (1891), whose notation is still in wide use in many fields of science, and Fedorov (1891), whose space-group diagrams–also rederived by Hilton (1903)–appear to be the basis of the diagrams we see in modern crystallographic symmetry tables. As soon as crystallography became a promising experimental science, it was decided to introduce a standard notation for space symmetry elements/operators, and to choose conventional space-group origins and basis vectors, so that the space-group information could be tabulated and used conveniently. The first such tables, entitled *Internationale Tabellen zur Bestimmung von Kristallstrukturen* (Hermann 1935) were edited by C. Hermann. The following series of tables, entitled *International Tables for X-ray Crystallography*, started with Volume I (Henry and Lonsdale 1952), edited by N. F. M. Henry and K. Lonsdale, which contains space-group symmetry information, and its application to structure factor and electron density calculations. Several corrected reprints of this volume followed. The present series of tables, entitled *International Tables for Crystallography*, started with Volume A (Hahn 1983), edited by Th. Hahn. This volume contains significantly enhanced information about various aspects of space-group and point-group symmetry, including a complete tabulation. The fourth revised edition of Volume A appeared in 1996. We intended here to provide a short introduction to the use of the most recent space-group tables given in Volume A, but before doing so, it is in order to discuss the problem of finding the location of symmetry elements, given the rotation and translation parts of the corresponding space-group operators. This is of major importance to the understanding of the space-group diagrams which appear in the tables.

3.6 Location of symmetry elements

Let us assume that the origin of the space group has been fixed according to some choice or convention, and that the basis vectors are those of the symmetric unit cell discussed in Chapter 2. Let us denote the rotation operator and translation vector by their symbolic matrix representations (bold italic). Before proceeding to a description of the *International Tables*, we wish to examine the problem of finding the location of the symmetry elements within the symmetric unit cell, in which an origin has been defined. The following considerations follow closely those given by Shmueli (1984).

Let the point $\boldsymbol{x}^{\mathrm{T}} = (x \;\; y \;\; z)$ coincide with some symmetry element, say that corresponding to the space group operator $(\boldsymbol{P},\ \boldsymbol{t}_{\mathrm{intr}} + \boldsymbol{t}_{\mathrm{loc}})$, where $\boldsymbol{t}_{\mathrm{intr}}$ is the *intrinsic* translation (for example a translation along a screw axis or parallel to

a glide plane) and $\boldsymbol{t}_{\rm loc}$ is the location-dependent component of the translation. If we apply the matrix of the space-group operator to the point $\boldsymbol{x}$, the point will remain unmoved if the intrinsic translation is a zero vector, and will be shifted by $\boldsymbol{t}_{\rm intr}$ if the latter is nonzero. In general, we have

$$\begin{aligned}(\boldsymbol{P}, \boldsymbol{t}_{\rm intr} + \boldsymbol{t}_{\rm loc})\boldsymbol{x} &= \boldsymbol{P}\boldsymbol{x} + \boldsymbol{t}_{\rm intr} + \boldsymbol{t}_{\rm loc} \\ &= \boldsymbol{x} + \boldsymbol{t}_{\rm intr},\end{aligned}$$

and, in any case

$$\boldsymbol{P}\boldsymbol{x} + \boldsymbol{t}_{\rm loc} = \boldsymbol{x} \quad \text{or} \quad (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} = \boldsymbol{t}_{\rm loc}, \tag{3.24}$$

where $\boldsymbol{I}$ is the unit matrix.

If $\boldsymbol{P}$ corresponds to the inversion, or to an axis of inversion $\overline{n}$, with $n > 2$, it is easily shown that $\boldsymbol{I} - \boldsymbol{P}$ is nonsingular and that the coordinates of a point located on the axis (or the center) are simply $\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{P})^{-1}\boldsymbol{t}_{\rm loc}$.

The simplest example is the location of a center of symmetry. We have here

$$\left[\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - \begin{pmatrix}\overline{1}&0&0\\0&\overline{1}&0\\0&0&\overline{1}\end{pmatrix}\right]^{-1} = \begin{pmatrix}1/2&0&0\\0&1/2&0\\0&0&1/2\end{pmatrix}$$

and hence the location of the center of symmetry is given by $\boldsymbol{x} = \boldsymbol{t}_{\rm loc}/2$. If we add a lattice vector, say $\boldsymbol{r}_{\rm L}$, such that $(\boldsymbol{t}_{\rm loc} + \boldsymbol{r}_{\rm L})/2$ lies within the unit cell, another translationally independent center is generated. For example, if $\boldsymbol{t}_{\rm loc}$ is a zero vector and the Bravais lattice is of type P, there are eight such centers and their positive fractional coordinates are (0, 0, 0), (1/2, 0, 0), (0, 1/2, 0), (0, 0, 1/2), (0, 1/2, 1/2), (1/2, 0, 1/2), (1/2, 1/2, 0) and (1/2, 1/2, 1/2). Similarly, there are 16 centers per cell in a singly centered lattice (types A, B, C, and I) and 32 centers in a symmetric unit cell of an F-type Bravais lattice (see the Exercises).

If $\boldsymbol{P}$ represents a twofold axis of proper rotation or a twofold axis of improper rotation (that is, a plane of reflection), the translation component $\boldsymbol{t}_{\rm loc}$ is perpendicular to the element and, in either case, the solution of eqn (3.24) can be written as $\boldsymbol{x} = \boldsymbol{t}_{\rm loc}/2 + \boldsymbol{x}_{\rm p}$, where $\boldsymbol{x}_{\rm p}$ is an indeterminate part of $\boldsymbol{x}$ and can be set equal to a zero vector. In general, for $\boldsymbol{P} \longrightarrow 2$, $\boldsymbol{x}_{\rm p}$ is parallel to the twofold axis and for $\boldsymbol{P} \longrightarrow m$, $\boldsymbol{x}_{\rm p}$ is parallel to the plane of reflection. For example, it is now easy to find the locations of the representative elements of the space group *Pbcn* (No. 60), considered above. Table 3.3 shows a summary of their derivation.

The above considerations suffice for the determination of the locations of the symmetry elements in all the space groups belonging to the triclinic, monoclinic, and orthorhombic systems, and are useful also for the twofold axes and reflection and glide planes in higher systems.

We are now left with axes of proper rotation of orders higher than 2. In standard representations of space groups (for example in ITA83), fourfold and sixfold axes is always parallel to one of the basis vectors, while a threefold axis

Table 3.3 Location of the representative elements of the space group *Pbcn*

Operator	$\boldsymbol{t}_{\text{intr}}^{\text{T}}$	$\boldsymbol{t}_{\text{loc}}^{\text{T}}$	$\boldsymbol{x}^{\text{T}}$
([1], **0**)			
([2_{001}], (**a**+**b**+**c**)/2)	(0 0 1/2)	(1/2 1/2 0)	(1/4 1/4 z)
([2_{010}], **c**/2)	(0 0 0)	(0 0 1/2)	(0 y 1/4)
([2_{100}], (**a**+**b**)/2)	(1/2 0 0)	(0 1/2 0)	(x 1/4 0)
([$\overline{1}$], **0**)	(0 0 0)	(0 0 0)	(0 0 0)
([m_{001}], (**a**+**b**+**c**)/2)	(1/2 1/2 0)	(0 0 1/2)	(x y 1/4)
([m_{010}], **c**/2)	(0 0 1/2)	(0 0 0)	(x 0 z)
([m_{100}], (**a**+**b**)/2)	(0 1/2 0)	(1/2 0 0)	(1/4 y z)

may be parallel to a basis vector or to a body diagonal of the unit cell. In either case the matrix corresponding to $\boldsymbol{I} - \boldsymbol{P}$ is singular, and the singularity is most obvious if the axis is parallel to a basis vector. If, for example, an axis of rotation is parallel to **c**, then the third coordinate (x_3) of the vector $\boldsymbol{x}$ is indeterminate, and the remaining two coordinates are solved for by deleting the third row and third column of the matrix corresponding to $\boldsymbol{I} - \boldsymbol{P}$ in eqn (3.24). Axes parallel to **a** and **b** are treated analogously.

The case where a threefold axis is parallel to a body diagonal of the unit cell is less obvious, since the singularity of the matrix corresponding to $\boldsymbol{I} - \boldsymbol{P}$ cannot be dealt with so simply as for the [100], [010] and [001] orientations above. The solution is:

1. Transform eqn (3.24) to a new system, such that one of its basis vectors coincides with a vector parallel to the axis of rotation.
2. Solve the transformed eqn (3.24) for the determinable components of the transformed $\boldsymbol{x}$, while equating the indeterminate component to zero.
3. Transform the solution vector back to the original system.

The general forms of the location vectors of the various symmetry elements can now be obtained as sums of the corresponding vectors $\boldsymbol{x}$ and vectors with variable coordinates (as in Table 3.3) which are based on lattice vectors parallel to the axes of rotation and planes of reflection (these include, of course, screw axes and glide planes). If, for example, a threefold axis in a cubic space group is parallel to $[1\overline{1}\overline{1}]$ and is found to pass through the point (1/6, -1/6, 1/3), a general location vector for this axis can be written as $(x+1/6, \overline{x}-1/6, \overline{x}+1/3)$. Of course, any vector with coordinates $(\delta, -\delta, -\delta)$ can be added to the general location vector without changing its geometrical significance.

The considerations outlined in this section are of fundamental importance in the construction of space-group diagrams, to be illustrated below. These diagrams, displayed in ITA83 for all the two- and three-dimensional space groups, play a most important role in the visualization of crystal structures.

3.7 Graphical representation of space groups

The printed symbols for symmetry planes and axes have already been given in Table 3.2. We now proceed to the graphical symbols for rotation and inversion axes, except those based on $m = \overline{2}$. These are given in Figs. 3.2, 3.3, and 3.4, which are reproduced from several parts of Table 1.4 in Volume A (ITA83). These symbols are used in ITA83 for axes perpendicular to the space-group diagrams–that is, tabulated projections of the unit cell belonging to a space group. Next we present in Fig. 3.5 the symbols used in space-group diagrams for mirror and glide planes perpendicular to the plane of projection (in three dimensions) and symmetry lines in the plane of the figure (in two dimensions). Finally, Fig. 3.6 presents the indication of symmetry planes parallel to the plane of projection, their height above the origin, and the direction of the gliding translation for glide planes. These symbols for axes and planes are almost universally used in the crystallographic literature and are presented and discussed in detail in ITA83. They are presented here as a geometrical realization of the analytical description of symmetry operators discussed earlier

Two examples of space-group diagrams are included. Figure 3.7 shows the page of Volume A (ITA83) pertaining to the standard representation of the monoclinic space group $P2_1/c$. It may be noted that a vast number of nonchiral crystals consisting of small and medium-sized organic molecules, the structures of which have been determined, belong to this space group. Let us describe this page in some detail.

The symbol $P2_1/c$, a Hermann–Mauguin or international symbol, tells us that the Bravais lattice of this space group is of type P and that the space group contains a twofold screw axis which is perpendicular to a glide plane of type c (see Table 3.1). As in the case of $Pbcn$, it can be seen that a center of inversion must be present, and the corresponding three operators with the identity are representative elements of a space group isomorphous with the point group $2/m$.

The next symbol in the first line is C_{2h}^5. This is the Schönflies notation for the space group considered. The symbol tells us that there is a twofold axis (C_2) in the isomorphic point group, as well as a "horizontal" (h) mirror plane, and the superscript "5" tells us that we have here the fifth space group isomorphic to C_{2h}. The geometrical information content carried by C_{2h}^5 is of course much lower than that of the Hermann–Mauguin symbol, and this is why the vast majority of crystallographic papers and books use the Hermann–Maguin notation.

Next, we have the point-group symbol $2/m$, which we have already encountered in Chapter 2, and we are told that the crystal system to which this space group belongs is monoclinic. The first item in the second line is the space-group number, here No. 14. This is a most important identifier within the compilation of space groups in ITA83.

The second item in the second line is the full space-group symbol. The first and third positions after P correspond to **a** and **c**, respectively, and are occupied by "1"; the second position is occupied by $2_1/c$. This means that the symmetry elements in the symbol are associated with the **b** axis only: the twofold screw

Symmetry axis or symmetry point	Graphical symbol‡
Identity	None
Twofold rotation axis Twofold rotation point (two dimensions)	
Twofold screw axis: '2 sub 1'	
Threefold rotation axis Threefold rotation point (two dimensions)	
Threefold screw axis: '3 sub 1'	
Threefold screw axis: '3 sub 2'	
Fourfold rotation axis Fourfold rotation point (two dimensions)	
Fourfold screw axis: '4 sub 1'	 
Fourfold screw axis: '4 sub 2'	
Fourfold screw axis: '4 sub 3'	 
Sixfold rotation axis Sixfold rotation point (two dimensions)	
Sixfold screw axis: '6 sub 1'	
Sixfold screw axis: '6 sub 2'	
Sixfold screw axis: '6 sub 3'	
Sixfold screw axis: '6 sub 4'	
Sixfold screw axis: '6 sub 5'	
Centre of symmetry, inversion centre: '1 bar' Reflection point, mirror point (one dimension)	
Inversion axis: '3 bar'	
Inversion axis: '4 bar'	
Inversion axis: '6 bar'	
Twofold rotation axis with centre of symmetry	
Twofold screw axis with centre of symmetry	
Fourfold rotation axis with centre of symmetry	
'4 sub 2' screw axis with centre of symmetry	
Sixfold rotation axis with centre of symmetry	
'6 sub 3' screw axis with centre of symmetry	

Fig. 3.2 *Symmetry axes perpendicular to the plane of projection.* Reproduced with copyright permission of the International Union of Crystallography (IUCr).

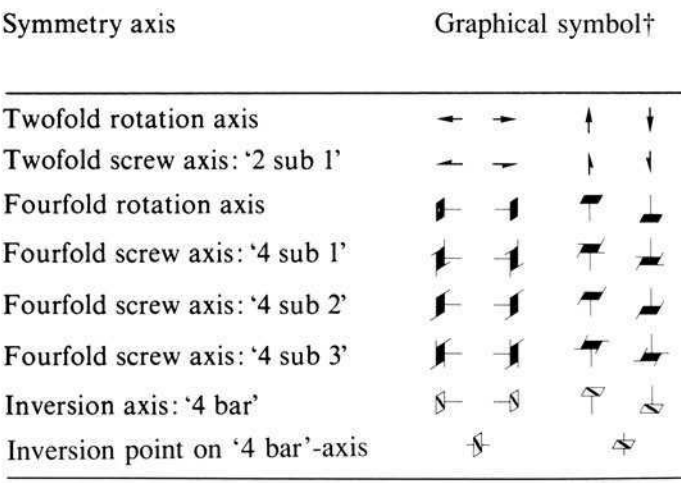


Fig. 3.3 *Symmetry axes parallel to the plane of projection.* Reproduced with copyright permission of the International Union of Crystallography (IUCr).

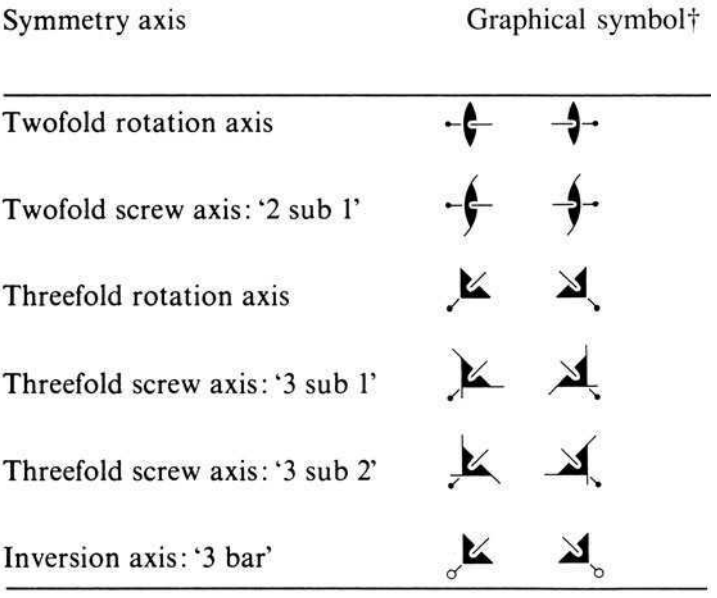


Fig. 3.4 *Symmetry axes inclined to the plane of projection (cubic only).* Reproduced with copyright permission of the International Union of Crystallography (IUCr).

Symmetry plane or symmetry line	Graphical symbol	Glide vector in units of lattice translation vectors parallel and normal to the projection plane
Reflection plane, mirror plane Reflection line, mirror line (two dimensions)	———	None
'Axial' glide plane Glide line (two dimensions)	– – – –	$\frac{1}{2}$ lattice vector along line in projection plane $\frac{1}{2}$ lattice vector along line in plane
'Axial' glide plane	.....	$\frac{1}{2}$ normal to projection plane
'Double' glide plane# (in centred cells only)	···–··–··–	*Two* glide vectors: $\frac{1}{2}$ along line parallel to projection plane, $\frac{1}{2}$ normal to projection plane
'Diagonal' glide plane	–·–·–·–	*One* glide vector with *two* components: $\frac{1}{2}$ along line parallel to projection plane, $\frac{1}{2}$ normal to projection plane
'Diamond' glide plane§ (pair of planes; in centred cells only)	–·–·–·– –·–·–·–	$\frac{1}{4}$ along line parallel to projection plane, combined with $\frac{1}{4}$ normal to projection plane (arrow indicates direction parallel to the projection plane for which the normal component is positive)

Fig. 3.5 *Symbols for mirror and glide planes (cf. Table 3.2).* Reproduced with copyright permission of the International Union of Crystallography (IUCr).

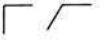
Symmetry plane	Graphical symbol†	Glide vector in units of lattice translation vectors parallel to the projection plane	Printed symbol
Reflection plane, mirror plane		None	*m*
'Axial' glide plane		$\frac{1}{2}$ lattice vector in the direction of the arrow	*a*, *b*, or *c*
'Double' glide plane# (in centred cells only)		*Two* glide vectors: $\frac{1}{2}$ in either of the directions of the two arrows	*e*
'Diagonal' glide plane		$\frac{1}{2}$ in the direction of the arrow	*n*
'Diamond' glide plane§ (pair of planes; in centred cells only)		$\frac{1}{2}$ in the direction of the arrow; the glide vector is always half of a centring vector, *i.e.* one quarter of a diagonal of the conventional face-centred cell	*d*

Fig. 3.6 *Symmetry planes parallel to plane of projection (cf. Table 3.2).* Reproduced with copyright permission of the International Union of Crystallography (IUCr).

axis is parallel to **b**, and the glide plane is perpendicular to it. The last item in the second line is the Patterson symmetry, which will be discussed in some detail in Chapter 6. Here too the **abc** notation is implied, where $2/m$ is associated with **b**. The first item in the third line says in words what we see in the full symbol, that is, that b is the "unique" axis associated with the symmetry elements.

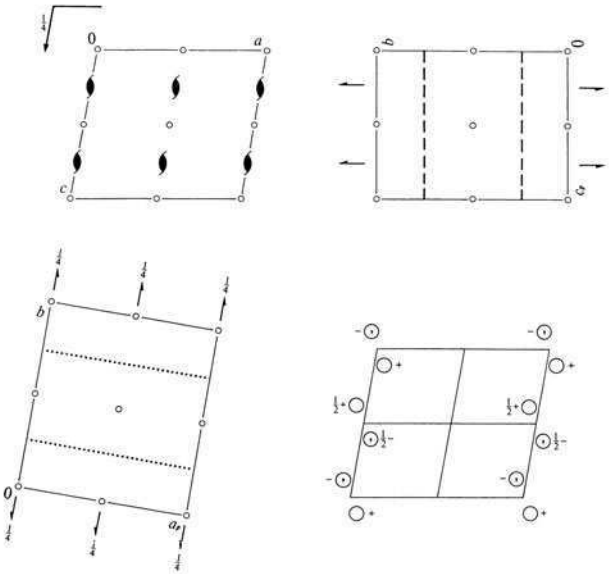
The attribute "CELL CHOICE 1" has to do with the designation of the glide plane in this monoclinic space group. As shown in Table 3.1, the possible glide planes in the monoclinic system are a, c and n. It can, however, be shown that these three glide planes are equivalent and can be related by a transformation of the basis vectors **a**, **b** and **c** or, in other words, that they correspond to different choices of the unit cell (only **a** and **c** are changed in the case of "unique axis b", and only **a** and **b** are changed in the case of "unique axis c"). The screw axis is always associated with **b** if b is unique, and with **c** if c is unique. Space group No. 14 is thus represented in ITA83 as $P12_1/c1$, $P12_1/n1$, and $P12_1/a1$, corresponding by convention to cell choices 1, 2 and 3, respectively with unique axis b, and as $P112_1/a$, $P112_1/n$, and $P112_1/b$, corresponding to cell choices 1, 2, and 3, respectively with a unique axis c. This redundancy is a most useful feature of ITA83. As we shall see in Chapter 5, the designation of the glide plane follows immediately from the interpretation of the diffraction pattern. Any of three cell choices for unique b or c may be found to be the desired one, and the once mandatory transformation to the formally standard cell choice 1 with unique b often seems to be an unnecessary constraint.

We now come to the graphical representation of the space group. The upper left drawing is the projection of the unit cell of the space group down the b axis. The translationally nonequivalent 2_1 axes are located at $(0, y, 1/4)$, $(1/2, y, 1/4)$, $(0, y, 3/4)$ and $(1/2, y, 3/4)$. The c glide is indicated to be located at a height of $1/4$ above the **ac** plane (fractional coordinates are employed throughout) and is parallel to that plane. The upper right and lower left drawings show that there is another c glide at a height of $3/4$ above the **ac** plane. There are eight translationally nonequivalent centers of symmetry at points $(p/2, q/2, r/2)$, where each of p, q and r is 0 or 1. One of these centers is of course at the origin of the unit cell. All these locations can be derived from the considerations given in Section 3.6. The lower left drawing is a projection of the cell down the c axis (note that the glide plane is dotted, indicating a translation normal to the plane of the projection; see Fig. 3.5). The edge a_p of the drawing is the projected a axis (that is, $a \sin \beta$). The lower right drawing has the same orientation as the upper left one. It shows the effect on the position of an object, say, the empty circle above the ac plane (indicated by a plus sign) adjacent to the origin, when the various space-group operations are applied to it. For example, the inversion through the center of symmetry at the origin produces a circle below the **ac** plane (indicated by a minus sign) but there is a comma at its center. This means that the "handedness" of the object has been changed, as well as its position. On the other hand, application of the screw rotation around an axis at $(0, y, 1/4)$ results in another empty circle, displaced by $1/2$ along the screw

$P\,2_1/c$ C_{2h}^5 $2/m$ Monoclinic

No. 14 $P\,1\,2_1/c\,1$ Patterson symmetry $P\,1\,2/m\,1$

UNIQUE AXIS b, CELL CHOICE 1



Origin at $\bar{1}$

Asymmetric unit $0\le x\le 1;\ 0\le y\le \frac{1}{4};\ 0\le z\le 1$

Symmetry operations

(1) 1 (2) $2(0,\frac{1}{4},0)$ $0,y,\frac{1}{4}$ (3) $\bar{1}$ 0,0,0 (4) c $x,\frac{1}{4},z$

Fig. 3.7 *Geometry of space group* $P2_1/c$ *(No. 14).* Reproduced with copyright permission of the International Union of Crystallography (IUCr).

axis. In general, proper rotations do not change the "handedness" of the object acted upon, whereas improper rotations (including reflections) produce such a change. This is readily seen if it is assumed that the object has attached to itself a right-handed system of coordinates: clearly, an inversion of the right-handed system transforms it into a left-handed one, and the same thing happens when the system is reflected (the rotation part of a glide-plane operator is a reflection in a mirror plane).

The asymmetric unit of the cell is chosen here chosen as a parallelepiped that has the same ac basis as that of the unit cell, its height being just 1/4 of the b

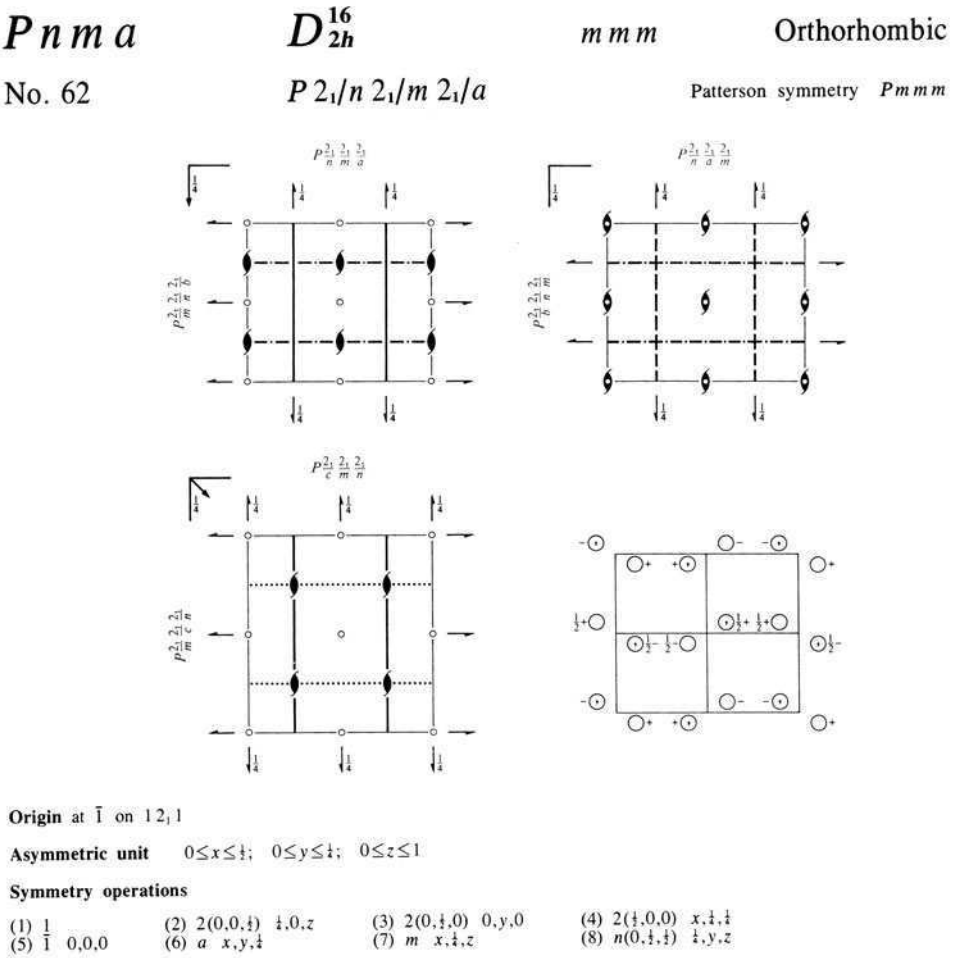


Fig. 3.8 *Geometry of space group Pnma (No. 62).* Reproduced with copyright permission of the International Union of Crystallography (IUCr).

axis. An application of the symmetry operations of the space group to this unit obviously fills the space exactly. Note, again, that the asymmetric unit can be chosen in many ways.

The page is concluded with a list of symmetry operations. The intrinsic translation components of the screw rotation are given in parentheses and those of the glide plane are implicit in its symbol. This list also presents the locations of the corresponding symmetry elements.

Our next example, shown in Fig. 3.8, is the page showing the standard representation of the orthorhombic space group $Pnma$ (No. 62). The structure of the header is similar to that for space group No. 14, except there is no "unique axis" and the concept behind the cell choice is different. This will become clear from the discussion given below. This space group, like the previously discussed $Pbcn$, is isomorphic to the point group $\frac{2}{m}\frac{2}{m}\frac{2}{m}$ or, briefly, mmm. The correspondence can be easily seen from the space-group diagrams, where the symbols are explained in Figs. 3.2–3.5. To each of the twofold axes of mmm there corresponds a family of parallel twofold screw axes and, as can be seen from the diagrams, these families of twofold screw axes are nonintersecting. To the first mirror plane in mmm there corresponds a family of diagonal (n) glide planes perpendicular to **a**, to the second mirror plane in mmm there corresponds a family of mirror planes (m) perpendicular to **b** and to the third mirror plane in mmm there corresponds a family of axial (a) glide planes perpendicular to **c**. Finally, to the center of symmetry in mmm there corresponds a family of such centers in $Pnma$. The upper left drawing is a projection of the cell down **c**, the a glide plane is parallel to the plane of projection, and the direction of the gliding translation (**a**/2) is indicated by the symbol in the upper left corner. If we place the origin in the upper-left corner of the drawing, the left edge of the drawing therefore corresponds to **a** (pointing down) and the upper edge of the drawing to **b** (pointing to the right). These three basis vectors form a right-handed triad–as they should. Let us now consider the upper right drawing. It follows from the above that this is a projection of the cell down **b**. If we choose the origin to be in the upper left corner of this drawing and assign to the left edge the vector **a** (pointing down the page), and to the upper edge the vector **c** (pointing to the right), in agreement with the glide-plane symbols, **a**, **b**, and **c** will form a left-handed triad, which is not allowed. One solution is to place the origin in the upper right corner, and associate **c** (pointing to the left) with the upper edge of the drawing and **a** with its right edge (pointing down). We end up with having $\mathbf{a} \perp n$ glide, $\mathbf{b} \perp m$ mirror, and $\mathbf{c} \perp a$ glide, with **a**, **b** and **c** forming a right-handed triad–as we should. A concise way of putting this is to say that the *setting* **abc** has been transformed into the setting $\mathbf{a}\overline{\mathbf{c}}\mathbf{b}$, or, in matrix notation:

$$\begin{pmatrix} \mathbf{a}' \\ \mathbf{b}' \\ \mathbf{c}' \end{pmatrix} = \begin{pmatrix} 1\,0\,0 \\ 0\,0\,\overline{1} \\ 0\,1\,0 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ \overline{\mathbf{c}} \\ \mathbf{b} \end{pmatrix}, \tag{3.25}$$

where $\mathbf{a}'$, $\mathbf{b}'$ and $\mathbf{c}'$ are the basis vectors corresponding to the new setting. There exist six settings for any orthorhombic space group, consistent with the permutations of the labels of the basis vectors and a right-handed system (see Chapter 2 of ITA83).

It can be seen in the diagram that six different full symbols for the space group appear near the upper and left edges of the three drawings. Let us try to understand one of them, namely that printed above the upper edge of the upper right drawing. If we look at the upper-right drawing from the top, we shall see

an n glide plane normal to $\mathbf{a}'$, an a glide plane normal to $\mathbf{b}'$ and a mirror plane m normal to $\mathbf{c}'$, where $\mathbf{a}'\mathbf{b}'\mathbf{c}'$ are given by eqn (3.25). The corresponding space group symbol is therefore $Pnam$, and as we can see in Fig. 3.8. there are six such symbols corresponding to the different settings of the same unit cell. This is of considerable practical importance; just as space group No. 14 may appear to one crystallographer as $P2_1/c$ and to another as $P2_1/n$, $Pnma$ and $Pnam$ two symbols that can be correctly assigned to the space group of the same crystal. The treatment of these multiple symbols is not as detailed in ITA83 as in the monoclinic case, but Table 4.3.1 in Chapter 4 of ITA83 (Bertaut 1983) contains all the necessary information.

3.8 Some computational aspects

Space-group symmetry accompanies all stages of crystal structure determination and refinement, and must therefore be readily dealt with by the relevant computer programs. A simple method which is used by many program packages is to input the symbolic coordinates of the general equivalent positions (see Section 3.4) from which the (matrix, vector) representations of the symmetry operators can be readily recovered (see the Exercises).

There are also programs which require as input all the components of the rotation and translation parts of the space-group operators. This is somewhat tedious, and is done when the performance of the program justifies the effort.

Another category of symmetry input involves the use of computer-readable space-group symbols. These can be stored in a small database, and the symmetry input can be confined to the space-group number and some of its attributes. The computer program interprets the symbols in terms of the generating elements of the space group, and these are used to compute the representative (matrix, vector) representations of the space-roup operators, either iteratively or directly. The space-group information given in Volume A (ITA83) was computed from a selection of generating elements for each space group (Fokkema 1983), presumably in an iterative manner.

A set of computer-readable symbols interpretable in terms of space-group generators was proposed by Hall (1981) for the 230 space groups; it was enlarged to include the symbols for the various choices of the monoclinic cell (Hall 1993), and was further extended to include the symbols for the orthorhombic space groups in their six (or sometimes fewer) different settings (Hall and Grosse-Kunstleve 2001). The representative elements of the space groups are generated from these symbols in an iterative manner. A detailed description of the Hall symbols is given in Hall's original paper (Hall 1981) and their further developments are described in the first and second editions of Volume B of the *International Tables for Crystallography* (Shmueli 1993a, 2001a).

Another approach was taken by Shmueli (1984), who proposed computer-readable space-group symbols containing explicitly a specific kind of generating element. These generators permit a direct (rather than iterative) derivation of

the space-group representatives according to the following scheme. Any space group may have one of the forms

$$\{(\mathbf{Q},\mathbf{u})\},$$
$$\{(\mathbf{Q},\mathbf{u})\} \times \{(\mathbf{R},\mathbf{v})\}, \text{ or}$$
$$\{(\mathbf{P},\mathbf{t})\} \times [\{(\mathbf{Q},\mathbf{u})\} \times \{(\mathbf{R},\mathbf{v})\}], \tag{3.26}$$

where $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{R}$ are point-group operators, and $\mathbf{t}$, $\mathbf{u}$, and $\mathbf{v}$ are zero vectors or translations not belonging to the lattice-translation subgroup. Each of the forms in eqn (3.26) enclosed in braces is evaluated as, for example,

$$\{(\mathbf{P},\mathbf{t})\} = \{(\mathbf{I},\mathbf{0}),(\mathbf{P},\mathbf{t}),(\mathbf{P},\mathbf{t})^2,\ldots,(\mathbf{P},\mathbf{t})^{g-1}\}, \tag{3.27}$$

where $\mathbf{I}$ is a unit operator and g is the order of the rotation operator $\mathbf{P}$ (that is, $\mathbf{P}^g = \mathbf{I}$). The representative elements of a space group are then given directly by the expansion in eqn (3.27) or by an ordered product of such expansions, as indicated in eqn (3.26). In the application of this formalism to automatic generation of geometric structure factors (Shmueli, 1993b; Shmueli, 2001c), the actual forms of the generators selected, based on the work of Zachariasen (1945), were adapted to all the three-dimensional space-groups given in ITA83.

The general structure of a three-generator symbol, corresponding to the last line of (3.26), as represented in Table 3.4 (Shmueli 2001c), is

$$\mathrm{LSC\$r_1Pt_1t_2t_3\$r_2Qu_1u_2u_3\$r_3Rv_1v_2v_3}, \tag{3.28}$$

where the symbols are defined as follows:

- L is the lattice type, which can be `P`, `A`, `B`, `C`, `I`, `F`, or `R`. The symbol `R` is used only for the seven rhombohedral space groups in their representations in terms of rhombohedral and hexagonal axes (the obverse setting in Henry and Lonsdale 1952).
- S is the crystal system, which can be `A` (triclinic), `M` (monoclinic), `O` (orthorhombic), `T` (tetragonal), `R` (trigonal), `H` (hexagonal) or `C` (cubic).
- C is the status of centrosymmetry, which can be `C` or `N` according to whether the space group is centrosymmetric or noncentrosymmetric, respectively.
- \$: this character is followed by six characters that define a generator of the space group.
- r_i is an indicator of the type of rotation that follows: r_i is `P` or `I` according to whether the rotation part of the ith generator is proper or improper, respectively.
- P, Q, R are two-character symbols for the matrix representations of the generators of the axial point groups (see Table 2.3) appearing in Table 3.4.
- $\mathrm{t_1t_2t_3}$, $\mathrm{u_1u_2u_3}$, $\mathrm{v_1v_2v_3}$ are the components of the translation parts of the generators, given in units of 1/12, for example, the translation part (0 1/2 3/4) is given in Table 3.4 as 069. An *exception*: (0 0 5/6) is denoted by 005 and not by 0010.

The two-character symbols for the matrices of rotation, which appear in the explicit space-group symbols in Table 3.3, are related to the generating point-group operators, shown in Table 2.3, as follows:

$$\begin{aligned}
&\texttt{1A}\rightarrow [\overline{1}], \qquad \texttt{2A}\rightarrow [2^{O,T,C}_{100}], \qquad [\texttt{2B}] \rightarrow [2^{M2,O,T,C}_{010}],\\
&\texttt{2C}\rightarrow [2^{all}_{001}], \quad \texttt{2D}\rightarrow [2^{T,C}_{110}], \qquad [\texttt{2E}] \rightarrow [2^{T,C}_{\overline{1}10}],\\
&\texttt{2F}\rightarrow [2^{H}_{100}], \quad \texttt{2G}\rightarrow [2^{H}_{210}], \qquad [\texttt{3Q}] \rightarrow [3^{C}_{111}],\\
&\texttt{3C}\rightarrow [3^{H}_{001}], \quad \texttt{4C}\rightarrow [4^{T,C}_{100}], \qquad [\texttt{6C}] \rightarrow [6^{H}_{001}].
\end{aligned}$$

The actual matrices are given in Table 2.3. Of course, only matrices of proper rotation are given (and required), since the corresponding matrices of improper rotation are created by the program for appropriate value of the indicator r_i. The first character of a symbol is the order of the axis of rotation, and the second character specifies its orientation: in terms of direct-space lattice vectors, we have

$$\begin{aligned}
&\texttt{A} = [100], \quad \texttt{B} = [010], \quad \texttt{C} = [001], \quad \texttt{D} = [110],\\
&\texttt{E} = [1\overline{1}0], \quad \texttt{F} = [100], \quad \texttt{G} = [210], \quad \text{and } \texttt{Q} = [111].
\end{aligned}$$

for the standard orientations of the axes of rotation. This is also seen from the subscripts on the orders of the axes in the symbols of the generating operators shown above.

In this scheme a space group is determined by one, two or at most three generators [see eqn (3.26). It should be pointed out that a convenient way of achieving a representation of the space group in any setting and relative to any origin is to start from the standard generators in Table 3.3 and let the computer program perform the appropriate transformation of the generators only. The subsequent expansion of the transformed generators and the formation of the required products (see eqns (3.26) and (3.27)) leads to the new representation of the space group.

In order to illustrate an explicit space-group symbol consider, for example, the symbol for the space group $Ia\overline{3}d$, as given in Table 3.4:

`ICC$I3Q000$P4C393$P2D933`.

The first three characters tell us that the Bravais lattice of this space group is of type `I`, that the space group is centrosymmetric, and that it belongs to the cubic system. We then see that the generators are (i) an improper threefold axis along [111] (`I3Q`) with a zero translation part, (ii) a proper fourfold axis along [001] (`P4C`) with a translation part (1/4, 3/4, 1/4), and (iii) a proper twofold axis along [110] (`P2D`) with a translation part (3/4, 1/4, 1/4).

If we make use of the above-outlined interpretation of the explicit symbol in eqn (3.28), the space-group symmetry transformations in direct space corresponding to the above three generators of the space group $Ia\overline{3}d$ become

$$\left[\begin{pmatrix} 0 & 0 & \overline{1} \\ \overline{1} & 0 & 0 \\ 0 & \overline{1} & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right] = \begin{pmatrix} \overline{z} \\ \overline{x} \\ \overline{y} \end{pmatrix},$$

Table 3.4 Explicit space-group symbols, according to Shmueli(2001c)

No.	H.-M. short	H.-M. var.	Explicit
1	$P1$		PAN$P1A000
2	$P\overline{1}$		PAC$I1A000
3	$P2$	$P121$	PMN$P2B000
3	$P2$	$P112$	PMN$P2C000
4	$P2_1$	$P12_11$	PMN$P2B060
4	$P2_1$	$P112_1$	PMN$P2C006
5	$C2$	$C121$	CMN$P2B000
5	$C2$	$A121$	AMN$P2B000
5	$C2$	$I121$	IMN$P2B000
5	$C2$	$A112$	AMN$P2C000
5	$C2$	$B112$	BMN$P2C000
5	$C2$	$I112$	IMN$P2C000
6	Pm	$P1m1$	PMN$I2B000
6	Pm	$P11m$	PMN$I2C000
7	Pc	$P1c1$	PMN$I2B006
7	Pc	$P1n1$	PMN$I2B606
7	Pc	$P1a1$	PMN$I2B600
7	Pc	$P11a$	PMN$I2C600
7	Pc	$P11n$	PMN$I2C660
7	Pc	$P11b$	PMN$I2C060
8	Cm	$C1m1$	CMN$I2B000
8	Cm	$A1m1$	AMN$I2B000
8	Cm	$I1m1$	IMN$I2B000
8	Cm	$A11m$	AMN$I2C000
8	Cm	$B11m$	BMN$I2C000
8	Cm	$I11m$	IMN$I2C000
9	Cc	$C1c1$	CMN$I2B006
9	Cc	$A1n1$	AMN$I2B606
9	Cc	$I1a1$	IMN$I2B600
9	Cc	$A11a$	AMN$I2C600
9	Cc	$B11n$	BMN$I2C660
9	Cc	$I11b$	IMN$I2C060
10	$P2/m$	$P12/m1$	PMC$I1A000$P2B000
10	$P2/m$	$P112/m$	PMC$I1A000$P2C000
11	$P2_1/m$	$P12_1/m1$	PMC$I1A000$P2B060
11	$P2_1/m$	$P112_1/m$	PMC$I1A000$P2C006
12	$C2/m$	$C12/m1$	CMC$I1A000$P2B000
12	$C2/m$	$A12/m1$	AMC$I1A000$P2B000
12	$C2/m$	$I12/m1$	IMC$I1A000$P2B000
12	$C2/m$	$A112/m$	AMC$I1A000$P2C000
12	$C2/m$	$B112/m$	BMC$I1A000$P2C000
12	$C2/m$	$I112/m$	IMC$I1A000$P2C000
13	$P2/c$	$P12/c1$	PMC$I1A000$P2B006
13	$P2/c$	$P12/n1$	PMC$I1A000$P2B606
13	$P2/c$	$P12/a1$	PMC$I1A000$P2B600
13	$P2/c$	$P112/a$	PMC$I1A000$P2C600
13	$P2/c$	$P112/n$	PMC$I1A000$P2C660
13	$P2/c$	$P112/b$	PMC$I1A000$P2C060
14	$P2_1/c$	$P12_1/c1$	PMC$I1A000$P2B066
14	$P2_1/c$	$P12_1/n1$	PMC$I1A000$P2B666
14	$P2_1/c$	$P12_1/a1$	PMC$I1A000$P2B660
14	$P2_1/c$	$P112_1/a$	PMC$I1A000$P2C606
14	$P2_1/c$	$P112_1/n$	PMC$I1A000$P2C666
14	$P2_1/c$	$P112_1/b$	PMC$I1A000$P2C066
15	$C2/c$	$C12/c1$	CMC$I1A000$P2B006
15	$C2/c$	$A12/n1$	AMC$I1A000$P2B606
15	$C2/c$	$I12/a1$	IMC$I1A000$P2B600
15	$C2/c$	$A112/a$	AMC$I1A000$P2C600
15	$C2/c$	$B112/n$	BMC$I1A000$P2C660
15	$C2/c$	$I112/b$	IMC$I1A000$P2C060
16	$P222$		PON$P2C000$P2A000
17	$P222_1$		PON$P2C006$P2A000
18	$P2_12_12$		PON$P2C000$P2A660
19	$P2_12_12_1$		PON$P2C606$P2A660
20	$C222_1$		CON$P2C006$P2A000
21	$C222$		CON$P2C000$P2A000
22	$F222$		FON$P2C000$P2A000
23	$I222$		ION$P2C000$P2A000
24	$I2_12_12_1$		ION$P2C606$P2A660
25	$Pmm2$		PON$P2C000$I2A000
26	$Pmc2_1$		PON$P2C006$I2A000
27	$Pcc2$		PON$P2C000$I2A006
28	$Pma2$		PON$P2C000$I2A600
29	$Pca2_1$		PON$P2C006$I2A606
30	$Pnc2$		PON$P2C000$I2A066
31	$Pmn2_1$		PON$P2C606$I2A000
32	$Pba2$		PON$P2C000$I2A660
33	$Pna2_1$		PON$P2C006$I2A666
34	$Pnn2$		PON$P2C000$I2A666
35	$Cmm2$		CON$P2C000$I2A000
36	$Cmc2_1$		CON$P2C006$I2A000
37	$Ccc2$		CON$P2C000$I2A006
38	$Amm2$		AON$P2C000$I2A000
39	$Abm2$		AON$P2C000$I2A060
40	$Ama2$		AON$P2C000$I2A600
41	$Aba2$		AON$P2C000$I2A660
42	$Fmm2$		FON$P2C000$I2A000
43	$Fdd2$		FON$P2C000$I2A333
44	$Imm2$		ION$P2C000$I2A000
45	$Iba2$		ION$P2C000$I2A660
46	$Ima2$		ION$P2C000$I2A600
47	$Pmmm$		POC$I1A000$P2C000$P2A000
48	$Pnnn$	O_1	POC$I1A666$P2C000$P2A000
48	$Pnnn$	O_2	POC$I1A000$P2C660$P2A066
49	$Pccm$		POC$I1A000$P2C000$P2A006
50	$Pban$	O_1	POC$I1A660$P2C000$P2A000
50	$Pban$	O_2	POC$I1A000$P2C660$P2A060
51	$Pmma$		POC$I1A000$P2C600$P2A600
52	$Pnna$		POC$I1A000$P2C600$P2A066
53	$Pmna$		POC$I1A000$P2C606$P2A000
54	$Pcca$		POC$I1A000$P2C600$P2A606
55	$Pbam$		POC$I1A000$P2C000$P2A660

Table 3.4 Explicit space-group symbols, according to Shmueli (2001c) (cont)

.

No.	H.-M. short	H.-M. var.	Explicit	No.	H.-M. short	H.-M. var.	Explicit
56	$Pccn$		POC$I1A000$P2C660$P2A606	101	$P4_2cm$		PTN$P4C006$I2A006
57	$Pbcm$		POC$I1A000$P2C006$P2A060	102	$P4_2nm$		PTN$P4C666$I2A666
58	$Pnnm$		POC$I1A000$P2C000$P2A666	103	$P4cc$		PTN$P4C000$I2A006
59	$Pmmn$	O_1	POC$I1A660$P2C000$P2A660	104	$P4nc$		PTN$P4C000$I2A666
59	$Pmmn$	O_2	POC$I1A000$P2C660$P2A600	105	$P4_2mc$		PTN$P4C006$I2A000
60	$Pbcn$		POC$I1A000$P2C666$P2A660	106	$P4_2bc$		PTN$P4C006$I2A660
61	$Pbca$		POC$I1A000$P2C606$P2A660	107	$I4mm$		ITN$P4C000$I2A000
62	$Pnma$		POC$I1A000$P2C606$P2A666	108	$I4cm$		ITN$P4C000$I2A006
63	$Cmcm$		COC$I1A000$P2C006$P2A000	109	$I4_1md$		ITN$P4C063$I2A666
64	$Cmca$		COC$I1A000$P2C066$P2A000	110	$I4_1cd$		ITN$P4C063$I2A660
65	$Cmmm$		COC$I1A000$P2C000$P2A000	111	$P\bar{4}2m$		PTN$I4C000$P2A000
66	$Cccm$		COC$I1A000$P2C000$P2A006	112	$P\bar{4}2c$		PTN$I4C000$P2A006
67	$Cmma$		COC$I1A000$P2C060$P2A000	113	$P\bar{4}2_1m$		PTN$I4C000$P2A660
68	$Ccca$	O_1	COC$I1A066$P2C660$P2A660	114	$P\bar{4}2_1c$		PTN$I4C000$P2A666
68	$Ccca$	O_2	COC$I1A000$P2C600$P2A606	115	$P\bar{4}m2$		PTN$I4C000$P2D000
69	$Fmmm$		FOC$I1A000$P2C000$P2A000	116	$P\bar{4}c2$		PTN$I4C000$P2D006
70	$Fddd$	O_1	FOC$I1A333$P2C000$P2A000	117	$P\bar{4}b2$		PTN$I4C000$P2D660
70	$Fddd$	O_2	FOC$I1A000$P2C990$P2A099	118	$P\bar{4}n2$		PTN$I4C000$P2D666
71	$Immm$		IOC$I1A000$P2C000$P2A000	119	$I\bar{4}m2$		ITN$I4C000$P2D000
72	$Ibam$		IOC$I1A000$P2C000$P2A660	120	$I\bar{4}c2$		ITN$I4C000$P2D006
73	$Ibca$		IOC$I1A000$P2C606$P2A660	121	$I\bar{4}2m$		ITN$I4C000$P2A000
74	$Imma$		IOC$I1A000$P2C060$P2A000	122	$I\bar{4}2d$		ITN$I4C000$P2A609
75	$P4$		PTN$P4C000	123	$P4/mmm$		PTC$I1A000$P4C000$P2A000
76	$P4_1$		PTN$P4C003	124	$P4/mcc$		PTC$I1A000$P4C000$P2A006
77	$P4_2$		PTN$P4C006	125	$P4/nbm$	O_1	PTC$I1A660$P4C000$P2A000
78	$P4_3$		PTN$P4C009	125	$P4/nbm$	O_2	PTC$I1A000$P4C600$P2A060
79	$I4$		ITN$P4C000	126	$P4/nnc$	O_1	PTC$I1A666$P4C000$P2A000
80	$I4_1$		ITN$P4C063	126	$P4/nnc$	O_2	PTC$I1A000$P4C600$P2A066
81	$P\bar{4}$		PTN$I4C000	127	$P4/mbm$		PTC$I1A000$P4C000$P2A660
82	$I\bar{4}$		ITN$I4C000	128	$P4/mnc$		PTC$I1A000$P4C000$P2A666
83	$P4/m$		PTC$I1A000$P4C000	129	$P4/nmm$	O_1	PTC$I1A660$P4C660$P2A660
84	$P4_2/m$		PTC$I1A000$P4C006	129	$P4/nmm$	O_2	PTC$I1A000$P4C600$P2A600
85	$P4/n$	O_1	PTC$I1A660$P4C660	130	$P4/ncc$	O_1	PTC$I1A660$P4C660$P2A666
85	$P4/n$	O_2	PTC$I1A000$P4C600	130	$P4/ncc$	O_2	PTC$I1A000$P4C600$P2A606
86	$P4_2/n$	O_1	PTC$I1A666$P4C666	131	$P4_2/mmc$		PTC$I1A000$P4C006$P2A000
86	$P4_2/n$	O_2	PTC$I1A000$P4C066	132	$P4_2/mcm$		PTC$I1A000$P4C006$P2A006
87	$I4/m$		ITC$I1A000$P4C000	133	$P4_2/nbc$	O_1	PTC$I1A666$P4C666$P2A006
88	$I4_1/a$	O_1	ITC$I1A063$P4C063	133	$P4_2/nbc$	O_2	PTC$I1A000$P4C606$P2A060
88	$I4_1/a$	O_2	ITC$I1A000$P4C933	134	$P4_2/nnm$	O_1	PTC$I1A666$P4C666$P2A000
89	$P422$		PTN$P4C000$P2A000	134	$P4_2/nnm$	O_2	PTC$I1A000$P4C606$P2A066
90	$P42_12$		PTN$P4C660$P2A660	135	$P4_2/mbc$		PTC$I1A000$P4C006$P2A660
91	$P4_122$		PTN$P4C003$P2A006	136	$P4_2/mnm$		PTC$I1A000$P4C666$P2A666
92	$P4_12_12$		PTN$P4C663$P2A669	137	$P4_2/nmc$	O_1	PTC$I1A666$P4C666$P2A666
93	$P4_222$		PTN$P4C006$P2A000	137	$P4_2/nmc$	O_2	PTC$I1A000$P4C606$P2A600
94	$P4_22_12$		PTN$P4C666$P2A666	138	$P4_2/ncm$	O_1	PTC$I1A666$P4C666$P2A660
95	$P4_322$		PTN$P4C009$P2A006	138	$P4_2/ncm$	O_2	PTC$I1A000$P4C606$P2A606
96	$P4_32_12$		PTN$P4C669$P2A663	139	$I4/mmm$		ITC$I1A000$P4C000$P2A000
97	$I422$		ITN$P4C000$P2A000	140	$I4/mcm$		ITC$I1A000$P4C000$P2A006
98	$I4_122$		ITN$P4C063$P2A063	141	$I4_1/amd$	O_1	ITC$I1A063$P4C063$P2A063
99	$P4mm$		PTN$P4C000$I2A000	141	$I4_1/amd$	O_2	ITC$I1A000$P4C393$P2A000
100	$P4bm$		PTN$P4C000$I2A660	142	$I4_1/acd$	O_1	ITC$I1A063$P4C063$P2A069

Table 3.4 Explicit space-group symbols, according to Shmueli (2001c) (cont)

No.	H.-M. short	H.-M. var.	Explicit
142	$I4_1/acd$	O_2	`ITC$I1A000$P4C393$P2A006`
143	$P3$		`PRN$P3C000`
144	$P3_1$		`PRN$P3C004`
145	$P3_2$		`PRN$P3C008`
146	$R3$	H	`RRN$P3C000`
146	$R3$	R	`PRN$P3Q000`
147	$P\overline{3}$		`PRC$I3C000`
148	$R\overline{3}$	H	`RRC$I3C000`
148	$R\overline{3}$	R	`PRC$I3Q000`
149	$P312$		`PRN$P3C000$P2G000`
150	$P321$		`PRN$P3C000$P2F000`
151	$P3_112$		`PRN$P3C004$P2G000`
152	$P3_121$		`PRN$P3C004$P2F008`
153	$P3_212$		`PRN$P3C008$P2G000`
154	$P3_221$		`PRN$P3C008$P2F004`
155	$R32$	H	`RRN$P3C000$P2F000`
155	$R32$	R	`PRN$P3Q000$P2E000`
156	$P3m1$		`PRN$P3C000$I2F000`
157	$P31m$		`PRN$P3C000$I2G000`
158	$P3c1$		`PRN$P3C000$I2F006`
159	$P31c$		`PRN$P3C000$I2G006`
160	$R3m$	H	`RRN$P3C000$I2F000`
160	$R3m$	R	`PRN$P3Q000$I2E000`
161	$R3c$	H	`RRN$P3C000$I2F006`
161	$R3c$	R	`PRN$P3Q000$I2E666`
162	$P\overline{3}1m$		`PRC$I3C000$P2G000`
163	$P\overline{3}1c$		`PRC$I3C000$P2G006`
164	$P\overline{3}m1$		`PRC$I3C000$P2F000`
165	$P\overline{3}c1$		`PRC$I3C000$P2F006`
166	$R\overline{3}m$	H	`RRC$I3C000$P2F000`
166	$R\overline{3}m$	R	`PRC$I3Q000$P2E000`
167	$R\overline{3}c$	H	`RRC$I3C000$P2F006`
167	$R\overline{3}c$	R	`PRC$I3Q000$P2E666`
168	$P6$		`PHN$P6C000`
169	$P6_1$		`PHN$P6C002`
170	$P6_5$		`PHN$P6C005`
171	$P6_2$		`PHN$P6C004`
172	$P6_4$		`PHN$P6C008`
173	$P6_3$		`PHN$P6C006`
174	$P\overline{6}$		`PHN$I6C000`
175	$P6/m$		`PHC$I1A000$P6C000`
176	$P6_3/m$		`PHC$I1A000$P6C006`
177	$P622$		`PHN$P6C000$P2F000`
178	$P6_122$		`PHN$P6C002$P2F000`
179	$P6_522$		`PHN$P6C005$P2F000`
180	$P6_222$		`PHN$P6C004$P2F000`
181	$P6_422$		`PHN$P6C008$P2F000`
182	$P6_322$		`PHN$P6C006$P2F000`
183	$P6mm$		`PHN$P6C000$I2F000`
184	$P6cc$		`PHN$P6C000$I2F006`
185	$P6_3cm$		`PHN$P6C006$I2F006`
186	$P6_3mc$		`PHN$P6C006$I2F000`
187	$P\overline{6}m2$		`PHN$I6C000$P2G000`
188	$P\overline{6}c2$		`PHN$I6C006$P2G000`
189	$P\overline{6}2m$		`PHN$I6C000$P2F000`
190	$P\overline{6}2c$		`PHN$I6C006$P2F000`
191	$P6/mmm$		`PHC$I1A000$P6C000$P2F000`
192	$P6/mcc$		`PHC$I1A000$P6C000$P2F006`
193	$P6_3/mcm$		`PHC$I1A000$P6C006$P2F006`
194	$P6_3/mmc$		`PHC$I1A000$P6C006$P2F000`
195	$P23$		`PCN$P3Q000$P2C000$P2A000`
196	$F23$		`FCN$P3Q000$P2C000$P2A000`
197	$I23$		`ICN$P3Q000$P2C000$P2A000`
198	$P2_13$		`PCN$P3Q000$P2C606$P2A660`
199	$I2_13$		`ICN$P3Q000$P2C606$P2A660`
200	$Pm\overline{3}$		`PCC$I3Q000$P2C000$P2A000`
201	$Pn\overline{3}$	O_1	`PCC$I3Q666$P2C000$P2A000`
201	$Pn\overline{3}$	O_2	`PCC$I3Q000$P2C660$P2A066`
202	$Fm\overline{3}$		`FCC$I3Q000$P2C000$P2A000`
203	$Fd\overline{3}$	O_1	`FCC$I3Q333$P2C000$P2A000`
203	$Fd\overline{3}$	O_2	`FCC$I3Q000$P2C330$P2A033`
204	$Im\overline{3}$		`ICC$I3Q000$P2C000$P2A000`
205	$Pa\overline{3}$		`PCC$I3Q000$P2C606$P2A660`
206	$Ia\overline{3}$		`ICC$I3Q000$P2C606$P2A660`
207	$P432$		`PCN$P3Q000$P4C000$P2D000`
208	$P4_232$		`PCN$P3Q000$P4C666$P2D666`
209	$F432$		`FCN$P3Q000$P4C000$P2D000`
210	$F4_132$		`FCN$P3Q000$P4C993$P2D939`
211	$I432$		`ICN$P3Q000$P4C000$P2D000`
212	$P4_332$		`PCN$P3Q000$P4C939$P2D399`
213	$P4_132$		`PCN$P3Q000$P4C393$P2D933`
214	$I4_132$		`ICN$P3Q000$P4C393$P2D933`
215	$P\overline{4}3m$		`PCN$P3Q000$I4C000$I2D000`
216	$F\overline{4}3m$		`FCN$P3Q000$I4C000$I2D000`
217	$I\overline{4}3m$		`ICN$P3Q000$I4C000$I2D000`
218	$P\overline{4}3n$		`PCN$P3Q000$I4C666$I2D666`
219	$F\overline{4}3c$		`FCN$P3Q000$I4C666$I2D666`
220	$I\overline{4}3d$		`ICN$P3Q000$I4C939$I2D399`
221	$Pm\overline{3}m$		`PCC$I3Q000$P4C000$P2D000`
222	$Pn\overline{3}n$	O_1	`PCC$I3Q666$P4C000$P2D000`
222	$Pn\overline{3}n$	O_2	`PCC$I3Q000$P4C600$P2D006`
223	$Pm\overline{3}n$		`PCC$I3Q000$P4C666$P2D666`
224	$Pn\overline{3}m$	O_1	`PCC$I3Q666$P4C666$P2D666`
224	$Pn\overline{3}m$	O_2	`PCC$I3Q000$P4C066$P2D660`
225	$Fm\overline{3}m$		`FCC$I3Q000$P4C000$P2D000`
226	$Fm\overline{3}c$		`FCC$I3Q000$P4C666$P2D666`
227	$Fd\overline{3}m$	O_1	`FCC$I3Q333$P4C993$P2D939`
227	$Fd\overline{3}m$	O_2	`FCC$I3Q000$P4C693$P2D936`
228	$Fd\overline{3}c$	O_1	`FCC$I3Q999$P4C993$P2D939`
228	$Fd\overline{3}c$	O_2	`FCC$I3Q000$P4C093$P2D930`
229	$Im\overline{3}m$		`ICC$I3Q000$P4C000$P2D000`
230	$Ia\overline{3}d$		`ICC$I3Q000$P4C393$P2D933`

$$\left[\begin{pmatrix} 0 & \overline{1} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \frac{1}{4} \\ \frac{3}{4} \\ \frac{1}{4} \end{pmatrix}\right] = \begin{pmatrix} \frac{1}{4} - y \\ \frac{3}{4} + x \\ \frac{1}{4} + z \end{pmatrix},$$

$$\left[\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \overline{1} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}\right] = \begin{pmatrix} \frac{3}{4} + y \\ \frac{1}{4} + x \\ \frac{1}{4} - z \end{pmatrix}.$$

Similarly, space-group symmetry transformations in reciprocal space can be constructed, which in fact give us the indices of symmetry-related reflections and phase shifts due to fractional translations (see Chapter 5 and Shmueli (2001*c*) for a more detailed analysis).

The explicit space-group symbols have been implemented in software which can be downloaded from the URL http://crystal.tau.ac.il/xtal/spgmic and run under the Windows system. Space-group operators can be obtained there in conventional, Cartesian, and general representations, and the conventional coordinates of general equivalent positions are a by-product of the computation. In fact, the explicit symbols given in Table 3.4 may serve as an input to the retrieval of fundamental information on the three-dimensional space-group settings treated in Volume A (ITA83), for its applications to direct and reciprocal space.

Finally, it must be mentioned that the fundamental space-group information, described above, as well as many group-theoretical computational treatments of space groups and their subgroups, are available online from the Bilbao Crystallographic Server (Aroyo *et al.*, 2006). The related theoretical background, which is outside the scope of this book, can be found in Volumes A (Hahn 1983,1996), A1 (Wondratschek and Müller 2004) and E (Kopsky and Litvin 2002) of the *International Tables for Crystallography* and in the literature quoted in these volumes.

3.8.1 Description of Table 3.4 - explicit space-group symbols

The first column of Table 3.4 contains the conventional space-group number. The second column shows the conventional short Hermann–Mauguin (H.-M.) or international, space-group symbol, and the third column, labeled "var.", shows the full international space-group symbol *only* for the various settings of the monoclinic space groups that are given in the main space-group tables of Volume A (ITA83). The other items appearing in the third column pertain to the choice of the space-group origin–where there are alternatives–and to axial systems. The Volume A designations "Origin 1" and "Origin 2" are abbreviated here to O_1 and O_2, respectively. Hexagonal axial system for trigonal space groups is indicated by H, and rhombohedral axial system by R. The fourth column shows the

explicit space-group symbols described above, for each of the settings considered in Volume A (ITA83). The descriptions of the first to the fourth column apply also to the fifth through the eighth columns,

Table 3.4 contains 306 entries/symbols which correspond to the 306 settings of the 230 three-dimensional space groups given in Volume A of the *international Tables for Crystallography* (Hahn 1983, 1996). The output of the software described above contains the symmetry operators/transformations in the same order as that given in Volume A.

3.9 Exercises for Chapter 3

1. Starting from eqn (3.7), show in detail that the equation

$$(\mathbf{P}_j, \mathbf{r}_{L_n})(\mathbf{P}'_j, \mathbf{r}'_{L_n}) = (\mathbf{I}, \mathbf{0})$$

leads to

$$(\mathbf{P}'_j, \mathbf{r}'_{L_n}) = (\mathbf{P}_j^{-1}, -\mathbf{P}_j^{-1}\mathbf{r}_{L_n}).$$

This completes the proof of the existence of an inverse space-group operator.

2. Convince yourself that for space-group operators with the composition law given by eqn (3.6), associativity holds true.

3. Analyze the orthorhombic space group *Abm*2: choose the origin to be on the twofold axis parallel to [001], and let the *b* glide plane, normal to [100], pass through the origin. Show that one of the mirror planes coincides with the plane $(x, 1/4, z)$. The *A*-centering is equivalent to the operator ([1], (**b**+**c**)/2). Find all the space-group operators derived from the centering, and show that the set of operators contains a double glide plane (cf. Table 3.2).

4. The space group *Pbcn* was derived in the text on the assumption that the origin was chosen at the point of intersection of three mutually perpendicular glide planes. Complete the transformation of all the representative operators to an origin chosen at a center of symmetry. The transformation of the operator of inversion is given in the text.

5. The space group *Pbcn* is referred to the basis (**a**, **b**, **c**). How does the space-group symbol change if the space group is referred to the basis (**c**, **a**, **b**)?

6. Look up in Volume A of the *International Tables for Crystallography* the coordinates of the general equivalent positions for the space group *Pnma* (No. 62) and retrieve from these coordinates the corresponding space-group operators in their (3×3 matrix, 3×1 matrix) representation.

This page intentionally left blank

4 X-ray diffraction techniques

4.1 Introduction

An integral part of the process of crystal structure determination is an experiment and the techniques used to collect the experimental data. The experiment consists of scattering radiation from crystalline matter, the radiations that are usually employed being X-rays, electrons and neutrons. Although all three radiations can be employed, the vast majority of structure determinations are based on X-ray diffraction data. This chapter will therefore be dedicated to X-ray techniques. It is, however, important to point out that all the other chapters are applicable to all three diffraction techniques. We begin this chapter by introducing the Laue and Bragg conditions, and their geometrical representation in terms of the Ewald sphere of reflection. These conditions are common to all the experimental techniques that will be discussed. We shall next discuss in some detail the principles of operation of traditional and modern sources of X-rays: the X-ray tube and its spectroscopic fundamentals, and the vastly more powerful synchrotron. Our discussion of sources of X-radiation will be followed by an outline of the operation of some area detectors for this radiation: (i) photographic film, which may be thought to have become of historical importance only but which has lasting underlying principles; (ii) the imaging plate, which is geometrically analogous to X-ray film but has different physical principles and greatly enhanced performance; and (iii) the charge-coupled device, which produces very efficiently a digital image of the diffraction pattern. We shall next discuss the instrumentation which is employed with the above sources and detectors for the generation of diffraction patterns. The rotation method will be treated in some detail, the principles of the Weissenberg, de Jong–Bouman and precession methods will be briefly indicated, and a detailed discussion of the four-circle diffractometer will be presented. The four-circle diffractometer is a very precise device, which enables one to measure very accurately one reflection at a time and uses a photon counter as a detector. The chapter is concluded with a discussion of the Laue method. This was the first method ever employed in the production of diffraction patterns, it had limited application, and was for a long time considered to be mainly of historical interest. However, with the advent of synchrotrons and highly linear and fast detectors, the Laue method has seen a major revival and is nowadays being applied to collect extensive amounts of data. We shall see how, in principle, a Laue pattern can be calculated–which provides a way to its interpretation.

4.2 Diffraction conditions

The purpose of this section is to describe the geometric conditions under which constructive interference of radiation scattered from a triply periodic arrangement of material units takes place. These conditions, known as diffraction conditions, are the basis of any experiment in which intensities of diffracted radiation are measured. These measurements have many purposes but two are outstandingly important: (i) to determine the periodicity, symmetry, and orientation of a crystal, and (ii) to obtain accurate estimates of the intensities of diffracted radiation, in order to elucidate from them the atomic arrangement within the asymmetric unit or, in other words, to determine its structure. Most usually, stage (ii) must be preceded by the completion of stage (i), but there are some important applications for which stage (i) is sufficient.

A more complete model of the structure of a material unit, and its effect on the diffraction pattern, will be considered in the next chapter. However, for the present purpose the simplest model, of classical electrons, each located at a lattice point and neutralized by a proton located nearby, will be considered. It will be seen in the next chapter how the mass of the proton makes its contribution negligible. This model is clearly a triply periodic arrangement of scatterers and suffices for the determination of the directions of the diffracted beams. No quantitative considerations of the intensities will be given.

4.2.1 The Laue and Bragg equations

Let us assume that the above "crystal" of point charges is irradiated with monochromatic X-radiation. Since X-rays are electromagnetic radiation, they can be described, at a large distance from the source, in terms of plane waves, with appropriate wavevectors. The electric field of the incident X-ray wave varies with time and therefore accelerates the point charges it encounters. Electromagnetic theory tells us that an accelerated charge emits energy in the form of electromagnetic radiation, with the same frequency as that of the incident wave. This may require some small correction for quantum effects, such as the Compton effect, but these will be neglected in the present treatment. We can therefore say that the incident X-rays are reemitted, with an unchanged wavelength, the efficiency of this reemission being determined by the scattering cross section of the electron for electromagnetic radiation. This topic will be dealt with in the next chapter.

We consider an X-ray wave, with wavevector $\mathbf{s}_0$, falling on a crystal, and a reemitted (or scattered) X-ray wave with wavevector $\mathbf{s}$. In view of the assumption of unchanged wavelength, the magnitudes of the wavevectors $\mathbf{s}_0$ and $\mathbf{s}$ will be identical, and we shall take them as

$$|\mathbf{s}| = |\mathbf{s}_0| = \frac{1}{\lambda}.$$

We assume further that the lattice of our crystal can be described in terms of the basis ($\mathbf{a}$ $\mathbf{b}$ $\mathbf{c}$). The question to be answered is: for what geometrical relationship

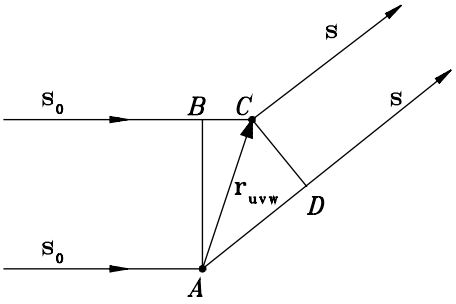


Fig. 4.1 Derivation of the Laue diffraction conditions.

between the wavevectors of the radiation and the basis vectors of the lattice will maximum constructive interference of the scattered X-ray waves occur? Because of the assumed strict periodicity of the arrangement, it is sufficient to consider two point charges related by translation through a lattice vector $\boldsymbol{r}_{uvw}$, as illustrated in Fig. 4.1.

The waves scattered from the charges at A and C will undergo maximum constructive interference if the difference between the optical paths passing through these charges is an integer multiple of the wavelength λ. It can be seen from Fig. 4.1 that the optical-paths difference is given by

$$\begin{aligned}\Delta &= AD - BC \\ &= \boldsymbol{r}_{uvw} \cdot \frac{\mathbf{s}}{|\mathbf{s}|} - \boldsymbol{r}_{uvw} \cdot \frac{\mathbf{s}_0}{|\mathbf{s}_0|} \\ &= \lambda \boldsymbol{r}_{uvw} \cdot (\mathbf{s} - \mathbf{s}_0).\end{aligned}$$

It follows that the required condition is

$$\boldsymbol{r}_{uvw} \cdot \mathbf{h} = \text{integer}, \tag{4.1}$$

where $\mathbf{h} = \mathbf{s} - \mathbf{s}_0$ is called the *diffraction vector*. This can also be rewritten as

$$(u\mathbf{a} + v\mathbf{b} + w\mathbf{c}) \cdot \mathbf{h} = \text{integer}.$$

Since the coefficients u, v, w can be any integers, eqn (4.1) is equivalent to the three equations

$$\mathbf{a} \cdot \mathbf{h} = h \tag{4.2}$$

$$\mathbf{b} \cdot \mathbf{h} = k \tag{4.3}$$

$$\mathbf{c} \cdot \mathbf{h} = l, \tag{4.4}$$

which have to be simultaneously satisfied for maximum interference to occur; here h, k, l are any integers. Equations (4.2)–(4.4) are known as the Laue equations (Laue 1914).

In order to find a possible expression for the vector $\mathbf{h}$, let us divide both sides of each of the Laue equation by its right-hand side, and then subtract the second equation from the first, and the third from the second. We obtain

$$\left(\frac{\mathbf{a}}{h} - \frac{\mathbf{b}}{k}\right) \cdot \mathbf{h} = 0, \tag{4.5}$$

$$\left(\frac{\mathbf{b}}{k} - \frac{\mathbf{c}}{l}\right) \cdot \mathbf{h} = 0. \tag{4.6}$$

It can be seen that the vector $\mathbf{h}$ is perpendicular to the vectors $\mathbf{a}/h - \mathbf{b}/k$ and $\mathbf{b}/k - \mathbf{c}/l$, and therefore also to the plane determined by these two vectors. It is also evident that the vector $\mathbf{h}$ has the same direction as the vector product of these two vectors, that is, it is proportional to this product. We then have

$$\begin{aligned}\mathbf{h} &= K\left[(\mathbf{a}/h - \mathbf{b}/k) \times (\mathbf{b}/k - \mathbf{c}/l)\right] \\ &= K\left[\frac{\mathbf{b}\times\mathbf{c}}{kl} + \frac{\mathbf{c}\times\mathbf{a}}{lh} + \frac{\mathbf{a}\times\mathbf{b}}{hk}\right] \\ &\equiv p(\mathbf{b}\times\mathbf{c}) + q(\mathbf{c}\times\mathbf{a}) + r(\mathbf{a}\times\mathbf{b}). \end{aligned} \tag{4.7}$$

If we form the scalar products of both sides of eqn (4.7) with the vector $\mathbf{a}$, we obtain

$$\mathbf{a}\cdot\mathbf{h} = p\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = pV,$$

where V is the volume of the unit cell. If we use eqns (4.2)–(4.4), we obtain $p = h/V$ and, similarly, $q = k/V$ and $r = l/V$. It follows that

$$\mathbf{h} = h\frac{\mathbf{b}\times\mathbf{c}}{V} + k\frac{\mathbf{c}\times\mathbf{a}}{V} + l\frac{\mathbf{a}\times\mathbf{b}}{V},$$

that is, $\mathbf{h}$ is a vector of the reciprocal lattice, as obtained from our discussion of lattice planes in Sections 1.2 and 1.3. This is a mathematical solution of the Laue equations.

Returning to the plane containing the vectors that appear in eqns (4.5) and (4.6), Fig. 4.2 shows that this is the plane whose intercepts on the coordinate axes are a/h, b/k, and c/l.

If we recall Section 1.2, we see that if h, k and l are relatively prime, this plane is just a lattice plane (hkl), adjacent to the origin. If, however, h, k, and l have a common factor, say n, the plane is parallel to the $(h/n, k/n, l/n)$ family of lattice planes and belongs to a family of parallel and equidistant planes, in which only a plane the serial number of which is an integer multiple of n is a lattice plane. That is, if we write the indices of such a plane as $nh'nk'nl'$, where $h'k'l'$ are relatively prime, we have for the interplanar distances

$$d_{nh'nk'nl'} = \frac{1}{n} d_{h'k'l'}. \tag{4.8}$$

The interplanar distance (given by the distance of the plane in Fig. 4.2 from the origin) is, analogously to the calculation in Section 1.2,

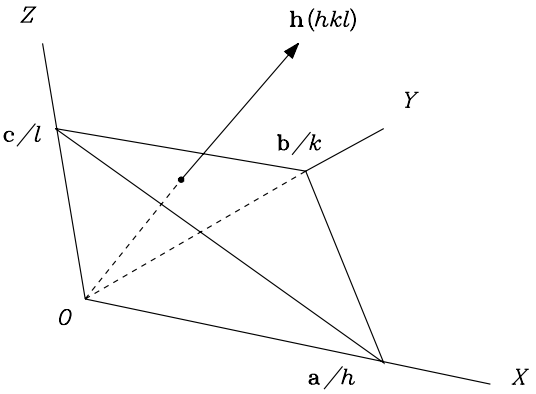


Fig. 4.2 *Geometrical interpretation of Laue's equations*

$$d_{hkl} = \frac{\mathbf{a}}{h} \cdot \frac{\mathbf{h}}{|\mathbf{h}|} = \frac{1}{|\mathbf{h}|}, \tag{4.9}$$

since it follows from eqn (4.2) that $\mathbf{a}\cdot\mathbf{h}/h = 1$. If we now denote the angle between $\mathbf{s}$ and $\mathbf{s}_0$ by 2θ, and construct the isosceles triangle obtained from $\mathbf{h} = \mathbf{s} - \mathbf{s}_0$, we have for $|\mathbf{h}|$

$$|\mathbf{h}| = |\mathbf{s}| \sin\theta + |\mathbf{s}_0| \sin\theta = \frac{2\sin\theta}{\lambda}. \tag{4.10}$$

From eqns (4.9) and (4.10), we have

$$\lambda = 2d_{hkl} \sin\theta, \tag{4.11}$$

where h, k, l can be any integers. If h, k, l are of the form nh', nk', nl', where h', k', l' are relatively prime and if we use eqn (4.8), we have

$$n\lambda = 2d_{h'k'l'} \sin\theta. \tag{4.12}$$

Equation (4.12) is the original version of the Bragg equation, and eqn (4.11) is the version most often encountered in practical applications. This is in fact a simpler interpretation of the Laue equations, in terms of a known wavelength, a single measurable scattering angle, and the interplanar spacing of the (hkl) family of planes. The interrelation of the geometry of lattice planes and the geometry of diffraction phenomena is quite remarkable.

Equations analogous to the Laue equations were already encountered in Exercise 3 in Chapter 1, in connection with the discussion of lattice planes and the reciprocal lattice. It is most significant that the diffraction vector $\mathbf{h}$ satisfying the Laue's equations can be represented as a reciprocal-lattice vector. This result is widely employed in the theory and practice of diffraction from crystals.

4.2.2 Ewald's sphere of reflection

A highly suggestive geometrical description of the conditions for diffraction was put forward by Ewald (1913), and is employed very extensively in the crystallographic literature, albeit in several differing representations.

Let us first find the maximum value of $|\mathbf{h}|$. From eqn (4.11),

$$|\mathbf{h}| = \frac{1}{d_{hkl}} = \frac{2}{\lambda}\sin\theta$$

and since the maximum value of $\sin\theta$ is 1, we must have

$$|\mathbf{h}|_{\max} = \frac{2}{\lambda}.$$

Hence, all the reciprocal-lattice vectors potentially satisfying the Laue or Bragg equations, must be enclosed in a sphere of radius $2/\lambda$, with the origin of the reciprocal lattice chosen at the center of this sphere. This sphere is called the *limiting sphere* (see Fig. 4.3 and also the Exercises).

The Ewald sphere fits into the limiting sphere as indicated in Fig. 4.3. This is a sphere tangent to the limiting sphere at the point A and passing through the point O, the origin of the reciprocal lattice. Its radius is $1/\lambda$, and the incident beam, with wavevector $\mathbf{s}_0$, is directed along the diameter AO. The wavevectors $\mathbf{s}$ of the scattered radiation propagate from the point C, the center of the Ewald sphere, to the surface of that sphere.

It is convenient to imagine the crystal to be associated with the point C; the fact that the spheres exist in reciprocal space and the crystal in the direct space should not give rise to difficulties, since all the reciprocal-lattice vectors involved have directions which are measurable in direct space, and magnitudes expressible in units of $(\text{length})^{-1}$ but calculable from direct-space quantities.

In a diffraction experiment carried out with monochromatic radiation, the Ewald sphere can either be fixed or be constrained to move within the limiting sphere while always passing through the point O and touching the limiting sphere from the inside. In either case, the triangle COP, built from the vectors $\mathbf{s}_0$, $\mathbf{s}$ and $\mathbf{s}-\mathbf{s}_0$ must be isosceles, but the vector $\mathbf{s}-\mathbf{s}_0$ is a diffraction vector *only* if it satisfies simultaneously the Laue equations or, equivalently, if the angle enclosed between $\mathbf{s}$, and $\mathbf{s}_0$ is twice the angle appearing in the Bragg equation. If this is the case, P is a point in the reciprocal lattice. Conversely, if a reciprocal lattice point comes into contact with the Ewald sphere, this point corresponds to a diffracted beam. This is the main idea of Ewald's most useful construction.

A diffracted beam is more often than not, called a reflection. The reason for this can be conveniently illustrated by the Ewald sphere. If the point P in Fig. 4.3 is a reciprocal-lattice point, the vector $\mathbf{h} = \mathbf{s}-\mathbf{s}_0$ is perpendicular to a family of lattice planes (hkl) in the crystal. The plane passing through the segment AP and perpendicular to the plane of the drawing is also parallel to the (hkl) family, since APO is of necessity a right angle (it is subtended by the diameter AO). If we shift this plane parallel to itself to the point C, it is seen that the angle

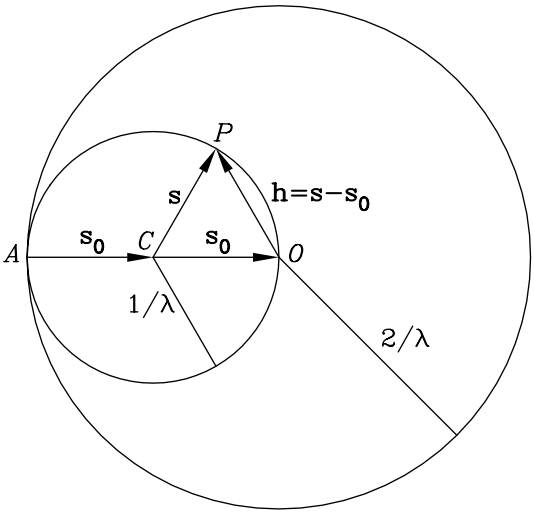


Fig. 4.3 The Ewald sphere.

formed by $\mathbf{s}_0$ and the plane is the same as the angle formed by the plane with $\mathbf{s}$ and that each of these angles equals θ, the Bragg angle. A diffracted beam can therefore be represented pictorially as a reflection from a family of lattice planes and its orders (the integer n on the left-hand side of the Bragg equation, eqn (4.12), is called the order of the reflection). For the above reason, the Ewald sphere is widely known as the *sphere of reflection*. Obviously, the analogy with reflection of electromagnetic radiation from a mirror is only qualitative.

4.3 Production of X-rays

4.3.1 The X-ray tube

The traditional method of producing X-rays in a crystallographic laboratory is by means of an X-ray tube. This device, originally invented by Roentgen in 1895, and improved technically during the following century, is still being used and has an interesting physical background that marks major scientific developments. The principle of its operation is also very instructive.

An X-ray tube (see Fig. 4.4) consists of a tungsten filament, and a water-cooled metallic cup, both components being enclosed in an evacuated housing. The filament is connected to a source of alternating current (of the order of 10 A) and is, accordingly, heated. As a result of this, thermionic emission of electrons from the surface of the filament takes place–initially in all directions. When, however, a high voltage (of the order of 50 kV) is placed between the filament and the metallic cup, where the cup is grounded, the electrons emitted from the

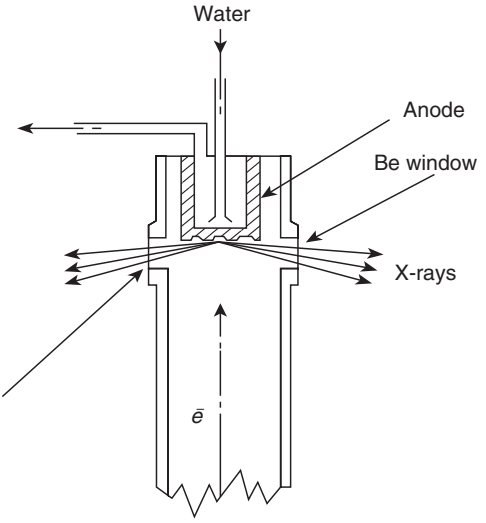


Fig. 4.4 Schematic drawing of an X-ray tube. Reproduced with copyright permission of the International Union of Crystallography (IUCr).

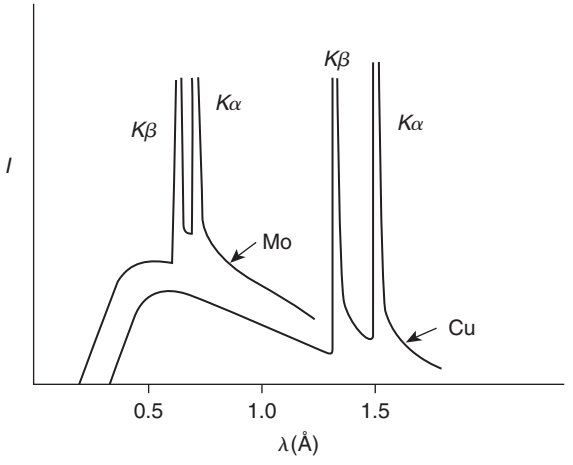


Fig. 4.5 Parts of the X-ray spectra from copper and molybdenum anodes. Reproduced with copyright permission of the International Union of Crystallography (IUCr).

filament are sharply focused in the direction of the cup. When these electrons collide with the cup (the anode), most of their kinetic energy is converted into heat–and hence the necessity for cooling the cup. The remaining part is converted into radiative energy, which was called *X-rays* by Roentgen, the "X" standing for something not understood. This radiation leaves the X-ray tube through thin

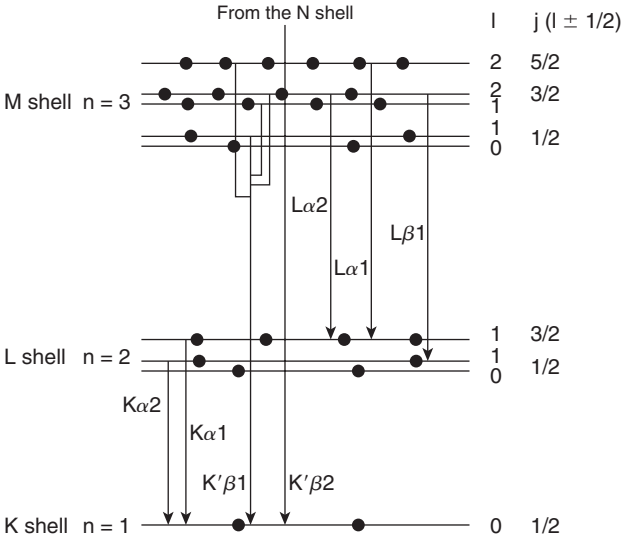


Fig. 4.6 *Electronic transitions and characteristic radiation.* Reproduced from the web site: http://ie.lbl.gov/xray/.

windows, usually made from beryllium.

Parts of the spectra of X-radiation obtained from collisions of electron beams with copper and molybdenum cups are illustrated in Fig. 4.5. In both cases we see a continuous broad "hill" and two sharp peaks superimposed on it, the second sharp peak appearing at wavelength of about 0.7 and 1.5 Å for molybdenum and copper, respectively. More sophisticated arrangements may show additional sharp peaks, which, however, are of little relevance for our purposes.

There obviously exists some shortest wavelength at which radiation is obtained. When an electron of charge e under a potential difference of V volts is brought to a halt at the surface of the anode, it is suddenly decelerated (negatively accelerated), and its energy eV is transfomed into the energy of an X-ray photon $h\nu$, where h is Planck's constant and ν is the photon's frequency. For the highest frequency of the photon, or its lowest wavelength, we have

$$eV = h\nu_{\max} = \frac{hc}{\lambda_{\min}} \quad \text{and} \quad \lambda_{\min} = \frac{hc}{eV} \approx \frac{12\,400}{V}.$$

For example, for $V = 50\,000$ volts we shall have $\lambda_{\min} \approx 0.25$ Å. An explanation of the intensity distribution within the broad hill will not be given here but it is clear that it depends on the accelerating voltage in the tube. This broad hill of radiation intensity, called *white radiation*, was thought for many years to be of little use but, as we shall see later, it is most useful with the Laue method.

Table 4.1 Some frequently applied wavelengths (in Ångstrom units).

Copper:	wavelength (Å)
$\lambda(K\beta)$	1.3922
$\lambda(K\alpha_1)$	1.5406
$\lambda(K\alpha_1)$	1.5444
$\lambda(K\alpha_{av})$	1.5418
Molybdenum:	**wavelength (Å)**
$\lambda(K\beta)$	0.6323
$\lambda(K\alpha_1)$	0.7093
$\lambda(K\alpha_1)$	0.7136
$\lambda(K\alpha_{av})$	0.7107

The sharp peaks are known as the *characteristic radiation* and the wavelengths at which they appear depend on the element from which the anode is constructed. They are, in fact, related to the energetic structure of the atoms of that element, as will be explained below. When an electron retains just enough energy to be able to ionize the lowest-lying shell in an atom of the anode, an electron with principal quantum number $n = 1$ is raised to the continuum and gives rise to transitions of electrons from higher energy levels to the lowest level. Upon such a transition, a photon is emitted with an energy related almost exactly to the levels involved. X-ray spectroscopy associates the quantum numbers $n = 1, 2, 3, 4$ etc. with the letters K,L,M,N etc, as seen in Fig. 4.6. Thus, the origin of the sharp peak labeled Kβ in Figure 4.5 is several transitions from the M shell and also some from the N shell. They appear as one sharp peak in all but very high-resolution measurements. The peak labeled Kα appears as a single peak at low or moderate scattering angles and as a doublet at high scattering angles: Kα_1 associated with a transition from the L shell, with $l = 1$ and spin $+1/2$, to the K shell, and Kα_2 associated with a transition from the L shell, with $l = 1$ and spin $-1/2$, to the K shell. This doublet is employed in crystallographic experiments aimed at very accurate determination of unit cell constants, but the most frequently employed wavelength is that corresponding to a weighted average of $\lambda(\mathrm{K}\alpha_1)$ and $\lambda(\mathrm{K}\alpha_2)$. Table 4.1 lists the wavelengths of special interest in routine crystallographic studies; these concern anodes made from copper and molybdenum. A comprehensive list of interesting wavelengths as well as detailed information on the properties of X-rays, is given in Chapter 4.2 of Volume C of the *International Tables for Crystallography* (Wilson and Prince 1999). The other transitions indicated in Fig. 4.6 give rise to radiation with longer wavelengths, lower intensity and which is much more readily absorbable.

More or less approximate monochromatization

Apart from the Laue method, to be discussed later, all diffraction techniques are based on the assumption that the radiation used is approximately monochro-

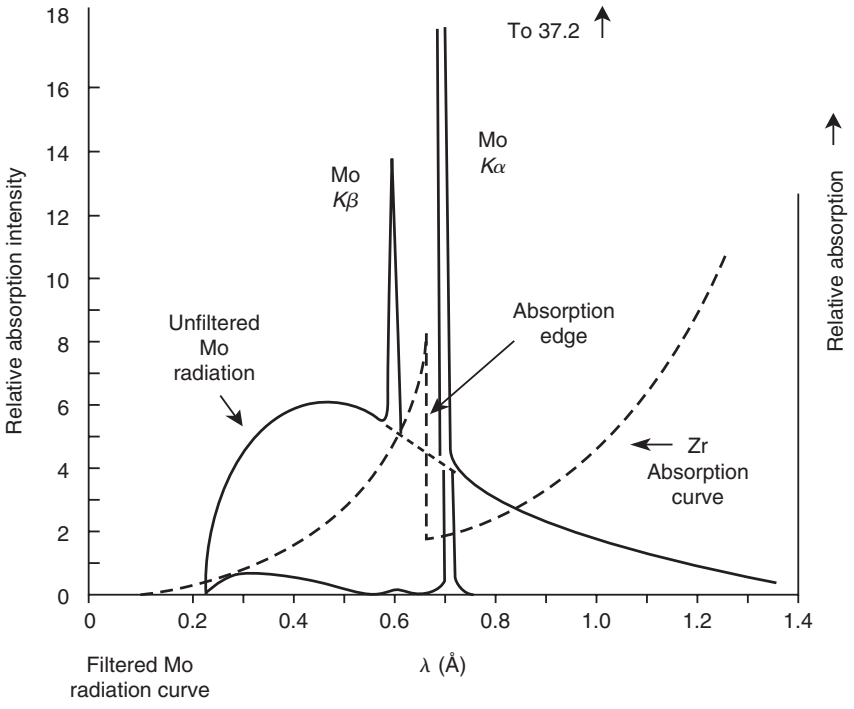


Fig. 4.7 Molybdenum radiation approximately monochromatized with a zirconium filter. Reproduced with copyright permission of the International Union of Crystallography (IUCr).

matic. The most intense characteristic radiation is Kα_1, or Kα_{av} if a limited angular range of the scattering is available. Therefore, one seeks to suppress the radiation at all wavelengths except in a narrow range around the required wavelength. This can be quite usefully, if not completely, done by using a thin foil of a material which has an absorption edge (see below) at a wavelength somewhat shorter than λ(Kα). If the atomic number of the metal from which the anode is made is Z, that of the filter material should be $Z-1$ or $Z-2$. Thus, nickel foil is used as a filter for copper radiation and zirconium for molybdenum radiation. Figure 4.7 shows the effect of a zirconium filter on the X-rays emitted from a molybdenum anode. Note the location of the absorption edge in the absorption spectrum of zirconium.

The above filtering method will now be briefly explained. X-rays are absorbed in matter according to Beer's law,

$$I = I_0 \exp(-\mu t),$$

where I_0 is the incident intensity of the X-ray beam, μ is the linear absorption

coefficient, t is the thickness of the irradiated specimen, and I is the intensity of the X-ray beam after it has passed through the specimen. This law is applicable to electromagnetic radiation in general. The absorption coefficient depends on the wavelength of the incident radiation, and in the case of X-rays, for a single atomic species (for example nickel atoms) it depends on the atomic number and on the third power of λ. Hence the absorption increases with increasing λ. However, when an energy is reached which corresponds exactly to the ionization energy of an atom in the absorber, the absorption falls abruptly and then continues to increase again as λ increases. The X-ray absorption spectrum of an atom has a sawtooth shape, each peak corresponding to the ionization of an electron from one of the atomic energy levels. The abrupt decrease of the absorption is called an absorption edge. For example, in the case of nickel the wavelength of the K absorption edge (corresponding to the ionization of the K shell) is $\lambda = 1.4882$ Å. Such a filter obviously decreases the Cu($K\alpha_{av}$) emission line but suppresses the Cu($K\beta$) and the white radiation to a much greater extent. Optimization of the filter thickness is of crucial importance here.

The above approximate monochromatization method is cheap, elegant, but rather imperfect and is only very infrequently used nowadays. A much more accurate method, is the use of a crystal monochromator. The principle is simple. A crystal is mounted in the incident X-ray beam so that its strongest reflection for a chosen wavelength, is active. The X-ray beam diffracted from the crystal is then used as the incident beam that falls on the sample to be examined. The wavelength of that radiation is just the above chosen wavelength. This is good but not entirely exact because of the width of the reflection profile from the monochromating crystal. There is another problem, that of harmonics: as we know, the Bragg equation is

$$n\lambda = 2d_{hkl} \sin\theta,$$

where $n = 1, 2, 3, \ldots$ is the order of the reflection from the lattice plane (hkl). Hence, together with radiation of wavelength λ, reflections corresponding to $\lambda/2$, $\lambda/3$, etc. may also be obtained. However, this can usually be taken care of either by an appropriate choice of the monochromator crystal or during the processing of the data.

4.3.2 Synchrotron radiation

All the diffraction techniques to be outlined below are of widespread availability; they can be found in crystallographic laboratories, and serve as the basic tools for the collection of diffracted-intensity data. A popular instrument is the four-circle diffractometer, because of its accuracy and sophisticated automation. Its main limitations, when a sealed X-ray tube is used, are the relatively low intensity of incident radation that can be obtained and the neccessity for collecting the diffracted intensities from one reflection at a time. The first of these results in time-consuming experiments, and the second adds the danger of crystal deterioration due to radiation damage. Ideally, therefore, one would like to be able to

collect a large number of diffracted intensities in a short time. The best answer to the latter requirement is offered by synchrotron radiation which is produced in special installations. The simultaneous collection of several items of intensity data is made possible by area detectors (see below).

The physical principle of synchrotron radiation goes back to classical electrodynamics: an accelerated moving charge emits a spectrum of electromagnetic energy, and if the magnitude of the velocity of its motion is comparable to the speed of light, very significant effects are predicted and, in fact, observed. The theory of synchrotron radiation and its application to crystallography have been discussed rather extensively in the literature (for example Koch 1983; Coppens 1992), and only a brief outline will be given in this chapter. If a charge e moves with a velocity $\mathbf{u}$, and has an acceleration vector $\dot{\mathbf{u}}$, then the power radiated by the charge is given by

$$P = \frac{e^2}{6\pi\epsilon_0 c^3} \frac{\dot{u}^2 - (\mathbf{u} \times \dot{\mathbf{u}})^2/c^2}{(1 - u^2/c^2)^3}, \tag{4.13}$$

(Schwinger 1949; Panofsky and Phillips 1956), where c is the speed of light in vacuo and ϵ_0 is the permittivity of free space. This general expression readily admits the basic ideal features of the synchrotron as a special case: a charge rotates in a circular orbit of radius R with speed u, caused by a strong magnetic field perpendicular to the plane of the orbit, and orbits with a constant circular frequency ω. At any instant the velocity vector is tangential to the orbit, and the acceleration vector is perpendicular to it. Hence the magnitude of the vector product $(\mathbf{u} \times \dot{\mathbf{u}})$ reduces to $u\dot{u}$ and eqn (4.13) can be rewritten as

$$P = \frac{e^2 \dot{u}^2}{6\pi\epsilon_0 c^3} \frac{1}{(1 - u^2/c^2)^2}. \tag{4.14}$$

It is readily seen that for speeds much lower than c, eqn (4.14) reduces to

$$P = \frac{e^2 \dot{u}^2}{6\pi\epsilon_0 c^3} \tag{4.15}$$

which is the total instantaneous power radiated by a nonrelativistic accelerated charge. Equation (4.15) is of importance in the description of the scattering of X-rays by electrons in a crystal, since relatively small speeds are involved. This will be discussed in some detail in the next chapter.

Returning to eqn (4.14) and the acceleration of a charge in circular motion, the magnitude of the acceleration is $R\omega^2$, and if we introduce the definitions $\beta' = u/c$ and $\gamma' = (1 - \beta'^2)^{-1/2}$, eqn (4.14) becomes

$$P = \frac{e^2 R^2 \omega^4 \gamma'^4}{6\pi\epsilon_0 c^3}. \tag{4.16}$$

Since, further, $\omega = u/R = c\beta'/R$, we can write

$$P = \frac{e^2 c}{6\pi\epsilon_0} \frac{\beta'^4 \gamma'^4}{R^2}. \tag{4.17}$$

The total instantaneous power radiated by an electron accelerated in this way is therefore approximately proportional to the fourth power of the energy of the relativistic electron. In fact, the parameter γ' can be written as the ratio of the energy of the moving electron to its rest energy, and the speed-dependent term β'^4 tends to unity as u tends to c. Very high speeds permit large values of the radius and hence an ample circumference of the orbit, which allows a large number of users to benefit from this radiating accelerator. As will be seen later, the radiated power is very much higher than that obtaineable from conventional sources, such as X-ray tubes with stationary or even rotating anodes.

The above description forms the theoretical basis of real synchrotron installations, which have led to major breakthroughs in structural studies. Detailed descriptions of the principles of operation of real synchrotrons and their application to crystallographic research are given in Coppens' (1992) and many other sources in the literature. We shall outline these principles briefly in what follows. We show in Fig. 4.8 a schematic view of an actual synchrotron installation.

- Electrons are injected by an electron gun into a linear accelerator (LINAC) in which they reach an energy of several hundred million electron volts (MeV).
- These energetic electrons are then injected into a synchrotron (BOOSTER), in which they circulate rapidly, while gaining an amount of energy in each revolution. This continues until the electrons reach an energy of several billion electron volts (GeV). At this point the speed of the electrons is very close to the speed of light c, and the parameter γ defined above becomes enormously large.
- These highly energetic electrons are then extracted from the synchrotron into the storage ring, where their motion is maintained, and they are therefore continuously accelerated and emit, tangentially to the ring, a spectrum of intense electromagnetic radiation.

A detailed description of the various experimental installations shown in Fig. 4.8 is outside the scope of this chapter. We shall just point out that many of them deal with extensive crystallographic research and they are well described on the web site http://www.aps.anl.gov/About/Research_Teams. It is also in order to point out that while APS is a major synchrotron installation, an increasing number of such installations can now be found in many countries.

A most important consideration is the spectral distribution of the synchrotron radiation, and specifically the achievement of high intensities of radiation in the interesting range of wavelengths–particularly those corresponding to X-rays. In practical installations, this is taken care of by suitable modifications of the path of the electron beam, and hence enhanced acceleration, with the aid of the insertion devices (see, for example, Coppens, 1992).

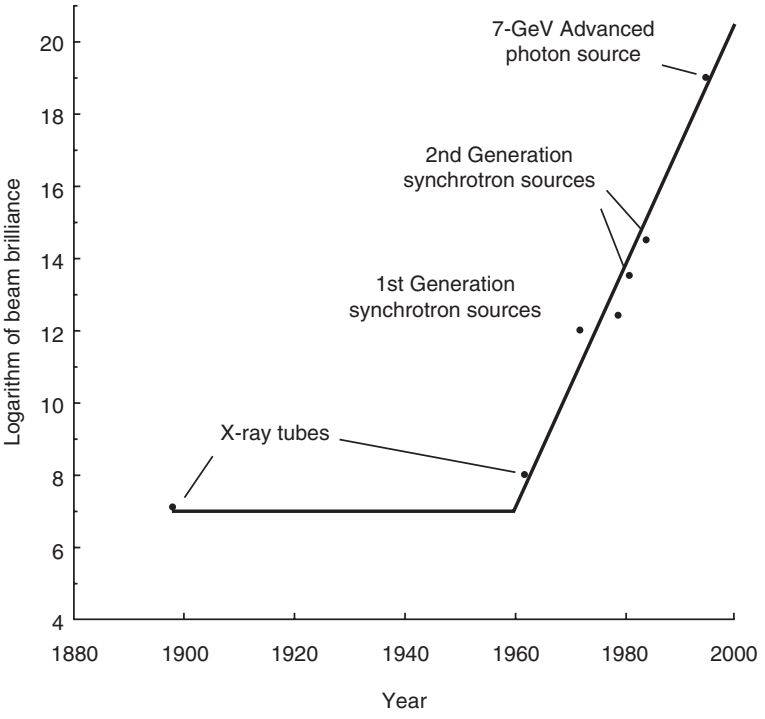


Fig. 4.9 *History of X-rays.* Reproduced by courtesy of Dr. R. Garrett from the web site: http://www.ansto.gov.au/natfac/asrp4.html.

Last but not least, a comparison of brilliance between conventional X-ray tubes and synchrotrons, as given in Fig. 4.9, brings out the main reason for the usefulness of this electron accelerator. (Brilliance is a quantity related to intensity (the average energy per unit time, per unit area) but also depends on the degree of spectral purity of the radiation, and has served until recently as a unit of comparison between various radiations.) It can be seen from Fig. 4.9 that no significant progress was made in the enhancement of the brilliance of X-ray sources from the invention of the X-ray tube by Roentgen in 1895 until about 1960. In the early 1960s, an X-ray tube with a rotating anode was introduced, which gave rise to an increase in the brilliance by an order of magnitude. This was regarded as a major development, but not for very long. The real breakthrough was afforded by particle accelerators in which *unwanted* synchrotron radiation from the electrons present was detected. This was accompanied by an increase in the brilliance by several orders of magnitude and marked the beginning of the so-called first-generation synchrotrons. It was soon realized that a program of construction of electron accelerators dedicated to the production of synchrotron radiation was indicated, and a second generation of synchrotrons appeared, with

a marked increase in brilliance. Further attempts were made at enhancing the brilliance by introducing various insertion devices, which contribute to increased acceleration of the electrons in the storage ring, and by improvong the maintenance of their energy–this led to the third-generation synchrotrons. Further research is in progress but even that accomplished some years ago has led to an increase in the brilliance by a factor of about 10^{12} as compared with a sealed X-ray tube with a rotating anode. This development is responsible for major advances in the structure determination of protein crystals and for the introduction of a variety of techniques which became feasible given the high brilliance of the radiation.

It should be pointed out that synchrotron radiation ranges throughout most of the useful spectrum of electromagnetic radiation, and thus constitutes a major stimulus to experimental science.

4.4 Detectors of X-rays

This section describes briefly the principles of operation of some detectors of X-rays that used to be or still are very popular. More detailed descriptions of these and other detectors, accompanied by graphical presentations and many references to relevant literature, are given in Volume C of the *International Tables for Crystallography* (Amemiya et al., 1999).

4.4.1 X-ray film

The oldest detector of X-rays, and one which is being used in many laboratories to this very day, is photographic film. Its principle of operation and processing are well known, but we shall recall them for the sake of completeness. X-ray film, unlike conventional photographic film, is coated on *both* sides with an emulsion, in which predominantly ionic silver halide crystals (usually AgBr) are dispersed.

- When an X-ray photon strikes the film, a small number of silver ions in an excited crystallite are converted to black metallic silver. So, upon completion of the experiment, a latent image of the diffraction pattern is stored in the film.
- The conversion process to metallic silver (only in the silver halide crystals exposed to X-rays) is completed by the developer solution, and the pattern of scattered X-rays which reached the film appears as appropriate blackenings on the emulsion. All this process is of course performed in a darkroom, to prevent exposure of the film to visible light.
- After the film has been washed in order to remove unwanted reaction products and traces of the developer solution, it is immersed in a fixer bath. The purpose of fixing is to remove the emulsion and all the siver halide crystals that were not exposed to X-rays. One now has (after washing and drying) a transparent film showing diffraction spots, the positions and relative intensities of which can be measured for the purpose of structure determination.

The positions of the spots are fairly accurate and can be used for a good determination of the unit cell parameters. However, the blackening of the film is proportional to the intensity of X-rays that caused it only in a relatively small range, called the linear range of optical density. If the whole range of optical densities was to be measured, it was customary to work with packs of several films: the weak reflections were measured on the film facing the incoming scattered radiation, and the intensity of the strongest reflections was reduced to the linear range in the last film of the pack. Although this procedure is rather tedious, this disadvantage is mainly technical. A more serious shortcoming is the very low quantum efficiency of X-ray film, which results in the need for very long exposures. The quantum efficiency of a detector is defined as the ratio of the number of detections to the number of incident photons. In the case of X-ray film, a detection can be taken as the excitation of a silver halide crystal, and the quantum efficiency amounts here to a few percent.

4.4.2 Imaging plate

The imaging plate is an area detector, qualitatively similar to the photographic film but operating on entirely different principles. It also consists of a support coated with an emulsion, which, however, contains crystallites of barium fluoride bromide or barium fluoride iodide with artificially introduced impurities of Eu^{2+} (doubly ionized europium). During the preparation of these crystals a large number of vacancies is created at the sites of fluoride and bromide (or iodide) negative ions, and these vacancies are essential to the process (see below).

- When a photon strikes the imaging plate, Eu^{2+} ions are ionized further to Eu^{3+} and the "detached" electrons are raised to the conduction band. When so excited, the electrons are trapped at the vacancies and thereby produce temporary color centers. Hence, the imaging plate changes color at the sites on which incident scattered radiation is falling.
- When the exposure has been completed, the diffraction pattern has been temporarily recorded on the imaging plate. The plate is then scanned by a He–Ne laser, and the trapped electrons are released, fall down to the valence band, and recombine with Eu^{3+} to Eu^{2+}. This transition is accompanied by a release of energy, which corresponds to the emission of blue light. The intensity of this luminescence, measured with a photomultiplier, is proportional to the intensity of the X-rays which gave rise to the color centers.
- The coordinates of the diffraction spots and the intensity of the luminescence are recorded online in a computer and constitute the required set of data. When the scanning process has been completed, the imaging plate is exposed to visible light, which erases all the remaining traces of color centers and the plate is suitable for further use.

Unlike the X-ray film discussed above, the imaging plate has a very large linearity range, and an excellent quantum efficiency, and is therefore a convenient and very fast detector of X-ray diffraction patterns. It is nowadays frequently

used in the collection of intensity data from protein crystals and is quite popular in other applications.

4.4.3 Charge-coupled device (CCD) detector

Another powerful area detector is based on a popular method of electronic imaging, which employs a two-dimensional array of small light-sensitive elements, known as a charge-coupled device; the elements are referred to as pixels. The CCD has a variety of applications, and their implementation in an X-ray detector is discussed by Amemiya et al. (1999) and in the literature referred to there. Let us see, in broad outline, how X-rays scattered from a crystal are converted into an image of a diffraction pattern.

- The detector itself has the shape of a truncated cone, at the large base of which is a phosphor screen, the purpose of which is to convert incident X-ray photons into visible light. The light emitted by the screen is conducted by a tapered bundle of optical fibers and strikes the array of pixels, each of which is a metal–oxide–semiconductor (MOS) capacitor.
- When a light photon strikes an MOS pixel, an electron is emitted owing to the photoelectric effect and stored in the capacitor (an electron–hole pair is produced). Therefore, the charge distribution throughout the whole CCD follows the distribution of radiation scattered from the crystal. The charge is subsequently transferred to an electronic circuit, and converted into an array of pulses the height of which is proportional to the intensity of X-rays that fell on the phosphor screen. This digital information is transferred to a computer, which records the pattern of diffracted intensity on a relative scale.
- The crystal is then rotated, new reciprocal-lattice vectors come into contact with the Ewald sphere, and a new charge frame is produced in the CCD. All this is repeated until the desired portion of the diffraction space has been covered by the motions imparted to the crystal. The required information on the distribution of diffracted intensity is now stored in the computer and available for further processing.

The CCD detector has a very large linearity range; it has a high quantum efficiency and a large dynamic range (the ratio between the maximum and minimum reliably measured intensities). It is not clear whether the imaging plate or the CCD detector is preferable, but both are certainly in the forefront of intensity data collection. The performance of the CCD detector also depends on the size of the pixel array. Typical values are 1.5 to about 4 million pixels. Interestingly, values of the same order are encountered in digital cameras, in which CCD arrays have replaced photographic film.

4.5 The rotating-crystal method

This is the oldest moving-crystal method. It was for many decades associated with photographic film and conventional X-ray tubes, but in modern research

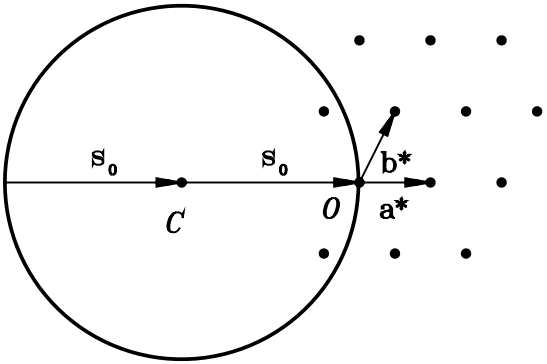


Fig. 4.10 Rotating crystal method. Schematic drawing.

the X-ray film is being replaced by imaging plates and–where feasible–the X-ray tube by synchrotron radiation. However, this has nothing to do with the geometrical considerations of this method, and an important variant of it, the oscillation method.

The diffraction condition is fulfilled when a point of the reciprocal lattice comes into contact with the sphere of reflection. If the radiation is monochromatic and the crystal is stationary, any occurrence of reflections is accidental and there may be none at all. Since the direct lattice can be represented in terms of families of parallel, equidistant lattice planes, and to each of these families there correspond collinear vectors in the reciprocal lattice, then if the crystal is rotated about some direction, a large number of reciprocal-lattice vectors will sweep through Ewald's sphere and give rise to reflections. This is the principle of the rotating crystal method, and is related to other methods in which the crystal is moved in order to bring reciprocal-lattice points into contact with the sphere of reflection. Let us consider Figure 4.10. The plane of the drawing contains the basis vectors $\mathbf{a}^*$ and $\mathbf{b}^*$ of the reciprocal lattice and linear combinations of them with integer coefficients, and the direct basis vector perpendicular to this lattice plane in the reciprocal lattice must be the vector $\mathbf{c}$ (see Section 1.3).

Consider the Laue equation $\mathbf{c} \cdot \mathbf{h} = l$, with $l = 0$. The locus of the vectors $\mathbf{h}$ (not only that of their endpoints) is a reciprocal-lattice plane perpendicular to $\mathbf{c}$, that is, a typical diffraction vector has the form $\mathbf{h} = h\mathbf{a}^* + k\mathbf{b}^* + 0\mathbf{c}^*$, and each of the reciprocal-lattice points in this plane has indices $hk0$. When the crystal is rotated about $\mathbf{c}$, each of the points with indices $hk0$ becomes a *reflection* $hk0$ as soon as the endpoint of the reciprocal-lattice vector $h\mathbf{a}^* + k\mathbf{b}^* + 0\mathbf{c}^*$ comes into contact with the surface of the sphere of reflection. The diffracted beams radiate from the point C and lie in the $(\mathbf{a}^*, \mathbf{b}^*)$ plane. They can also be regarded

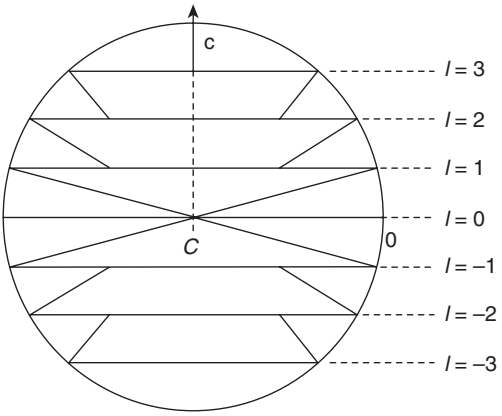


Fig. 4.11 Projected Laue cones limited by the Ewald sphere.

as lying on the surface of a flat cone, with the apex at the point C.

We now proceed to the Laue equations $\mathbf{c} \cdot \mathbf{h} = l$ with $l \neq 0$. If we divide both sides of this equation by $c = |\mathbf{c}|$, we obtain

$$\frac{\mathbf{c}}{c} \cdot \mathbf{h} = \frac{l}{c}, \tag{4.18}$$

that is, the projection of $\mathbf{h}$ on the direction of $\mathbf{c}$ is constant and equals l/c. The locus of the endpoints of $\mathbf{h}$, satisfying eqn (4.18), is a plane in the reciprocal lattice in which each point has indices hkl. As the crystal rotates about $\mathbf{c}$, the vectors $\mathbf{h}_{hkl}$, for l not exceeding a certain maximum value, cross the sphere of reflection and give rise to corresponding diffracted beams. These beams, or reflections, are located on the envelope of a cone with its apex at the point C, its axis parallel to the axis of rotation, and its half-opening angle given by

$$\alpha_l = \cos^{-1} \frac{l\lambda}{c}.$$

Clearly, the maximum value of l is c/λ, truncated to the nearest integer.

Each Laue equation therefore corresponds to a family of planes in the reciprocal lattice which, upon rotation of the crystal, intersect the sphere of reflection and form a series of reflection cones, known as *Laue cones*. Figure 4.11 illustrates this statement.

The rotating crystal method, the oldest technique employing monochromatic radiation, is related in a simple manner to the above description. Suppose that a single crystal is irradiated, with $\mathbf{s}_0$ perpendicular to $\mathbf{c}$, while it is rotating about the direction of $\mathbf{c}$ at a uniform angular speed. Let us now surround the crystal with a cylindrical photographic film, suitably protected from exposure to light, so that the axis of rotation coincides with the axis of the cylinder. The Laue cones, or the loci of the reflections, will intersect the cylinder in circles

on the circumference of which the X-ray reflections (assumed to penetrate the protecting medium) will give rise to latent sharp spots. After the experiment has been performed and the cylindrical film or other imaging medium has been flattened out, we obtain a series of straight rows of sharp spots, of varying degree of blackening.

The spots in the central row correspond to the Laue equation $\mathbf{c} \cdot \mathbf{h} = 0$ and therefore to indices $hk0$. The first row above the center has indices $hk1$, the first below the center has indices $hk\overline{1}$, and so on. A complete interpretation of the photograph would involve the assignment of indices h and k to each spot in the row of $hk0$ reflections (this turns out to be sufficient) and a quantitative estimation of the intensity of the spots in the photograph. Such an assignment of indices, or *indexing*, requires a knowledge of unit cell parameters of the direct or reciprocal cell, and a single rotation photograph furnishes only one such parameter, as will be seen below. Let d be the distance of the lth circle on the film from the central row, and let R be the distance of the film from the crystal. Therefore,

$$2\theta = \tan^{-1}\left(\frac{d}{R}\right).$$

The distance from the base of the lth Laue cone from the flat cone, in the sphere of reflection, is l/c. We thus have

$$\frac{l/c}{1/\lambda} = \frac{l\lambda}{c} = \sin(2\theta)$$

and hence

$$c = \frac{l\lambda}{\sin[\tan^{-1}(d/R)]}$$

and only the length of the vector $\mathbf{c}$ can be determined.

It should be pointed out that a rotation photograph of a cubic crystal can be readily indexed, since only one parameter is needed. However, for lower symmetries other types of information are usually required. The rotating-crystal method has been described here mainly in order to introduce the reader to the basics of the formation of a diffraction pattern. The actual experimental technique involved and further details of the interpretation are very clearly described for example, by Buerger (1941) and by Stout and Jensen (1968), and the interested reader is referred to these works.

4.6 Moving-crystal–moving-film methods

We shall now mention briefly some photographic methods that served as the crystallographer's tool for several decades, and some of which are still used, albeit not frequently. However, their revival is possible in view of the development of new, highly efficient, detectors which can replace the classical photographic film, as well as in view of the possibility of computerized indexing.

4.6.1 The Weissenberg method

Consider the arrangement described in the previous section, with two modifications: (i) only one Laue cone is allowed to reach the film, for example by the use of a sliding cylindrical metallic absorber with a circular slit that can be positioned so that only a desired cone is transmitted, and (ii) the cylindrical film is allowed to move back and forth, while remaining coaxial with the axis of rotation of the crystal; the movement of the film and the rotation of the crystal are synchronous (for example, Buerger, 1941). If, for example, only the $hk0$ Laue cone is allowed to pass, the $hk0$ reflections will be spread throughout the film in a regular manner and this turns out to permit the determination of the parameters a^*, b^* and the angle γ^*. So, from a single setting of the crystal, four out of the six possible parameters can be determined. Further experimental details are provided in the references quoted.

4.6.2 The de Jong–Bouman method

In this method, very elegantly illustrated by Woolfson (1997), also only one Laue cone is allowed to pass. This is done by placing a flat metallic absorber with a circular ring aperture in a plane perpendicular to the axis of rotation of the crystal, say the **c** axis, so that the axis is directed towards the center of the circular absorber. The Laue cone is selected by setting the inclination of $\mathbf{s}_0$ with respect to the **c** axis, and the distance of the absorber from the crystal. A flat film is rotated about an axis parallel to the axis of rotation of the crystal at the same angular speed as the crystal, the plane of the film being perpendicular to the axis of rotation. In our example, the reflections $hk0$ are spread all over the film. However, if a crystal is rotated about an axis at a certain angular speed, the reciprocal lattice is rotated about a parallel axis with the same speed. Hence, the distribution of the spots on the de Jong–Bouman photograph will follow the geometry of the reciprocal-lattice plane based on $\mathbf{a}^*$ and $\mathbf{b}^*$. This is the first example of the so-called "undistorted reciprocal-lattice photography". In fact, the Weissenberg method also produces "photographs of the reciprocal lattice", but seriously distorted ones owing to the cylindrical geometry.

4.6.3 The Buerger precession method

This method is described in great detail by Buerger (1964) and is dealt with more briefly in most crystallographic texts. In this method, again only one Laue cone is allowed to pass in a given experiment, and its result is an undistorted image of a reciprocal-lattice plane. However, the mechanical design is based on a precession– rather than rotation–of a direct lattice vector, a corresponding precession of the transmitted Laue cone and the absorber involved, and a rather complicated motion of the film involving a combination of precession and translation of the film parallel to itself. While the de Jong-Bouman principle illustrates reciprocal-lattice photography neatly, Buerger's precession camera–although complicated– is versatile and much more frequently used. This is especially true for preliminary examinations of protein crystals.

4.7 The four-circle diffractometer

4.7.1 Geometrical considerations

An outstandingly important instrument, allowing one to determine the full set of unit-cell parameters as well as to measure accurately the intensities of all the accessible reflections–all with a single setting of the crystal–is the four-circle diffractometer. This instrument is equipped with a photon-counting device and a mechanical system which can be programmed (i) to bring the crystal into an orientation in which the wavevector of the incident radiation forms the Bragg angle θ with the desired plane hkl, and (ii) to bring the slit of the detector to a position in which it can receive the scattered radiation, with a wavevector *also* forming the Bragg angle θ with the plane hkl. This very general description indicates that the diffractometer can be programmed to measure automatically the intensities of a large range of reflections, which is an obvious asset. However, it also indicates that the reflections are measured one at a time, which is a disadvantage if the number of reflections is very large and the intensity of the scattered radiation deteriorates upon prolonged exposure of the crystal to X-rays. For this reason, and in order to perform the data collection more expediently, the slit that accepts one reflection at a time is being gradually replaced with an area detector such as,for example, the CCD device discussed above. The problem of crystal deterioration is encountered most often in studies of protein crystals, and is less acute in crystals of small and medium-sized molecules. Of course, the replacement of the slit with an area detector radically changes the computational aspects of the data collection, a detailed treatment of which is outside the scope of this book. We shall, by way of an introduction, analyze the classical four-circle diffractometer which measures one reflection at a time. The present analysis is based on the article by Hamilton (1974). A schematic drawing of a four-circle diffractometer is shown in Fig. 4.7.1.

Since the control of the four-circle diffractometer is the precursor of that of most modern diffraction techniques, we shall describe here the geometrical details involved in the Eulerian cradle variant of the single-crystal diffractometer. The instrument can be described as follows: three points (i) the center of the source of the radiation (S), (ii) the center of the crystal (C), and (iii) the center of the receiving slit of the detector (D) define a plane, which call the *diffraction plane.* The axis passing through the crystal and perpendicular to the diffraction plane is called the *principal axis* of the instrument. Its direction remains fixed throughout the experiment (perpendicular to the table on which the instrument is mounted), and hence the diffraction plane is horizontal. The detector is therefore constrained to rotate about the principal axis only. The angle SCD equals $180^\circ - 2\theta$,where 2θ is the angle between the incident and the diffracted beam. The diffraction vector corresponding to the Bragg angle θ is parallel to the bisector of the angle SCD. The other axes of rotation are:

- *The χ axis.* This is an axis passing through the crystal and lying in the diffraction plane. In a conventional diffractometer, this is the symmetry axis

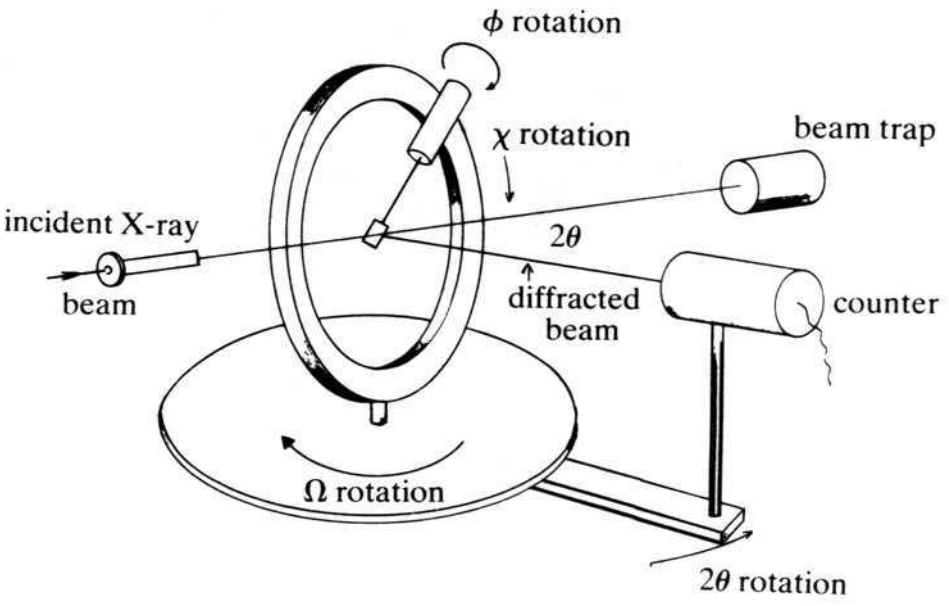


Fig. 4.12 A schematic drawing of a four-circle diffractometer. Reproduced with copyright permission of the International Union of Crystallography (IUCr).

of a ring on whose internal cylindrical surface the device to which the crystal is rigidly attached can be displaced by a predetermined angle, called the χ angle. The center of the crystal must coincide with the center of the χ ring, throughout the experiment, and the plane of the χ ring is perpendicular to the diffraction plane.

- *The ϕ axis.* This is an axis about which the crystal, together with the device to which the crystal is rigidly attached, can be rotated through a predetermined angle, called the ϕ angle. The device carrying the crystal is called the goniometer head. During the rotation about the ϕ axis the center of the crystal must remain at the center of the χ ring and the orientation of the axis of rotation of the crystal within the plane of the ring is determined by the χ angle.
- *The Ω axis.* This axis passes through the crystal and through the plane of the χ ring, and is perpendicular to the diffraction plane. By definition, the Ω axis coincides with the principal axis of the instrument. Physically, however, there are two independent rotations associated with this axis: the Ω motor rotates the χ ring (with everything it carries) and does not affect the position of the detector, and the 2θ motor rotates the detector without affecting the orientation of the crystal with respect to the incident beam.

- *The ψ axis.* This is (usually) a virtual axis, the direction of which coincides with the direction of the diffraction vector. If the crystal is very small, or ground to a sphere, rotation of the crystal about the ψ axis will not cause appreciable fluctuations in the diffracted intensity corresponding to this vector. If, on the other hand, the crystal is strongly anisotropic (for example, if the crystal has the form of a platelet or a needle), the intensity of diffracted radiation will in general vary as the crystal is rotated about the ψ axis, because of varying absorption. The ψ rotation can be realized by a suitable combination of the angles χ, ϕ,and Ω.

In a diffraction experiment performed with the aid of a four-circle diffractometer, the diffraction vector is represented in terms of several sets of basis vectors:

1. The conventional basis of the reciprocal lattice, the coordinates of the diffraction vector are simply the integers appearing in the Laue equations. This representation can be written as

$$\mathbf{h} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \equiv \mathbf{H}^{\mathrm{T}}\mathbf{A}^*, \tag{4.19}$$

   where $\mathbf{H}^{\mathrm{T}} = (h\ k\ l)$ and $\mathbf{A}^{*\mathrm{T}} = (\mathbf{a}^*\ \mathbf{b}^*\ \mathbf{c}^*)$.
2. An orthonormal basis attached to the diffraction vector and the diffractometer. Such a basis is needed for the construction of the laboratory working system. We represent $\mathbf{h}$ in this system as

$$\mathbf{h} = \mathbf{X}_{\mathrm{D}}^{\mathrm{T}}\mathbf{E}_{\mathrm{D}}, \tag{4.20}$$

   where $\mathbf{X}_{\mathrm{D}}^{\mathrm{T}} = (x_{\mathrm{D}}^1\ x_{\mathrm{D}}^2\ x_{\mathrm{D}}^3)$ and $\mathbf{E}_{\mathrm{D}}^{\mathrm{T}} = (\mathbf{e}_{1\mathrm{D}}\ \mathbf{e}_{2\mathrm{D}}\ \mathbf{e}_{3\mathrm{D}})$, and where the basis vectors $\mathbf{e}_{j\mathrm{D}}$, $j = 1, 2, 3$, form a right-handed set of orthonormal (unit) vectors. These vectors are defined as follows:
   - The vector $\mathbf{e}_{2\mathrm{D}}$ is parallel to the diffraction vector and therefore bisects the complementary angle $180° - 2\theta$ between the incident and diffracted beams.
   - The vector $\mathbf{e}_{1\mathrm{D}}$ lies in the diffraction plane, is perpendicular to $\mathbf{e}_{2\mathrm{D}}$ and points to the source of radiation when $\theta = 0$.
   - The vector $\mathbf{e}_{3\mathrm{D}}$ coincides with the principal axis of the instrument and is directed so as to make the system of basis vectors right handed.
3. An orthonormal basis attached to the crystal and the diffractometer. This is a necessary mediator between the crystal system and the laboratory system. The diffraction vector is given in this system by

$$\mathbf{h} = \mathbf{X}_{\mathrm{G}}^{\mathrm{T}}\mathbf{E}_{\mathrm{G}}, \tag{4.21}$$

   where $\mathbf{X}_{\mathrm{G}}^{\mathrm{T}} = (x_{\mathrm{G}}^1\ x_{\mathrm{G}}^2\ x_{\mathrm{G}}^3)$ and $\mathbf{E}_{\mathrm{G}}^{\mathrm{T}} = (\mathbf{e}_{1\mathrm{G}}\ \mathbf{e}_{2\mathrm{G}}\ \mathbf{e}_{3\mathrm{G}})$, where the basis vectors $\mathbf{e}_{j\mathrm{G}}$, $j = 1, 2, 3$, form a right-handed set of orthonormal (unit) vectors. The basis vectors in eqn (4.21) are defined so that the $\mathbf{E}_{\mathrm{G}}$ and $\mathbf{E}_{\mathrm{D}}$ sets

of basis vectors coincide when $\chi = \phi = \Omega = 0$. Also, the unit vector $\mathbf{e}_{3G}$ always coincides with the ϕ axis (the axis about which the goniometer head rotates).

For any values of the angles χ, ϕ, and Ω, the two orthonormal bases described above are related by a rotation matrix depending on these three angles, which correspond to a known triplet of Eulerian angles (see, for example, Goldstein, 1956; note, however, that the meaning of Golstein's symbols is different from the present usage). This rotation matrix is obtained as a product of three rotation matrices about the corresponding axes. That is,

$$\mathbf{E}_{\mathrm{G}} = \mathbf{F}\mathbf{E}_{\mathrm{D}}, \tag{4.22}$$

where

$$\begin{aligned}\mathbf{F} &= \boldsymbol{r}_{\phi}\boldsymbol{r}_{\chi}\boldsymbol{r}_{\Omega} \\ &= \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & & \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\chi & \sin\chi \\ 0 & -\sin\chi & \cos\chi \end{pmatrix} \begin{pmatrix} \cos\Omega & \sin\Omega & 0 \\ -\sin\Omega & \cos\Omega & 0 \\ 0 & & \end{pmatrix} \\ &= \begin{pmatrix} \cos\phi\cos\Omega - \sin\phi\sin\Omega\cos\chi & \cos\phi\sin\Omega - \sin\phi\cos\Omega\cos\chi & \sin\phi\sin\chi \\ -\sin\phi\cos\Omega - \cos\phi\sin\Omega\cos\chi & -\sin\phi\sin\Omega + \cos\phi\cos\Omega\cos\chi & \cos\phi\sin\chi \\ \sin\chi\sin\Omega & -\sin\chi\cos\Omega & \cos\chi \end{pmatrix}\end{aligned} \tag{4.23}$$

Since, however,

$$\mathbf{h} = \mathbf{X}_{\mathrm{G}}^{\mathrm{T}}\mathbf{E}_{\mathrm{G}} = \mathbf{X}_{\mathrm{D}}^{\mathrm{T}}\mathbf{E}_{\mathrm{D}} = \mathbf{X}_{\mathrm{D}}^{\mathrm{T}}\mathbf{F}^{-1}\mathbf{F}\mathbf{E}_{\mathrm{D}}$$

we must have

$$\mathbf{X}_{\mathrm{G}}^{\mathrm{T}} = \mathbf{X}_{\mathrm{D}}^{\mathrm{T}}\mathbf{F}^{-1}$$

or

$$\mathbf{X}_{\mathrm{G}} = \mathbf{F}\mathbf{X}_{\mathrm{D}}, \tag{4.24}$$

because a matrix of rigid rotation is orthogonal, and for such a matrix its inverse and transpose are identical.

4.7.2 The orientation matrix

It is very useful to define a matrix $\mathbf{V}$ that satisfies the relation

$$\mathbf{A}^{*} = \mathbf{V}\mathbf{E}_{\mathrm{G}}. \tag{4.25}$$

Each row of $\mathbf{V}$ contains the Cartesian components of a basis vector of the reciprocal lattice, in the system linked to the crystal and diffractometer. If the orientation matrix is known, the unit cell dimensions can be obtained in a straightforward manner. It can be shown that

$$\mathbf{A}^{*} \cdot \mathbf{A}^{*\mathrm{T}} = \mathbf{g}^{-1},$$

where $\mathbf{g}$ is the matrix of the direct metric tensor (see Appendix B). Indeed

$$\begin{aligned}\mathbf{A}^* \cdot \mathbf{A}^{*\mathrm{T}} &= \mathbf{V}\mathbf{E}_\mathrm{G} \cdot \mathbf{E}_\mathrm{G}^\mathrm{T}\mathbf{V}^\mathrm{T} \\ &= \mathbf{V}\mathbf{V}^\mathrm{T}\end{aligned}$$

since $\mathbf{E}_\mathrm{G} \cdot \mathbf{E}_\mathrm{G}^\mathrm{T}$ is a unit matrix. Therefore, the product $\mathbf{V}\mathbf{V}^\mathrm{T}$ is indentically equal to the matrix of the metric tensor of the basis of the reciprocal lattice. By inverting this matrix, we obtain the matrix of the metric tensor of the basis of the direct lattice, and hence the direct unit cell parameters. The orientation matrix is of central importance in planning diffraction experiments by diffractometric methods as well as other methods.

4.7.3 Coordinates and angles

Recall that the diffraction vector $\mathbf{h}$ is always parallel to the unit vector $\mathbf{e}_{2\mathrm{D}}$ of the orthonormal basis linked to $\mathbf{h}$ and the diffractometer. For any reflection, we can therefore write $\mathbf{h} = |\mathbf{h}|\mathbf{e}_{2\mathrm{D}}$, or

$$\mathbf{X}_\mathrm{D} = \begin{pmatrix} 0 \\ |\mathbf{h}| \\ 0 \end{pmatrix}. \tag{4.26}$$

If we premultiply the right-hand side of eqn (4.26) by the rotation matrix $\mathbf{F}$ given by eqn (4.23), we obtain the Cartesian coordinates of the diffraction vector in the system linked to the crystal and diffractometer,

$$\mathbf{X}_\mathrm{G} = |\mathbf{h}| \begin{pmatrix} \cos\phi\sin\Omega + \sin\phi\cos\chi\cos\Omega \\ -\sin\phi\sin\Omega + \cos\phi\cos\chi\cos\Omega \\ -\sin\phi\cos\Omega \end{pmatrix}. \tag{4.27}$$

We now obtain, from eqns (4.19), (4.21) and (4.25),

$$\mathbf{H}^\mathrm{T}\mathbf{A}^* = \mathbf{H}^\mathrm{T}\mathbf{V}\mathbf{E}_\mathrm{G} = \mathbf{X}_G^\mathrm{T}\mathbf{E}_\mathrm{G},$$

from which it follows that

$$\mathbf{H}^\mathrm{T}\mathbf{V} = \mathbf{X}_\mathrm{G}^\mathrm{T}. \tag{4.28}$$

If we know $\mathbf{H}^\mathrm{T} = (h\ k\ l)$ and the orientation matrix $\mathbf{V}$, we can compute the components of $\mathbf{X}_\mathrm{G}$ and solve eqn (4.27) for the values of the angles χ, ϕ, Ω which are required in order to bring the crystal to an orientation at which the intensity of the reflection $\mathbf{h}$ can be measured. Equations (4.27) and (4.28) are of value in programming a diffractometer to carry out intensity measurements for given ranges of reflection indices.

If, finally, we know the angular settings of a given reflection , for which the Cartesian coordinates of the corresponding diffraction vector can be computed from eqn (4.27) and the orientation matrix is also given, then the indices of this reflection are in principle found as

$$\mathbf{H} = (\mathbf{V}^{\mathrm{T}})^{-1}\mathbf{X}_{\mathrm{G}} \tag{4.29}$$

Equation (4.29) is of importance for the indexing of reflections in the preliminary stages of the work.

It remains to show how the orientation matrix can be determined, or alternatively, how the Cartesian components of the basis vectors of the reciprocal lattice can be obtained. This can be usefully preceded by some comments on the experimental strategies employed.

4.7.4 Comments on the experiment

There are several methods of determining the orientation matrix, all of them requiring some preliminary experimental work. If some information about the crystal is already available, for example from photographic work, this can be of value in the determination of the orientation matrix and the unit-cell parameters. It is, however, more common to "put the crystal on the diffractometer" in an arbitrary orientation and carry out an experiment. An important prerequisite is to bring the crystal to the center of the diffractometer system, at which all the axes of rotation intersect. This can be done manually by bringing the crystal to the center of the field of a properly aligned optical microscope. Once this is done, the computer-controlled operation of the diffractometer takes over. With the radiation on, the crystal is systematically scanned over the $\theta, \chi, \phi, \Omega$ space, until a significantly diffracting orientation is reached. Once it is there, the computer makes minor adjustments of the various axes until a maximum of a diffraction peak is obtained. The values of all four angles corresponding to the diffraction maximum are automatically recorded, and the magnitude of the diffraction vector is computed as $2\sin\theta/\lambda$. The coordinates of the diffraction vector in the "G" Cartesian system can now be computed from eqn (4.27) and its direction is also defined. The automatic search for diffraction maxima continues until some 20 or so reflections have been recorded and the corresponding diffraction vectors defined. For better accuracy, the available reflections are checked together and recentered. An often useful alternative to the above systematic search is a rotation photograph taken on the diffractometer, on which the coordinates of some 20 or so reflections are measured and serve as an input to a program which locates the reflections in the $\theta, \chi, \phi, \Omega$ space. This procedure also involves a restricted search but is superior to a full systematic search, since the exposure of the crystal to radiation is significantly reduced. The next stage is an automatic indexing procedure. All the sums and differences of the available diffraction vectors are now sorted according to increasing magnitude, and the three shortest vectors which form intervector angles as close as possible to 90° are chosen as the basis vectors of the reciprocal lattice. Since their coordinates in the "G" system are available, a first approximation to the orientation matrix is immediately obtained, and so are the metric tensors of the reciprocal and direct bases, and the corresponding unit-cell dimensions. All the reflections which have so far been located are now indexed, and the unit-cell parameters are refined by a least-squares procedure, which also provides their standard deviations.

At this point the ranges of the indices hkl are specified and the collection of the intensity data collection is planned. There are several modes of scanning the diffraction space around each reflection, most of them being implemented on particular instruments. The process of data collection is usually automatic and its reliability can be monitored in a number of ways. A frequently applied way is to choose two or three strong reflections and remeasure their intensities at regular time intervals. The intensities of these so-called "standard" reflections give useful indications of the stability of the system (crystal + diffractometer). The actual measurement of the intensity of a reflection consists of (i) bringing the crystal and counter to an orientation corresponding to the maximum intensity of the reflection; (ii) performing a scan of the diffraction space around the reflection according to the mode chosen, where the intensity is measured at each real or virtual step of the scan, thus creating an intensity profile; and (iii) integrating the intensity profile, including its background tails, and obtaining the net integrated intensity of the reflection along with the standard deviation of this intensity. Stage (iii) may range from a straightforward summation of intensity and background counts to more sophisticated profile analysis.

The results of this experiment, which are really the "raw material" for the determination of the crystal structure, are (i) the unit-cell dimensions and partial or complete information on the symmetry of the crystal (this subject will be discussed in the next chapter), and (ii) a number of records containing, for each reflection, its indices hkl, its net integrated intensity (or simply intensity) $I(hkl)$ and its standard deviation $\sigma_I(hkl)$. The actal planning of the experiment usually involves considerations of the accuracy which is aimed at and the number of structural parameters to be determined.

4.8 The Laue method

4.8.1 Principle of the method

The first diffraction pattern obtained by irradiating a stationary single crystal with a continuous spectrum of X-rays was observed by Friedrich, Knipping, and Laue (1912). Its interpretation was fully consistent with the existence of a periodic arrangement of material units within the crystal (Laue 1912), an idea which had been put forward in the eighteenth century and which gave rise to the theory of crystal symmetry–which dealt mainly with its microscopic aspects. It is most remarkable that the above experiment and all later diffraction experiments–even the most recent ones–bore out fully the theories of crystallographic lattices, point groups, and space groups, which were based mainly on macroscopic observations and on sound reasoning.

The technique based on the above experiment came to be known as the Laue method. The experimental arrangement is rather simple. The source of radiation is a continuous spectrum of X-rays, which fall directly on the crystal after passing through a "collimator". The crystal is stationary, and the diffracted radiation is usually collected by a flat radiation-sensitive plate, perpendicular to

the direction of the incident beam. In practice, the source can be a stationary-anode X-ray tube, a significantly more powerful rotating-anode X-ray tube, or, finally, a suitable range of X-rays selected from the spectrum of synchrotron radiation, the intensity of which is higher by several orders of magnitude than that emitted from laboratory X-ray tubes. The collector of diffracted radiation can be a photographic film or, as discussed above, an imaging plate, which is much more sensitive..

As we shall see in what follows, Laue patterns are not easily indexed, and were believed until recently to be of little or no use for structure determination. However, the most fruitful combination of a synchrotron source and an imaging-plate detector showed that a vast amount of information can be obtained in a very short time, and thus stimulated a search for indexing algorithms. Applications of the Laue method to structure determination of protein crystals may be encountered in the recent literature (for example, Helliwell 1992).

A well-known property of a Laue pattern is that it allows one to determine the orientation of the crystal, say in terms of the coordinates of a reciprocal-lattice vector which is perpendicular to an irradiated crystal face. This has many application to metallurgy and materials science in general. The symmetry of the Laue pattern is very sensitive to deviations of the direction of the incident beam from the normal to the irradiated crystal face, and hence its use in orienting crystals. The symmetry of the weighted reciprocal lattice will be dealt with in the next chapter.

Let us now see what the origins of the Laue pattern are. Since a stationary crystal is irradiated here with polychromatic radiation, then instead of a single Ewald sphere, as in the case of monochromatic radiation, we have now a range of such spheres, the largest one corresponding to the shortest wavelength and the smallest to some arbitrarily chosen longest wavelength. Both spheres pass through the origin of the reciprocal lattice, and *any* reciprocal-lattice point lying within the large sphere and outside the small one corresponds to a possible reflection. This is how a stationary crystal can give rise to a large number of simultaneously produced reflections. We shall now show, expanding the derivation given by Rabinovich and Lourie (1987), how a Laue pattern can be computed or, perhaps, simulated. This procedure leads to the possibility of indexing the pattern.

4.8.2 Calculation of the Laue pattern

Let us assume that a polychromatic ("white") X-ray beam is perpendicular to a circular flat plate that acts as a detector (a photographic film or an imaging plate), and passes through a stationary single crystal with known unit-cell dimensions. Let the crystal-to-plate distance be d centimeters, the radius of the plate be R_{m} centimeters, the X-ray tube operate at a high voltage of V volts, and the absolute maximum values of the diffraction indices be $h_{\max}$, $k_{\max}$ and $l_{\max}$. Under these conditions, the shortest wavelength is given by

$$\lambda_{\min} = \frac{12\,398}{V} \quad \text{Å},$$

the largest recordable Bragg angle is

$$\theta_{\max} = 0.5 \tan^{-1}\left(\frac{R_{\mathrm{m}}}{d}\right),$$

and the largest magnitude of the diffraction vector is

$$|\mathbf{h}|_{\max} = \frac{2 \sin \theta_{\max}}{\lambda_{\min}}.$$

As indicated in Appendix B, a knowledge of the unit-cell dimensions enables us to obtain the matrix of the metric tensor of the direct lattice, with components g_{ij}, and inversion of the latter matrix leads to the metric tensor of the reciprocal lattice, with components g^{ij}. This is of use in the calculation of the magnitude of the diffraction vector as

$$|\mathbf{h}| = \left(\sum_{i=1}^{3}\sum_{j=1}^{3} h_i h_j g^{ij}\right)^{1/2}, \tag{4.30}$$

where $-|h_{\max}| \leq h_1 \equiv h \leq h_{\max}$, $-|k_{\max}| \leq h_2 \equiv k \leq k_{\max}$, and $-|l_{\max}| \leq h_3 \equiv l \leq l_{\max}$ (see also Appendix B).

It is also convenient to define a Cartesian system in the diffraction device so that the unit vector $\mathbf{e}_3$ points towards the X-ray source (is antiparallel to $\mathbf{s}_0$), $\mathbf{e}_1$ is horizontal, $\mathbf{e}_2$ is vertical, and the three orthonormal basis vectors form a right-handed triad. (see Fig. 4.13). The diffraction vector in the reciprocal and Cartesian bases can then be written as

$$\mathbf{h} = \sum_{j=1}^{3} h_j \mathbf{a}^j = \sum_{i=1}^{3} q^i \mathbf{e}_i. \tag{4.31}$$

If we relate the basis vectors of the reciprocal lattice to the Cartesian basis by means of an orientation matrix, say $\mathbf{D}$, the diffraction vector can be written as

$$\mathbf{h} = \sum_{j=1}^{3} h_j \sum_{i=1}^{3} D^{ji} \mathbf{e}_i, \tag{4.32}$$

and the Cartesian components of the diffraction vector are therefore given by

$$q^i = \sum_{j=1}^{3} h_j D^{ji}. \tag{4.33}$$

The components of the wavevectors $\mathbf{s}$ and $\mathbf{s}_0$ are now

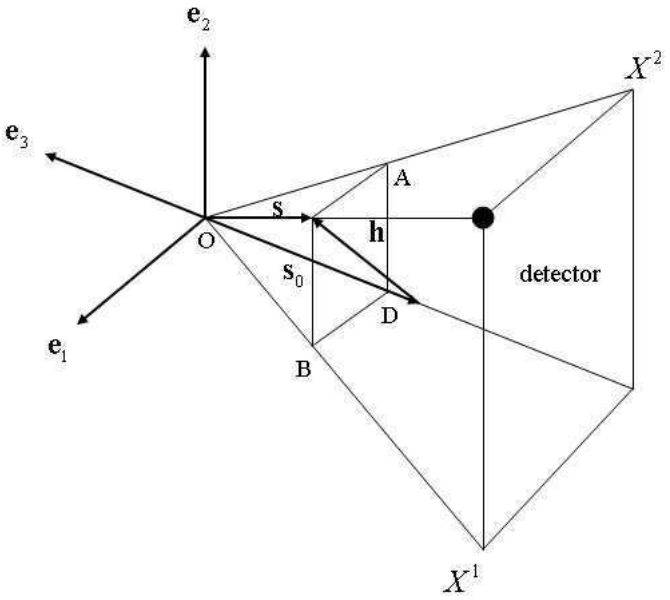


Fig. 4.13 *To the indexing of the Laue pattern.*

$$\mathbf{s}_0 : (0,\ 0,\ -1/\lambda) \tag{4.34}$$

and

$$\mathbf{s} = \mathbf{h} + \mathbf{s}_0 : (q^1\lambda,\ q^2\lambda,\ q^3\lambda - 1)/\lambda. \tag{4.35}$$

Let us now consider a plane in reciprocal space, perpendicular to the wavevector $\mathbf{s}_0$ and passing through the endpoint of $\mathbf{s}$ (see Fig. 4.13). The distance of the plane from the center of the sphere corresponding to the current diffraction vector is the projection of $\mathbf{s}$ onto the direction of $\mathbf{s}_0$,

$$OD = \mathbf{s} \cdot \frac{\mathbf{s}_0}{|\mathbf{s}_0|} = \frac{1 - q^3\lambda}{\lambda} = \frac{\cos(2\theta)}{\lambda}, \tag{4.36}$$

which readily leads to an expression for the wavelength corresponding to the current diffraction vector. Indeed,

$$q^3\lambda = 1 - \cos(2\theta) = 2\sin^2\theta$$

and

$$q^3 = \frac{2\sin\theta}{\lambda}\sin\theta = |\mathbf{h}|\sin\theta.$$

On the other hand, we always have $|\mathbf{h}| = (2\sin\theta)/\lambda$. It follows that

$$\lambda = \frac{2q^3}{|\mathbf{h}|^2}. \tag{4.37}$$

The other components of the wavevector $\mathbf{s}$ are

$$AD = \mathbf{s} \cdot \mathbf{e}_1 = q^1 \text{ and } BD = \mathbf{s} \cdot \mathbf{e}_2 = q^2. \tag{4.38}$$

If the plane ABD is projected onto the flat detector, the coordinates of the diffraction spot, measured in centimeters from the center of the circular plate, can be obtained from the similar triangles shown in Fig. 4.13, as

$$X^1 = d\frac{AD}{OD} = d\frac{q^1\lambda}{1 - q^3\lambda} \tag{4.39}$$

and

$$X^2 = d\frac{BD}{OD} = d\frac{q^2\lambda}{1 - q^3\lambda}. \tag{4.40}$$

Given the orientation matrix $\mathbf{D}$, the coordinates of the diffraction spot can now be related to the diffraction indices $h_1h_2h_3$ for the wavelength given by eqn (4.37).

A possible practical realization of the above algorithm would be to mount a crystal on a four-circle diffractometer, obtain its orientation matrix and unit-cell dimensions, record the Laue pattern (either on the diffractometer or after transferring the crystal to a Laue device), and compare it with the pattern computed as indicated above.

4.9 Exercises for Chapter 4

1. The Laue equations can be written as

$$\mathbf{a}_1 \cdot \mathbf{h} = h_1, \qquad \mathbf{a}_2 \cdot \mathbf{h} = h_2, \qquad \mathbf{a}_3 \cdot \mathbf{h} = h_3$$

and the vector $\mathbf{h}$ then becomes

$$\mathbf{h} = h_1\mathbf{a}_1^* + h_2\mathbf{a}_2^* + h_3\mathbf{a}_3^*,$$

where $\mathbf{a}_1^*\mathbf{a}_2^*\mathbf{a}_3^*$ are basis vectors of the reciprocal lattice. However, the vector $\mathbf{h}$ can also be referred to the direct basis vectors $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$ as

$$\mathbf{h} = q_1\mathbf{a}_1 + q_2\mathbf{a}_2 + q_3\mathbf{a}_3.$$

Find and interpret the transformation matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ relating the coordinates of $\mathbf{h}$ in the above two representations in accordance with

$$h_i = \sum_{j=1}^{3} Q_{ij}q_j \quad \text{and} \quad q_i = \sum_{j=1}^{3} P_{ij}h_j.$$

2. A crystal was investigated by scientists I and II. Each of them assigned to it a different unit cell and the relation between the basis vectors they chose is:

$$\mathbf{a}_{\mathrm{II}} = \mathbf{a}_{\mathrm{I}} - \mathbf{b}_{\mathrm{I}}$$
$$\mathbf{b}_{\mathrm{II}} = \mathbf{a}_{\mathrm{I}} + \mathbf{b}_{\mathrm{I}}$$
$$\mathbf{c}_{\mathrm{II}} = \mathbf{c}_{\mathrm{I}}$$

What is the relation between the diffraction indices hkl that I and II assigned to the reflections they observed? Can this be generalized to any linear transformation of the basis bectors?

3. For a given crystal, consider the transformation

$$\mathbf{a}' = (\mathbf{b} + \mathbf{c})/2, \quad \mathbf{b}' = (\mathbf{c} + \mathbf{a})/2, \quad \mathbf{c}' = (\mathbf{a} + \mathbf{b})/2,$$

where **abc** are the basis vectors of a Bravais lattice of type F. Find the relation between the diffraction indices $h'k'l'$ and hkl and hence the condition for possible reflections hkl from this crystal. Show that the unit cell based on the vectors $\mathbf{a}'$, $\mathbf{b}'$, and $\mathbf{c}'$ is primitive.

(Note: The primed indices correspond to primed basis vectors, etc).

4. The volume of a primitive unit cell of a certain crystal is $V = 1564$ Å^3. What are the approximate total numbers of reflections which can be obtained from this crystal when it is irradiated with copper and molybdenum radiation?

(Assume that $\lambda(\mathrm{K}\alpha_{av})$ is being used.)

5 The structure factor and the electron density

5.1 Introduction

The purpose of this chapter is to introduce the reader to the model underlying the interpretation of X-ray diffraction intensities, which is the basis of the methods for crystal structure determination to be described in later chapters. The next section deals in some detail with a simple description of the physics of the scattering of X-rays. That section makes use of results provided by classical electrodynamics, and presents a derivation of the Thomson scattering cross section, which is most usually quoted without proof in textbooks of crystallography. We shall then turn to the construction of the structure factor as the Fourier transform of the electron density function within the unit cell, first as a Fourier integral and later as a sum of atomic contributions, and proceed to the electron density function which is reexpressed as a Fourier synthesis, the coefficients of which are the structure factors. The last section of this chapter introduces relationships between the space-group symmetry in the direct space and the symmetry of the weighted reciprocal lattice, and their application to the determination of lattice type, point-group symmetry, and non-symmorphic space-group symmetry elements of the crystal investigated.

5.2 Scattering by a free charged particle

X-rays belong to the spectrum of electromagnetic radiation, and those applied in crystallographic diffraction studies have wavelengths of the order of 1 Å(10^{-10} m). Their speed in vacuo is c, about 3×10^8 m/s, and thus their frequency is of the order of 10^{18} cycles per second. When an X-ray beam encounters a charged particle, for example, an electron or a proton, the rapidly alternating electric field of the X-ray wave forces the particle to execute an oscillatory motion, that is, to become an oscillating dipole. The particle is thus accelerated, and electrodynamic theory tells us that an accelerated charge emits energy in the form of electromagnetic radiation. The radiation reemitted by the oscillating charge has the same frequency as the incident radiation. Summing up, irradiation of charged particles by X-rays accelerates them and gives rise to a reemission of X-rays. This is known as the scattering of X-rays by charged particles. Let us now turn to a relevant result from classical electrodynamics. If a charge e of mass m is moving with velocity vector $\mathbf{u}$, and its acceleration vector is $\ddot{\mathbf{r}}$, the charge emits an electromagnetic wave, in which the radiation component of the

electric field has the form

$$\mathbf{E}_{\text{rad}} = \frac{e}{4\pi\epsilon_0 c^2 s^3}\left\{\mathbf{r}\times\left[\left(\mathbf{r}-\frac{|\mathbf{r}|\mathbf{u}}{c}\right)\times\ddot{\mathbf{r}}\right]\right\}, \tag{5.1}$$

where the constant ϵ_0 equals $10^7/(4\pi c^2)$ F/m in the SI system of units, c is the velocity of light, $\mathbf{r}$ is the position vector of the point charge at the time the wave was emitted, and $s = |\mathbf{r}| - \mathbf{u}\cdot\mathbf{r}/c$ (see, for example, Panofsky and Phillips 1956). For a slowly moving charge, relative to c, which is the case for an electron in a crystal, s reduces to $|\mathbf{r}|$, and the $|\mathbf{r}|\mathbf{u}/c$ term in eqn (5.1) can be neglected. We then have

$$\mathbf{E}_{\text{rad}} = \frac{e}{4\pi\epsilon_0 c^2|\mathbf{r}|^3}\left[\mathbf{r}\times(\mathbf{r}\times\ddot{\mathbf{r}})\right]. \tag{5.2}$$

If we now define the angle between $\mathbf{r}$ and $\ddot{\mathbf{r}}$ as α, the magnitude of the triple vector product in eqn (5.2) reduces to

$$\begin{aligned}|\mathbf{r}\times(\mathbf{r}\times\ddot{\mathbf{r}})| &= |\mathbf{r}||\mathbf{r}\times\ddot{\mathbf{r}}| \\ &= |\mathbf{r}|^2|\ddot{\mathbf{r}}|\sin\alpha\end{aligned}$$

and the magnitude of the radiation component of the electric field becomes

$$|\mathbf{E}_{\text{rad}}| = \frac{e|\ddot{\mathbf{r}}|\sin\alpha}{4\pi\epsilon_0 c^2|\mathbf{r}|}, \tag{5.3}$$

which can be evaluated in terms of $|\mathbf{r}|$ if the acceleration $|\ddot{\mathbf{r}}|$ is given. We need not repeat the above considerations for the radiation part of the magnetic field, since that effect can be neglected because of the assumed slow motion. The rate at which the reradiated energy crosses a unit area can be expressed as

$$|\mathbf{N}| = \sqrt{\frac{\epsilon_0}{\mu_0}}|\mathbf{E}_{\text{rad}}|^2, \tag{5.4}$$

where the constant μ_0 equals $4\pi\times10^{-7}$ H/m in the SI system of units.

Since we wish to obtain an expression for the acceleration, let us consider the equation of motion of an oscillating charge, say an electron. In the general case of harmonic oscillation we have the Newtonian acceleration, a damping term proportional to the velocity, a restoring force (the binding to the environment), and an external force, which in our case is the electric field of the incident electromagetic wave. This equation of motion will have the form

$$\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + \omega_0^2\mathbf{r} = \frac{e}{m}\mathbf{E}_0\exp(i\omega t) \tag{5.5}$$

and we seek a solution of the form $\mathbf{x} = \mathbf{x}_0\exp(i\omega t)$.Substitution of such a solution in eqn (5.5) gives

$$\mathbf{r} = \frac{1}{\omega_0^2 - \omega^2 + i\gamma\omega}\frac{e}{m}\mathbf{E}, \tag{5.6}$$

where $\mathbf{E} = \mathbf{E}_0 \exp(i\omega t)$ is the electric field of the incoming wave. The acceleration is

$$\ddot{\mathbf{r}} = \frac{-\omega^2}{\omega_0^2 - \omega^2 + i\gamma\omega} \frac{e}{m} \mathbf{E}, \tag{5.7}$$

and its magnitude can be used with eqns (5.3) and (5.4). This is of relevance when the presence of dispersive scatterers must be considered. However, in many cases of practical importance it is possible to assume that the electrons are weakly bound (that is, $\omega_0 \ll \omega$) and appreciable damping is absent. This reduces the equation of motion to that for a free electron, and the acceleration for this simple case is

$$\ddot{\mathbf{r}} = \frac{e}{m} \mathbf{E}. \tag{5.8}$$

If we substitute the magnitude of the acceleration obtained from eqn (5.8) in eqn (5.3), we obtain for eqn (5.4)

$$|\mathbf{N}| = \sqrt{\frac{\epsilon_0}{\mu_0}} \left[\frac{e^2 E \sin\alpha}{4\pi\epsilon_0 m c^2 r} \right]^2 \tag{5.9}$$

$$\equiv r_0^2 \sin^2\alpha \left(\frac{I_0}{r^2} \right), \tag{5.10}$$

where

$$r_0 = \frac{e^2}{4\pi\epsilon_0 m c^2} \tag{5.11}$$

is the classical electron radius (2.8178×10^{-15} m) and

$$I_0 = \sqrt{\frac{\epsilon_0}{\mu_0}} E^2$$

is the intensity of the incident radiation.

We see from eqns (5.10) and (5.11) that the reradiated or scattered power is inversely proportional to m^2. It follows that the power radiated by a proton will be $(1837)^2$ times smaller than that radiated by an electron, and can be safely neglected; indeed, only the electrons are of importance in the present case.

Figure 5.1 shows schematically the quantities of interest. A free electron is located at the origin of a Cartesian coordinate system $\xi\eta\zeta$. The plane of the electromagnetic wave propagates so that its normal is parallel to ζ, and is indicated here while passing through the origin (that is, interacting with the electron) and coinciding with the $\xi\eta$ plane. The end of the vector $\mathbf{r}_-$ denotes the field point, at which the radiation part of the electric field of the scattered wave is measured. The system is oriented so that $\mathbf{r}_-$ lies in the $\xi\zeta$ plane, and forms an angle 2θ with ζ. The electric field of the incident wave, $\mathbf{E}$, lies in the $\xi\eta$ plane and forms an angle φ with the ξ axis. The angles φ and 2θ are called the polarization and scattering angles, respectively. The angle between $\mathbf{r}_-$ and $\mathbf{E}$ is denoted by α (note that the direction of the acceleration vector is the same as that of the vector of

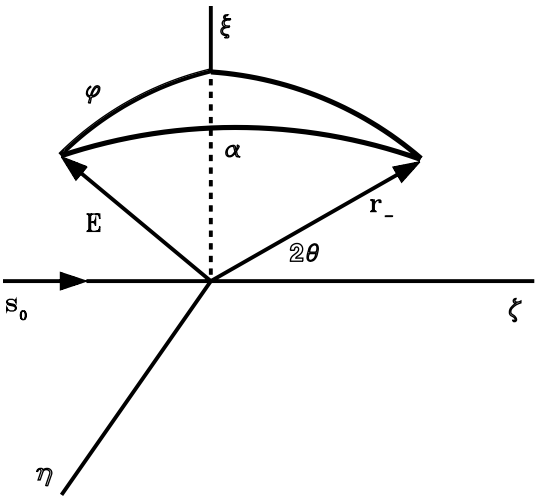


Fig. 5.1 Scattering by a free electron.

the electric field in the incident wave). The angles α, φ, and $90^\circ - 2\theta$ subtend a right-angled spherical triangle, for which the following relation is valid:

$$\cos\alpha = \cos\varphi \sin 2\theta.$$

It follows that

$$\sin^2\alpha = 1 - \cos^2\varphi(1 - \cos^2 2\theta).$$

In the usual experimental arrangements the incident wave is unpolarized, that is, the electric field vector may assume any orientation in the plane of the wave. We must then average $\sin^2\alpha$ over all values of φ. We obtain

$$\langle \sin^2\alpha \rangle_\varphi = \frac{1 + \cos^2 2\theta}{2}, \tag{5.12}$$

since the average of $\cos^2\varphi$ is $1/2$. The right-hand side of eqn (5.12) is called the *polarization correction*, which is routinely applied to the θ-dependent intensity data. The polarization correction may, of course, take a different functional form if the electric field of the incident wave is partially polarized. We shall not deal with such corrections in this book.

The intensity scattered into a unit solid angle at an angle 2θ, on the assumption that the incident wave is unpolarized, is given by

$$I(2\theta) = I_0 \frac{r_0^2}{r^2} \frac{1 + \cos^2 2\theta}{2}. \tag{5.13}$$

If we want to obtain the total scattered power, we have to integrate eqn (5.13) over the surface of a sphere of radius r and centered at the origin, as follows:

$$I_{\text{scatt}} = I_0 \frac{r_0^2}{r^2} \int_0^{\pi} \frac{1+\cos^2 2\theta}{2} 2\pi r^2 \sin 2\theta \, \mathrm{d}(2\theta)$$
$$= 2\pi I_0 r_0^2 \int_0^{\pi} \frac{(1+\cos^2 2\theta)}{2} \sin 2\theta \, \mathrm{d}(2\theta)$$
$$= \frac{8\pi I_0 r_0^2}{3}. \tag{5.14}$$

If we now divide the total integrated power by the intensity of the incident radiation, we obtain the scattering cross section per electron,

$$\sigma_0 = \frac{8\pi r_0^2}{3} = \frac{e^4}{6\pi\epsilon_0^2 m^2 c^4}. \tag{5.15}$$

Equation (5.14) is known as the Thomson scattering formula, and eqn (5.15) is called the cross section for Thomson scattering. The value of this cross section is about 6.65×10^{-29} m^2 per electron. It is essential for the calculation of absolute values of scattered intensity (along with corrections for quantum effects, such as Compton scattering). However, we shall try to express the scattering from a crystal in terms of the scattering from a single classical electron at the origin, since such relative intensities are sufficient for our purpose.

5.3 Scattering by a material unit

Let us now try to find the wave scattered by a distribution of electron density within a single material unit. This unit can be an atom, a molecule, or a group of molecules of arbitrary complexity. We shall assume that the unit, as illustrated schematically in Fig. 5.2, (i) is kept fixed with respect to the incoming radiation, and (ii) has a well-defined function of electron density (number of electrons per unit volume, the unit being chosen here as Å^3) at each point. We shall also assume that the process of scattering is not accompanied by any loss of energy (that is, elastic scattering is assumed), and define the wavevectors of the incident and scattered radiation as $\mathbf{s}_0$ and $\mathbf{s}$ respectively, where $|\mathbf{s}_0| = |\mathbf{s}| = 1/\lambda$ and λ is the wavelength of the radiation. Finally, all the scattering will be taken as relative to the scattering by a single classical electron located at the origin, and consideration of the detailed mechanism of scattering will therefore not be needed. We denote the electron density at the position $\mathbf{r}$ by $\rho(\mathbf{r})$ and proceed to relate the wave scattered from the volume element $\mathrm{d}V$ to that scattered from a classical electron at the origin O. The relative amplitude of the wave scattered from $\mathrm{d}V$ is just the number of electrons in this volume element, this number being given by $\rho(\mathbf{r})\,\mathrm{d}V$. The spatial part of the phase of the wave scattered from the electron content of dV, relative to that scattered from the electron at the origin, depends on the difference between the optical paths traversed by the two waves. Since the relation between the phase difference φ and difference of optical paths Δ in any given direction is

$$\frac{\varphi}{2\pi} = \frac{\Delta}{\lambda}$$

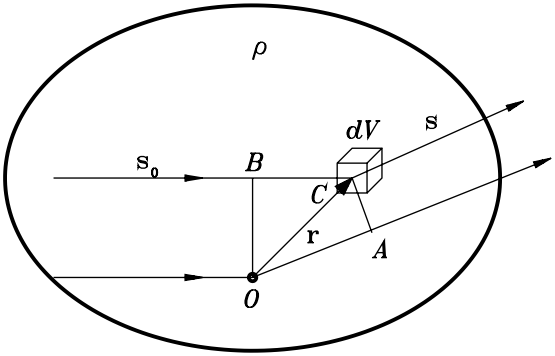


Fig. 5.2 Scattering by a material unit.

the wave scattered from the density distribution within the volume element $\mathrm{d}V$ relative to that scattered from the reference electron at the origin is

$$\mathrm{d}F(\Delta) = \rho(\mathbf{r})\,\mathrm{d}V \exp\left(\frac{2\pi i\Delta}{\lambda}\right). \tag{5.16}$$

The time-dependent part of the phase, $\exp(i\omega t)$, has been omitted, since the frequencies of the incident and scattered waves are the same. The path difference Δ, is given (see Fig. 5.2) by

$$\begin{aligned}\Delta &= OA - BC \\ &= \mathbf{r}\cdot\frac{\mathbf{s}}{|\mathbf{s}|} - \mathbf{r}\cdot\frac{\mathbf{s}_0}{|\mathbf{s}_0|} \\ &= \lambda\mathbf{r}\cdot(\mathbf{s}-\mathbf{s}_0),\end{aligned} \tag{5.17}$$

and we can now express the wave in terms of the field point and wavevectors as

$$\mathrm{d}F(\mathbf{h}) = \rho(\mathbf{r})\,\mathrm{d}V \exp(2\pi i\mathbf{h}\cdot\mathbf{r}), \tag{5.18}$$

where the abbreviation $\mathbf{h} = \mathbf{s} - \mathbf{s}_0$ is called the *scattering vector* or *diffraction vector* of the wave. The wave scattered from the whole material unit is now obtained by integrating both sides of eqn (5.18) over the volume occupied by the material unit, that is, over the whole three-dimensional space. We obtain

$$F(\mathbf{h}) = \int \rho(\mathbf{r}) \exp(2\pi i\mathbf{h}\cdot\mathbf{r})\,\mathrm{d}V. \tag{5.19}$$

The wave scattered from the unit and corresponding to the diffraction vector $\mathbf{h}$ turns out to be the Fourier transform of the electron density function that defines the structure of the unit. The function $F(\mathbf{h})$ is therefore called the structure factor corresponding to the diffraction vector $\mathbf{h}$. Of course, the structure of the unit

could in principle be specified if the values of the function $F(\mathbf{h})$ were available in all of the $\mathbf{h}$ space, since the electron density $\rho(\mathbf{r})$ could then be obtained as an inverse Fourier transform of $F(\mathbf{h})$. There are, however, two main reasons that prevent us from carrying out a scattering experiment on a single material unit which would lead to its complete structure determination:

1. Our requirement of a well-defined electron density function within the unit examined is not consistent with the chaotic motion that isolated molecular units execute in the gaseous or liquid state. We can then only think of an average electron density over a vast range of configurations, which is not likely to be useful. We therefore need a device which keeps the material unit fixed or very nearly so.
2. The scattering cross section for Thomson scattering, as discussed in Section 5.2, is of the order of 10^{-29} m^2 per electron. The intensity scattered from a single unit is unlikely to be detectable unless the intensity of the incident radiation is very high and the unit is is very large (possesses many electrons). These requirements have been made less severe by the availability of powerful synchrotron radiation.

A good solution to the above two problems is provided by a single crystal composed of such material units, playing the role of its fundamental structural units. These units are arranged on a triply periodic array and, owing to their closely packed arrangement, maintain closely similar orientations, or a periodic electron density function.

5.4 Scattering by a periodic array of material units

We must, however, consider the effect that a triply periodic arrangement of identical units has on the pattern of the diffracted radiation. Let us first take two parallel identical units, say α and β, with their centers of gravity separated by the vector $\mathbf{a}$. Thus, the unit β can be generated from the unit α by a translation by $\mathbf{a}$, parallel to itself. In terms of their electron densities, we have

$$\rho_\alpha(\mathbf{r}) = \rho_\beta(\mathbf{r}+\mathbf{a}) \equiv \rho(\mathbf{r}),$$

where the origin is fixed in the unit α. Assuming the same geometry of the wavevectors as in Fig. 5.2, the wave scattered from the pair of units is given by

$$F_{\alpha+\beta}(\mathbf{h}) = \int \rho(\mathbf{r}) \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\,\mathrm{d}V + \int \rho(\mathbf{r}) \exp[2\pi i \mathbf{h}\cdot(\mathbf{r}+\mathbf{a})]\,\mathrm{d}V \quad (5.20)$$

$$= F(\mathbf{h})[1+\exp(2\pi i \mathbf{h}\cdot\mathbf{a})]. \quad (5.21)$$

The intensity of the radiation scattered from the two units is proportional to the squared magnitude of their combined structure factor, that is,

$$\begin{aligned} I_{\alpha+\beta}(\mathbf{h}) &\propto |F_{\alpha+\beta}(\mathbf{h})|^2 \\ &= |F(\mathbf{h})|^2[1+\exp(2\pi i \mathbf{h}\cdot\mathbf{a})][1+\exp(-2\pi i \mathbf{h}\cdot\mathbf{a})] \\ &= 2|F(\mathbf{h})|^2[1+\cos(2\pi \mathbf{h}\cdot\mathbf{a})]. \end{aligned} \quad (5.22)$$

We see that, independently of the structure of the scattering material unit, the intensity attains its maximal value when $\mathbf{h}\cdot\mathbf{a}$ is an integer, and vanishes when $\mathbf{h}\cdot\mathbf{a}$ is an odd multiple of one-half. It attains values between zero and $4|F(\mathbf{h})|^2$ for other values of the scalar product $\mathbf{h}\cdot\mathbf{a}$. These are examples of constructive and destructive interference.

Let us now consider the intensity of radiation scattered from a row of N_a identical, parallel material units, in which the electron density is periodic, with a period $\mathbf{a}$,

$$\rho(\mathbf{r}) = \rho(\mathbf{r} + u\mathbf{a}),$$

where $0 \le u \le N_a - 1$. The structure factor for the diffraction vector $\mathbf{h}$ for this chain of units is

$$\begin{aligned} F_a(\mathbf{h}) &= \sum_{u=0}^{N_a-1} \left\{ \int \rho(\mathbf{r}) \exp[2\pi i \mathbf{h}\cdot(\mathbf{r}+u\mathbf{a})]\,\mathrm{d}V \right\} \\ &= F(\mathbf{h}) \sum_{u=0}^{N_a-1} \exp(2\pi i u \mathbf{h}\cdot\mathbf{a}) \qquad (5.23) \\ &= F(\mathbf{h}) \frac{1-\exp(2\pi i N_a \mathbf{h}\cdot\mathbf{a})}{1-\exp(2\pi i \mathbf{h}\cdot\mathbf{a})}, \qquad (5.24) \end{aligned}$$

since the sum in eqn (5.23) is that of a geometric series in which the first term is 1, the number of terms is N_a and the ratio between successive terms is $\exp(2\pi i \mathbf{h}\cdot\mathbf{a})$. The intensity scattered from the above chain follows as

$$\begin{aligned} I_a(\mathbf{h}) &\propto |F(\mathbf{h})|^2 \frac{1-\exp(2\pi i N_a \mathbf{h}\cdot\mathbf{a})}{1-\exp(2\pi i \mathbf{h}\cdot\mathbf{a})} \frac{1-\exp(-2\pi i N_a \mathbf{h}\cdot\mathbf{a})}{1-\exp(-2\pi i \mathbf{h}\cdot\mathbf{a})} \\ &= |F(\mathbf{h})|^2 \frac{1-\cos(2\pi N_a \mathbf{h}\cdot\mathbf{a})}{1-\cos(2\pi \mathbf{h}\cdot\mathbf{a})} \\ &= |F(\mathbf{h})|^2 \frac{\sin^2(\pi N_a \mathbf{h}\cdot\mathbf{a})}{\sin^2(\pi \mathbf{h}\cdot\mathbf{a})}. \qquad (5.25) \end{aligned}$$

The function of the form $\sin^2 Nx/\sin^2 x$ which appears in eqn (5.25) is often encountered in problems of diffraction from periodic arrays, and is called the interference function, since it accounts for interference effects that are due to the periodicity of the diffracting medium. We present a graph of it in Fig. 5.3, for a modest value of $N = 10$. It is seen that the function attains its largest maxima for the values of x which are integer multiples of π, and has small subsidiary maxima within its period. As N increases, the largest subsidiary maxima move towards the main ones, and for very large N it becomes a comb function with nonzero values at $x = n\pi$ only. If we consider eqn (5.25), we see that the intensity may be nonzero only if the scalar product $\mathbf{h}\cdot\mathbf{a}$ is an integer, and must vanish if it is not. The values of the interference function at $x = n\pi$ cannot be obtained by direct substitution, but follow readily by applying l'Hôpital's rule twice:

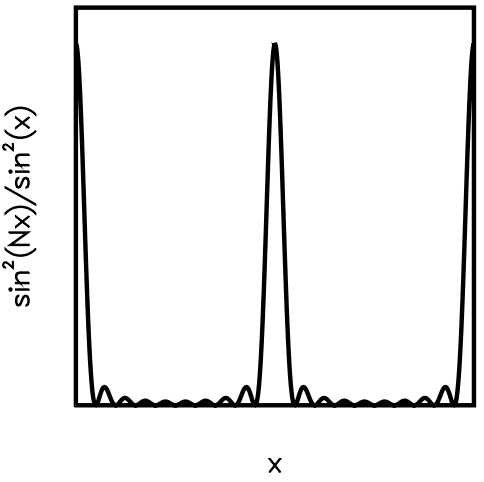


Fig. 5.3 The interference function.

$$\begin{aligned}\lim_{x\to n\pi}\frac{\sin^2 Nx}{\sin^2 x} &= \lim_{x\to n\pi}\frac{2N\sin Nx\cos Nx}{2\sin x\cos x}\\ &= \lim_{x\to n\pi}\frac{2N^2\cos^2 Nx - 2N^2\sin^2 Nx}{2\cos^2 x - 2\sin^2 x}\\ &= N^2.\end{aligned}$$

Equation (5.25) now becomes

$$I_a(\mathbf{h}) \propto \begin{cases} N_a^2|F_a(\mathbf{h})|^2 & \text{if } \mathbf{h}\cdot\mathbf{a} = n \\ 0 & \text{if } \mathbf{h}\cdot\mathbf{a} \neq n \end{cases} \tag{5.26}$$

where n is an integer. If we now consider a block of crystalline material containing $N_aN_bN_c$ unit cells, where N_a, N_b, and N_c are the numbers of cells along **a**, **b** and **c**, respectively, a closely similar calculation to that leading to eqn (5.25) gives the result

$$I(\mathbf{h}) \propto |F(\mathbf{h})|^2 \frac{\sin^2(\pi N_a\mathbf{h}\cdot\mathbf{a})}{\sin^2(\pi\mathbf{h}\cdot\mathbf{a})}\frac{\sin^2(\pi N_b\mathbf{h}\cdot\mathbf{b})}{\sin^2(\pi\mathbf{h}\cdot\mathbf{b})}\frac{\sin^2(\pi N_c\mathbf{h}\cdot\mathbf{c})}{\sin^2(\pi\mathbf{h}\cdot\mathbf{c})}. \tag{5.27}$$

Since each of the three interference functions in eqn (5.27) leads to a condition analgous to that expressed in eqn (5.26), their product must require that the three conditions be simultaneously satisfied. We thus have for the intensity diffracted by a crystal

$$I(\mathbf{h}) \propto |F(\mathbf{h})|^2, \tag{5.28}$$

and this intensity may be nonzero only if the equations

$$\begin{aligned} \mathbf{h} \cdot \mathbf{a} &= h, \\ \mathbf{h} \cdot \mathbf{b} &= k, \\ \mathbf{h} \cdot \mathbf{c} &= l, \end{aligned} \tag{5.29}$$

where h, k and l are integers, are simultaneously satisfied. As we already know from Chapter 4, eqns (5.29) are the Laue equations, formulated by Laue during the interpretation of the first diffraction pattern obtained from a crystal (Laue 1912, 1914).

Note that the intensity of the diffracted radiation is proportional to that diffracted by a single material unit, independently of the number of units involved. The proportionality factor preceding the squared magnitude of the structure factor appears to include the squared number of unit cells in the crystal. This is, however, true only up to a certain limit. Matter also absorbs electromagnetic radiation, and when the thickness of the absorber exceeds a certain critical dimension the scattered intensity starts decreasing, rather than increasing as would be indicated by the N^2 proportionality. In fact, corrections for X-ray absorption are needed in many instances and are an important part of the reduction of intensity data. Methods for doing this were recently reviewed by Maslen (1999).

5.5 The atomic scattering factor

We have obtained the structure factor as a Fourier transform of the electron density, as given by eqn (5.19), which we rewrite for convenience as

$$F(\mathbf{h}) = \int \rho(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}) \, \mathrm{d}V. \tag{5.30}$$

This is an idealization, since we have not allowed for possible small oscillations of the atoms about their equilibrium positions, nor for departures from perfect periodicity or other kinds of disorder that are present in real structures. We shall deal with these topics in later chapters. The first simplification of the above important expression is obtained by assuming that the electron density at the point $\mathbf{r}$ is a superposition of atomic electron densities, each of which is centered at the mean position $\mathbf{r}_j$ of the jth atom. This can be expressed as

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} \rho_j(\mathbf{r} - \mathbf{r}_j), \tag{5.31}$$

where N is the number of atoms in the unit cell. If we now introduce the variable $\mathbf{u} = \mathbf{r} - \mathbf{r}_j$, insert eqn (5.31) into eqn (5.30), and interchange the order of the integration and summation, we obtain

$$F(\mathbf{h}) = \sum_{j=1}^{N} \left\{ \int \rho_j(\mathbf{u}) \exp(2\pi i \mathbf{h}\cdot\mathbf{u})\, \mathrm{d}V \right\} \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_j) \tag{5.32}$$

$$\equiv \sum_{j=1}^{N} f_j(\mathbf{h}) \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_j), \tag{5.33}$$

where $f_j(\mathbf{h})$ is called the scattering factor of the jth atom. This factor is assumed to contain the effect of atomic displacements. We shall now assume that the atomic electron density possesses spherical symmetry. Although this assumption appears to be a strong one, especially for covalently bonded light atoms, whose bonding electron density departs significantly from sphericity, it is sufficient in order to determine an essentially correct structure of the crystal. Once this is accomplished, the final stages of the refinement of the structural parameters allow one to take account of some features of the bonding density. There are several approaches to this problem, one of them being reviewed by Coppens, Su, and Becker (1999).

The above assumption leads to a considerable simplification of the integral in the right-hand side of eqn (5.32). Spherical symmetry of the atomic density means that $\rho(\mathbf{u})$ approximates to $\rho(u)$, and if we choose $\mathbf{h}$ as the polar axis of the system of spherical coordinates we have

$$f_j = \int \rho_j(\mathbf{u}) \exp(2\pi i \mathbf{h}\cdot\mathbf{u})\, \mathrm{d}V \tag{5.34}$$

$$\simeq \int_0^\infty \int_0^{2\pi} \int_0^{\pi} \rho_j(u) \exp(2\pi i |\mathbf{h}| u \cos\vartheta) u^2 \sin\vartheta \,\mathrm{d}\vartheta \,\mathrm{d}\varphi \,\mathrm{d}u$$

$$= 2\pi \int_0^\infty u^2 \rho_j(u) \left\{ \int_0^{\pi} \exp(2\pi i |\mathbf{h}| u \cos\vartheta) \sin\vartheta \,\mathrm{d}\vartheta \right\} \mathrm{d}u$$

$$= 2\pi \int_0^\infty u^2 \rho_j(u) \left\{ \frac{1}{2\pi|\mathbf{h}|u} \int_{-2\pi|\mathbf{h}|u}^{2\pi|\mathbf{h}|u} \exp(ix)\, \mathrm{d}x \right\} \mathrm{d}u$$

$$= 4\pi \int_0^\infty u^2 \rho_j(u) \frac{\sin(2\pi|\mathbf{h}|u)}{2\pi|\mathbf{h}|u}\, \mathrm{d}u. \tag{5.35}$$

Tables of atomic scattering factors, presented as functions of $|\mathbf{h}|/2 = \sin\theta/\lambda$, have been computed from eqn (5.35) for all the elements and most ions by quantum mechanical methods and can be found in Volume C of the *International Tables for Crystallography* (Wilson 1992). Scattering factors of several elements have also been determined experimentally from absolute intensity measurements (for example, James 1965).

Note that the value of the atomic scattering factor for $\mathbf{h} = \mathbf{0}$, whether given by the exact eqn (5.34) or by the approximate eqn (5.35), equals exactly the number of the electrons in the atom, or the atomic number, Z.

It is useful, for many practical purposes, to represent the structure factor in terms of the components of the vectors involved. The mean position vector of the jth atom is given by $\mathbf{r}_j = x_j\mathbf{a} + y_j\mathbf{b} + z_j\mathbf{c}$, and the scalar product $\mathbf{h} \cdot \mathbf{r}_j$ appearing in eqn (5.33) is therefore

$$\mathbf{h} \cdot \mathbf{r}_j = x_j(\mathbf{h} \cdot \mathbf{a}) + y_j(\mathbf{h} \cdot \mathbf{b}) + z_j\mathbf{h} \cdot \mathbf{c}.$$

If we now make use of the Laue equations, eqns (5.29), the scalar product reduces to

$$\mathbf{h} \cdot \mathbf{r}_j = hx_j + ky_j + lz_j.$$

Of course, the same result would be obtained if we expressed $\mathbf{h}$ in terms of the basis of the reciprocal lattice, and evaluated the scalar product in a direct manner. Since the diffraction vector $\mathbf{h}$ is uniquely associated with the triplet of integers hkl appearing in the Laue equations, we can rewrite the structure factor, in the spherical-atom approximation, as

$$F(hkl) = \sum_{j=1}^{N} f_j \left(\frac{\sin\theta}{\lambda}\right) \exp[2\pi i(hx_j + ky_j + lz_j)]. \tag{5.36}$$

5.6 Space-group symmetry and the structure factor

Incorporation of space-group symmetry into the expression for the structure factor leads to simplifications in most stages of crystallographic computing. Let us first view the process in rather general terms, and then conclude with some examples.

Suppose that the crystal belongs to a P-type Bravais lattice, and all the atoms are located in general positions of the space group the representative elements of which are $\{(\boldsymbol{P}_i, \boldsymbol{t}_i)\}$, $i = 1, \ldots, g$. Here, $(\boldsymbol{P}_i, \boldsymbol{t}_i)$ is the (matrix, vector) representation of the ith space-group operator. Let $\boldsymbol{x}_j$, $j = 1, \ldots, N$ be the column vectors of the atomic coordinates of the N atoms within the unit cell and let $\boldsymbol{h}^{\mathrm{T}}$ denote a row matrix of diffraction indices (hkl). The general form of the structure factor is

$$F(\boldsymbol{h}) = \sum_{j=1}^{N} f_j \exp(2\pi i \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}_j). \tag{5.37}$$

In view of the above, the unit cell contains g asymmetric units, with N/g atoms in each of them. There are also N/g possibly different atomic scattering factors. Equation (5.37) can therefore be rewritten as follows:

$$F(\boldsymbol{h}) = \sum_{j=1}^{N/g} f_j \sum_{i=1}^{g} \exp[2\pi i \boldsymbol{h}^{\mathrm{T}}(\mathbf{P}_i \boldsymbol{x}_j + \boldsymbol{t}_i)]. \tag{5.38}$$

The inputs required for the computation of such a structure factor, for any space group, are (i) the coordinates of the atoms within the asymmetric unit, (ii) the

Table 5.1 Some simplified expressions for trigonometric structure factors

Space group	Trigonometric structure factor	Parity
$P\overline{1}$	$2\cos[2\pi(hx+ky+lz)]$	
$P2_1/c$	$4\cos[2\pi(hx+lz)]\cos(2\pi ky)$	$k+l=2n$
	$-4\sin[2\pi(hx+lz)]\sin(2\pi ky)$	$k+l=2n+1$
$Pbcn$	$8\cos(2\pi hx)\cos(2\pi ky)\cos(2\pi lz)$	$h+k=2n; l=2n$
	$-8\cos(2\pi hx)\sin(2\pi ky)\sin(2\pi lz)$	$h+k=2n; l=2n+1$
	$-8\sin(2\pi hx)\cos(2\pi ky)\sin(2\pi lz)$	$h+k=2n+1; l=2n$
	$-8\sin(2\pi hx)\sin(2\pi ky)\cos(2\pi lz)$	$h+k=2n+1; l=2n+1$

atomic scattering factors for these N/g atoms (with appropriate displacement parameters), and (iii) the g space-group symmetry operators. The latter can be supplied as matrix-column items or as symbolic coordinates of the general equivalent positions, or computed from a computer-readable space-group symbol (see Chapter 3).

Another method is based on the use of a tabulation of the inner summation in eqn (5.38), known as the trigonometric structure factor. This is of use in cases of high symmetry (large values of g) and is indispensable if an analytic expression for the symmetry-dependent structure factor is required. Trigonometric structure factors are tabulated in Volume I of the *International Tables for X-ray Crystallography* (Henry and Lonsdale 1952, 1965) and in Volume B of *International Tables for Crystallography* (Shmueli 1993a, 2001a) for all 230 space groups and are valid for atoms in general positions and for general reflections. Some simplified expressions for trigonometric structure factors are shown in Table 5.1.

Only centrosymmetric cases are given in Table 5.1. For noncentrosymmetric space groups we would expect the real and imaginary parts of the trigonometric structure factor to be listed. It can also be seen that the functional form of the trigonometric structure factor depends on the parity of the reflection indices involved. This is so, however, only for space groups containing screw axes and/or glide planes, that is, for non-symmorphic space groups.

5.7 The Fourier synthesis of the electron density

One of the principal aims of our study is to elucidate the internal structure of the repeating material unit. With X-ray diffraction, this internal structure is given by the features of the electron density function within this unit. We can express the structure factor as a Fourier transform of the electron density, but the calculation of the inverse transform is not straightforward, since we do not measure the (generally complex) structure-factor function but, rather, only its magnitude. As we shall see in later chapters, this problem can be solved, and let us in the meantime prepare the formalism that has to be used once the

values of the structure-factor function have been retrieved from their observed magnitudes. The electron density function in the crystal is triply periodic, and as such it can be expanded in a triple Fourier series:

$$\rho(xyz) = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} C_{pqr} \exp[-2\pi i(px + qy + rz)], \tag{5.39}$$

where xyz are the fractional coordinates of the point at which the electron density is to be calculated. We wish to substitute eqn (5.39) in the structure-factor equation, eqn (5.19), and carry out the integration, but we must first represent eqn (5.19) as a triple integral over the variables xyz. We choose the range of the integration as a single primitive unit cell, which we know to contain just one repeating material unit. The volume element $\mathrm{d}V$ is chosen as an infinitesimal parallelepiped, constructed from the vectors $\mathrm{d}\mathbf{a}$, $\mathrm{d}\mathbf{b}$ and $\mathrm{d}\mathbf{c}$. We thus have: $\mathrm{d}V = \mathrm{d}\mathbf{a} \cdot (\mathrm{d}\mathbf{b} \times \mathrm{d}\mathbf{c}) = [\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})]\, dx\, dy\, dz = V\, dx\, dy\, dz$, where V is the volume of the unit cell. The scalar product $\mathbf{h} \cdot \mathbf{r}$ in eqn (5.19) is given by $hx + ky + lz$ (see the derivation of eqn (5.36)) and the structure factor can therefore be written as

$$F(hkl) = V \int_0^1 \int_0^1 \int_0^1 \rho(xyz) \exp[2\pi i(hx + ky + lz)]\, dx\, dy\, dz. \tag{5.40}$$

We now substitute eqn (5.39) into eqn (5.40), and interchange the order of the summations and the integrations:

$$F(hkl) = V \sum_p \sum_q \sum_r C_{pqr} \times \int_0^1 \exp[2\pi i(h-p)x]\, dx \int_0^1 \exp[2\pi i(k-q)y]\, dy \int_0^1 \exp[2\pi i(l-r)z]\, dz. \tag{5.41}$$

It is easily seen that each of the integrals in eqn (5.41) equals unity when the argument of the exponential is zero (that is, when $h = p, k = q$, or $l = r$) and vanishes when the argument of the exponential is nonzero. Therefore

$$F(hkl) = V \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} C_{pqr} \delta_{hp} \delta_{kq} \delta_{lr} \tag{5.42}$$

$$= V C_{hkl}, \tag{5.43}$$

where δ_{ij} is the Kronecker delta which equals 1 or 0 according as $i = j$ or $i \neq j$, respectively. Equation (5.39) can now be rewritten as

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp[-2\pi i(hx + ky + lz)]. \tag{5.44}$$

Equation (5.44) is of fundamental importance in reconstructing the structure of the unit cell from crystallographic information. We should point out, however,

that while in theory the summation indices hkl should range from minus infinity to infinity, they never do so in practice, predominantly because the range of indices is restricted by the wavelength of the radiation employed.

5.7.1 Computational considerations

The computation of Fourier series used to be a very laborious task, prior to the discovery of the Fast Fourier Transform technique. This was true for crystallography, as well as for other sciences which make use of this tool. Therefore, major effort was directed towards simplifications of eqn (5.44), and these are described in great detail in the crystallographic literature. We shall mention these simplifications briefly at this point, with appropriate references, and present an elementary introduction to the Fast Fourier Transform technique in Appendix D. The oldest simplification, still in some use for space groups of high symmetry, is the explicit introduction of space-group symmetry into the functional form of the electron density. These simplified forms are described, for example, by Lipson and Cochran (1966), in Volume I of the *International Tables for X-ray Crystallography* (Henry and Lonsdale 1952, 1965) and in Volume B of the *International Tables for Crystallography* (Shmueli 1993a, 2001a), and the relevant symmetry-dependent coefficients are tabulated for all the three-dimensional space groups in these volumes of *International Tables*.

Another simplification of eqn (5.44) is a reduction of it to a succession of one-dimensional summations (Lipson and Cochran 1966). This is described in detail in that reference for a summation with real terms, and can be hinted at as follows. If we omit the limits, eqn (5.44) can be written as

$$\rho(xyz) = \frac{1}{V} \sum_l \left[\sum_k \left\{ \sum_h F(hkl) \exp(-2\pi ihx) \right\} \exp(-2\pi iky) \right] \exp(-2\pi ilz). \tag{5.45}$$

For given indices k and l, the one-dimensional summation in the braces is a function of x, k, and l, say $C(x,k,l)$. Equation (5.45) then becomes

$$\rho(xyz) = \frac{1}{V} \sum_l \left[\sum_k C(x,k,l) \exp(-2\pi iky) \right] \exp(-2\pi ilz). \tag{5.46}$$

For given l, the one-dimensional summation in brackets is a function of x, y and l, say $D(x,y,l)$. Equation (5.46) finally reduces to the one-dimensional summation

$$\rho(xyz) = \frac{1}{V} \sum_l D(x,y,l) \exp(-2\pi ilz). \tag{5.47}$$

Of course, there are many one-dimensional summations that have to be carried out, but it is clear that *only* one-dimensional summations have to be computed.

The problem now reduces to the computation of a one-dimensional summation. An ingenious method of doing this was proposed by Beevers and Lipson (1936), and is described in detail by Lipson and Cochran (1966). However, this

has been superseded by the Fast Fourier Transform technique, which quite drastically reduces the computing time. The principle of this powerful technique is to separate a one-dimensional summation into several short ones, thereby complicating the functional form of the summation but reducing the number of the operations required. We present in Appendix D a brief introduction to the discrete Fourier transform and show how it leads to the Fast Fourier Transform method. The rich underlying theory of Fourier methods (see, for example, Bricogne 2001) is outside the scope of this book.

5.8 The phase problem

We have already indicated that the experiment furnishes a quantity proportional to the magnitude of the structure factor, whereas the previous section shows that it is the structure factor itself which is needed in order to evaluate the electron density within the unit cell of the crystal. The relation between the *available* $|F(\boldsymbol{h})|$ and the *desired* $F(\boldsymbol{h})$ can be usefully formulated as follows.

The general form of the structure factor can be written as

$$\begin{aligned} F(\boldsymbol{h}) &= \sum_{j=1}^{N} f_j \exp(2\pi i \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}_j) \\ &= \sum_{j=1}^{N} f_j [\cos(2\pi \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}_j) + i \sin(2\pi \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}_j)] \\ &\equiv A(\boldsymbol{h}) + iB(\boldsymbol{h}), \end{aligned}$$

where

$$A(\boldsymbol{h}) = \sum_{j=1}^{N} f_j \cos(2\pi \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}_j)$$

and

$$B(\boldsymbol{h}) = \sum_{j=1}^{N} f_j \sin(2\pi \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}_j)$$

are the real and imaginary parts, respectively, of the (generally) complex number $F(\boldsymbol{h})$. If we think of $F(\boldsymbol{h})$ as a vector in the complex plane, its squared length (related to the observed intensity) is

$$|F(\boldsymbol{h})|^2 = F(\boldsymbol{h})F^*(\boldsymbol{h}) = [A(\boldsymbol{h}) + iB(\boldsymbol{h})][A(\boldsymbol{h}) - iB(\boldsymbol{h})] = A(\boldsymbol{h})^2 + B(\boldsymbol{h})^2.$$

The phase of $F(\boldsymbol{h})$, which we shall denote by $\varphi(\boldsymbol{h})$, is defined as the angle between $F(\boldsymbol{h})$ and the real axis or between $F(\boldsymbol{h})$ and $A(\boldsymbol{h})$. $A(\boldsymbol{h})$, $B(\boldsymbol{h})$, and $F(\boldsymbol{h})$ form a right-angled triangle, in which

$$A(\boldsymbol{h}) = |F(\boldsymbol{h})| \cos(\varphi(\boldsymbol{h}))$$

and

$$B(\boldsymbol{h}) = |F(\boldsymbol{h})| \sin(\varphi(\boldsymbol{h})).$$

Consequently,

$$F(\boldsymbol{h}) = |F(\boldsymbol{h})|(\cos(\varphi(\boldsymbol{h})) + i \sin(\varphi(\boldsymbol{h}))) = |F(\boldsymbol{h})| \exp(i\varphi(\boldsymbol{h})).$$

It can also be easily shown that $\varphi(\boldsymbol{h}) = \tan^{-1}[B(\boldsymbol{h})/A(\boldsymbol{h})]$.

The above relations and considerations lead to the phase problem of crystallography: *In order to determine the crystal structure we need both the magnitudes and the phases of the structure factors, whereas only quantities related to the magnitudes are furnished by the experiment.*
Chapters 6 and 8 are dedicated to indirect and direct solutions of the phase problem.

5.9 Effects of space-group symmetry

We shall now examine the effect of the space-group symmetry of the crystal on the structure-factor function. These effects are of importance in the determination of crystal symmetry and can be conveniently discussed at this stage. If $(\mathbf{P}, \mathbf{t})$ is a representative operator of the space group of the crystal, then, by definition,

$$\rho(\boldsymbol{x}) = \rho(\boldsymbol{P}\boldsymbol{x} + \boldsymbol{t}),$$

where $\rho(\boldsymbol{x})$ is the electron density function at the point $\boldsymbol{x}$. Before comparing the electron density at these two equivalent locations, let us consider the effect of the point-group symmetry of the direct lattice on its associated reciprocal lattice. If the application of the operator $\mathbf{P}$ to the direct lattice brings it into self-coincidence, this operation will also bring families of crystal planes (with integer indices hkl that are not necessarily relatively prime) into equivalent families of crystal planes. That is, the point-group symmetry operation acts on direct-lattice vectors as well as on the normals to the crystal planes (the reciprocal-lattice vectors), bringing both sets of vectors into self-coincidence. This means that the direct lattice and the reciprocal lattice have the same point-group symmetry. We shall see this more quantitatively in what follows.

The electron density at the point $\boldsymbol{x}$ is given by

$$\rho(\boldsymbol{x}) = \frac{1}{V} \sum_{\boldsymbol{h}} F(\boldsymbol{h}) \exp(-2\pi i \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}), \tag{5.48}$$

where, in this and the following equations $\boldsymbol{h}^{\mathrm{T}}$ is the row matrix $(h \quad k \quad l)$ and $\boldsymbol{x}$ is a column matrix containing x, y, and z in the first, second, and third row, respectively. Of course, $\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}$ is simply equivalent to the scalar product $\mathbf{h} \cdot \mathbf{r}$ used above. We introduce this matrix notation to allow the convenient inclusion of matrices of symmetry operators in the expressions. Similarly, the electron density at the equivalent point $\boldsymbol{P}\boldsymbol{x} + \boldsymbol{t}$ is given by

$$\begin{aligned}\rho(\boldsymbol{P}\boldsymbol{x}+\boldsymbol{t}) &= \frac{1}{V}\sum_{\boldsymbol{h}} F(\boldsymbol{h})\exp[-2\pi i\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{P}\boldsymbol{x}+\boldsymbol{t})],\\ &= \frac{1}{V}\sum_{\boldsymbol{h}}[F(\boldsymbol{h})\exp(-2\pi i\boldsymbol{h}^{\mathrm{T}}\boldsymbol{t})]\exp(-2\pi i\boldsymbol{h}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x})\\ &= \frac{1}{V}\sum_{\boldsymbol{h}}[F(\boldsymbol{h})\exp(-2\pi i\boldsymbol{h}^{\mathrm{T}}\boldsymbol{t})]\exp[-2\pi i(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{h})^{\mathrm{T}}\boldsymbol{x}]. \qquad (5.49)\end{aligned}$$

Of course, the summations in eqns (5.48) and (5.49) are identical and range over the same terms, albeit in a different order. It follows that the coefficient of $\exp[-2\pi i(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{h})^{\mathrm{T}}\boldsymbol{x}]$ in (eqn 5.49) must be the structure factor of the reflection $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{h}$. Hence,

$$F(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{h}) = F(\boldsymbol{h})\exp(-2\pi i\boldsymbol{h}^{\mathrm{T}}\boldsymbol{t}). \qquad (5.50)$$

Equation (5.50) is the fundamental relationship between symmetry-related reflections. If we write $F(\boldsymbol{h}) = |F(\boldsymbol{h})|\exp[i\varphi(\boldsymbol{h})$, eqn (5.50) leads to the following relationships:

$$|F(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{h})| = |F(\boldsymbol{h})| \qquad (5.51)$$

and

$$\varphi(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{h}) = \varphi(\boldsymbol{h}) - 2\pi\boldsymbol{h}^{\mathrm{T}}\boldsymbol{t}. \qquad (5.52)$$

Equation (5.51) indicates the equality of the intensities of truly symmetry-related reflections, while eqn (5.52) relates the phases of the corresponding structure factors. The latter equation is a most important phase relationship and will be discussed further in connection with direct methods of phase determination (see Chapter 8).

Returning to eqn (5.51), if, for example, the crystal is monoclinic and belongs to the point group $2/m$, the rotational parts of its representative space-group operators–the second setting being assumed–are:

$$\boldsymbol{P}_1 = \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}, \quad \boldsymbol{P}_2 = \begin{pmatrix}\overline{1}&0&0\\0&\overline{1}&0\\0&0&\overline{1}\end{pmatrix}, \quad \boldsymbol{P}_3 = \begin{pmatrix}\overline{1}&0&0\\0&1&0\\0&0&\overline{1}\end{pmatrix}, \quad \boldsymbol{P}_4 = \begin{pmatrix}1&0&0\\0&\overline{1}&0\\0&0&1\end{pmatrix},$$

the row matrices of the corresponding symmetry-related diffraction vectors, $(\boldsymbol{P}_i^{\mathrm{T}}\boldsymbol{h})^{\mathrm{T}}$ are $(h \quad k \quad l)$, $(\overline{h} \quad \overline{k} \quad \overline{l})$, $(\overline{h} \quad k \quad \overline{l})$, and $(h \quad \overline{k} \quad l)$, and the intensities of any four such reflections are therefore predicted to be equal. In other words, the point-group symmetry of the crystal must be contained in the symmetry of the intensity distribution in the diffraction pattern. The converse is, unfortunately, not true. If, for example, the intensity distribution displays a $2/m$ point-group symmetry, we cannot immediately conclude that this is also the point-group symmetry of the crystal. In fact, the diffraction pattern appears in the majority of cases to possess a center of symmetry, whether or not such a center is also

present in the crystal, and, in the above example, the symmetry of the diffraction pattern will be apparently the same for all monoclinic crystals,that is, for crystals belonging to the point groups 2, m, and $2/m$. This phenomenon is associated with the fundamentals of the process of scattering of X-rays by electrons, as outlined in Section 5.2. If the electrons behave as if they are free, the atomic scattering factor is an approximately real quantity, and the structure factor and its complex conjugate are given by

$$F(\boldsymbol{h}) = \sum_{j=1}^{N} f_j \exp(2\pi i \boldsymbol{h}^{\mathrm{T}} \boldsymbol{r}) \quad \text{and}$$

$$F^*(\boldsymbol{h}) \cong \sum_{j=1}^{N} f_j \exp(-2\pi i \boldsymbol{h}^{\mathrm{T}} \boldsymbol{r}).$$

It can be seen from the above that $F^*(\boldsymbol{h}) \cong F(-\boldsymbol{h})$ and $F(\boldsymbol{h}) \cong F^*(-\boldsymbol{h})$. Consequently, under the above assumption, $|F(\boldsymbol{h})|^2 \cong |F(-\boldsymbol{h})|^2$ or $I(\boldsymbol{h}) \cong I(-\boldsymbol{h})$. The latter frequently observed relationship, is known as Friedel's law (Friedel 1913). Table 5.2 summarizes the apparent symmetries of the diffraction pattern along with the possible point-group symmetries of the crystal, under the assumption that Friedel's law is valid.

As pointed out above, Friedel's law is based on the assumption that all the electrons behave as if they were free–only under such an assumption are the atomic scattering factors real quantities. In fact, electrons which are close to the nucleus are screened by the outer electrons in all but the lightest atoms, and they can be assumed to scatter X-rays like bound, damped, oscillating charges. This additional contribution (as compared with free electrons) is often called "anomalous scattering", or "anomalous dispersion" in the literature, but there is really nothing anomalous about it, and the terms "resonant scattering" or "dispersion correction" are certainly more appropriate. Such effects are stronger the higher is the atomic number, that is, the heavier the atom, the more strongly a dispersive scatterer it is.

The atomic scattering factor, allowing for resonant scattering, can be approximated as

$$f(|\mathbf{h}|, \lambda) = f^0(|\mathbf{h}|) + f'(|\mathbf{h}|, \lambda) + i f''(|\mathbf{h}|, \lambda),$$

where f^0 is the part of the scattering factor based on the assumption of free electrons, and f' and f'' are the real and imaginary parts, respectively, of the dispersion correction. It is still assumed that the atom is spherically symmetric, and that effects of atomic displacements are included. However, the dependence of the dispersion correction on the diffraction vector is very weak and is usually neglected. It is the dependence on the wavelength that is of importance.

If we define $f^{\mathrm{R}} \equiv f^0 + f'$, the structure factor can be written as

$$F(\mathbf{h}) = \sum_{j=1}^{N} f_j^{\mathrm{R}} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) + i \sum_{j=1}^{N} f_j'' \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

Table 5.2 Apparent symmetries of diffraction patterns

Point-group symmetries of crystal	Apparent symmetry of $I(\mathbf{h})$
$1, \overline{1}$	$\overline{1}$
$2, m, 2/m$	$2/m$
$222, mm2, mmm$	mmm
$4, \overline{4}, 4/m$	$4/m$
$422, 4mm, \overline{4}2m, 4/mmm$	$4/mmm$
$3, \overline{3}$	$\overline{3}$
$32, 3m, \overline{3}m$	$\overline{3}m$
$6, \overline{6}, 6/m$	$6/m$
$622, 6mm, \overline{6}2m, 6/mmm$	$6/mmm$
$23, m\overline{3}$	$m\overline{3}$
$432, \overline{4}3m, m\overline{3}m$	$m\overline{3}m$

and if we write the first sum on the right-hand side as $A + iB$ and the second as $a + ib$ it can easily be shown that Friedel's law breaks down for noncentrosymmetric crystals but holds true for centrosymmetric ones (see the Exercise).

With the advent of synchrotron radiation and the ability to collect intensity data at several chosen wavelengths, dispersion corrections became an important tool in several branches of crystallography (see, for example, Giacovazzo 1998 and Vijayan 2001). There are several extensive reviews of the classical and quantum mechanical theories of the dispersion correction and its experimental determination (see, for example, James 1965 and Creagh 1999).

We shall now resume the discussion of the effects of crystallographic symmetry on the diffraction pattern. First, the apparent symmetry (when Friedel's law is approximately valid) of the diffracted intensity is related to the possible point-group symmetries in Table 5.2.

The eleven centrosymmetric point groups appearing in the right column of Table 5.2 are called the Laue groups. These are furnished directly by the symmetry of the intensity distribution. The Laue group of a given diffraction pattern tells us first of all tells us to which system the crystal, from which the pattern was obtained, belongs to. These corresponding relationships are summarized in Table 5.3.

We shall return to the problem of resolving possible point-group ambiguities after the discussion of the effects of screw axes and glide planes on the diffraction pattern.

5.9.1 Effect of the lattice type

The next item of information we require is the type of Bravais lattice that characterizes the periodicity of the crystal investigated. Since a lattice translation is also a space-group operator, it can be written as $([1], \boldsymbol{t}_{\mathrm{L}})$, where $\boldsymbol{t}_{\mathrm{L}}$ is a column

Table 5.3 Relationship between Laue group and crystal system

Laue symmetries	Crystal system
$\overline{1}$	Triclinic
$2/m$	Monoclinic
mmm	Orthorhombic
$4/m, 4/mmm$	Tetragonal
$\overline{3}, \overline{3}m$	Trigonal
$6/m, 6/mmm$	Hexagonal
$m\overline{3}, m\overline{3}m$	Cubic

consisting of the components of a lattice vector. Implementing this operator in eqn (5.50), that equation reduces to

$$F(\boldsymbol{h}) = F(\boldsymbol{h}) \exp(-2\pi i \boldsymbol{h}^{\mathrm{T}} \boldsymbol{t}_{\mathrm{L}}) \tag{5.53}$$

and $F(\boldsymbol{h})$ can be nonzero only if $\exp(-2\pi i \boldsymbol{h}^{\mathrm{T}} \boldsymbol{t}_{\mathrm{L}}) = 1$. This leads to conditions for possible nonzero reflections, depending only on the lattice type. For example, if the components of $\mathbf{t}_{\mathrm{L}}$ are all integers, which is the case for a P-type lattice, the above condition is fulfilled for all $\boldsymbol{h}$ and this lattice type does not impose any restrictions. If the lattice is a type I, there are two lattice points in the unit cell, at say $(0 \quad 0 \quad 0)$ and $(1/2 \quad 1/2 \quad 1/2)$. The first of these does not lead to any restrictions on possible reflections. The second, however, requires that $\exp[-\pi i(h + k + l)]$ be equal to unity. Since $\exp(\pi i n) = (-1)^n$, where n is an integer, the possible reflections from a crystal with an I lattice must have indices such that their sum is an even integer; those adding up to an odd integer are *systematically absent*. We can examine all lattice types in this manner and find the restrictions that they impose on the parity of the indices hkl or their sum. The results are summarized in Table 5.4.

The conditions shown in Table 5.4 apply to all reflections. Conversely, when systematic absences consistent with any of these conditions are detected, the corresponding lattice type is determined. For example, if the intensity distribution in the diffraction pattern obeys the operation of the point group $2/m$ we know that the crystal belong to the monoclinic systems. If, in addition, the intensities of the reflections with $h + k = 2n + 1$ are systematically absent this indicates that the lattice type of our crystal is C. With this information, and assuming that Friedel's law holds true, the possible space groups of the crystal are $C2$, Cm, Cc, $C2/m$ and $C2/c$. This list can be narrowed down by considering the systematic absences that are induced by the presence of a c glide plane. We shall now turn to these and similar considerations.

5.9.2 Effect of screw axes and glide planes

The intensities that are systematically affected by the presence of nonsymmorphic symmetry elements,that is, screw axes and glide planes, correspond to

Table 5.4 Effects of lattice type on conditions for possible reflections

Lattice type	$\mathbf{t}_{\mathrm{L}}^{\mathrm{T}}$	$\boldsymbol{h}^{\mathrm{T}}\mathbf{t}_{\mathrm{L}}$	Conditions for possible reflections
P	$(0\ \ 0\ \ 0)$	0	none
A	$(0\ \ \frac{1}{2}\ \ \frac{1}{2})$	$(k+l)/2$	hkl: $k+l=2n$
B	$(\frac{1}{2}\ \ 0\ \ \frac{1}{2})$	$(h+l)/2$	hkl: $h+l=2n$
C	$(\frac{1}{2}\ \ \frac{1}{2}\ \ 0)$	$(h+k)/2$	hkl: $h+k=2n$
I	$(\frac{1}{2}\ \ \frac{1}{2}\ \ \frac{1}{2})$	$(h+k+l)/2$	hkl: $h+k+l=2n$
F	$(0\ \ \frac{1}{2}\ \ \frac{1}{2})$	$(k+l)/2$	h, k and l are all even or all odd
	$(\frac{1}{2}\ \ 0\ \ \frac{1}{2})$	$(h+l)/2$	(simultaneous fulfillment of the
	$(\frac{1}{2}\ \ \frac{1}{2}\ \ 0)$	$(h+k)/2$	conditions for types A, B and C).
R	$(\frac{2}{3}\ \ \frac{1}{3}\ \ \frac{1}{3})$	$(2h+k+l)/3$	hkl: $-h+k+l=3n$
	$(\frac{1}{3}\ \ \frac{2}{3}\ \ \frac{2}{3})$	$(h+2k+2l)/3$	

diffraction vectors that are eigenvectors of the rotation part of the space-group operator involved. Specifically, if the space-group operator is $(\mathbf{P}, \mathbf{t})$, the diffraction vectors involved in absences must satisfy the equation $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{h} = \boldsymbol{h}$. Equation (5.50) then becomes

$$F(\boldsymbol{h}) = F(\boldsymbol{h}) \exp(-2\pi i \boldsymbol{h}^{\mathrm{T}}\boldsymbol{t}) \tag{5.54}$$

and can be satisfied only if the exponential factor on the right-hand side of eqn (5.54) equals unity. Let us consider first the monoclinic c glide as a specific example. Assuming the second setting, that is, the normal to the glide plane is parallel to the basis vectors $\mathbf{b}$ and $\mathbf{b}^*$, the indices of the reflections that remain unchanged under the operation of the mirror component of the glide operator must have the form $(h\ \ 0\ \ l)$. These are indeed the eigenvectors of the reflection operator, since the mirror plane is parallel to the plane containing the vectors $\mathbf{a}^*$ and $\mathbf{c}^*$. The translation part of the c-glide operator has the form $(0\ \ y\ \ 1/2)$, where $y = 0$ corresponds to the plane passing through the origin. Hence, for any value of y, the scalar product $\boldsymbol{h}^{\mathrm{T}}\mathbf{t}$ is simply $l/2$, and the necessary condition for a nonzero value of the intensity of an $h0l$ relflection is $l = 2n$. Intensities of $h0l$ reflections with odd l will be systematically absent.

Suppose we have detected in the diffraction pattern of the crystal with a C-type lattice considered above, such an absence of $h0l$ reflections with odd l index. This indicates the presence of a c glide plane in the crystal, and the possible space groups are now Cc and $C2/c$. It remains to find out whether or not the crystal is centrosymmetric, in order to determine the space group unequivocally. This can often be easily done, and methods for doing this will be discussed in Chapter 7. A somewhat more difficult situation arises when we see that there is no c glide in our crystal, and it has been shown to be noncentrosymmetric. We are then left with the problem of distinguishing between the two noncentrosymmetric space

Table 5.5 Effects of screw axes on conditions for possible reflections

Screw axis	$\mathbf{t}^{\mathrm{T}}$	$\boldsymbol{h}^{\mathrm{T}}$	Conditions for possible reflections
$2_1 \parallel [100]$	$(1/2 \quad y \quad z)$	$(h \quad 0 \quad 0)$	$h = 2n$
$2_1 \parallel [010]$	$(x \quad 1/2 \quad z)$	$(0 \quad k \quad 0)$	$k = 2n$
$2_1 \parallel [001]$	$(x \quad y \quad 1/2)$	$(0 \quad 0 \quad l)$	$l = 2n$
$2_1 \parallel [110]$	$(1/2 \quad 1/2 \quad z)$	$(h \quad h \quad 0)$	None
$3_1 \parallel [001]$	$(x \quad y \quad 1/3)$	$(0 \quad 0 \quad l)$	$l = 3n$
$3_1 \parallel [111]$	$(1/3\ 1/3\ 1/3)$	$(h \quad h \quad h)$	None
$4_1 \parallel [100]$	$(1/4 \quad y \quad z)$	$(h \quad 0 \quad 0)$	$h = 4n$
$4_1 \parallel [010]$	$(x \quad 1/4 \quad z)$	$(0 \quad k \quad 0)$	$k = 4n$
$4_1 \parallel [001]$	$(x \quad y \quad 1/4)$	$(0 \quad 0 \quad l)$	$l = 4n$
$4_2 \parallel [001]$	$(x \quad y \quad 1/2)$	$(0 \quad 0 \quad l)$	$l = 2n$
$6_1 \parallel [001]$	$(x \quad y \quad 1/6)$	$(0 \quad 0 \quad l)$	$l = 6n$
$6_2 \parallel [001]$	$(x \quad y \quad 1/3)$	$(0 \quad 0 \quad l)$	$l = 3n$
$6_3 \parallel [001]$	$(x \quad y \quad 1/2)$	$(0 \quad 0 \quad l)$	$l = 2n$

groups $C2$ and Cm. We shall also indicate how such situations can be dealt with.

Let us now assume a monoclinic crystal with a P-type Bravais lattice, in the second setting, which contains a twofold screw axis as well as a c glide plane. The eigenvectors of the rotational part of the screw-axis operator, in reciprocal space, must be parallel to $\mathbf{b}^*$ and the corresponding triplets of diffraction indices must be of the form: $(0k0)$. The translation part of the screw axis operation must have the form: $(x \quad 1/2 \quad z)$, where the values of x and z depend on the location of the origin. Hence for any values of x and z the scalar product $\boldsymbol{h}^{\mathrm{T}}\mathbf{t}$ is simply $k/2$, and the necessary condition for a non-zero value of the intensity of a $0k0$ reflection is $k = 2n$. The intensities of $0k0$ reflectons with odd k will be systematically absent. Of course, a monoclinic crystal with P-type lattice, which contains a twofold screw axis as well as a c glide plane must belong to the space group $P2_1/c$.

A brief summary of the effects of the various screw axes on the conditions for possible reflections from the corresponding special subsets of hkl is given in Table 5.5. Complete reference tables of such conditions, for screw axes and glide planes, with specification of the relevant crystal systems, can be found in Chapter 2 of Part 1 of Volume A of *International Tables for Crystallography* (Hahn 1983, 1996).

The components of $\mathbf{t}^{\mathrm{T}}$which are represented by simple fractions in Table 5.5 are the intrinsic translations of the screw-axis operators and are obviously independent of the choice of the origin of the space group, whereas those denoted by the symbols x, y and z depend on the location of that origin. It is interesting to note that some "diagonal" screw axes, which are so conspicuous in the space-

group diagrams seen in Volume A of the *International Tables for Crystallography* (Hahn, 1983, 1996) do not give rise to any conditions for possible reflections.

5.10 Exercises for Chapter 5

1. Consider a material unit within which an electron density function is defined, and three more such units related to the first by the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{a}+\mathbf{b}$. That is, if the first unit were displaced parallel to itself by each of these vectors, it would come into coincidence with each of the other three. The four units are irradiated by X-rays of wavelength λ and it is assumed that there is no loss of energy upon scattering. Calculate the structure factor of this assembly. For what values of $\mathbf{h}\cdot\mathbf{a}$ and $\mathbf{h}\cdot\mathbf{b}$ will the intensity of the scattered radiation be greatest? For what values will it be smallest?

   [It is possible to answer the questions without calculating the complete expression for $|F(\mathbf{h})|^2$.]

2. Assuming spherical symmetry, the electron density of a hydrogen atom is given by

$$\rho_{\mathrm{H}}(u) = R_{10}^2(u), \quad \text{where} \quad R_{10}(u) = \left(\frac{1}{\pi a_0^3}\right)^{1/2} \exp\left(-\frac{u}{a_0}\right)$$

   is the normalized radial wavefunction of the electron and a_0 is the Bohr radius. Show that the scattering factor of the hydrogen atom in the free-electron approximation is given by

$$f_{\mathrm{H}}(|\mathbf{h}|) = (1+\pi^2 a_0^2 |\mathbf{h}|^2)^{-2}.$$

3. The coordinates of sodium and chlorine in the unit cell of the NaCl crystal are the following.

$$\begin{array}{lllll} \mathrm{Na}: & (0,0,0), & (0,1/2,1/2), & (1/2,0,1/2), & (1/2,1/2,0) \\ \mathrm{Cl}: & (0,0,1/2), & (0,1/2,0), & (1/2,0,0), & (1/2,1/2,1/2). \end{array}$$

   (a) Show that the structure factor of any reflection hkl from an NaCl crystal with possibly nonvanishing intensity, is given by

$$F_{\mathrm{NaCl}}(hkl) = \begin{cases} 4(f_{\mathrm{Na}} + f_{\mathrm{Cl}}) & \text{if} \quad h+k+l = 2n \\ 4(f_{\mathrm{Na}} - f_{\mathrm{Cl}}) & \text{if} \quad h+k+l = 2n+1 \end{cases}$$

   (b) Crystals of KCl have a structure analogous to that of NaCl (apart from the unit cell parameter) and hence a similar expression for the structure factor $F_{\mathrm{KCl}}(hkl)$. However, in the diffraction pattern of KCl, the reflections with $h+k+l = 2n+1$ are too weak to be measured and seem systematically absent. Explain this phenomenon.

4. Convince yourself that in the presence of a significant dispersion correction, Friedel's law breaks down for noncentrosymmetric crystals but holds true for centrosymmetric ones.

6 The Patterson function

6.1 Introduction

Early approaches to crystal structure analysis included extensive use of space-group symmetry, and trial-and-error methods. These approaches proved satisfactory when highly symmetric or very simple structures, characterized by a small number of structural parameters were of interest. However, quite apart from the obvious scientific aspects, it was soon realized that the real structures of interest were becoming more and more complex, and systematic approaches were clearly indispensable. The first such systematic approach to structure determination was put forward by Patterson (1934), and methods based on this approach are still in use to this very day, especially in the solution of structures of macromolecules. This chapter is devoted to an introduction to Patterson techniques and their recent ramifications.

6.2 The Patterson function

The principle of the Patterson method is the easily demonstrated dependence of the diffracted intensity, and hence the squared magnitude of the structure factor, on the set of interatomic vectors. In fact, if we assume that the atomic scattering factors are real, we have

$$\begin{aligned} I(\mathbf{h}) &\propto |F(\mathbf{h})|^2 \\ &= F(\mathbf{h})F^*(\mathbf{h}) \\ &= \left(\sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_j)\right)\left(\sum_{k=1}^{N} f_k \exp(-2\pi i \mathbf{h}\cdot\mathbf{r}_k)\right) \\ &= \sum_j \sum_k f_j f_k \exp[2\pi i \mathbf{h}\cdot(\mathbf{r}_j - \mathbf{r}_k)], \end{aligned} \tag{6.1}$$

where N is the number of atoms in the unit cell, f_j and f_k are the scattering factors of atoms j and k, $\mathbf{r}_j$ and $\mathbf{r}_k$ are their respective position vectors, and $\mathbf{h}$ is the diffraction vector. Hence, the measurable $|F|^2$ depends on the set of interatomic vectors $\{\mathbf{r}_j - \mathbf{r}_k\}$. In words, whereas the structure factor $F(\mathbf{h})$ is a sum of atomic contributions, each depending on the scattering factor and position vector of an atom, its squared magnitude $|F(\mathbf{h})|^2$ is a sum of pairs of atomic contributions, each depending on the product of the scattering factors of the two atoms involved and the difference between their position vectors (that is, on the interatomic vector). The two summations have similar mathematical structures.

They differ merely in the number of terms, in the nature of the coefficients of the exponentials, and in the arguments of the latter. More specifically, a term in the structure factor summation depends on a position vector whereas a term in $|F(\mathbf{h})|^2$ depends on the *difference* between two position vectors. The position vector is meaningful only if it has been referred to some origin, whereas the interatomic vector is clearly independent of the (arbitrary) choice of such an origin.

We saw in the previous chapter that the electron density at the point $\mathbf{r}$ in the crystal space is given by

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{r}). \tag{6.2}$$

We also know that $\rho(\mathbf{r})$ is a peaked function, and the peaks of this function correspond to atomic position vectors. Let us now consider a Fourier synthesis analogous to eqn (6.2), but with $F(\mathbf{h})$ replaced by $|F(\mathbf{h})|^2$. In view of the similarity of the two types of summation, it is reasonable to expect that the peaks of the new Fourier synthesis, say $P(\mathbf{u})$, will correspond to interatomic vectors. This can, however, be easily verified:

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{u}) \tag{6.3}$$

$$= \frac{1}{V}\sum_{\mathbf{h}} \left\{ \sum_j \sum_k f_j f_k \exp[2\pi i \mathbf{h}(\mathbf{r}_j - \mathbf{r}_k)] \right\} \exp(-2\pi i \mathbf{h}\cdot\mathbf{u})$$

$$= \sum_j \sum_k \left\{ \frac{1}{V}\sum_{\mathbf{h}} f_j f_k \exp\{-2\pi i \mathbf{h}\cdot[\mathbf{u} - (\mathbf{r}_j - \mathbf{r}_k)]\} \right\} \tag{6.4}$$

$$\equiv \sum_j \sum_k P_{jk}[\mathbf{u} - (\mathbf{r}_j - \mathbf{r}_k)]. \tag{6.5}$$

If we compare eqns (6.4) and (6.5) we see that the term P_{jk} attains its maximum value when the argument of the exponential in eqn (6.4) vanishes, that is, when $\mathbf{u} = \mathbf{r}_j - \mathbf{r}_k$. It follows that P_{jk} is a peaked function, with maximum height

$$P_{jk}^{\max} = \frac{1}{V}\sum_{\mathbf{h}} f_j f_k, \tag{6.6}$$

and the new Fourier synthesis $P(\mathbf{u})$ given by eqn (6.3) is a superposition of such peaked functions, each associated with an interatomic vector, this is what was to be verified. The function $P(\mathbf{u})$ is known as the Patterson function. It should be noted that for each interatomic vector $(\mathbf{r}_j - \mathbf{r}_k)$ there is also a vector $(\mathbf{r}_k - \mathbf{r}_j)$ in the summation and the Patterson function therefore posesses a center of symmetry–provided of course Friedel's law is valid. The original derivation

of the Patterson function follows from considering the electron density at the points $\mathbf{r}$ and $\mathbf{r}+\mathbf{u}$, where $\mathbf{u}$ is a constant vector, and the endpoint of $\mathbf{r}$ ranges over the whole unit cell. The integral of the product of the electron densities at the above two points over the variable components of the vector $\mathbf{r}$, that is, the autocorrelation function of the electron density, with a correlation vector $\mathbf{u}$, is expected to attain its maximum value when the vector $\mathbf{u}$ is an interatomic vector. Indeed, as verified in the Exercises,

$$P(\mathbf{u}) = \int_{\text{cell}} \rho(\mathbf{r})\rho(\mathbf{r}+\mathbf{u})\,\mathrm{d}^3\mathbf{r} = \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{u}). \tag{6.7}$$

It is perhaps in order to reemphasize that the argument of the electron density function is a position vector of a point in the *point* space, and that of the Patterson function is a difference between two such position vectors–or, simply, an element of a *vector* space. The Patterson vector is thus independent of the choice of the origin in the point space.

6.2.1 Periodicity of the Patterson function

As we know, the periodicity of the electron density can be represented as

$$\rho(\mathbf{r}+u\mathbf{a}+v\mathbf{b}+w\mathbf{c}) = \rho(\mathbf{r}),$$

where $\mathbf{r}$ is the position vector of the point at which the density is calculated; u, v, and w are any integers; and $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are basis vectors of the direct lattice, such that the unit cell constructed from these vectors is primitive. In view of the analogous structure of the Fourier syntheses for the electron density and the Patterson function, it is expected that the periodicities of the two functions will also be the same. Indeed,

$$\begin{aligned} P(\mathbf{u}+u\mathbf{a}+v\mathbf{b}+w\mathbf{c}) &= \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp[-2\pi i \mathbf{h}\cdot(\mathbf{u}+u\mathbf{a}+v\mathbf{b}+w\mathbf{c})] \\ &= \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{u}) \exp[2\pi i(hu+kv+lw)] \\ &= \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{u}) \\ &= P(\mathbf{u}), \end{aligned}$$

as was expected. Note that the first line of the above equation reduces to the second if we use the Laue equations or, equivalently, the definitions of the mutually reciprocal bases. This periodicity of the Patterson function has some far-reaching consequences that affect its applicability. Since the periodicity of $P(\mathbf{u})$ is the same as that of $\rho(\mathbf{r})$, both functions are characterized by the same unit cell. The volume of this cell is filled with N atomic peaks in the case of $\rho(\mathbf{r})$, and with N^2 interatomic peaks in the case of $P(\mathbf{u})$. It follows that if the atoms have

approximately similar scattering factors, if there is no substantial overlap of interatomic vectors (which would decrease the number of peaks and enhance their heights–see below), and N is not too small (say, larger than 20), then the unit cell of the Patterson function will be overcrowded to an extent that will usually make the recognition of individual Patterson peaks impossible. In other words, the Patterson function will then be featureless. This seems to be a rather serious drawback, but there is a way out of the difficulty, and this will be described in the next subsection.

6.2.2 The heavy atom method

Equation (6.6) indicates that the height of a Patterson peak is roughly proportional to the product of the atomic numbers of the two atoms involved, that is, $P_{jk}^{\max} \propto Z_j Z_k$.

Consider for example two identical molecules, each containing 20 carbon atoms and one bromine atom, and the Patterson peaks corresponding to this assembly. We expect, first, a single peak corresponding to the 42 zero vectors in the system (vectors relating each atom to itself), of height proportional to $40Z_{\mathrm{C}}^2 + 2Z_{\mathrm{Br}}^2 = 3890$. This peak does not carry any structural information, but is important insofar it defines the origin of the vector space. We then have 1560 carbon-carbon peaks, with heights proportional to $Z_{\mathrm{C}}^2 = 36$. These peaks occupy roughly the same volume as the 40 carbon atoms and are therefore so heavily overcrowded that their probable contribution is to raise the background of the Patterson function. We next have 160 bromine-carbon and carbon-bromine peaks, with heights proportional to $Z_{\mathrm{Br}} Z_{\mathrm{C}} = 215$. Here, the chance of resolving the peaks is better than in the carbon-carbon case; however, they still seem to be rather numerous. Finally, we have just two bromine-bromine peaks, with height proportional to $Z_{\mathrm{Br}}^2 = 1225$. These two peaks are outstandingly high and should be easily located in a map of the Patterson function. Let us assume further that the two molecules are related by a center of symmetry, and choose the origin of the assembly at this center. One bromine – bromine vector, $\mathbf{u}_{\mathrm{Br-Br}}$, now relates bromine atoms at $\mathbf{r}_{\mathrm{Br}}$ and $-\mathbf{r}_{\mathrm{Br}}$, and the position vector of the bromine can be taken as one-half of $\mathbf{u}_{\mathrm{Br-Br}}$. In a real example, this would lead immediately to the possible fractional coordinates of the bromine. We shall show in a later section, in a more general manner, how to apply space-group symmetry to the location of a heavy atom. Let us, try, however, to continue the above argument and see how the coordinates of a single heavy atom in the asymmetric unit can be used in the structure determination of a crystal. Let the structure factor be written as

$$F(\mathbf{h}) = F_{\mathrm{known}}(\mathbf{h}) + F_{\mathrm{unknown}}(\mathbf{h}), \tag{6.8}$$

where, in our case, $F_{\mathrm{known}}(\mathbf{h})$ is the known contribution of the heavy atoms (the single heavy atom plus those generated by symmetry operations) in the unit cell. We can use this contribution in calculating an approximation to the phase of the structure factor,

$$\varphi_{\text{approx}}(\mathbf{h}) = \tan^{-1} \frac{B_{\text{known}}}{A_{\text{known}}}, \tag{6.9}$$

for a range of reflections, where $F = A + iB$, and compute a Fourier synthesis for an approximation to the electron density. We use as Fourier coefficients the magnitudes of the structure factors derived from the experiment, multiplied by approximate phase factors given by eqn (6.9). This density is therefore given by

$$\rho_{\text{approx}}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h}} |F_{\text{obs}}(\mathbf{h})| \exp[i\varphi_{\text{approx}}(\mathbf{h})] \exp(-2\pi \mathbf{h} \cdot \mathbf{r}). \tag{6.10}$$

When we examine this approximate electron density, several peaks appear. These include of course the heavy atoms which were input, but most usually, additional peaks, corresponding to the hitherto unknown light atoms, are found. These new peaks are then included in the known part of the structure, by recalculating $F_{\text{known}}(\mathbf{h})$. Equations (6.8), (6.9), and (6.10) are then used in an iterative manner until all or most of the atoms have been located.

The method described above was first applied to the determination of structures that contained outstandingly heavy atoms, and hence the name "heavy atom method". It is, however, clear that *any* substantial known part of the structure may be used as a starting point, and the heavy atom was just an example, but one which happens to be very pertinent to a discussion of the basic applications of the Patterson technique.

6.2.3 Overlap of interatomic vectors

It is sometimes possible to take advantage of a known overlap of interatomic vectors and determine the structure without the need for introducing a heavy atom. Because of the powerful direct methods of phase determination (to be dealt with in a later chapter), this is nowadays very seldom done. It is, however, an elegant approach and deserves to be mentioned.

A typical example is one where there is a small number of planar polycyclic aromatic molecules in the unit cell, of completely or partly known structural formula, where the aim of the study is the determination of their packing arrangement.

The principle is simple: if several, say n, interatomic vectors in an "equal-atom" structure have the same length and are identically oriented, they will give rise to a single peak in the Patterson map. The height of this peak will of course be n times that expected for a single interatomic vector. The overall effect is favorable: the number of Patterson peaks decreases and their height increases, thus enhancing the chance of their recognition. Indeed, many structures of organic aromatics, investigated in the 1950s, 1960s, and also somewhat later were solved using this principle (see, for example, any issue of *Acta Crystallographica* from that period).

6.3 Classification of Patterson vectors

We shall now consider the various kinds of Patterson vector, determined by the (known) space-group symmetry of the crystal. Let $\mathbf{r}_j$ and $\mathbf{r}_k$ be general position vectors of two atoms within the asymmetric unit, and let $(\boldsymbol{P}_s, \mathbf{t}_s)$ and $(\boldsymbol{P}_u, \mathbf{t}_u)$ be space-group operations. Let N and g be the number of atoms and number of asymmetric units, respectively, in the unit cell. The general form of an interatomic vector is therefore

$$\begin{aligned}\mathbf{u} &= (\mathbf{P}_s, \mathbf{t}_s)\mathbf{r}_j - (\mathbf{P}_u, \mathbf{t}_u)\mathbf{r}_k \\ &= \mathbf{P}_s\mathbf{r}_j - \mathbf{P}_u\mathbf{r}_k + \mathbf{t}_s - \mathbf{t}_u. \end{aligned} \tag{6.11}$$

We can now distinguish the following four cases:

1. *Case*1. Let $s = u$ and $j = k$ in eqn (6.11). We obtain zero vectors (vectors between the atoms and themselves) $\mathbf{u}$=$\mathbf{0}$, and their number is the number of asymmetric units in the unit cell times the number of atoms in the asymmetric unit, or $g \times N/g = N$.
2. *Case*2. Let $s \neq u$ and $j = k$ in eqn (6.11). We obtain vectors between equivalent atoms in different asymmetric units. Such a vector has the form

$$\mathbf{u}_{\text{Harker}} = (\mathbf{P}_s - \mathbf{P}_u)\mathbf{r}_j + \mathbf{t}_s - \mathbf{t}_u. \tag{6.12}$$

   This is a most useful kind of interatomic vector, since it is directly related to an atomic position vector. As we shall see in an example, this is the crystallographer's tool for the location of a heavy atom in a wide range of space groups. The use of the vectors given by eqn (6.12) was first proposed by Harker (1936), and the corresponding Patterson peaks are known as Harker peaks. The reason for their usefulness will be evident from the example given below. The number of Harker vectors is the number of combinations of different asymmetric units times the number of atoms in an asymmetric unit, or $g(g-1) \times N/g = (g-1)N$.
3. *Case*3. Let $s = u$ and $j \neq k$ in eqn (6.11). We obtain, for a given s, the vectors between nonequivalent atoms within the sth asymmetric unit, also known as *intraunit* Patterson vectors. Such a vector has the form

$$\mathbf{u}_{\text{intra}} = \mathbf{P}_s(\mathbf{r}_j - \mathbf{r}_k). \tag{6.13}$$

   The intraunit vectors are independent of the translation parts of the space-group operations. They do not carry any information about the location of the molecule but fully define its orientation. The latter property is the basis of the rotation function (Rossmann and Blow 1962), which will be described briefly in the last section of this chapter. The intraunit vectors are also known as self-Patterson vectors. The number of intraunit vectors is the number of asymmetric units in the unit cell times the number of interatomic vectors within an asymmetric unit, less those counted in Case 1, or $g \times (N/g)[(N/g) - 1] = N^2/g - N$.

4. *Case*4. Let $s \neq u$ and $j \neq k$ in eqn (6.11). We obtain the vectors between nonequivalent atoms in different asymmetric units, also known as *interunit* vectors. Such a vector has the form given by (6.11). These vectors depend on both the orientations and the relative locations of the asymmetric units. They are, in general, not easily interpretable by visual inspection, but computerized search methods may provide information about the locations of the units if their orientations are known. This is done by computing the translation function (see, for example, Crowther and Blow 1967), and will be discussed briefly in the last section of this chapter. The interunit vectors are also known as cross-Patterson vectors. The number of interunit (non-Harker) vectors is the number of combinations of different asymmetric units times the number of interatomic interunit vectors between atoms in two asymmetric units, or $g(g-1) \times (N/g)[(N/g)-1] = N^2 - gN - N^2/g + N$.

The above numbers of interatomic vectors of course add up to N^2. The above classification leads us to the most important applications of the Patterson function. We shall present below an example of some detailed considerations of the use of Harker vectors, and the last section of this chapter contains a survey of other applications. We shall, however, consider first the symmetry of the Patterson function.

6.4 Symmetry of the Patterson function

We have shown above that the Patterson function has the same periodicity as the function for the electron density. What, then, is the space group of the Patterson function, given that of the electron density?

The general expression for the Patterson function is given by

$$P(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h} \cdot \mathbf{u}), \tag{6.14}$$

where $\mathbf{h}$ extends over all the available part of the reciprocal lattice. We have also shown, in the previous chapter, that the structure factors of symmetry-related reflections obey the equation

$$F(\mathbf{h}^{\mathrm{T}}\mathbf{P}_s) = F(\mathbf{h}^{\mathrm{T}}) \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{t}_s), \tag{6.15}$$

where $(\mathbf{P}_s, \mathbf{t}_s)$ is the sth representative operator of the space group of the crystal considered. If we multiply each side of eqn (6.15) by its complex conjugate we obtain

$$|F(\mathbf{h}^{\mathrm{T}}\mathbf{P}_s)|^2 = |F(\mathbf{h}^{\mathrm{T}})|^2 \tag{6.16}$$

and so on, for all the operators of the point group $\{\mathbf{P}_i\}$ of the crystal. The translation part $\mathbf{t}_s$ of the space-group operation is clearly irrelevant. This simply confirms that the distribution of the diffracted intensity has the full symmetry of the point group of the crystal, as shown by the diffraction experiment. Note that we do not assume the validity of Friedel's law here.

If the point group of the crystal is of order g, the summation in eqn (6.14) can be split into g summations, each taken over an asymmetric unit of the available reciprocal lattice. Equation (6.14) can thus be rewritten as

$$\begin{aligned} P(\mathbf{u}) &= \frac{1}{V}\sum_{s=1}^{g}{\sum_{\mathbf{h}}}' |F(\mathbf{h}^{\mathrm{T}}\mathbf{P}_s)|^2 \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{P}_s\mathbf{u}) \\ &= \frac{1}{V}\sum_{s=1}^{g}{\sum_{\mathbf{h}}}' |F(\mathbf{h}^{\mathrm{T}})|^2 \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{P}_s\mathbf{u}) \\ &= \frac{1}{V}{\sum_{\mathbf{h}}}' |F(\mathbf{h}^{\mathrm{T}})|^2 \sum_{s=1}^{g}\exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{P}_s\mathbf{u}), \end{aligned} \tag{6.17}$$

where the primed summation over $\mathbf{h}$ extends over the asymmetric unit of the reciprocal lattice.

It is now easily shown that the Patterson function $P(\mathbf{u})$ is symmetric with respect to the operations of the symmorphic space group $\{\mathbf{P}_i, \mathbf{r}_{\mathrm{L}}\}$, where $\{\mathbf{P}_i\}$ is the point group of the crystal and $\{\mathbf{r}_{\mathrm{L}}\}$ is its Bravais lattice. We have

$$P(\mathbf{P}_j\mathbf{u}+\mathbf{r}_L) = \frac{1}{V}{\sum_{\mathbf{h}}}' |F(\mathbf{h}^{\mathrm{T}})|^2 \sum_{s=1}^{g}\exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{P}_s\mathbf{P}_j\mathbf{u}) \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{P}_s\mathbf{r}_{\mathrm{L}}) \tag{6.18}$$

$$= \frac{1}{V}{\sum_{\mathbf{h}}}' |F(\mathbf{h}^{\mathrm{T}})|^2 \sum_{s'=1}^{g}\exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{P}_{s'}\mathbf{u}) \tag{6.19}$$

$$= P(\mathbf{u}). \tag{6.20}$$

Equation (6.19), identical to eqn (6.17), is obtained from eqn (6.18) if we observe that

- $\exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{P}_s\mathbf{r}_{\mathrm{L}}) = 1$, since the point group $\{\mathbf{P}_i\}$ of the crystal leaves its Bravais lattice invariant and hence $\mathbf{P}_s\mathbf{r}_L$ is also a lattice vector, and the second exponential in eqn (6.18) equals unity for any $\mathbf{h}$; and
- the product $\mathbf{P}_s\mathbf{P}_j$ is also a point-group operation, say $\mathbf{P}_{s'}$.

The space group of the Patterson function is the symmorphic group obtained from the space group of the crystal by deleting screw axes and glide planes, if any, and retaining the Bravais lattice. So, for example, $P2_1/c$ and $C2/c$ in the crystal become $P2/m$ and $C2/m$ respectively, in the Patterson function. However, this function may be either centrosymmetric or noncentrosymmetric if Friedel's law has not been assumed. If Friedel's law is assumed to be valid, the space group of the Patterson function is always centrosymmetric (just like the apparent diffraction pattern, of which the Patterson function is a Fourier transform), and its point group is one of the 11 Laue groups. It is observed in practice, that the deviations of the Patterson function from centrosymmetry are often small and not readily detectable. This may not be so, however, if significant dispersion corrections are present.

Table 6.1 Patterson–Harker vectors for the example discussed in the text

	x, y, z	$\frac{1}{2}+\overline{x}, \frac{1}{2}+y, \frac{1}{2}+\overline{z}$	$\overline{x}, \overline{y}, \overline{z}$	$\frac{1}{2}+x, \frac{1}{2}+\overline{y}, \frac{1}{2}+z$
x, y, z	0 , 0 , 0	$\frac{1}{2}+2x, \frac{1}{2}, \frac{1}{2}+2z$	$2x, 2y, 2z$	$\frac{1}{2}, \frac{1}{2}+2y, \frac{1}{2}$
$\frac{1}{2}+\overline{x}, \frac{1}{2}+y, \frac{1}{2}+\overline{z}$	$\frac{1}{2}+\overline{2x}, \frac{1}{2}, \frac{1}{2}+\overline{2z}$	0 , 0 , 0	$\frac{1}{2}, \frac{1}{2}+2y, \frac{1}{2}$	$\overline{2x}, 2y, \overline{2z}$
$\overline{x}, \overline{y}, \overline{z}$	$\overline{2x}, \overline{2y}, \overline{2z}$	$\frac{1}{2}, \frac{1}{2}+\overline{2y}, \frac{1}{2}$	0 , 0 , 0	$\frac{1}{2}+\overline{2x}, \frac{1}{2}, \frac{1}{2}+\overline{2z}$
$\frac{1}{2}+x, \frac{1}{2}+\overline{y}, \frac{1}{2}+z$	$\frac{1}{2}, \frac{1}{2}+\overline{2y}, \frac{1}{2}$	$2x, \overline{2y}, 2z$	$\frac{1}{2}+2x, \frac{1}{2}, \frac{1}{2}+2z$	0 , 0 , 0

6.5 Harker peaks

We shall now present a detailed example of the use of Harker peaks. Let us assume, first, that the space group of the crystal is $P2_1/n$ (No. 14), with its origin at a center of symmetry and in the second setting (with the b axis unique), and construct a table of differences between the coordinates of the general equivalent positions of $P2_1/n$, as given in ITA83. Note that a Patterson vector will be denoted here by $\mathbf{u} = u_1\mathbf{a} + u_2\mathbf{b} + u_3\mathbf{c}$ or a triple $\mathbf{u} = (u_1, u_2, u_3)$. A commonly used designation is $\mathbf{u} = (u, v, w)$, but these symbols have been used here for another purpose. The Patterson-Harker vectors are given in Table 6.1.

In Table 6.1, the table entry (i, j) is a difference between the ith position vector in the leftmost column and the jth position vector in the uppermost row. All the fractions 1/2 were made positive by adding a basis vector where appropriate. Assuming that (x, y, z) are the fractional coordinates of an atom, all the off-diagonal table entries are coordinates of Harker vectors, given by eqn (6.15). The vector $(2x, 2y, 2z)$ results from the coordinates of two atoms related by a center of symmetry, and is located in a general position of the space group $P2/m$, which is the space group of the Patterson function in the present example. The three positions related to $(2x, 2y, 2z)$ by symmetry operations of $P2/m$ are also seen in the table. Of greater importance are the Harker vectors $(\frac{1}{2}+2x, \frac{1}{2}, \frac{1}{2}+2z)$ and $(\frac{1}{2}, \frac{1}{2}+\overline{2y}, \frac{1}{2})$, and their symmetry-related ones. Each of these vectors appears in the table twice, and hence the corresponding peaks are approximately twice as high as the peak due to inversion, which appears only once. The Harker peaks corresponding to $\pm(\frac{1}{2}+2x, \frac{1}{2}, \frac{1}{2}+2z)$ are located on a mirror plane of the space group $P2/m$ and are found by computing the section of the Patterson function at $u_2 = \frac{1}{2}$, that is, $P(u_1, \frac{1}{2}, u_3)$. The Harker peaks corresponding to $\pm(\frac{1}{2}, \frac{1}{2}+\overline{2y}, \frac{1}{2})$ are located on a twofold axis of $P2/m$ and are found by computing the line section $P(\frac{1}{2}, u_2, \frac{1}{2})$ of the Patterson function. Indeed, these Harker peaks are of outstanding importance because (i) we know where to look for them, (ii) their locations are simply related to the coordinates of the atom at (x, y, z) and (iii) when that atom is a significantly heavy one, these Harker peaks are usually readily seen.

It is not easy to say what is meant by "significantly heavy". However, a conservative criterion for this situation is that the ratio

$$r = \frac{\sum Z^2_{\text{heavy}}}{\sum Z^2_{\text{light}}} \tag{6.21}$$

be close to unity. This usually guarantees a successful determination of the heavy-atom coordinates, but there are examples of determinations with $r < 1$. It is interesting to mention vitamin B_{12}, with formula $C_{62}H_{88}CoO_{14}P$, in which r is well below 0.2 for the cobalt atom alone (Hodgkin et al. 1957).

Some computational considerations are in order. The underlying derivations will be given in the next chapter, and we summarize their results here. Briefly, the following calculations are needed:

- The observed magnitudes of the structure factors are brought to an absolute scale by Wilson's (1942) method. This is needed for the calculation of magnitudes of normalized structure factors (see below) and easy removal of the Patterson origin peak, which is very high and not informative.
- The Patterson function is computed from modified squared magnitudes of the normalized structure factors, the modification consisting mainly of the removal of the origin peak. Fourier maps computed from normalized structure factors have much sharper peaks than those computed from conventional ones, as explained in Chapter 8. The sharpening procedures normally used are therefore not necessary.

6.6 Overview of Advanced Applications

6.6.1 Superposition methods

The above discussion indicates that the problem underlying the interpretation of the Patterson function is to recover the set of atomic position vectors, referred to some arbitrary origin, from the observable set of interatomic vectors. Of course, the intrinsically high density of Patterson peaks makes such a proposition of doubtful applicability–when *all* the peaks are considered–but a consideration of the problem is of interest for its own sake, as well as in view of related applications. It was shown by Wrinch (1939) that the problem can indeed be solved if it is assumed that all the peaks in the Patterson function are completely resolved. The principle of Wrinch's method is the recognition of the Patterson function as a superposition of images of each atom in the rest of the structure. In fact, the set of Patterson vectors consists of all the vectors from atom 1 to all the atoms in the structure, all the vectors from atom 2 to all the atoms in the structure, and so on, until all the atoms in the asymmetric unit are exhausted. Proceeding along these lines, the set of Patterson peaks is the superposition of the image of atom 1 in all the atoms of the structure, the image of atom 2 in all the atoms of the structure and so on. The interpretation of the Patterson function here consists of finding the image that corresponds exactly to the underlying structure. Several "image seeking" functions have been proposed in the early 1950s, and these are rather extensively reviewed by Buerger (1959). For other discussions of the subject, and many pertinent references, the reader is referred to Lipson

and Cochran (1966) and Rossmann and Arnold (2001). Superposition methods have been largely superseded by computerized direct methods, to be discussed in a later chapter, but the principle is important for a discussion of rotation and translation functions and their synthesis in the method of molecular replacement, to be dealt with below.

6.6.2 Rotation functions

In our classification of Patterson vectors we saw that there exists a subset of vectors, called $\mathbf{u}_{\text{intra}}$ or self-Patterson vectors, which relate atoms within the same asymmetric unit–for all the asymmetric units present in the unit cell. The set of self-Patterson vectors within a given unit (not necessarily an asymmetric unit) specifies its orientation with respect to the basis vectors **a**, **b**, and **c** and this or related information can often be extracted with the aid of a rotation function, to be discussed in this subsection. We shall follow closely the formalism of Rossmann and Blow (1962), as presented in a much more recent review (Rossmann and Arnold 2001*b*), in an occasionally different notation. Let us imagine a structure consisting of two identical but differently oriented units. The sets of self-Patterson vectors of these units are also differently oriented, because of the different orientations of the units. In addition, we have the cross-Patterson vectors which, since they are interunit vectors, are usually longer than the self-Patterson ones. If we now consider the Patterson function of this structure, in a volume surrounding the origin and limited by the overall dimensions of a unit, it is likely that this volume will contain mainly self-Patterson vectors. Let us now denote the original Patterson function by $P_1(\mathbf{u})$ and superimpose on it a rotated version, or a self-Patterson function based on a model, denoted by $P_2(\mathbf{Cu})$ where **C** is a finite rotation operator (see, for example, Section A.3), that rotates the vector **u** through a certain angle and about a certain direction. No particular agreement or "overlap" is expected except when the two sets of self-vectors corresponding to the two structural units have the same orientation. Rossmann and Blow define the measure of agreement, for a general superposition of mutually rotated Patterson functions, by the integral

$$R = \int_U P_1(\mathbf{u}) P_2(\mathbf{Cu})\, \mathrm{d}^3\mathbf{u}, \tag{6.22}$$

where U is an envelope centered on the superimposed origins, and call this integral the *rotation function*. The rotation function attains its maximum value when the agreement is exact and the corresponding rotation measures the relative orientation of the sets of self-Patterson vectors involved. This rotation also yields the relative orientation of the two units.

As explained by Rossmann and Arnold (2001*b*), a computation of the rotation function which is mathematically simple and reasonably efficient makes use of a Fourier representation of the Patterson function. Omitting constants, and introducing the standard representation of the Patterson function, we can rewrite eqn (6.22) as

$$\begin{aligned} R &= \int_U \left[\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{u})\right] \\ &\times \left[\sum_{\mathbf{h}'} |F_{\mathbf{h}'}|^2 \exp(-2\pi i \mathbf{h}'^{\mathrm{T}}\mathbf{C}\mathbf{u})\right] \mathrm{d}^3\mathbf{u} \\ &= U \sum_{\mathbf{h}} \sum_{\mathbf{h}'} |F_{\mathbf{h}}|^2 |F_{\mathbf{h}'}|^2 G_{\mathbf{h}\mathbf{h}'}, \end{aligned} \tag{6.23}$$

where

$$\begin{aligned} UG_{\mathbf{h}\mathbf{h}'} &= \int_U \exp[-2\pi i(\mathbf{h}^{\mathrm{T}} + \mathbf{h}'^{\mathrm{T}}\mathbf{C})\mathbf{u}]\, \mathrm{d}^3\mathbf{u} \\ &\equiv \int_U \exp(-2\pi i \mathbf{g}^{\mathrm{T}}\mathbf{u})\, \mathrm{d}^3\mathbf{u}, \end{aligned} \tag{6.24}$$

and where $\mathbf{g}^{\mathrm{T}} = \mathbf{h}^{\mathrm{T}} + \mathbf{h}'^{\mathrm{T}}\mathbf{C}$. If U is the volume of a sphere of radius R, which is often a convenient assumption, we can solve the integral in eqn (6.24) in spherical coordinates as follows:

$$\begin{aligned} UG_{\mathbf{h}\mathbf{h}'} &= \int_0^R \int_0^\pi \int_0^{2\pi} \exp(-2\pi i g u \cos\theta) u^2 \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\varphi\, \mathrm{d}u \\ &= 2\pi \int_0^R u^2 \left\{ \int_0^\pi \exp(-2\pi i g u \cos\theta) \sin\theta \mathrm{d}\theta \right\} \mathrm{d}u \\ &= \frac{2}{g} \int_0^R u \sin(2\pi g u)\, \mathrm{d}u. \end{aligned} \tag{6.25}$$

A straightforward solution of the last integral in eqn (6.25) leads to the following expression for G:

$$G_{\mathbf{h}\mathbf{h}'} = \frac{3(\sin\alpha - \alpha\cos\alpha)}{\alpha^3}, \tag{6.26}$$

where $\alpha = 2\pi g R$ and $U = (4\pi R^3)/3$. The right-hand side of eqn (6.26) is an interference function for a spherical envelope of radius R at a distance H from the origin of the reciprocal space. As shown by Rossmann and Arnold (1993), only the terms with $HR \leq 1$ need be computed, the others being small enough to be neglected. The number of significant terms in the summation in eqn (6.23) is further reduced by the requirement that $\mathbf{C}^{\mathrm{T}}\mathbf{h}'$ be close to a reciprocal-lattice vector. This approach is probably fast enough if the number of terms is not unduly large. For problems involving macromolecules, a much faster but mathematically more demanding method called the *fast rotation function* (Crowther, 1972), is usually recommended.

The matrix-algebraic considerations pertinent to the Rossmann and Blow (1962) type of rotation function can be found in Chapter 2.3 of Volume B of the *International Tables for Crystallography* (Rossmann and Arnold, 2001*b*) and in Appendix A of this book.

6.6.3 Translation functions

The rotation function discussed above may furnish the correct orientations of the molecules or molecular fragments present in the structure. The vectors which participate in the computation of the rotation function are, as pointed out above, the self-Patterson vectors. The remaining problem of determination of the structure is to obtain the locations of the molecules or molecular fragments, the orientations of which are already known. In fact, soon after the establishment of the rotation function, the problem of finding the location of the molecules was addressed for some special situations (see, for example, Rossmann *et al.* 1964) The solution of this problem involves location-dependent Patterson vectors, and these are the cross-Patterson and Harker vectors, corresponding to cases 4 and 2 in Section 6.3. Both will be termed cross-Patterson vectors in what follows. The present overview of the above problem follows the article by Crowther and Blow (1967) on "A Method of Positioning a Known Molecule in an Unknown Crystal Structure". By a "known molecule" these authors mean that the atomic coordinates relative to some local origin, fixed in the molecule, are known and that the molecule has the same orientation as one of the molecules in the unknown crystal structure. Briefly, the method consists of superimposing cross-Patterson vectors obtained from a model on the observed Patterson function until a good agreement is found and a vector relating the local origins of two symmetry-related molecules can be obtained. We shall omit constant multipliers in the following derivation.

The cross-Patterson vectors relating molecule 0 with molecule 1 can be written as

$$P_{01}(\mathbf{u}) = \int_{\text{unit cell}} \rho_0(\mathbf{r})\rho_1(\mathbf{r}+\mathbf{u})\,\mathrm{d}^3\mathbf{r}, \tag{6.27}$$

where ρ_0 and ρ_1 are the electron density functions of two molecules related by the space-group operator $(\mathbf{P}, \mathbf{t})$, that is, the coordinates of equivalent atoms are related by $\boldsymbol{x}_j(1) = \boldsymbol{P}\boldsymbol{x}_j(0) + \boldsymbol{t}$. If we assume that the local origin of molecule 0 is located at position $\mathbf{d}$, and make use of the relation between the symmetry of the electron density in its function space and in the underlying coordinate point space,

$$\overline{(\mathbf{P},\mathbf{t})}\rho(\mathbf{y}) = \rho[(\mathbf{P},\mathbf{t})^{-1}\mathbf{y}]$$

(see, for example, Bricogne 2001, section 1.3.1.1.2), the expression in eqn (6.27) for the cross-Patterson vectors can be represented as

$$P_{01}(\mathbf{u},\mathbf{d}) = \int_{\text{unit cell}} \rho_{\mathrm{M}}(\mathbf{r}-\mathbf{d})\rho_{\mathrm{M}}[\mathbf{P}^{-1}(\mathbf{r}+\mathbf{u}-\mathbf{P}\mathbf{d}-\mathbf{t})]\,\mathrm{d}^3\mathbf{r}, \tag{6.28}$$

where ρ_{M} denotes the molecular density referred to a local origin.

If we now expand ρ_{M} as a Fourier series in terms of the structure factors F_{M}, of the known molecule, calculated relative to the local origin at $\mathbf{d}$, we obtain

$$
\begin{aligned}
P_{01}(\mathbf{u},\mathbf{d}) &= \int_{\text{unit cell}} \sum_{\mathbf{h}} F_{\mathrm{M}}(\mathbf{h}) \exp[-2\pi i \mathbf{h}^{\mathrm{T}}(\mathbf{r}-\mathbf{d})] \\
&\quad \times \sum_{\mathbf{h}'} F_{\mathrm{M}}(\mathbf{h}') \exp[-2\pi i \mathbf{h}'^{\mathrm{T}} \mathbf{P}^{-1}(\mathbf{r}+\mathbf{u}-\mathbf{P}\mathbf{d}-\mathbf{t})]\, \mathrm{d}^3\mathbf{r} \\
&= \sum_{\mathbf{h}} \sum_{\mathbf{h}'} F_{\mathrm{M}}(\mathbf{h}) F_{\mathrm{M}}(\mathbf{h}') \exp[2\pi i(\mathbf{h}^{\mathrm{T}}\mathbf{d} + \mathbf{h}'^{\mathrm{T}}\mathbf{P}^{-1}(\mathbf{P}\mathbf{d}+\mathbf{t}))] \\
&\quad \times \exp(-2\pi i \mathbf{h}'^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{u}) \int_{\text{unit cell}} \exp[-2\pi i(\mathbf{h}^{\mathrm{T}} + \mathbf{h}'^{\mathrm{T}}\mathbf{P}^{-1})\mathbf{r}]\, \mathrm{d}^3\mathbf{r}.
\end{aligned}
\tag{6.29}
$$

Since $\mathbf{P}$ is a crystallographic symmetry operator, the integral in the last line of eqn (6.29) vanishes, unless $\mathbf{h}^{\mathrm{T}} + \mathbf{h}'^{\mathrm{T}}\mathbf{P}^{-1} = \mathbf{0}^{\mathrm{T}}$ (see eqn (5.41)), in which case it equals the volume of the unit cell (a constant). Since in that case $-\mathbf{h}^{\mathrm{T}} = \mathbf{h}'^{\mathrm{T}}\mathbf{P}^{-1}$ and $\mathbf{h}'^{\mathrm{T}} = -\mathbf{h}^{\mathrm{T}}\mathbf{P} = -\mathbf{P}^{\mathrm{T}}\mathbf{h}$, we have

$$
P_{01}(\mathbf{u},\mathbf{d}) = \sum_{\mathbf{h}} F_{\mathrm{M}}(\mathbf{h}) F_{\mathrm{M}}(-\mathbf{P}^{\mathrm{T}}\mathbf{h}) \exp[2\pi i \mathbf{h}^{\mathrm{T}}(\mathbf{d}-\mathbf{P}\mathbf{d}-\mathbf{t})] \exp(2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{u}). \tag{6.30}
$$

Denoting the vector linking the local origins of the two molecules by $\mathbf{s} = (\mathbf{P},\mathbf{t})\mathbf{d} - \mathbf{d}$, and assuming Friedel's law to be valid, we have

$$
P_{01}(\mathbf{u},\mathbf{s}) = \sum_{\mathbf{h}} F_{\mathrm{M}}(\mathbf{h}) F_{\mathrm{M}}^{*}(\mathbf{P}^{\mathrm{T}}\mathbf{h}) \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{s}) \exp(2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{u}). \tag{6.31}
$$

The translation function is now defined by the integral

$$
T(\mathbf{t}) = \int_{\text{unit cell}} P_{01}(\mathbf{u},\mathbf{t}) P_{\text{obs}}(\mathbf{u})\, \mathrm{d}^3\mathbf{u}, \tag{6.32}
$$

where $P_{\text{obs}}(\mathbf{u})$ is the observed Patterson function of the crystal. When the computed cross-Patterson vectors P_{01} fit correctly the observed Patterson function P, the translation function is likely to have a large positive value. If we substitute in the standard Fourier representation of the Patterson function into eqn (6.32), we obtain

$$
\begin{aligned}
T(\mathbf{s}) &= \int_{\text{unit cell}} \sum_{\mathbf{h}} F_{\mathrm{M}}(\mathbf{h}) F_{\mathrm{M}}^{*}(\mathbf{P}^{\mathrm{T}}\mathbf{h}) \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{s}) \exp(2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{u}) \\
&\quad \times \sum_{\mathbf{h}'} |F_{\text{obs}}(\mathbf{h}')|^2 \exp(-2\pi i \mathbf{h}'^{\mathrm{T}}\mathbf{u})\, \mathrm{d}^3\mathbf{u}.
\end{aligned}
\tag{6.33}
$$

The integral vanishes unless $\mathbf{h} - \mathbf{h}' = \mathbf{0}$, and we finally have

$$
T(\mathbf{s}) = \sum_{\mathbf{h}} |F_{\text{obs}}(\mathbf{h})|^2 F_{\mathrm{M}}(\mathbf{h}) F_{\mathrm{M}}^{*}(\mathbf{P}^{\mathrm{T}}\mathbf{h}) \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{s}). \tag{6.34}
$$

The superposition of cross-Patterson peaks on the observed Patterson function has thus been expressed in terms of a conventional crystallographic Fourier series,

which can be rapidly evaluated by a standard Fast Fourier Transform routine and in which the interorigin vector should appear as a conspicuous peak. The two molecules can of course be related by an arbitrary space-group symmetry operation.

This translation function is subject to some modifications. The important ones are subtraction of the origin peak and self-Patterson vectors from $|F_{\rm obs}(\mathbf{h})|^2$ in eqn (6.34), and peak sharpening. We shall see that reasonable or good peak sharpening can be achieved if normalized, rather than conventional, structure factors are used in such computations (see Chapter 8). It is also possible to modify the Fourier coefficients of the translation function so that the summation will have peaks for all possible intermolecular vectors in the unknown structure (for details, see Crowther and Blow 1967), and not only for one such vector as in the above derivation. There are also a variety of alternative translation functions, described, for example, by Rossmann and Arnold (2001*b*), and by Beurskens *etal*. (1987), and in the references quoted therein.

6.6.4 Molecular replacement

This is a very nice technique for structure determination, which can be applied, for example, to unknown macromolecules consisting of known structural fragments. A typical example is the structure of a virus consisting of RNA which is folded in some unknown manner but surrounded by a shell of protein molecules of known structure. The main idea is to exploit the structural information available about the protein, from a crystal the structure of which has been solved, and apply it to the structure determination of the virus in which this protein appears in a variety of orientations.

We shall introduce this subject following, in part, Main and Rossmann (1966) but with a somehat different notation which will be consistent with that used in this book. Suppose that the structure of a molecule has been determined from a set of phased diffraction data $\{F(\mathbf{h}\}$, which we shall call the $(\mathbf{h})$ crystal for short. Suppose further that the same molecule appears N times in the asymmetric unit of another crystal, and that a set of diffraction amplitudes $\{|F(\mathbf{g})|\}$ has been collected from that crystal, which we shall call the $(\mathbf{g})$crystal. We also know the internal symmetry of the asymmetric unit of the $(\mathbf{g})$ crystal, that is, we can generate all N molecules given the orientation and location of one of them. We assume that these have been obtained by the use of the rotation and translation functions discussed in the previous subsections. The purpose of molecular replacement, in this context, is to express the *approximately phased* structure factors of the $(\mathbf{g})$ crystal in terms of the *known* structure factors of the $(\mathbf{h})$ crystal. For simplicity, let us assume that both crystals belong to the space group $P1$.

The structure factor for the $(\mathbf{g})$ crystal is given by

$$F(\mathbf{g}) = \int_{V_g} \rho_{\mathbf{g}}(\mathbf{r}) \exp(2\pi i \mathbf{g}^{\mathrm{T}}\mathbf{r})\, \mathrm{d}^3\mathbf{r}$$
$$= \sum_{n=1}^{N} \int_{U} \rho_{\mathbf{g}M}(\mathbf{r}_n) \exp(2\pi i \mathbf{g}^{\mathrm{T}}\mathbf{r}_n)\, \mathrm{d}^3\mathbf{r}_n, \tag{6.35}$$

where V_g is the volume of the unit cell of the (**g**) crystal, U is the volume of any one of the N identical molecules in this unit cell, $\rho_{\mathbf{g}}(\mathbf{r})$ is the electron density at point **r** (with coordinates $\mathbf{r}(x,\ y,\ z)$) in the unit cell of the (**g**) crystal and $\rho_{\mathbf{g}\mathrm{M}}(\mathbf{r}_n)$ is the electron density at point $\mathbf{r}_n$ (with coordinates $\mathbf{r}_n(x_n,\ y_n,\ z_n)$) in the nth molecule, in the (**g**) crystal.

The representations of the known non-crystallographic symmetry operators, relating molecules $1, 2, \ldots, N$, will be defined as $(\mathbf{C}_1, \mathbf{d}_1), (\mathbf{C}_2, \mathbf{d}_2), \ldots, (\mathbf{C}_N, \mathbf{d}_N)$, where $\mathbf{C}_1$ is a unit operator and $\mathbf{d}_1$ is a zero vector. We thus have

$$\mathbf{r}_n = \mathbf{C}_n\mathbf{r}_1 + \mathbf{d}_n, \qquad n = 1, 2, \ldots, N$$

for the relations between density-equivalent positions within the (**g**) unit cell.

The structure factor for the (**g**) crystal therefore becomes

$$F(\mathbf{g}) = \sum_{n=1}^{N} \int_{U} \rho_{\mathbf{g}\mathrm{M}}(\mathbf{r}_1) \exp[2\pi i \mathbf{g}^{\mathrm{T}}(\mathbf{C}_n\mathbf{r}_1 + \mathbf{d}_n)]\, \mathrm{d}^3\mathbf{r}_1 \tag{6.36}$$

since, by equivalence, $\rho_{\mathbf{g}\mathrm{M}}(\mathbf{r}_1) = \rho_{\mathbf{g}\mathrm{M}}(\mathbf{r}_2) = \cdots = \rho_{\mathbf{g}\mathrm{M}}(\mathbf{r}_N)$.

We now turn to the (**h**) crystal, about which everything is supposed to be known. The electron density at the point **r** in the (**h**) crystal is of course given by

$$\rho(\mathbf{r}) = \frac{1}{V_h} \sum_{\mathbf{h}} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}^{\mathrm{T}}\mathbf{r}). \tag{6.37}$$

Here comes the crucial step: we bring the origins to which the molecule in (**h**) and the molecule 1 in (**g**) are referred into coincidence, and find the rotation matrix, say **Y**, and the translation vector, say **y**, relating the equivalent points $\mathbf{r}_1$ (in (**g**)) and **r** (in (**h**)). Then, by definition, $\rho(\mathbf{r})$ is equivalent to $\rho_{\mathbf{g}\mathrm{M}}(\mathbf{r}_1)$. We have

$$\mathbf{r} = \mathbf{Y}\mathbf{r}_1 + \mathbf{y} \tag{6.38}$$

and from eqn (6.38) and the above argument it follows that

$$\rho_{\mathbf{g}\mathrm{M}}(\mathbf{r}_1) = \frac{1}{V_h} \sum_{\mathbf{h}} F(\mathbf{h}) \exp[-2\pi i \mathbf{h}^{\mathrm{T}}(\mathbf{Y}\mathbf{r}_1 + \mathbf{y})]. \tag{6.39}$$

If we now substitute eqn (6.39) into eqn (6.36), we obtain

$$
\begin{aligned}
F(\mathbf{g}) &= \sum_{n=1}^{N} \int_U \frac{1}{V_h} \sum_{\mathbf{h}} F(\mathbf{h}) \exp[-2\pi i \mathbf{h}^{\mathrm{T}} (\mathbf{Y}\mathbf{r}_1 + \mathbf{y})] \\
&\times \exp[2\pi i \mathbf{g}^{\mathrm{T}} (\mathbf{C}_n \mathbf{r}_1 + \mathbf{d}_n)] \, \mathrm{d}^3\mathbf{r}_1 \\
&= \frac{1}{V_h} \sum_{\mathbf{h}} \sum_{n=1}^{N} F(\mathbf{h}) \exp[2\pi i (\mathbf{g}^{\mathrm{T}} \mathbf{d}_n - \mathbf{h}^{\mathrm{T}} \mathbf{y})] \\
&\times \int_U \exp[2\pi i (\mathbf{g}^{\mathrm{T}} \mathbf{C}_n - \mathbf{h}^{\mathrm{T}} \mathbf{Y}) \mathbf{r}_1] \, \mathrm{d}^3\mathbf{r}_1. \qquad (6.40)
\end{aligned}
$$

It can already be seen from the above preliminary formulations that approximately phased structure factors for the (**g**) crystal can be expressed in terms of the known structure factors for the (**h**) crystal and other quantities assumed to be known. Various phase extensions and refinements can now be applied, the simplest being based on the principle of the "heavy-atom" method discussed above. This technique is not trivial to implement in practice, but recent advances in instrumentation and computer technology have given rise to remarkable progress in its applications. The interested reader is referred to Chapter 2.3 in Volume B of the *International Tables for Crystallography* (Rossmann and Arnold 2001*b*) and to more extensive discussions of the subject in Volume F (Rossmann and Arnold 2001*a*).

6.7 Exercises for Chapter 6

1. The convolution of two functions f and g is defined as

$$f(\mathbf{r}) * g(\mathbf{r}) = \int_V f(\mathbf{r}) g(\mathbf{u} - \mathbf{r}) \, \mathrm{d}^3\mathbf{r}.$$

A special kind of convolution, where $f(\mathbf{r}) = \rho(\mathbf{r})$ and $g(\mathbf{r}) = \rho(-\mathbf{r})$, is given by

$$P(\mathbf{u}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_V \rho(\mathbf{r}) \rho(\mathbf{u} + \mathbf{r}) \, \mathrm{d}^3\mathbf{r}, \qquad (6.41)$$

and is called the self-convolution, autoconvolution or autocorrelation function of the electron density; it is known as the Patterson function. Let us first represent the volume element $\mathrm{d}^3\mathbf{r}$ in terms of infinitesimals of fractional coordinates, so that the Patterson function integral can be written as a triple integral over the fractional coordinates. The volume element $\mathrm{d}^3\mathbf{r}$ can be taken as the volume of a parallelepiped constructed from the vectors $\mathrm{d}\mathbf{a}$, $\mathrm{d}\mathbf{b}$, and $\mathrm{d}\mathbf{c}$, that is,

$$
\begin{aligned}
\mathrm{d}^3\mathbf{r} &= \mathrm{d}\mathbf{a} \cdot (\mathrm{d}\mathbf{b} \times \mathrm{d}\mathbf{c}) \\
&= \mathrm{d}x \, \mathbf{a} \cdot (\mathrm{d}y \, \mathbf{b} \times \mathrm{d}z \, \mathbf{c}) \\
&= [\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})] \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \\
&= V \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z, \qquad (6.42)
\end{aligned}
$$

where V is the volume of the unit cell. The electron density function at a point with fractional coordinates xyz is given by

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F(hkl)\exp[2\pi i(hx+ky+lz)]. \tag{6.43}$$

Substitute $\rho(x,y,z)$ and $\rho(x+u_1, y+u_2, z+u_3)$ in eqn (6.41), using the volume element given by eqn (6.42), assume that Friedel's law is valid, and show that

$$P(u_1,u_2,u_3) = \frac{1}{V}\sum_h \sum_k \sum_l |F(hkl)|^2 \exp[2\pi i(hu_1+ku_2+lu_3)], \tag{6.44}$$

as indicated by eqn (6.7) in the text.
[Hint: show first that

$$\int_0^1 \exp[2\pi i(h+h')x]\,\mathrm{d}x = \delta_{h,\overline{h'}},$$

where δ_{ij} equals 1 or 0 according as $i=j$ or $i\neq j$, respectively.]

7 Structure-factor statistics

7.1 Introduction

Most of the procedures involved in structure determination, except perhaps the use of the Patterson function, are rather heavily dependent on probabilistic considerations. A brief summary of the underlying principles, oriented towards the applications encountered in this book, is presented in Appendix C.

Before we embark on the present description, it is interesting to try to answer to a question which suggests itself: why are probabilistic considerations needed in such a well-ordered system as a crystal, with such highly predictable directions of the diffracted beams? An answer to this not altogether trivial question can be found in the crystallographic literature (for example, Hauptman and Karle 1953b, Giacovazzo 1980, Shmueli and Weiss 1995), and we shall state here the most important part of the argument.

The structure factor has the form

$$F = \sum_{j=1}^{N} f_j \exp(i\theta_j), \tag{7.1}$$

where

$$\theta = 2\pi(hx_j + ky_j + lz_j). \tag{7.2}$$

Let us consider two situations: (i) an unknown structure and known, fixed hkl, and (ii) a fixed, known structure and each of hkl varying over a large range of integers. In situation (i), each of $x_j y_j z_j$ may assume any value in the range $(0, 1)$ with approximately the same probability, the approximation being due to the neglect of effects of excluded volume (atoms cannot interpenetrate each other). In other words, each of $x_j y_j z_j$ can be regarded as a random variable approximately uniformly distributed in the range$(0, 1)$. The same must hold for the fractional part of the scalar product $\mathbf{h} \cdot \mathbf{r}_j$, and hence θ_j can be taken as uniform in the range $(0, 2\pi)$. The functions $\cos\theta_j$ and $\sin\theta_j$ are therefore also random variables, and the structure factor in eqn (7.1) can be taken as a pair of sums of random variables. An analogous result is obtained for situation (ii) if we invoke a theorem first developed in the context of number theory (usually known as Weyl's theorem). Broadly, it states that if x, y, and z are rationally independent real numbers, and each of h, k, and l ranges over a wide range of integers, then the fractional part of the sum $hx + ky + lz$ is uniformly distributed in the range $(0, 1)$. Strictly, the variables x, y, and z are rationally independent if there does not exist a set of integers (m_1, m_2, m_3, m_4) such that $m_1 x + m_2 y +$

$m_3 z = m_4$. In practice, however, it is sufficient to require that x, y, and z do not approximate to small fractions.

In either case, the structure factor can be regarded as a pair of sums of random variables, and several methods of the classical theory of probability can be profitably applied to its statistics.

7.2 Wilson's method of scaling the intensities

Probabilistic considerations entered crystallography in the context of the fundamental problem of bringing the experimental results to the scale of the model that is meant to describe them. Such a procedure for scaling the diffracted intensities, already corrected for Lorentz, polarization, and other effects, and therefore known as reduced intensities, to the squared magnitude of the structure factor was first introduced by Wilson (1942) and will be described below.

If all the experimentally obtained diffracted intensities are on the same relative scale, the scale factor can be defined as

$$K = \frac{\langle |F_{\rm obs}(\mathbf{h})|^2 \rangle_{\mathbf{h}}}{\langle |F_{\rm model}(\mathbf{h})|^2 \rangle_{\mathbf{h}}}, \tag{7.3}$$

where $|F_{\rm obs}(\mathbf{h}|^2$ is the reduced intensity of the reflection corresponding to the diffraction vector $\mathbf{h}$, and $|F_{\rm model}(\mathbf{h})|^2$ is the reduced intensity of this reflection calculated from the available model of the structure, that is, depends on the available positional and displacement parameters. However, none of these parameters is known from the outset in the structure determination. Let us consider the denominator of eqn (7.3), which is just the average of the squared magnitude of the structure factor, as introduced in Chapter 5:

$$\begin{aligned} |F_{\rm model}(\mathbf{h})|^2 \equiv |F(\mathbf{h})|^2 &= \sum_{j=1}^{N} \sum_{k=1}^{N} f_j f_k \exp[2\pi i \mathbf{h} \cdot (\mathbf{r}_j - \mathbf{r}_k)] \\ &= \sum_{j=1}^{N} f_j^2 + 2 \sum_{j>k} f_j f_k \cos[2\pi \mathbf{h} \cdot (\mathbf{r}_j - \mathbf{r}_k)], \end{aligned} \tag{7.4}$$

where the scattering factor f_j contains the effects of atomic displacements. If we assume that the components of the interatomic vectors appearing in the double summation in eqn (7.4) are rationally independent it follows from the previous section that the fractional part of the scalar product $\mathbf{h} \cdot (\mathbf{r}_j - \mathbf{r}_k)$ is uniformly distributed over the range $(0, 1)$, provided a large number of diffraction vectors are involved. It follows that the double summation in the right-hand side of eqn (7.4) vanishes upon averaging over $\mathbf{h}$, and we have

$$\langle |F_{\rm model}(\mathbf{h})|^2 \rangle_{\mathbf{h}} \simeq \left\langle \sum_{j=1}^{N} f_j^2 \right\rangle_{\mathbf{h}}. \tag{7.5}$$

Following Wilson (1942), we assume the simplest model of atomic displacements, a single displacement parameter common to all the atoms in the structure. That

is, each atom is displaced isotropically from its mean position, with the same amplitude. As will be shown in Chapter 9, the atomic scattering factor can then be written as

$$\begin{aligned} f_j &= f_j^{(0)} \exp\left(-8\pi^2 \langle u^2 \rangle \frac{\sin^2\theta}{\lambda^2}\right) \\ &\equiv f_j^{(0)} \exp\left(-B\frac{\sin^2\theta}{\lambda^2}\right), \end{aligned} \tag{7.6}$$

where $f_j^{(0)}$is the scattering factor of an atom at rest, $\langle u^2 \rangle$ is the mean square displacement of an atom, and $B = 8\pi^2 \langle u^2 \rangle$. The averaging of the right-hand side of eqn (7.5) poses a problem, since the summand depends on the Bragg angle in a nonlinear manner. Wilson's (1942) approximate solution of this problem proceeds as follows:

1. Subdivide the available limiting sphere into shells of approximately equal occupancy, and such that the number of reflections in each shell is sufficiently large to make averaging meaningful and that dependence of the scattering factor on the Bragg angle is as linear as possible.
2. Perform the averaging of the numerator of eqn (7.3) over each shell, and denote such a shell average by $\langle |F_{\rm obs}(\mathbf{h})|^2 \rangle_{\mathbf{h},r}$.
3. Assuming a linear dependence of the scattering factor on the Bragg angle, the value of θ at the center of a shell can be taken to be a representative value for that shell; let us denote it by θ_r. Calculate the average in eqn (7.5) for the rth shell as

$$\langle |F_{\rm model}(\mathbf{h})|^2 \rangle_{\mathbf{h},r} \simeq \exp\left(-2B\frac{\sin^2\theta_r}{\lambda^2}\right) \left\langle \sum_{j=1}^{N} (f_j^{(0)})^2 \left(\frac{\sin\theta}{\lambda}\right) \right\rangle_{\mathbf{h},r}. \tag{7.7}$$

Finally, if we combine eqn (7.3) with eqn (7.7), we have for the rth shell the equation

$$\frac{\langle |F_{\rm obs}(\mathbf{h})|^2 \rangle_{\mathbf{h},r}}{\left\langle \sum_{j=1}^{N} (f_j^{(0)})^2 (\sin\theta/\lambda) \right\rangle_{\mathbf{h},r}} = K \exp\left(-2B\frac{\sin^2\theta_r}{\lambda^2}\right), \tag{7.8}$$

or, in logarithmic form,

$$\ln \frac{\langle |F_{\rm obs}(\mathbf{h})|^2 \rangle_{\mathbf{h},r}}{\left\langle \sum_{j=1}^{N} (f_j^{(0)})^2 (\sin\theta/\lambda) \right\rangle_{\mathbf{h},r}} = \ln K - 2B\frac{\sin^2\theta_r}{\lambda^2}. \tag{7.9}$$

If we now plot the left-hand side of eqn (7.9) against $\sin^2\theta_r/\lambda^2$, we obtain a straight line, the intercept of which is the logarithm of the scale factor K, and its slope yields the displacement parameter B (or $\langle u^2 \rangle$). This is widely known as the *Wilson plot*, which forms part of all practical procedures for structure determination. There exist variants of Wilson's method which enable one to incorporate partial structural information into the scaling procedure (see, for example, Main, 1975).

7.3 Basic Wilson statistics

A problem which is often encountered in the initial stages of the solution of a structure is to determine whether or not the crystal is centrosymmetric. As pointed out in Chapter 5, the observed diffraction intensities do not readily provide this information, and use must be made of complementary methods. Among the oldest are measurements of physical properties of crystals, which, when reliably observed, can provide an indication that a crystal is *not* centrosymmetric. Some important types of crystals are composed of asymmetric molecules of the same kind and of known asymmetry (for example protein molecules) and therefore *cannot* be centrosymmetric. However, none of these methods and considerations can provide positive evidence of the existence of a center of symmetry in the crystal. Wilson's (1949) pioneering work on probability distributions of the structure factor proved to be a most important step towards the resolution of the above ambiguity, as described below. If all the atoms are located in general positions and their scattering powers are not widely different, it is convenient to apply the central limit theorem, similarly to its first application by Wilson (1949). A simple (and useful for our purpose) form of this theorem states the following.

If $x_1, x_2, \ldots, x_n$ are independent, identically distributed random variables, each of them having the same mean μ and variance σ^2, then the sum

$$S_n = \sum_{j=1}^{n} x_j \tag{7.10}$$

tends to be normally distributed–independently of the distribution(s) of the individual random variables–with mean $M_n = n\mu$ and variance $\Sigma_n = n\sigma^2$, provided n is sufficiently large. This means that the probability density function (hereafter abbreviated to "pdf") of the sum S_n tends to the expression

$$p(S_n) \simeq \frac{1}{\sqrt{2\pi\Sigma_n}} \exp\left[-\frac{S_n - M_n}{2\Sigma_n}\right]. \tag{7.11}$$

Proofs of this version of the central limit theorem can be found in most textbooks on the theory of probability (for example, Cramér 1951). The theorem is also applicable when the means and variances of the random variables are not strictly identical, but there are no outstanding disparities among them. In this situation we shall write $M_n = \sum_{j=1}^{n} \mu_j$ and $\Sigma_n = \sum_{j=1}^{n} \sigma_j^2$, where μ_j and σ_j^2 are the mean and variance of the jth random variable.

Let us now derive the pdf of the magnitude of the structure factor when the underlying structure is, in turn, centrosymmetric and noncentrosymmetric. If the structure is centrosymmetric, and the origin is chosen at a center of symmetry, then to each atom at $\mathbf{r}_j$ there corresponds an atom at $-\mathbf{r}_j$ and the structure factor can be written as

$$F(\mathbf{h}) = \sum_{j=1}^{N/2} f_j[\exp(2\pi i\mathbf{h}\cdot\mathbf{r}_j) + \exp(-2\pi i\mathbf{h}\cdot\mathbf{r}_j)]$$
$$= 2\sum_{j=1}^{N/2} f_j \cos(2\pi\mathbf{h}\cdot\mathbf{r}_j), \tag{7.12}$$

where N is the number of atoms in the unit cell. Regarding $F(\mathbf{h})$ as a sum of random variables, the jth such variable is now $x_j = 2f_j\cos(2\pi\mathbf{h}\cdot\mathbf{r}_j)$, its mean is $\mu = \langle x_j\rangle = 2f_j\langle\cos(2\pi\mathbf{h}\cdot\mathbf{r}_j)\rangle = 0$, where the brackets denote angular averaging over the range (0,2π) and its variance is

$$\sigma_j^2 = 4f_j^2\langle\cos^2(2\pi\mathbf{h}\cdot\mathbf{r}_j)\rangle = 2f_j^2,$$

since the average of $\cos^2\theta$ when θ ranges from 0 to 2π is just 1/2. The pdf of F is therefore

$$p_{\rm c}(F) \simeq \frac{1}{\sqrt{2\pi\Sigma}}\exp\left(-\frac{F^2}{2\Sigma}\right), \tag{7.13}$$

and since the probabilities of F being positive and negative are the same, the pdf of the magnitude of F is

$$p_{\rm c}(|F|) \simeq \sqrt{\frac{2}{\pi\Sigma}}\exp\left(-\frac{|F^2|}{2\Sigma}\right). \tag{7.14}$$

If we now consider the normalized structure factor, defined as

$$E(\mathbf{h}) = \sum_{j=1}^{N} n_j \exp(2\pi i\mathbf{h}\cdot\mathbf{r}_j), \tag{7.15}$$

where

$$n_j = \frac{f_j}{\sqrt{\sum_{k=1}^{N} f_k^2}},$$

we obtain the following for the pdf of $|E|$, derived from a centrosymmetric crystal:

$$p_{\rm c}(|E|) \simeq \sqrt{\frac{2}{\pi}}\exp\left(-\frac{|E|^2}{2}\right). \tag{7.16}$$

The probabilities $p_{\rm c}(|F|)\,{\rm d}|F|$ and $p_{\rm c}(|E|)\,{\rm d}|E|$ are known as *centric* distributions. A structure factor derived from a noncentrosymmetric structure is given in its most general form by

$$F(\mathbf{h}) = A(\mathbf{h}) + iB(\mathbf{h}), \tag{7.17}$$

where

$$A(\mathbf{h}) = \sum_{j=1}^{N} f_j \cos(2\pi\mathbf{h}\cdot\mathbf{r}_j) \tag{7.18}$$

and

$$B(\mathbf{h}) = \sum_{j=1}^{N} f_j \sin(2\pi \mathbf{h} \cdot \mathbf{r}_j). \tag{7.19}$$

We now wish to find the joint pdf of A and B, which–on the assumption that A and B are uncorrelated–is given by

$$\begin{aligned} p_{\mathrm{a}}(A,B) &= p(A)p(B) \\ &\simeq \left[\frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{A^2}{2\sigma_A^2}\right)\right] \left[\frac{1}{\sqrt{2\pi\sigma_B^2}} \exp\left(-\frac{B^2}{2\sigma_B^2}\right)\right], \end{aligned} \tag{7.20}$$

since A and B can be regarded as sums of random variables, the means $\langle A \rangle$ and $\langle B \rangle$ are zero (see eqns (7.18) and (7.19)), and the central limit theorem is assumed to apply. We now have

$$\sigma_A^2 = \sum_{j=1}^{N} f_j \langle \cos^2(2\pi \mathbf{h} \cdot \mathbf{r}_j) \rangle = \frac{1}{2} \sum_{j=1}^{N} f_j = \frac{1}{2}\Sigma, \tag{7.21}$$

and the same value is obtained for σ_B^2. Equation (7.20) thus becomes

$$p_{\mathrm{a}}(A,B) \simeq \frac{1}{\pi\Sigma} \exp\left(-\frac{A^2+B^2}{\Sigma}\right) = \frac{1}{\pi\Sigma} \exp\left(-\frac{|F|^2}{\Sigma}\right). \tag{7.22}$$

If we now introduce the change of variables: $A = |F| \cos\varphi$ and $B = |F| \sin\varphi$, we see that the Jacobian resulting from this change of variables is just $|F|$. Hence

$$p_{\mathrm{a}}(A,B)\,\mathrm{d}A\,\mathrm{d}B = p_{\mathrm{a}}(|F|,\varphi)|F|\,\mathrm{d}|F|\,\mathrm{d}\varphi,$$

and if we finally integrate out the unknown phase φ, we obtain the following expression for the pdf of $|F|$:

$$p_{\mathrm{a}}(|F|) \simeq \frac{2|F|}{\Sigma} \exp\left(-\frac{|F|^2}{\Sigma}\right). \tag{7.23}$$

The corresponding expression for the magnitude of the normalized structure factor is

$$p_{\mathrm{a}}(|E|) \simeq 2|E| \exp(-|E|^2). \tag{7.24}$$

Equations (7.16) and (7.24) are plotted in Fig. 7.1 for values of $|E|$ ranging from 0 to 3 (most of the observed intensities lie in this range). Analogously to the centric case, the probabilities $p_{\mathrm{a}}(|F|))\,\mathrm{d}|F|$ and $p_{\mathrm{a}}(|E|)\,\mathrm{d}|E|$ are known as *acentric distributions*.

Discrimination between centrosymmetric and noncentrosymmetric structures based on a comparison of theoretical and experimental pdfs of $|E|$ is in principle possible but rather rarely done. Much greater popularity is enjoyed by cumulative distribution functions and moments of $|E|$, which are obtained with the aid of the pdfs derived above.

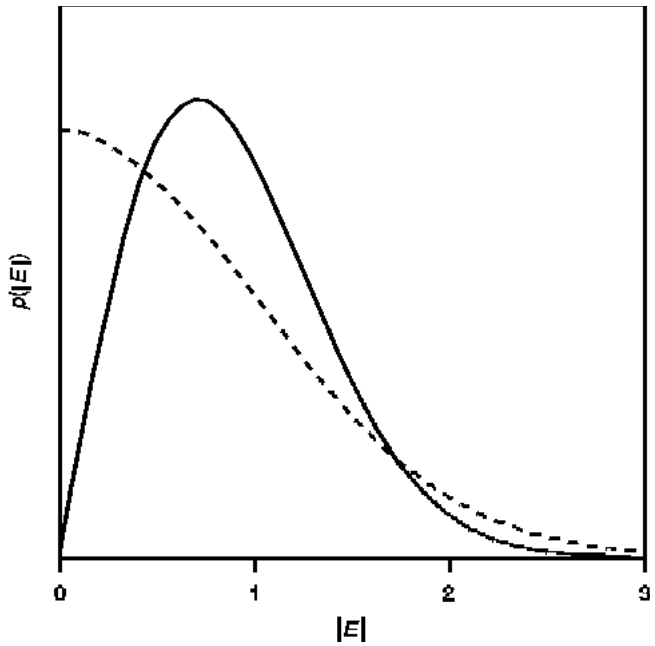


Fig. 7.1 Pdfs of $|E|$ based on the central limit theorem: dashed line, centric; solid line, acentric.

7.3.1 Cumulative distributions

The cumulative distribution function of some given $|E_{\rm m}|$ is the probability that $|E| \leq |E_{\rm m}|$. This is just the fraction of values of $|E|$ that do not exceed $|E_{\rm m}|$. So, the cumulative distribution function of $|E_{\rm m}|$, for a centric distribution, is given by

$$\begin{aligned} N_{\rm cen} &= \int_0^{|E_{\rm m}|} p_{\rm c}(|E|)\,{\rm d}|E| \\ &\simeq \sqrt{\frac{2}{\pi}} \int_0^{|E_{\rm m}|} \exp\left(-\frac{|E|^2}{2}\right)\,{\rm d}|E|. \end{aligned}$$

If we put $x = |E|/\sqrt{2}$, then ${\rm d}|E| = \sqrt{2}\,{\rm d}x$ and the upper limit becomes $|E_{\rm m}|/\sqrt{2}$. We therefore obtain

$$N_{\rm cen}(|E_{\rm m}|) \simeq \frac{2}{\sqrt{\pi}} \int_0^{|E_{\rm m}|/\sqrt{2}} \exp(-x^2)\,{\rm d}x. \tag{7.25}$$

However, the right-hand side of eqn (7.25) is the well-tabulated error function of the upper limit of the integral. Hence, in conventional notation, we have

$$N_{\rm cen}(|E_{\rm m}|) \simeq {\rm erf}\left(\frac{|E_{\rm m}|}{\sqrt{2}}\right) \tag{7.26}$$

(see, for example, Abramowitz and Stegun, 1972).

The cumulative distribution of $|E_{\rm m}|$, for an acentric distribution, is similarly obtained as

$$N_{\rm acen}(|E_{\rm m}|) = \int_0^{|E_{\rm m}|} p_{\rm a}(|E|)\,{\rm d}|E| \simeq 2\int_0^{|E_{\rm m}|} |E|\exp(-|E|^2)\,{\rm d}|E|. \tag{7.27}$$

If we put $x = |E|^2$, then ${\rm d}x = 2|E|\,{\rm d}|E|$ and it follows readily that

$$N_{\rm acen}(|E_{\rm m}|) \simeq 1 - \exp(-|E_{\rm m}|^2). \tag{7.28}$$

It is customary to express the cumulative distributions in terms of the normalized intensity, given by $z = |E|^2$. The corresponding expressions are

$$N_{\rm cen}(z) \simeq {\rm erf}(\sqrt{z/2})$$

and

$$N_{\rm acen}(z) \simeq 1 - \exp(-z).$$

The cumulative distributions of $|E|$ for centric and acentric distributions are compared in Fig. 7.2.

7.3.2 Moments of $|E|$

The status of centrosymmetry is often indicated by even and odd moments of the magnitude of the normalized structure factor. The general expressions are

$$\langle |E|^{2n}_{\rm cen}\rangle \simeq \sqrt{\frac{2}{\pi}}\int_0^\infty |E|^{2n}\exp\left(-\frac{|E|^2}{2}\right)\,{\rm d}|E| = (2n-1)!! = \frac{(2n)!}{2^n n!}, \tag{7.29}$$

$$\langle |E|^{2n+1}_{\rm cen}\rangle \simeq \sqrt{\frac{2}{\pi}}\int_0^\infty |E|^{2n+1}\exp\left(-\frac{|E|^2}{2}\right)\,{\rm d}|E| = \frac{2^{n+1/2}n!}{\sqrt{\pi}}, \tag{7.30}$$

$$\langle |E|^{2n}_{\rm acen}\rangle \simeq 2\int_0^\infty |E|^{2n+1}\exp(-|E|^2)\,{\rm d}|E| = n!, \tag{7.31}$$

$$\langle |E|^{2n+1}_{\rm acen}\rangle \simeq 2\int_0^\infty |E|^{2(n+1)}\exp(-|E|^2)\,{\rm d}|E| = \Gamma\left(n+\frac{3}{2}\right). \tag{7.32}$$

The expressions in eqns (7.29)–(7.32) can be derived by use of known definite integrals, given by Gradshteyn and Ryzhik (1980) and straightforward substitutions.

In the following definite integrals, the label GR;n denotes the number of the integral in Gradshteyn and Ryzhik (1980). For the derivation of eqn (7.29), we use

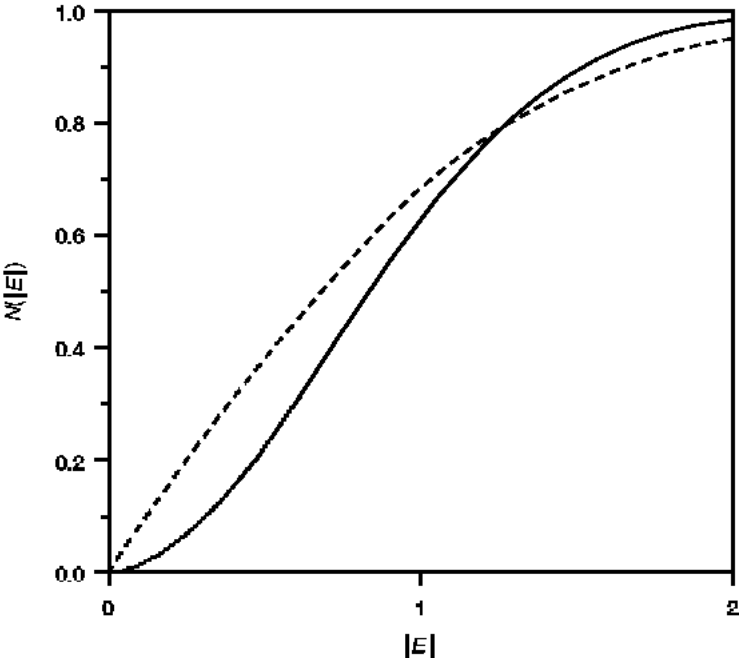


Fig. 7.2 Cumulative distributions of $|E|$ based on the central limit theorem: dashed line, centric; solid line, acentric.

$$\int_0^\infty x^{2n} e^{-px^2}\,\mathrm{d}x = \frac{(2n-1)!!}{2^{n+1}p^n}\sqrt{\frac{\pi}{p}}. \qquad [\mathrm{GR};3.461(2)]$$

For the derivation of eqns (7.30) and (7.31), we use:

$$\int_0^\infty x^{2n+1} e^{-px^2}\,\mathrm{d}x = \frac{n!}{2p^{n+1}}. \qquad [\mathrm{GR};3.461(3)]$$

For the derivation of eqn (7.32), we use

$$\int_0^\infty x^{\nu-1} e^{-\mu x^p}\,\mathrm{d}x = \frac{1}{p}\mu^{-\nu/p}\Gamma\left(\frac{\nu}{p}\right). \qquad [\mathrm{GR};3.478(1)]$$

In order to get an idea of the value of such moments in problems involving centric/acentric ambiguities, the numerical values of some low-order moments of $|E|$ are summarized in Table 7.1.

The second moment of $|E|$ equals unity for both centric and acentric distributions, and thus contains no information on the status of centrosymmetry. However, it is a valuable indicator of a correct computation of normalized structure factor amplitudes. The discrepancies between the values of the moments

Table 7.1 Some low-order moments of $||E|$

Average	Acentric	Centric
$\langle\|E\|\rangle$	0.886	0.798
$\langle\|E\|^2\rangle$	1.000	1.000
$\langle\|E\|^3\rangle$	1.329	1.596
$\langle\|E\|^4\rangle$	2.000	3.000
$\langle\|E\|^5\rangle$	3.323	6.383
$\langle\|E\|^6\rangle$	6.000	15.000

for the centric and acentric distributions increase with increasing order of the moment. However, there is a corresponding increase in the effects of errors of measurement and therefore the higher the moment, the less reliable it tends to be.

7.4 Non-ideal structure-factor statistics

The pdfs of $|E|$ given by eqns (7.16) and (7.24) depend on $|E|$ alone, and do not depend on the atomic composition and the space-group symmetry. It is well known, however, that the experimental intensity statistics depend on these two factors, and often rather strongly. The dependence on atomic composition is the more obvious one: if the asymmetric unit contains an outstandingly heavy atom among several similar light ones, the atomic contributions to the structure factor are no longer identically distributed, although they may still be independent. This impairs the validity of the version of the central limit theorem stated above, and may well lead to a pdf which is not Gaussian. As will be seen later, this is in fact so in the case of the space group $P\overline{1}$, but the pdf is hardly affected by the presence of an outstandingly heavy atom in the asymmetric unit of the space group P222. To account for this situation, we therefore need a pdf of $|F|$ or $|E|$ which takes into account both an arbitrary atomic composition and the possible space-group symmetry of the crystal under investigation. In contrast to pdfs obeying the central-limit theorem, which are sometimes termed *ideal pdfs*, we shall denote these composition- and symmetry-dependent pdfs as *nonideal pdfs*. This is not a new problem. Solutions were proposed as early as in 1953 by Hauptman and Karle (Karle and Hauptman 1953; Hauptman and Karle 1953*b*), by Klug (1958), and by Shmueli and Wilson (1981). Without going into detail, these solutions have in common a generalized pdf of the form

$$p_{\rm gen}(|E|) = p_{\rm CLT}(|E|) \sum_k c_k g_k(|E|), \tag{7.33}$$

where $p_{\rm CLT}(|E|)$ is a centric or acentric pdf based on the central limit theorem, c_k are coefficients dependent on symmetry and composition, and $g_k(|E|)$ are suitable basis functions. Briefly, the generalized pdf is taken as the corresponding ideal pdf multiplied by a correction factor which accounts for the departure of the generalized pdf from that predicted by the central limit theorem. For a

detailed description of the method and its successes, as well as its limitations, the interested reader is referred to Shmueli and Weiss (1995).

A different approach to the above problem is offered by a representation of the pdf in terms of a Fourier series (see Shmueli and Weiss (1995) and references therein). As will be seen later, this approach is, in general, a much more accurate one. We shall describe briefly its principle, illustrate a derivation of a pdf of $|E|$ for the space group $P\overline{1}$, and compare the various pdfs discussed with a reliably simulated distribution of $|E|$.

Consider a normalized structure factor for the space group $P\overline{1}$. This is given by

$$E(\mathbf{h}) = 2\sum_{j=1}^{N/2} n_j \cos\theta_j, \tag{7.34}$$

where $\theta_j = 2\pi\mathbf{h}\cdot\mathbf{r}_j$ is assumed to be uniformly distributed in the range $(-\pi, \pi)$. We also assume that the atomic contributions to the structure factor are statistically independent, that there is no noncrystallographic symmetry, and dispersion can be neglected. Under the latter assumption, the maximum value of the structure factor is just the sum of the normalized scattering factors, that is,

$$E_{\max} = \sum_{j=1}^{N} n_j. \tag{7.35}$$

Since the phase of E in eqn (7.34) is restricted to 0 or π only, that is, E can only be positive or negative, the minimum value of E is of course $-E_{\max}$. The value of E is thus confined to the closed interval $(-E_{\max}, E_{\max})$ and, therefore, $p(E) \equiv 0$ for any E lying outside this interval. Owing to this property, the function $p(E)$ may be expanded in a Fourier series with the understanding that only the values of $p(E)$ with $-E_{\max} \le E \le E_{\max}$ are relevant.

The Fourier representation of the pdf of E for $P\overline{1}$ (and, indeed, for any centrosymmetric space group) can be written as

$$p(E) = \frac{\alpha}{2}\sum_{m=-\infty}^{\infty} C_m \exp(-\pi i m E\alpha), \tag{7.36}$$

where $\alpha = 1/E_{\max}$. The Fourier coefficients of this series are given by

$$C_m = \int_{-1/\alpha}^{1/\alpha} p(E)\exp(\pi i m E\alpha)\mathrm{d}E \tag{7.37}$$

$$= \int_{-\infty}^{\infty} p(E)\exp(\pi i m E\alpha)\mathrm{d}E \tag{7.38}$$

$$\equiv \langle\exp(\pi i m E\alpha)\rangle \tag{7.39}$$

Equation (7.37), the basic form of the Fourier coefficient for the series in eqn (7.36), can be rewritten as eqn (7.38) since $p(E)$ is zero outside the interval

$(-E_{\max}, E_{\max})$. This, in turn, is the standard form of the expectation value of the function $\exp(\pi i m E \alpha)$, which is the characteristic function of the pdf $p(E)$ (see Appendix C). This function is further evaluated as follows.

$$\langle \exp(\pi i m E \alpha) \rangle = \left\langle \exp\left[2\pi i m \alpha \sum_{j=1}^{N/2} n_j \cos\theta_j \right] \right\rangle \tag{7.40}$$

$$= \left\langle \prod_{j=1}^{N/2} \exp(2\pi i m \alpha n_j \cos\theta_j) \right\rangle \tag{7.41}$$

$$= \prod_{j=1}^{N/2} \langle \exp(2\pi i m \alpha n_j \cos\theta_j) \rangle \tag{7.42}$$

$$= \prod_{j=1}^{N/2} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(2\pi i m \alpha n_j \cos\theta)\, \mathrm{d}\theta \right\} \tag{7.43}$$

$$\equiv \prod_{j=1}^{N/2} J_0(2\pi m \alpha n_j). \tag{7.44}$$

A direct substitution of E from eqn (7.34) leads to eqn (7.40), which is of course equal to eqn (7.41). Now, since the atomic contributions have been assumed to be statistically independent, the average of their product equals the product of their averages, given by eqn (7.42). Since, in addition, the angular variable θ_j has been assumed to be uniform in the range $(-\pi, \pi)$, the average of the contribution of the jth atom can be replaced by the expression in braces appearing in eqn (7.43). This expression, in turn, is the integral representation of the Bessel function of the first kind of order zero (see, for example, Abramowitz and Stegun 1972). The Fourier coefficient C_m thus becomes a product of such Bessel functions, as given by eqn (7.44).

Since $J_0(0) = 1$ we have $C_0 = 1$, and since $J_0(-x) = J_0(x)$ we have $C_{-m} = C_m$. Therefore, eqn (7.36) simplifies to

$$p(E) = \frac{\alpha}{2} \left(1 + 2 \sum_{m=1}^{\infty} C_m \cos(\pi m E \alpha) \right), \tag{7.45}$$

and in view of the fact that $p(E)$ is symmetric in E, the expression for the pdf of $|E|$ for the space group $P\overline{1}$ becomes

$$p(|E|) = \alpha \left\{ 1 + 2 \sum_{m=1}^{\infty} \left[\prod_{j=1}^{N/2} J_0(2\pi m n_j \alpha) \right] \cos(\pi m |E| \alpha) \right\}. \tag{7.46}$$

Fourier representations of the pdfs of $|E|$ have been derived for all space groups except the highest 24 cubic space groups (Rabinovich *etal*. 1991). The

latter can be adequately dealt with by the correction-factor method outlined above even if the atomic heterogeneity is rather drastic, and attempts at derivation of Fourier representations of pdfs of $|E|$ for these space groups have not been made. However, the Fourier method shows its power in the cases of lowest symmetry, and one such case is illustrated by the following simulation, where the asymmetric unit was assumed to contain fourteen carbon atoms and one uranium. This is certainly a highly heterogeneous composition, and it is unlikely that the central-limit theorem with "identically distributed random variables" will be valid in this case. The normalized scattering factors were computed as

$$n_j = \frac{f_j}{\left(\sum_{k=1}^{N} f_k^2\right)^{1/2}} \simeq \frac{Z_j}{\left(\sum_{k=1}^{N} Z_k^2\right)^{1/2}},$$

where Z_j is the atomic number of atom j. Using these approximate scattering factors, and taking each θ_j in eqn (7.34) as a computer-generated pseudorandom number multiplied by 2π, a computation of 100 000 values of $|E|$ was carried out. During this computation, a frequency histogram of $|E|$ was constructed by simply counting the number of values of $|E|$ falling between 0 and 0.05, 0.05 and 0.1, etc. Altogether, the range $(0, 3)$ was subdivided into 60 equal intervals. Such a frequency histogram gives a representation of the pdf of $|E|$ which depends correctly on the atomic composition and space-group symmetry. Pdfs based on Wilson statistics, the correction-factor method, and the Fourier method were then computed at the values of $|E|$ corresponding to the midpoints of the histogram channels, and all three pdfs and the histogram were brought to the same scale. The result of this computation is shown in Fig. 7.3. The superior performance of the Fourier method and the breakdown of the central limit theorem are evident. It should be noted, however, that for an equal-atom structure all three pdfs would agree very well with the simulated histogram (and with each other).

Figure 7.3 shows clearly that the Fourier pdf takes account very well of the drastic atomic heterogeneity inherent in the simulation. The obvious breakdown of the central limit theorem in this case is due mainly to the assumed presence of the outstandingly heavy uranium atom which is associated with a distribution very different from those of the carbon atoms. Similar discrepancies from the ideal pdf are seen when the the number of (equal) atoms in the asymmetric unit is very small.

7.5 Exercises for Chapter 7

Probabilistic considerations arise in various aspects of crystallography owing to the frequent validity of the assumption that the fractional part of the scalar product $\mathbf{h}\cdot\mathbf{r}$ is uniformly distributed in the range $[0, 1]$. The purpose of these (computational) exercises is to simulate probability density functions of the magnitude of the normalized structure factor for the triclinic space groups $P1$ and $P\overline{1}$. The case of identical atoms as well as that of heterogeneous atomic composition will be considered. The reader should thereby gain a better understanding of the equations derived in the text.

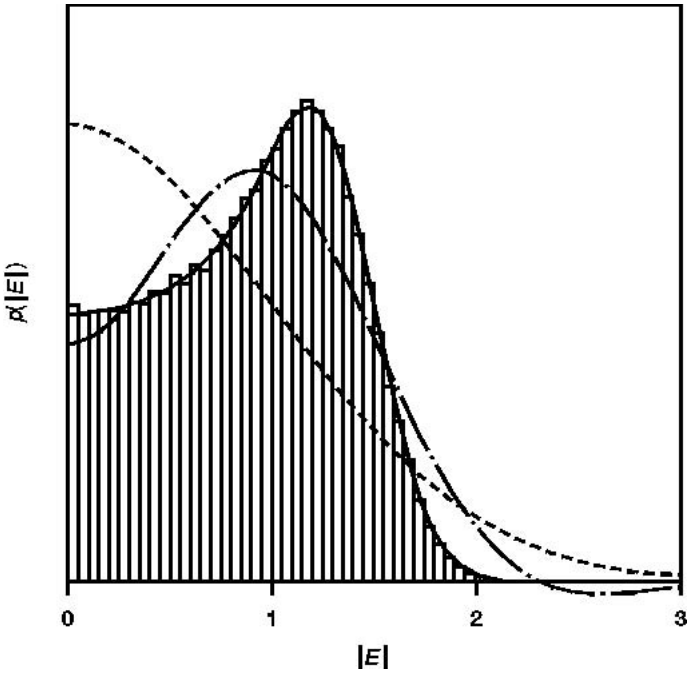


Fig. 7.3 Pdfs of $|E|$ for the space group $P\overline{1}$. Dashed line, based on the central limit theorem; dashed-dotted line, based on a four-term correction-factor expansion; solid line, based on the Fourier method. The contents of the asymmetric unit are assumed to be 14 carbon atoms and one uranium. All three pdfs are brought to the scale of the simulated histogram (see text).

1. *Statistical properties of centric and acentric diffraction patterns.* Assuming that all the atoms are identical, the expressions for the normalized structure factors are given by

$$E(\mathbf{h}) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{r_j}) \tag{7.47}$$

and

$$E(\mathbf{h}) = \frac{2}{\sqrt{N}} \sum_{j=1}^{N/2} \cos(2\pi \mathbf{h} \cdot \mathbf{r_j}) \tag{7.48}$$

for $P1$ and $P\overline{1}$, respectively. Equation (7.47) can be written as

$$E(\mathbf{h}) = A(\mathbf{h}) + iB(\mathbf{h}),$$

where

$$A(\mathbf{h}) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \cos(2\pi \mathbf{h} \cdot \mathbf{r_j}),$$

and

$$B(\mathbf{h}) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \sin(2\pi \mathbf{h} \cdot \mathbf{r_j}),$$

and hence the magnitude $|E|$ for $P1$ can be computed as $(A^2 + B^2)^{1/2}$. The magnitude $|E|$ for $P\overline{1}$ is simply the magnitude of the right-hand side of (7.48).

(a) Compute and store 6000 values of $|E|$ for one of the space groups, while assuming $N = 30$ and replacing the scalar product $\mathbf{h} \cdot \mathbf{r}$ with a computer-generated pseudorandom number (uniform in the $(0, 1)$ range). Most computer languages offer this facility. Remember that the "random number" must be drawn separately for each atomic contribution to each structure factor.
(b) Construct a histogram from the stored values, for $|E|$ ranging from 0 to 3 in 30 equal channels (count the number of values falling between 0.0 and 0.1, between 0.1 and 0.2 etc.) and arrange it in tabular form. A graphic representation can be used but is not necessary.
(c) Repeat the calculation for the second space group and compare the two histograms. This is the basis of the resolution of space-group ambiguities.

2. *Comparison of simulation and theory.* Normalize the histograms obtained in Exercise 1 to unity and compute the approximate theoretical values of the probability density functions

$$p(|E|) \simeq \begin{cases} 2|E| \exp(-|E|^2) & \text{for } P1 \\ \sqrt{\frac{2}{\pi}} \exp(-|E|^2/2) & \text{for } P\overline{1} \end{cases}$$

at the midpoints of the corresponding histogram channels. Compute, for each pdf–histogram pair, a discrepancy factor:

$$R = \left\{ \frac{\sum_{i=1}^{M} [p_i(|E|) - h_i(|E|)]^2}{\sum_{i=1}^{M} h_i^2(|E|)} \right\}^{1/2},$$

where h_i is the (normalized) histogram count in the ith channel, p_i is the pdf of $|E|$ at its midpoint, and M is the number of channels in the histogram. Would the discrepancy be much smaller if you were to repeat the calculation for a histogram containing twice as many values of $|E|$?

3. *Effect of atomic heterogeneity on the pdf of $|E|$ for the space group $P\overline{1}$.* Assume that the asymmetric unit of $P\overline{1}$ contains one (heavy) atom of atomic number Z_{H} and $N/2 - 1$ (equal light) atoms of atomic number Z_{L}.

(a) Make use of the approximation

$$n_j = \frac{f_j}{\left(\sum_{k=1}^{N} f_k^2\right)^{1/2}} \simeq \frac{Z_j}{\left(\sum_{k=1}^{N} Z_k^2\right)^{1/2}}$$

and express the normalized scattering factors n_{H} and n_{L} in terms of the ratio $\rho = Z_{\mathrm{H}}/Z_{\mathrm{L}}$ and the number of atoms, N, in the unit cell of $P\overline{1}$.
(b) The normalized structure factor for the above model is given by

$$E(\mathbf{h}) = 2n_{\mathrm{H}} \cos(2\pi \mathbf{h} \cdot \mathbf{r}_1) + 2 \sum_{j=2}^{N/2} n_{\mathrm{L}} \cos(2\pi \mathbf{h} \cdot \mathbf{r}_j).$$

Follow the procedure of Exercise 1 and construct simulated distributions of the magnitude $|E|$, for several choices of the ratio ρ (say, $\rho = 5, 10, 15$) and $N = 30$. Perform the simulation in the range of $|E|$ $[0, 3]$, in 30 equal channels.

(c) The Fourier pdf for the space group $P\overline{1}$ is given by eqn (7.46), which was derived in Section 7.4. Express $\alpha = E_{\max}^{-1}$ in terms of ρ and N and compute the Fourier pdf at the midpoints of the histogram channels obtained in (b), for all the values of ρ for which simulated histograms were computed. For each value of ρ, bring the the pdf and the histogram to a common scale (either normalize the histogram or scale up the pdf) and evaluate the discrepancy factor as defined in Exercise 2. A Fortran function, from which the Bessel function, $J_0(x)$, can be computed, is given below.

A routine for the computation of $J_0(x)$.

```
FUNCTION BESSJ0(X)
REAL*8 Y,P1,P2,P3,P4,P5,Q1,Q2,Q3,Q4,Q5,R1,R2,R3,R4,R5,R6
REAL*8 S1,S2,S3,S4,S5,S6,A,B,F,G
DATA P1,P2,P3/1.D0,-.1098628627D-2.2734510407D-4/
DATA P4,P5/-.2073370639D-5,.2093887211D-6/
DATA Q1,Q2/-.1562499995D-1,.1430488765D-3/
DATA Q3,Q4,Q5/-.6911147651D-5,.7621095161D-6,-.934945152D-7/
DATA R1,R2,R3/57568490574.D0,-13362590354.D0,651619640.7D0/
DATA R3,R4,R5/-11214424.18D0,77392.33017D0,-184.9052456D0/
DATA S1,S2,S3/57568490411.D0,1029532985.D0,9494680.718D0/
DATA S4,S5,S6/59272.64853D0,267.8532712D0,1.D0/
IF(ABS(X).LT.8.)THEN
Y=X**2
A=(R1+Y*(R2+Y*(R3+Y*(R4+Y*(R5+Y*R6)))))
B=(S1+Y*(S2+Y*(S3+Y*(S4+Y*(S5+Y*S6)))))
BESSJ0=A/B
ELSE
AX=ABS(X)
Z=8./AX
Y=Z**2
XX=AX-.785398164
F=(P1+Y*(P2+Y*(P3+Y*(P4+Y*P5))))
G=(Q1+Y*(Q2+Y*(Q3+Y*(Q4+Y*Q5))))
BESSJ0=SQRT(.636619772/AX)*(COS(XX)*F-Z*SIN(XX)*G)
ENDIF
RETURN
END
```

8 Direct methods

8.1 Introduction

The phase problem of crystallography was stated and defined in Section 5.8 and will be restated here for convenience: *in order to determine the crystal structure we need both the magnitudes and phases of the structure factors, $F(\mathbf{h})$, whereas only quantities related to their magnitudes are furnished by the experiment.*

In Chapter 6, we saw an indirect method for solution of this problem, which makes use of relationships between the reduced intensity and the interatomic vectors, the Patterson method. This method is often powerful but relies on known structural features such as the presence of a heavy atom or a known molecular geometry. The methods discussed in this chapter aim at a direct determination of the phases of the structure factors and hence at a structure determination without any prior information. These methods, known as *direct methods*, are responsible for the great majority of the known crystal and molecular structures of small and medium-sized molecular compounds. On the other hand, the crystal and molecular structures of proteins are, as of today, most usually elucidated with the aid of Patterson techniques. However, the phasing of structure factors obtained from macromolecules is subject to very active research, and the methodology is likely to change in the not too distant future.

The notation employed in this chapter will be consistent with the rest of the book, as well as with the literature. So, the symbol $\mathbf{h}$ will stand for $h\mathbf{a}^*+k\mathbf{b}^*+l\mathbf{c}^*$ as usual, but in some places it will be understood as (h,k,l), $(h+k,l)$, etc. It will then be referred to as the "array $\mathbf{h}$" rather than the "vector $\mathbf{h}$", where possible. Before any further discussion, the term "direct phase determination" must be qualified. The structure factor is given by

$$F(\mathbf{h}) = \sum_{j=1}^{N} f_j(|\mathbf{h}|) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) = |F(\mathbf{h})| \exp[i\varphi(\mathbf{h})],$$

where $\mathbf{r}_j$ is the position vector of the jth atom, referred to the unit-cell origin, and $\varphi(\mathbf{h})$ is the phase of the structure factor corresponding to the diffraction vector $\mathbf{h}$ (cf. Section 5.8). Suppose that the origin is shifted by some vector, say $\boldsymbol{\rho}$. The position vector of the jth atom relative to the new origin is

$$\mathbf{r}'_j = \mathbf{r}_j - \boldsymbol{\rho},$$

and the structure factor becomes

$$\begin{aligned} F'(\mathbf{h}) &= \sum_{j=1}^{N} f_j(|\mathbf{h}|) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}'_j) \\ &= \exp(-2\pi i \mathbf{h} \cdot \boldsymbol{\rho}) F(\mathbf{h}) \\ &= |F(\mathbf{h})| \exp\{i[\varphi(\mathbf{h}) - 2\pi \mathbf{h} \cdot \boldsymbol{\rho}]\}. \end{aligned} \tag{8.1}$$

It is evident from eqn (8.1) that $|F'(\mathbf{h})| = |F(\mathbf{h})|$ and therefore

$$F'(\mathbf{h}) = |F'(\mathbf{h})| \exp[i\varphi'(\mathbf{h})]$$

where

$$\varphi'(\mathbf{h}) = \varphi(\mathbf{h}) - 2\pi \mathbf{h} \cdot \boldsymbol{\rho}. \tag{8.2}$$

We thus see that the phase of the structure factor depends on the location or choice of the unit-cell origin, and that an origin shift by a vector $\boldsymbol{\rho}$ gives rise to a change in the phase by an amount $2\pi \mathbf{h} \cdot \boldsymbol{\rho}$. We shall give an introduction to the origin problem in Section 8.3.

In the next section we shall consider combinations of phases which are independent of the choice of the origin, called structure invariants, and combinations of phases which are invariant with respect to the choice of the origin if it is in any of the allowable locations for a given space group; these are called structure seminvariants. These considerations are of the utmost importance to the development of direct methods and their application to phase determination.

The direct methods originated from exact relationships first put forward by Harker and Kasper in 1947, and generalized further by other authors, notably Karle and Hauptman. We shall devote a section to this interesting topic.

We then proceed to the derivation of the Sayre equation and the tangent formula, which are the basis of the applications of the classic direct methods. These relationships are firmly based on the concepts of structure invariants and seminvariants, as well as on the specification of the origin. Although they are exact, the applications of these relationships to phase determination call for probabilistic considerations. We shall first describe in some detail a probability density function (pdf) of a structure invariant based on the central limit theorem, and comment on an exact pdf of such an invariant, based on the Fourier method presented in Chapter 7. The next section will illustrate with some examples the route from the deterministic Harker–Kasper inequalities to applications of probability methods in cases where the magnitudes of the structure factors are not large enough for these inequalities.

We conclude this chapter with some remarks on the practical aspects of the solution of the phase problem.

At least in the case of X-ray diffraction, we seek representations of the structure factor in which the angular dependence due to the scattering factor of an atom at rest, as well as that due to the Debye–Waller factor, is eliminated insofar as possible. One such representation is the normalized structure factor, $E(\mathbf{h})$,

already introduced and defined in Chapter 7. We shall only point out that the normalized scattering factor, defined as

$$n_j = \frac{f_j}{\left(\sum_{k=1}^{N} f_j^2\right)^{1/2}},$$

reduces, for an equal-atom structure, to

$$n_j = \frac{f}{[(Nf^2)^{1/2}]} = \frac{1}{\sqrt{N}}.$$

Another such representation is afforded by the unitary structure factor, defined as

$$U(\mathbf{h}) = \sum_{j=1}^{N} u_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

where

$$u_j = \frac{f_j}{\sum_{k=1}^{N} f_k}$$

is the unitary scattering factor. When all the atoms are identical, u_j obviously reduces to $1/N$. This representation will be used in the discussion of the Harker–Kasper inequalities in Section 8.4.

8.2 Phase invariants and seminvariants

8.2.1 Phase invariants

We have already seen that if the unit-cell origin is shifted by a vector $\boldsymbol{\rho}$, the phase of the structure factor changes, as given by eqn (8.2), which is restated below:

$$\varphi'(\mathbf{h}) = \varphi(\mathbf{h}) - 2\pi \mathbf{h} \cdot \boldsymbol{\rho}. \tag{8.3}$$

Following Hauptman (1980), consider any finite linear combination of both sides of eqn (8.3) with integer coefficients $A_{\mathbf{h}}$:

$$\sum_{\mathbf{h}} A_{\mathbf{h}} \varphi'_{\mathbf{h}} = \sum_{\mathbf{h}} A_{\mathbf{h}} \varphi_{\mathbf{h}} - 2\pi \left(\sum_{\mathbf{h}} A_{\mathbf{h}} \mathbf{h} \right) \cdot \boldsymbol{\rho}. \tag{8.4}$$

It is evident that if

$$\sum_{\mathbf{h}} A_{\mathbf{h}} \mathbf{h} = 0, \tag{8.5}$$

then

$$\sum_{\mathbf{h}} A_{\mathbf{h}} \varphi'_{\mathbf{h}} = \sum_{\mathbf{h}} A_{\mathbf{h}} \varphi_{\mathbf{h}}, \tag{8.6}$$

no matter what the origin shift $\boldsymbol{\rho}$ may be. Hence the linear combination in eqn (8.6) has the same value for every choice of the origin, and is a structure

invariant. If each of the coefficients equals +1, the sum of the phases is a structure invariant and so is the product of the corresponding structure factors.

Let us consider the problem in a less formal manne. For a single structure factor $F(\mathbf{h})$, what is the vector $\mathbf{h}$ for which $F(\mathbf{h})$ is independent of the choice of the origin? It follows from eqn (8.3) that the answer is: for $\mathbf{h} = \mathbf{0}$. We have

$$F(\mathbf{0}) = \sum_{j=1}^{N} f_j(0) = \sum_{j=1}^{N} Z_j$$

and the sum of atomic numbers is certainly a structure invariant.

Considering the product $F(\mathbf{h})F(\mathbf{k})$, for what vectors $\mathbf{h}$ and $\mathbf{k}$ is the product independent of the choice of the origin? If we carry out the same calculation that led to eqn (8.1), we have

$$F'(\mathbf{h})F'(\mathbf{k}) = |F(\mathbf{h})F(\mathbf{k})| \exp\{i[\varphi(\mathbf{h}) + \varphi(\mathbf{k}) - 2\pi(\mathbf{h}+\mathbf{k})\cdot\boldsymbol{\rho}]\}. \tag{8.7}$$

Both sides of eqn (8.7) can be the same for any vector $\boldsymbol{\rho}$ only if

$$\mathbf{h} + \mathbf{k} = \mathbf{0} \text{ or } \mathbf{k} = -\mathbf{h}.$$

Our product now becomes $F(\mathbf{h})F(-\mathbf{h})$, which is a structure invariant. Moreover, if Friedel's law holds the product becomes

$$F(\mathbf{h})F(-\mathbf{h}) = F(\mathbf{h})F^*(\mathbf{h}) = |F(\mathbf{h})|^2,$$

which, as a quantity related to the diffracted intensity, is certainly independent of the choice of the origin.

Let us now consider the triple product $F(\mathbf{h})F(\mathbf{k})F(\mathbf{l})$ and ask an analogous question: for what vectors $\mathbf{h}$, $\mathbf{k}$, and $\mathbf{l}$ will the triple product be independent of the choice of the origin? Following a calculation analogous to those which led us to eqns (8.1) and (8.7), we obtain

$$F'(\mathbf{h})F'(\mathbf{k})F'(\mathbf{l}) = |F(\mathbf{h})F(\mathbf{k})F(\mathbf{l})| \exp\{i[\varphi(\mathbf{h})+\varphi(\mathbf{k})+\varphi(\mathbf{l})-2\pi(\mathbf{h}+\mathbf{k}+\mathbf{l})\cdot\boldsymbol{\rho}]\}. \tag{8.8}$$

Again, both sides of eqn (8.8) can be the same for any vector $\boldsymbol{\rho}$ only if

$$\mathbf{h} + \mathbf{k} + \mathbf{l} = \mathbf{0} \text{ or } \mathbf{l} = -\mathbf{h} - \mathbf{k}.$$

It follows that

$$\varphi(\mathbf{h}) + \varphi(\mathbf{k}) + \varphi(-\mathbf{h} - \mathbf{k}) \tag{8.9}$$

is independent of the choice of the origin. This lowest nontrivial phase invariant, of great importance, is known as the *three-phase invariant*. Similar considerations lead to higher phase invariants (see, for example, Giacovazzo 1980, 1998), which, however, will not be treated in any detail in this book.

8.2.2 Structure seminvariants

In the discussion of structure invariants, the choice of the origin was completely unrestricted. We know, however, that apart from the space group $P1$, the origin of the unit cell is chosen at some symmetry element or intersection of such elements. For example, in the space group $P\overline{1}$ there are eight independent centers of symmetry in the unit cell, and the origin is chosen at one of them. A translation of the origin from one center to another does not change the functional form of the structure factor. In the general case, origins which maintain the same functional form of the structure factor are called *equivalent origins* and are referred to as *permissible* or *allowed* origins. Also, translations between allowed origins are called *allowed translations*. The literature on direct methods observes the standard choices of allowed origins, as given in Volume A of the *International Tables for Crystallography* (Hahn 1996).

Consider again eqn (8.3)

$$\varphi'(\mathbf{h}) = \varphi(\mathbf{h}) - 2\pi\mathbf{h}\cdot\boldsymbol{\rho}.$$

If no restrictions are placed on $\boldsymbol{\rho}$, the phase of the structure factor will, in general, undergo a change as a result of the shift of the origin by this vector. If, however, $\boldsymbol{\rho}$ is an allowed translation, there exist classes of vectors $\mathbf{h}$ for which the scalar product $\mathbf{h}\cdot\boldsymbol{\rho}$ is an integer. When this is so, the phase of the structure factor changes by an integer multiple of 2π, which is a trivial phase change that can be ignored. Such phases are called structure seminvariants. Summing up, the phase $\varphi(\mathbf{h})$ is called a structure seminvariant if it does not change (except by an integer multiple of 2π) when the origin is shifted by an allowed translation.

Let us return to the space group $P\overline{1}$. An allowed translation vector is of the form

$$\boldsymbol{\rho} = p\mathbf{a} + q\mathbf{b} + r\mathbf{c}$$

where each of p, q, or r can be 0 or 1/2. The scalar product $\mathbf{h}\cdot\boldsymbol{\rho}$ is therefore given by $hp+kq+lr$ and can be an integer for all the allowed translations if and only if h, k, and l are all even integers. Hence, for this space group, $\varphi(2\mathbf{h})$ is a structure seminvariant for an array $\mathbf{h}$ with components of any parity.

We have seen above an example of a single phase in $P\overline{1}$ which is a structure seminvariant. Consider, in the same space group, two arrays $\mathbf{h}_1$ and $\mathbf{h}_2$, the corresponding indices of which have the same parity (for example the reflections 136 and 514). The components of $\mathbf{h}_1+\mathbf{h}_2 \equiv 2\mathbf{h}$ are then all even. The combination

$$\varphi(\mathbf{h}_1) + \varphi(\mathbf{h}_2) + \varphi(-2\mathbf{h})$$

is a structure invariant and, since $\varphi(-2\mathbf{h})$ is itelf a structure seminvariant, the combination

$$\mathbf{h}_1 + \mathbf{h}_2$$

is also a structure seminvariant–a two-phase seminvariant.

A completely general presentation of the invariant and seminvariant properties of all the 230 space groups was given by Hauptman and Karle in their

pioneering series of papers (Hauptman and Karle 1953a, 1956, 1959), it was also studied by Giacovazzo (1974) and its latest presentation was given by Giacovazzo (2001) in Chapter 2.2 of Volume B of *International Tables for Crystallography*. We shall add here, following Giacovazzo (1980), an algebraic consideration which facilitates this extension. It was shown earlier, in Section 3.8, that if the space-group origin is shifted by a vector $\boldsymbol{\rho}$, the rotation part of the space-group operator $(\mathbf{P}_s, \mathbf{t}_s)$ remains unchanged, while the new translation part becomes

$$\mathbf{t}'_s = (\mathbf{P}_s - \mathbf{I})\boldsymbol{\rho} + \mathbf{t}_s.$$

where $\mathbf{I}$ is a unit operator. The increment of the translation part is the vector

$$\mathbf{V}_s = (\mathbf{P}_s - \mathbf{I})\boldsymbol{\rho}, \tag{8.10}$$

and the corresponding increment of the phase factor is

$$2\pi\mathbf{h} \cdot \mathbf{V}_s.$$

If the coordinates of $\mathbf{V}_s$ are integers, for all the relevant space-group operators, we have $\exp(2\pi i\mathbf{h} \cdot \mathbf{V}_s) = 1$, and the functional form of the structure factor is preserved. This allows us to determine the coordinates of the allowed $\boldsymbol{\rho}$ for each space group or, in other words, to determine the permissible origins of a space group.

Example: the general equivalent positions in the space group $P3$, with the origin on a threefold axis, are

$$(x, y, z), \qquad (\overline{y}, x - y, z), \qquad (y - x, \overline{x}, z).$$

The $\boldsymbol{P}_s - \boldsymbol{I}$ matrices can then be obtained as

$$\boldsymbol{P}_1 - \boldsymbol{I} = \begin{pmatrix} 0\ 0\ 0 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{pmatrix}, \quad \boldsymbol{P}_2 - \boldsymbol{I} = \begin{pmatrix} \overline{1}\ \overline{1}\ 0 \\ 1\ \overline{2}\ 0 \\ 0\ 0\ 0 \end{pmatrix}, \quad \boldsymbol{P}_3 - \boldsymbol{I} = \begin{pmatrix} \overline{2}\ 1\ 0 \\ \overline{1}\ \overline{1}\ 0 \\ 0\ 0\ 0 \end{pmatrix}. \tag{8.11}$$

The components of the translations satisfying eqn (8.10), that is the allowed translations, are given by

$$\begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}, \qquad \begin{pmatrix} 2/3 \\ 1/3 \\ z \end{pmatrix}, \qquad \begin{pmatrix} 1/3 \\ 2/3 \\ z \end{pmatrix}.$$

This example is taken from Giacovazzo (1980). Note that the components of the above allowed translations are the coordinates of points located on the three independent threefold axes in the unit cell of $P3$. It can also be readily verified that for crystals belonging to this space group, the phases $\varphi(hh0)$ are structure seminvariants. This can be generalized as shown in the next section.

Table 8.1 An example of the specification of the origin

No.	p_x	p_y	p_z	*eee*	*oee*	*eoe*	*eeo*	*eoo*	*oeo*	*ooe*	*ooo*
1	0	0	0	+	+	+	+	+	+	+	+
2	1/2	0	0	+	−	+	+	+	−	−	−
3	0	1/2	0	+	+	−	+	−	+	−	−
4	0	0	1/2	+	+	+	−	−	−	+	−
5	0	1/2	1/2	+	+	−	−	+	−	−	+
6	1/2	0	1/2	+	−	+	−	−	+	−	+
7	1/2	1/2	0	+	−	−	+	−	−	+	+
8	1/2	1/2	1/2	+	−	−	−	+	+	+	−

8.3 Specification of the origin

We shall first see how the origin of the unit cell may be is uniquely specified for the simplest centrosymmetric space group $P\overline{1}$. As we know, there are eight independent centers of symmetry in the unit cell of this space group, and we require that the origin be fixed at one of them. The fractional coordinates of these centers are p_x, p_y, p_z, where each coordinate may be equal to 0 or $\frac{1}{2}$. Assuming that all the atoms are located in general positions, the structure factor (assumed to be referred to an origin at $(0, 0, 0)$) is given by

$$F(hkl) = 2\sum_{j=1}^{N/2} f_j \cos[2\pi(hx_j + ky_j + lz_j)].$$

It is easy to show that if x, y, z are replaced by $x - p_x$, $y - p_y$, $z - p_z$, the new structure factor becomes

$$F'(hkl) = F(hkl)\cos[2\pi(hp_x + kp_y + lp_z)].$$

The structure factor will remain unchanged if the multiplier of π in the argument of the cosine is an even integer, and will change its sign if the multiplier is an odd integer. We recall that positive and negative signs correspond here to $\varphi(hkl)$ being 0 and π, respectively. This depends, of course, on the parities of hkl. We denote even parity of an index by e and odd parity by o, so that *eee* means that h, k, and l are all even, etc., and construct a table of centers versus parity groups so that each entry of the table will tell us whether or not F changes its sign. The symbols "+" and "−" in Table 8.1 mean that $F(hkl)$ remains unchanged and changes its sign, respectively, when for the given parity group the origin was shifted to the given center.

We see that if h, k, and l are all even, that is, for a reflection belonging to the parity group *eee*, the sign of the structure factor does not change, irrespective of the center to which the origin has been shifted. This is not surprising, because we know that for $P\overline{1}$, the phase $\varphi(2\mathbf{h})$ is a structure seminvariant. It is therefore

impossible to assign arbitrarily a positive (or negative) sign to an *eee* reflection– its sign has been determined by the (so far unknown) structure. Let us assign arbitrarily a positive sign to an *oee* reflection. This will remain unchanged upon shifting the origin to centers nos. 1, 3, 4, and 5 only, and the rows corresponding to the other four centers can be deleted from the table. Let us assign further a positive sign to an *eeo* reflection. This will restrict the possible candidates for shifting the origin without changing the sign of the structure factor, to centers nos. 1 and 3 only. If we repeat the above procedure for an *eoo* reflection, only no. 1 remains, and the origin is fixed at $(0, 0, 0)$. There are other possibilities; however, we *cannot* choose *oeo* as the third reflection, because the parities of the three reflections would then add up to *eee* which corresponds to a seminvariant.

We thus have a fixed origin of the unit cell, and phases with which all the others, except seminvariants, must be consistent. As we shall see later, we shall also require that the origin-fixing phases be associated with structure factors that are of large magnitude and are involved in many three-phase-invariant relationships.

A much more general approach is afforded by tables of seminvariants for all the space groups (Hauptman and Karle 1953*a*, 1956, 1959; Giacovazzo 1974, 2001). A prerequisite for an understanding of these tables is some background in modern algebra and we shall try to present and explain the definitions which are necessary for an understanding of the parameters involved.

We first define the concept of congruence. Given three integers a, b, and m, a is said to be congruent to b modulo m if the ratio $(a-b)/m$ is an integer. This is written symbolically as

$$a \equiv b \ (\text{mod } m),$$

where the integer m is called the *modulus*. For example,

$$10 \equiv 15 \ (\text{mod } 5)$$

is a congruence relationship in which the above ratio is -1. A more relevant example is

$$n \equiv 0 \ (\text{mod } 2),$$

from which it is seen that n can be any even integer, positive, negative or zero. Note, however, that the congruence $a \equiv b \ (\text{mod } 0)$ means $a = b$. For example $x \equiv 0 \ (\text{mod } 0)$ means $x = 0$.

Congruences possess many interesting algebraic properties that can be found in any text on modern algebra, but we shall content ourselves with the above definition and examples. The basic definition is in terms of single integers; however, we can also speak of an array (of integers) being congruent to another array (of integers) modulo a third array (of integers). For example

$$(h, k, l) \equiv (0, 0, 0) \ (\text{mod } (2,2,2)) \tag{8.12}$$

is interpreted as the three separate congruences

$$h \equiv 0 \ (\text{mod } 2) \quad k \equiv 0 \ (\text{mod } 2) \quad l \equiv 0 \ (\text{mod } 2),$$

which mean that h, k, and l are all even. Hence, eqn (8.12) defines the coordinates of a diffraction vector which is associated with a phase seminvariant. This is certainly so in the space group $P\overline{1}$ but, as shown in the tables of seminvariants, it is also valid for many other space groups. A general form of eqn (8.12) is written symbolically (for any space group) as

$$\mathbf{h}_\mathrm{s} \equiv \mathbf{0} \ (\mathrm{mod}\ \boldsymbol{\omega}_\mathrm{s}), \tag{8.13}$$

where the array $\mathbf{h}_\mathrm{s}$ is referred to as the vector *seminvariantly associated* with vector $\mathbf{h}$=(h,k,l) (in our case: $\mathbf{h}_\mathrm{s} = (h,k,l)$), and the array $\boldsymbol{\omega}_\mathrm{s}$ is called the *seminvariant modulus* (in our case: $(2,2,2)$).

In the space group $P4$, the permissible origins must lie on either of the two fourfold axes. Hence, $\boldsymbol{\rho}$ must be of one the forms $z\mathbf{c}$ or $\frac{1}{2}\mathbf{a}+\frac{1}{2}\mathbf{b}+z\mathbf{c}$. If the phase $\varphi(\mathbf{h})$ is to be a structure seminvariant we must have

$$2\pi\mathbf{h}_\mathrm{s}\cdot\boldsymbol{\rho} = 2\pi \times \text{ integer}$$

and the following equations must be satisfied:

$$lz = \text{ integer} \quad \text{and} \quad h\frac{1}{2}+k\frac{1}{2}+lz = \text{ integer}.$$

These hold for all values of z only if $h+k$ is even and $l=0$. The seminvariantly associated vector is thus $(h+k,l)$ and the seminvariant modulus is (2,0).

The seminvariant tables present these quantities for all the space groups.

A condition that must be obeyed in order that the origin be uniquely specified is that the determinant formed from the indices of the origin-fixing reflections, when reduced modulo the components of the seminvariant modulus, must be equal to ±1. This is called the *primitivity condition*. For example, in $P\overline{1}$ the three reflections (346), (685), and (431) (conforming to the *oee*, *eeo* and *eoo* chosen above) form the determinant

$$\begin{vmatrix} 3 & 4 & 6 \\ 6 & 8 & 5 \\ 4 & 3 & 1 \end{vmatrix}$$

which reduces rowwise (mod (2,2,2)) to the determinant

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -1$$

and satisfies the above condition.

In general, the number of phases/signs to be used in the specification of the origin equals the number of components of the seminvariant modulus. It follows from the above that we need only two phases for defining the origin of the unit

cell of the space group $P4$. For example, are the reflections (743) and (241) suitable for this purpose? The required determinant is

$$\begin{vmatrix} 7+4 & 3 \\ 2+4 & 1 \end{vmatrix} = \begin{vmatrix} 11 & 3 \\ 6 & 1 \end{vmatrix}$$

and when the rows are reduced modulo $(2, 0)$ we obtain

$$\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix} = +1$$

which satisfies the primitivity condition.

The seminvariant tables are so arranged that once we know the space group all the other parameters required for the specification of the origin are readily available and have to be supplemented by the primitivity test and other considerations involving the magnitude of the relevant structure factors and their appearance in three-phase invariants. The process of origin fixing is implemented in appropriate software; the only input required is the space-group information.

8.4 Inequalities

The first attempt at determining phases of structure factors from their magnitudes was made by Harker and Kasper and appeared in 1947 as a Letter to the Editor of The *Journal of Chemical Physics*, entitled "Phases of Fourier Coefficients from Crystal Diffraction Data" (Harker and Kasper 1947). Although, as will be seen from what follows, it is a long way from the contents of the above Letter to its title, that publication nevertheless constitutes the first indication that the phase problem is not as insoluble as it was thought to be in those years. It should be noted that when reference is made to Harker and Kasper's work on this problem, the reference is usually given to their paper published in the first volume of *Acta Crystallographica* a year later.

The approach of Harker and Kasper is based on the well-known inequalities of Schwarz and Cauchy. Cauchy's inequality (which applies to sums) is

$$\left| \sum_{j=1}^{N} a_j b_j \right|^2 \leq \left(\sum_{j=1}^{N} |a_j|^2 \right) \left(\sum_{j=1}^{N} |b_j|^2 \right), \tag{8.14}$$

where a_j and b_j are any numbers, real or complex. Let us write the unitary structure factor of the reflection $\mathbf{h}$ as

$$U(\mathbf{h}) = \sum_{j=1}^{N} u_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j), \tag{8.15}$$

and assume

$$a_j = \sqrt{u_j}$$

and

$$b_j = \sqrt{u_j}\exp(2\pi i\mathbf{h}\cdot\mathbf{r}_j).$$

The above substitutions are subject to the assumption that the unitary scattering factors are positive. If we use these substitutions with eqns (8.15) and (8.14), we obtain

$$\left|\sum_{j=1}^{N} a_j b_j\right|^2 = |U(\mathbf{h})|^2 \leq \left(\sum_{j=1}^{N} u_j\right)\left(\sum_{j=1}^{N} u_j|\exp(2\pi i\mathbf{h}\cdot\mathbf{r}_j)|^2\right) \tag{8.16}$$

but

$$|\exp(ix)|^2 = \exp(ix)\exp(-ix) = 1 \quad \text{for any} \quad x$$

and since

$$\left(\sum_{j=1}^{N} u_j\right) = 1,$$

we have from eqn (8.16) and the above identities

$$|U(\mathbf{h})|^2 \leq \left(\sum_{j=1}^{N} u_j\right)^2 = 1. \tag{8.17}$$

This is generally true for the unitary structure factor under the assumptions that the unitary scattering factors are positive and do not vary with the scattering angle (see Section 8.1) but does not provide any further information.

Let us consider a centrosymmetric structure with the origin at $(0,0,0)$. The unitary structure factor is then

$$U(\mathbf{h}) = \sum_{j=1}^{N} u_j \cos(2\pi\mathbf{h}\cdot\mathbf{r}_j), \tag{8.18}$$

and we assume

$$a_j = \sqrt{u_j}$$

and

$$b_j = \sqrt{u_j}\cos(2\pi\mathbf{h}\cdot\mathbf{r}_j).$$

The left-hand side of eqn (8.14) is $|U(\mathbf{h})|^2$, and the right-hand side of this inequality is

$$\left(\sum_{j=1}^{N} u_j\right)\left(\sum_{j=1}^{N} u_j\cos^2(2\pi\mathbf{h}\cdot\mathbf{r}_j)\right) = \frac{1}{2}\sum_{j=1}^{N} u_j[1+\cos(2\pi 2\mathbf{h}\cdot\mathbf{r}_j)]$$
$$= \frac{1}{2}[1+U(2\mathbf{h})]$$

since $\cos^2(x) = [1 + \cos(2x)]/2$. We thus have

$$|U(\mathbf{h})|^2 \leq \frac{1}{2}[1 + U(2\mathbf{h})]. \tag{8.19}$$

This inequality formally links the magnitude of $U(\mathbf{h})$ to the sign of $U(2\mathbf{h})$; however, it can only show that $U(2\mathbf{h})$ is positive. For example, if $|U(\mathbf{h})| = 0.6$ and $|U(2\mathbf{h})| = 0.5$, we obtain from eqn (8.19)

$$0.36 \leq \frac{1}{2}(1 \pm 0.5).$$

It is obvious that the inequality is true only if $U(2\mathbf{h})$ is positive, and it fails for the negative sign of $U(2\mathbf{h})$.

Many inequalities have been derived for a wide range of space groups by several authors, and a number of very simple structures were solved by the use of such inequalities. These studies were reviewed by Giacovazzo (1980), who also presents several relevant references. As a practical tool for phase determination, inequalities involving unitary structure factors are largely of historical interest. Nevertheless, this was the first step in applying a direct method to the determination of signs of structure factors. The most advanced inequalities, which lead in a way to the application of probability theory to phase determination and thus to the very successful direct methods, were put forward by Karle and Hauptman in their famous 1950 paper.

We shall omit the details of the derivation of the Karle–Hauptman method and just state the final result in terms of unitary structure factors: *The determinant*

$$D_n = \begin{vmatrix} U_{\mathbf{h}_1-\mathbf{h}_1} & U_{\mathbf{h}_2-\mathbf{h}_1} & U_{\mathbf{h}_3-\mathbf{h}_1} & \ldots & U_{\mathbf{h}_n-\mathbf{h}_1} \\ U_{\mathbf{h}_1-\mathbf{h}_2} & U_{\mathbf{h}_2-\mathbf{h}_2} & U_{\mathbf{h}_3-\mathbf{h}_2} & \ldots & U_{\mathbf{h}_n-\mathbf{h}_2} \\ U_{\mathbf{h}_1-\mathbf{h}_3} & U_{\mathbf{h}_2-\mathbf{h}_3} & U_{\mathbf{h}_3-\mathbf{h}_3} & \ldots & U_{\mathbf{h}_n-\mathbf{h}_3} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ U_{\mathbf{h}_1-\mathbf{h}_n} & U_{\mathbf{h}_2-\mathbf{h}_n} & U_{\mathbf{h}_3-\mathbf{h}_n} & \ldots & U_{\mathbf{h}_n-\mathbf{h}_n} \end{vmatrix}$$

it is non-negative if and only if the electron density is nonnegative at each point in the unit cell. This was a necessary condition in the derivation of Harker–Kasper inequalities, and Karle and Hauptman (1950) were the first to show that the nonnegativity of the above determinant is also a sufficient condition for the nonnegativity of the electron density.

Clearly, each such determinant gives rise to inequalities which become increasingly complex with increasing order of the determinant. Let us take $\mathbf{h}_1 = \mathbf{0}, \mathbf{h}_2 = \mathbf{h}, \mathbf{h}_3 = \mathbf{k}$ and consider the determinants D_2 and D_3.

$$D_2 = \begin{vmatrix} 1 & U_{\mathbf{h}} \\ U_{-\mathbf{h}} & 1 \end{vmatrix} \geq 0,$$

which, assuming $U_{-\mathbf{h}} = U_{\mathbf{h}}^*$, gives rise to the known result $|U_{\mathbf{h}}|^2 \leq 1$. For D_3, we obtain

$$D_3 = \begin{vmatrix} 1 & U_{\mathbf{h}} & U_{\mathbf{k}} \\ U_{-\mathbf{h}} & 1 & U_{\mathbf{k}-\mathbf{h}} \\ U_{-\mathbf{k}} & U_{\mathbf{h}-\mathbf{k}} & 1 \end{vmatrix} \geq 0$$

If we expand this determinant, we have

$$D_3 = 1 - |U_{\mathbf{h}}|^2 - |U_{\mathbf{k}}|^2 - |U_{\mathbf{h}-\mathbf{k}}|^2 + 2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}| \cos \alpha_{\mathbf{h},\mathbf{k}} \geq 0, \tag{8.20}$$

where

$$2 \cos \alpha_{\mathbf{h},\mathbf{k}} = \exp[i(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}})] + \exp[-i(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}})].$$

If we add to both sides of the inequality in eqn (8.20) $|U_{\mathbf{h}}U_{-\mathbf{k}}|^2$ and $|U_{\mathbf{h}-\mathbf{k}}|^2$, and subtract from both sides $U_{\mathbf{h}}U_{-\mathbf{k}}U_{-(\mathbf{h}-\mathbf{k})} + U_{-\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}$ this inequality can be rearranged to

$$(1 - |U_{\mathbf{h}}|^2)(1 - |U_{\mathbf{k}}|^2) \geq |U_{\mathbf{h}-\mathbf{k}} - U_{\mathbf{h}}U_{-\mathbf{k}}|^2, \tag{8.21}$$

since

$$\begin{aligned} |U_{\mathbf{h}-\mathbf{k}} - U_{\mathbf{h}}U_{-\mathbf{k}}|^2 &= (U_{\mathbf{h}-\mathbf{k}} - U_{\mathbf{h}}U_{-\mathbf{k}})(U_{\mathbf{h}-\mathbf{k}} - U_{\mathbf{h}}U_{-\mathbf{k}})^* \\ &= U_{\mathbf{h}-\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}^* + U_{\mathbf{h}}U_{-\mathbf{k}}U_{\mathbf{h}}^*U_{-\mathbf{k}}^* - U_{\mathbf{h}}U_{-\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}^* - U_{\mathbf{h}}^*U_{-\mathbf{k}}^*U_{\mathbf{h}-\mathbf{k}} \\ &= |U_{\mathbf{h}-\mathbf{k}}|^2 + |U_{\mathbf{h}}U_{-\mathbf{k}}|^2 - U_{\mathbf{h}}U_{-\mathbf{k}}U_{-(\mathbf{h}-\mathbf{k})} - U_{-\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}. \end{aligned}$$

If we make the substitutions $\mathbf{H} = \mathbf{h} - \mathbf{k}$ and $\mathbf{K} = -\mathbf{k}$, the inequality in eqn (8.21) becomes

$$|U_{\mathbf{H}} - U_{\mathbf{H}-\mathbf{K}}U_{\mathbf{K}}|^2 \leq (1 - |U_{\mathbf{H}-\mathbf{K}}|^2)(1 - |U_{\mathbf{K}}|^2)$$

or, following Karle and Hauptman (1950) and Giacovazzo (1980)

$$|U_{\mathbf{H}} - \delta_{\mathbf{H},\mathbf{K}}| \leq r_{\mathbf{H},\mathbf{K}}, \tag{8.22}$$

where

$$\delta_{\mathbf{H},\mathbf{K}} = U_{\mathbf{H}-\mathbf{K}}U_{\mathbf{K}} \quad \text{and} \quad r_{\mathbf{H},\mathbf{K}} = \sqrt{(1 - |U_{\mathbf{H}-\mathbf{K}}|^2)(1 - |U_{\mathbf{K}}|^2)}.$$

It is readily seen that this simple substitution provides a qualitatative explanation of how deterministic inequalities lead to a probabilistic treatment of the phase problem: the smaller $r_{\mathbf{H},\mathbf{K}}$ in eqn (8.22) is, the closer are $U_{\mathbf{H}}$ and $U_{\mathbf{K}}U_{\mathbf{H}-\mathbf{K}}$ both in phase and in magnitude. Hence the probability that

$$\varphi_{\mathbf{H}} \quad \text{approximates} \quad \varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}}$$

is likely to increase with increasing magnitudes of $|U_{\mathbf{K}}|$ and $|U_{\mathbf{H}-\mathbf{K}}|$.

This argument appears to have been a major breakthrough in the development of the theory of direct methods, and it is largely because of their above quoted 1950 article that Hauptman and Karle were awarded a Nobel Prize in Chemistry.

8.5 The Sayre equation and the tangent formula

Sayre (1952) took an entirely different approach, which he termed "the squaring method" and which led to most significant advances. Sayre's formalism can be derived very concisely with the aid of fundamental Fourier transform theory but it can also be arrived at by more elementary means. The derivation rests on two assumptions:

1. The crystal is composed of identical atoms with spherical symmetry and identical isotropic displacement parameters. This assumption is roughly satisfied by organic compounds in which the hydrogen atoms have been neglected at the stage of phase/sign determination.
2. We consider a hypothetical structure in which the electron density has been squared at each point. Clearly, the positions of the atomic peaks will be the same in the real crystal and in the "squared" one.

Since all the atoms are the same, they have the same scattering factors and the structure factor can be written as

$$F(\mathbf{h}) = f \sum_{j=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

where f is the common scattering factor. We can define the structure factor of the squared structure as

$$F^{\mathrm{sq}}(\mathbf{h}) = f^{\mathrm{sq}} \sum_{j=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

where f^{sq} is the scattering factor of the hypothetical squared structure. We therefore have the following relationship between the structure factors of the reflection $\mathbf{h}$ in the two structures:

$$F^{\mathrm{sq}}(\mathbf{h}) = \frac{f^{\mathrm{sq}}}{f} F(\mathbf{h}). \tag{8.23}$$

Let us now consider the expressions for the electron density. For the real crystal, the electron density at point $\mathbf{r}$ is given by

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h}} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}) \tag{8.24}$$

and the squared electron density is given by

$$\rho^2(\mathbf{r}) = \frac{1}{V^2} \sum_{\mathbf{h}} \sum_{\mathbf{h}'} F(\mathbf{h}) F(\mathbf{h}') \exp[-2\pi i (\mathbf{h} + \mathbf{h}') \cdot \mathbf{r}]. \tag{8.25}$$

If we substitute $\mathbf{H} = \mathbf{h} + \mathbf{h}'$ and $\mathbf{K} = \mathbf{h}'$, we obtain

$$\rho^2(\mathbf{r}) = \frac{1}{V^2} \sum_{\mathbf{H}} \sum_{\mathbf{K}} F_{\mathbf{K}} F_{\mathbf{H}-\mathbf{K}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}). \tag{8.26}$$

On the other hand, the squared electron density is given, in terms of the structure factor of the squared structure, by

$$\rho^2(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} F^{\mathrm{sq}}_{\mathbf{H}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}). \tag{8.27}$$

If we require that eqns (8.26) and (8.27) be identical, as they should, we must have

$$F^{\mathrm{sq}}_{\mathbf{H}} = \frac{1}{V} \sum_{\mathbf{K}} F_{\mathbf{K}} F_{\mathbf{H}-\mathbf{K}} \tag{8.28}$$

for all diffraction vectors $\mathbf{H}$. This is Sayre's first exact relationship between structure factors. The summation on the right-hand side ranges over all the values of $\mathbf{K}$, but in practice only over all the values of $\mathbf{K}$ that are present in the available triplets of reflections with diffraction vectors $\mathbf{H}$, $\mathbf{K}$, and $\mathbf{H} - \mathbf{K}$.

If we make use of eqn (8.23), eqn (8.28) becomes

$$F_{\mathbf{H}} = \frac{f}{V f^{\mathrm{sq}}} \sum_{\mathbf{K}} F_{\mathbf{K}} F_{\mathbf{H}-\mathbf{K}} \tag{8.29}$$

which is the same as equation (1.3) in Sayre's (1952) article, in a different notation. Since the "shape factor" f/f^{sq} is a positive real quantity, it is evident that the phase of $F(\mathbf{H})$ is the same as the phase of the summation on the right-hand side of eqn (8.29). Indeed, recalling that

$$F_{\mathbf{H}} = A_{\mathbf{H}} + iB_{\mathbf{H}} \quad \text{and} \quad \tan(\varphi_{\mathbf{H}}) = \frac{B_{\mathbf{H}}}{A_{\mathbf{H}}}$$

we have

$$\tan(\varphi_{\mathbf{H}}) = \frac{\mathcal{I}(F_{\mathbf{H}})}{\mathcal{R}(F_{\mathbf{H}})} = \frac{\mathcal{I}\left(\sum_{\mathbf{K}} F_{\mathbf{K}} F_{\mathbf{H}-\mathbf{K}}\right)}{\mathcal{R}\left(\sum_{\mathbf{K}} F_{\mathbf{K}} F_{\mathbf{H}-\mathbf{K}}\right)},$$

where $\mathcal{I}$ and $\mathcal{R}$ denote the imaginary part and real part respectively, or

$$\tan(\varphi_{\mathbf{H}}) = \frac{\sum_{\mathbf{K}} |F_{\mathbf{K}} F_{\mathbf{H}-\mathbf{K}}| \sin(\varphi_{\mathbf{K}}) + \varphi_{\mathbf{H}-\mathbf{K}})}{\sum_{\mathbf{K}} |F_{\mathbf{K}} F_{\mathbf{H}-\mathbf{K}})| \cos(\varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}}}. \tag{8.30}$$

This is the very important tangent formula, which relates the phase of $F_{\mathbf{H}}$ to the phase of the summation in the right-hand side of eqn (8.29). It can also be derived by other methods but we can see that it follows quite naturally from Sayre's equation. We shall comment on its application to phase determination in the last section of this chapter.

It is interesting to note that the triplet $\mathbf{H}$, $\mathbf{K}$,and $\mathbf{H}-\mathbf{K}$ appears both in the Karle–Hauptman inequality and in Sayre's equality, in spite of the quite different approaches that were adopted in the work of these authors. It was shown at the end of the last section that if the structure factors of the reflections $\mathbf{K}$ and $\mathbf{H}-\mathbf{K}$ are very large, the sum of their phases may approximate the phase of reflection $\mathbf{H}$. A similar conclusion may follow from an inspection of the tangent formula. Suppose that the product $|F_{\mathbf{K}}F_{\mathbf{H}-\mathbf{K}}|$ is a leading term in the summation in eqn (8.30) and the phases of reflections $\mathbf{K}$ and $\mathbf{H}-\mathbf{K}$ are known (say, from the fixing of the origin) or have been assumed. Their sum then leads to an approximation to the phase of the reflection $\mathbf{H}$, which now becomes an approximately known phase that can participate in the approximation of new phases.

We can consider this "leading term" argument directly from Sayre's equation. Multiply both sides of eqn (8.29) by $F_{-\mathbf{H}}$ (assumed to be equal to $F^*_{\mathbf{H}}$). We have

$$|F_{\mathbf{H}}|^2 = \frac{f}{Vf^{\mathrm{sq}}}\sum_{\mathbf{K}} F_{-\mathbf{H}}F_{\mathbf{K}}F_{\mathbf{H}-\mathbf{K}}. \tag{8.31}$$

The left-hand side of eqn (8.31) is positive, and the same applies to the summation. If $F_{-\mathbf{H}}F_{\mathbf{K}}F_{\mathbf{H}-\mathbf{K}}$ is now the leading term in the summation on the right-hand side of eqn (8.31), it is also likely to be positive and its phase is likely to be close to zero. Hence, the approximate relation:

$$\varphi_{-\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}} \approx 0,$$

or, interchanging $\mathbf{H}$ with $-\mathbf{H}$,

$$\varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}} \approx 0. \tag{8.32}$$

The left-hand side of the approximate equality in eqn (8.32) is the three-phase invariant which we derived in Section 8.2. This approximate equality plays a very important role in routine applications of direct methods but it must be put on a firmer basis by probabilistic considerations, such as those given in the next section.

8.6 Conditional probability density of a three-phase invariant

The qualitative indications regarding the approximate value of a three-phase invariant presented in the previous sections will now be put on a more quantitative basis with the aid of an appropriate probabilistic treatment. We redefine a three-phase invariant as

$$\Phi \equiv \varphi(\mathbf{h}_1) + \varphi(\mathbf{h}_2) + \varphi(\mathbf{h}_3) \;\text{ subject to }\; \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = \mathbf{0}$$

and rederive in this section a probability density function of such an invariant in terms of the normalized structure factors nowadays invariably employed. This

derivation is based on the work of Cochran (1955), who derived a conditional pdf of the phase $\varphi_{\mathbf{h}}$, given the phases $\varphi_{\mathbf{k}}$, $\varphi_{\mathbf{h}-\mathbf{k}}$ and the magnitudes of the unitary structure factors $|U_{\mathbf{h}}|$, $|U_{\mathbf{k}}|$ and $|U_{\mathbf{h}-\mathbf{k}}|$. Rederivations of this pdf in terms of normalized structure factors have been presented by several authors (for example Giacovazzo (1980) and Shmueli and Weiss(1995)). The derivations are based on the central limit theorem.

Logically, we would have to start from the pdf of the complex normalized structure factor $E_{\mathbf{h}}$. However, the pdf of a complex random variable is equivalent to the joint pdf of its real and imaginary parts, and hence

$$p(E_{\mathbf{h}}|E_{\mathbf{k}}, E_{\mathbf{h}-\mathbf{k}}) \equiv p(A_{\mathbf{h}}, B_{\mathbf{h}}|A_{\mathbf{k}}, B_{\mathbf{k}}, A_{\mathbf{h}-\mathbf{k}}, B_{\mathbf{h}-\mathbf{k}}),$$

which, owing to the assumed independence of $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$, can be represented as

$$p(A_{\mathbf{h}}|A_{\mathbf{k}}, B_{\mathbf{k}}, A_{\mathbf{h}-\mathbf{k}}, B_{\mathbf{h}-\mathbf{k}})p(B_{\mathbf{h}}|A_{\mathbf{k}}, B_{\mathbf{k}}, A_{\mathbf{h}-\mathbf{k}}, B_{\mathbf{h}-\mathbf{k}}).$$

The derivation rests on the following assumptions:

1. The diffraction vector $\mathbf{h}$ is fixed.
2. The diffraction vector $\mathbf{k}$ ranges uniformly over those vectors for which $E_{\mathbf{k}}$ and $E_{\mathbf{h}-\mathbf{k}}$ have well-defined magnitudes and phases.
3. All the atoms are identical, their contributions are statistically independent, and their coordinates are unifomly distributed in the interval $(0, 1)$.

The normalized structure factor $E_{\mathbf{h}}$ can be represented as follows:

$$E_{\mathbf{h}} = N^{-1/2} \sum_{j=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) \tag{8.33}$$

$$= N^{-1/2} \sum_{j=1}^{N} \exp(2\pi i \mathbf{k} \cdot \mathbf{r}_j) \exp(2\pi i (\mathbf{h} - \mathbf{k}) \cdot \mathbf{r}_j) \tag{8.34}$$

$$\equiv N^{-1/2} \sum_{j=1}^{N} (\xi_{\mathbf{h}j} + i\eta_{\mathbf{h}j}) \tag{8.35}$$

$$\equiv A_{\mathbf{h}} + iB_{\mathbf{h}}. \tag{8.36}$$

According to the central limit theorem, which applies to real random variables (see Section 7.3), $A_{\mathbf{h}}$ is normally distributed with mean

$$\langle A_{\mathbf{h}} \rangle = N^{-1/2} \sum_{j=1}^{N} \langle \xi_{\mathbf{h}j} \rangle \tag{8.37}$$

and variance

$$\sigma_A^2 = N^{-1} \sum_{j=1}^{N} [\langle \xi_{\mathbf{h}j}^2 \rangle - (\langle \xi_{\mathbf{h}j} \rangle)^2], \tag{8.38}$$

and similarly for $B_{\mathbf{h}}$. The expressions for the conditional pdfs of $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ are

$$p(A_{\mathbf{h}}|...) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{(A_{\mathbf{h}} - \langle A_{\mathbf{h}}\rangle)^2}{2\sigma_A^2}\right) \quad \text{and}$$
$$p(B_{\mathbf{h}}|...) = \frac{1}{\sqrt{2\pi\sigma_B^2}} \exp\left(-\frac{(B_{\mathbf{h}} - \langle B_{\mathbf{h}}\rangle)^2}{2\sigma_B^2}\right). \tag{8.39}$$

However, in view of assumption 2 it will be seen that, unlike in Section 7.3, the means $\langle A_{\mathbf{h}}\rangle$ and $\langle B_{\mathbf{h}}\rangle$ do not vanish. Similarly to Woolfson's (1954) treatment of a centrosymmetric sign invariant, consider the product

$$E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}} = N^{-1}\sum_{j=1}^{N}\sum_{m=1}^{N} \exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j)\exp(2\pi i(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_m), \tag{8.40}$$

or

$$\mathcal{R}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}) = N^{-1}\sum_{j=1}^{N}\sum_{m=1}^{N} \mathcal{R}[\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j)\exp(2\pi i(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_m)] \tag{8.41}$$

and

$$\mathcal{I}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}) = N^{-1}\sum_{j=1}^{N}\sum_{m=1}^{N} \mathcal{I}[\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j)\exp(2\pi i(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_m)]. \tag{8.42}$$

In view of assumption 3, each term on the right-hand side of eqns (8.41) and (8.42) is a random variable, and its average can be approximated by the corresponding left-hand side divided by the number of terms in the summation:

$$\langle\mathcal{R}[\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j)\exp(2\pi i(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_m]\rangle = \frac{\mathcal{R}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})}{(1/N)N^2} = \frac{\mathcal{R}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})}{N} \tag{8.43}$$

and

$$\langle\mathcal{I}[\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j)\exp(2\pi i(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_m]\rangle = \frac{\mathcal{I}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})}{(1/N)N^2} = \frac{\mathcal{I}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})}{N}. \tag{8.44}$$

Since the terms with $j \neq m$ and $j = m$ in eqns (8.43) and (8.44) are approximately equivalent, if we combine these equations with the representations of $E_{\mathbf{h}}$ given by eqns (8.34) and (8.35) we have

$$\langle\xi_{\mathbf{h}j}\rangle = \frac{\mathcal{R}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})}{N} = \frac{|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}})}{N} \tag{8.45}$$

and

$$\langle A_{\mathbf{h}}\rangle = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\langle\xi_{\mathbf{h}j}\rangle = \frac{|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}})}{N^{-1/2}}. \tag{8.46}$$

For the calculation of the variance we use the representation of $E_{\mathbf{h}}$ in eqn (8.33). Here, also using eqn (8.45), we obtain

$$\begin{aligned}\sigma_A^2 &= N^{-1}\sum_{j=1}^{N}\left[\langle\cos^2(2\pi\mathbf{h}\cdot\mathbf{r}_j)\rangle - \mathcal{R}(\langle\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j)\exp(2\pi i(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_m)\rangle)^2\right]\\ &= N^{-1}\sum_{j=1}^{N}\left[\frac{1}{2} - \frac{|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|^2\cos^2(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})}{N^2}\right]\\ &= \frac{1}{2} - \frac{|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|^2\cos^2(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})}{N^2} \qquad (8.47)\\ &\approx \frac{1}{2}, \qquad (8.48)\end{aligned}$$

provided N is sufficiently large (a condition of the central limit theorem). We obtain analogously the mean of B, and its variance is also $\frac{1}{2}$.

We now substitute these means and variances in eqn (8.39), and since $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ are independent, we obtain the conditional joint pdf of $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$,

$$\begin{aligned}&p(A_{\mathbf{h}}, B_{\mathbf{h}}|E_{\mathbf{k}}, E_{\mathbf{h}-\mathbf{k}})\\ &= \frac{1}{\pi}\exp\left[-\left(A_{\mathbf{h}} - \frac{|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\cos(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})}{N^{-1/2}}\right)^2\right]\\ &\times \exp\left[-\left(B_{\mathbf{h}} - \frac{|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\sin(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})}{N^{-1/2}}\right)^2\right]. \qquad (8.49)\end{aligned}$$

We now make a change of variables, $A_{\mathbf{h}} = |E_{\mathbf{h}}|\cos(\varphi_{\mathbf{h}})$ and $B_{\mathbf{h}} = |E_{\mathbf{h}}|\sin(\varphi_{\mathbf{h}})$. The Jacobian of this transformation,

$$\begin{vmatrix} \dfrac{\partial A_{\mathbf{h}}}{\partial|E_{\mathbf{h}}|} & \dfrac{\partial A_{\mathbf{h}}}{\partial\varphi_{\mathbf{h}}} \\ \dfrac{\partial B_{\mathbf{h}}}{\partial|E_{\mathbf{h}}|} & \dfrac{\partial B_{\mathbf{h}}}{\partial\varphi_{\mathbf{h}}} \end{vmatrix}.$$

readily evaluates to $|E|$ and we have to multiply the transformed eqn (8.49) by it. When we take into account that $|E_{\mathbf{h}}|$ is also known, as it is given by the experiment, eqn (8.49) reduces to

$$\begin{aligned}&p(\varphi_{\mathbf{h}}||E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h}-\mathbf{k}}|, \varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}})\\ &= \frac{|E_{\mathbf{h}}|}{\pi}\exp(-W + \kappa_{\mathbf{hk}}\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}), \qquad (8.50)\end{aligned}$$

where

$$W = |E_{\mathbf{h}}|^2 + \frac{|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|^2}{N},$$

$$\kappa_{\mathbf{hk}} = \frac{2}{\sqrt{N}}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h-k}}|, \tag{8.51}$$

and use has also been made of the trigonometric identity

$$\cos(x - y) = \cos(x)\cos(y) + \sin(x)\sin(y).$$

The conditional pdf of $\varphi_{\mathbf{h}}$, normalized to unity, is obtained as

$$p(\varphi_{\mathbf{h}}||E_{\mathbf{h}}|, E_{\mathbf{k}}, E_{\mathbf{h-k}}) = \frac{\exp[\kappa_{\mathbf{hk}}\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}})]}{\int_{-\pi}^{\pi}\exp[\kappa_{\mathbf{hk}}\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}})]\mathrm{d}\varphi_{\mathbf{h}}} \tag{8.52}$$

and if we make use of the integral representation of the modified Bessel function of the first kind of zero order, $I_0(x)$, eqn (8.52) reduces to

$$p(\varphi_{\mathbf{h}}||E_{\mathbf{h}}|, E_{\mathbf{k}}, E_{\mathbf{h-k}}) = \frac{1}{2\pi I_0(\kappa_{\mathbf{hk}})}\exp[\kappa_{\mathbf{hk}}\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}})]. \tag{8.53}$$

This is the equation that was derived by Cochran (1955), presented here in terms of normalized structure factors. Since $\varphi_{-\mathbf{h}} = -\varphi_{\mathbf{h}}$ in the noncentrosymmetric and dispersionless case, it is obvious that the argument of the cosine in eqn (8.51) is a three-phase invariant: indeed $\Phi_{\mathbf{hk}} \equiv \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}} = \varphi_{\mathbf{h}} + \varphi_{-\mathbf{k}} + \varphi_{\mathbf{k-h}}$ and the diffraction vectors in the latter phase triplet add up to a zero vector. Cochran's equation can be rewritten as a conditional pdf of the three-phase invariant as follows:

$$p(\Phi_{\mathbf{hk}}||E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h-k}}|) = \frac{1}{2\pi I_0(\kappa_{\mathbf{hk}})}\exp(\kappa_{\mathbf{hk}}\cos\Phi_{\mathbf{hk}}). \tag{8.54}$$

It can easily be seen that this conditional pdf attains its maximum value for $\Phi_{\mathbf{hk}} = 0$ and is symmetric with respect to the maximum. The height of this maximum depends on the parameter $\kappa_{\mathbf{hk}}$ which is given by eqn (8.51). The smaller N is and the greater the magnitudes of the normalized structure factors involved, the greater is the maximum of the pdf and the better is the approximation

$$\varphi_{\mathbf{h}} \approx \varphi_{\mathbf{k}} + \varphi_{\mathbf{h-k}}, \tag{8.55}$$

to which reference has been made in previous sections of this chapter and which plays a crucial role in the early stages of phase determination by direct methods (see, for example, the review by Giacovazzo (2001)). The shape of this pdf is illustrated in Fig. 8.1 for several values of $\kappa_{\mathbf{hk}}$. Cochran's pdf and its further developments have contributed greatly to the success of direct methods in the solution of the phase problem for small and medium-sized molecules, but the very high values of N encountered in protein crystals make the pdf too shallow (nearly uniform) to be useful as it stands. Direct phasing of protein crystals is being extensively investigated as a stand-alone approach, as well as in combination with Patterson techniques.

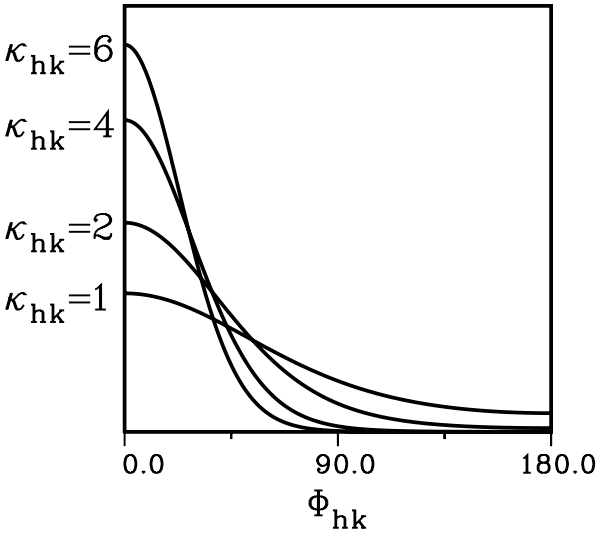


Fig. 8.1 Conditional probability density functions of the three-phase invariant $\Phi_{\mathbf{hk}}$ computed from eqn (8.54) for the values of $\kappa_{\mathbf{hk}}$ shown in the figure. Thus, the higher the values of $\kappa_{\mathbf{hk}}$ are, the more sharply peaked is the pdf and the better is the approximation in eqn (8.55). The functions are symmetric about $\Phi_{\mathbf{hk}} = 0$.

As has been shown, the derivation of Cochran's pdf is subject to approximations and simplifying assumptions. An exact conditional pdf of a three-phase invariant was derived by Shmueli, Rabinovich, and Weiss (1989), who used the Fourier method (as illustrated in Section 7.4) and no approximations or assumptions. The result is computationally demanding and is not intended for routine applications. However, test calculations show some interesting results. For example, for $N = 40$ and moderate magnitudes of the Es, the exact pdf and Cochran's pdf agree almost completely. For small N, Cochran's pdf underestimates the exact one, as would be expected from the assumptions underlying the central limit theorem (the exact pdf is much more sharply peaked). In the presence of an outstandingly heavy atom, which would lead to the breakdown of the central limit theorem, the exact pdf is also more sharply peaked than Cochran's. However, these disagreements only improve the quality of the approximation in eqn (8.55). The latter disagreement explains why is it generally easier to solve the phase problem for an asymmetric unit containing a heavy atom attached to not too many equal light ones than for an "equal-atom" structure. This is discussed further by Shmueli and Weiss (1995), who also show that their exact

pdf of the three-phase invariant reduces to Cochran's (1955) pdf when terms of lowest order in their expansion of the characteristic function are retained and an appropriate Fourier inversion is performed.

8.7 Other probabilistic considerations

We have seen that, for certain magnitudes of the unitary structure factors of two reflections with diffraction vectors **h** and 2**h**, in the centrosymmetric case, the sign of the structure factor of the reflection 2**h** is determined as positive by the simplest Harker–Kasper inequality

$$|U_{\mathbf{h}}^2| \leq \frac{1}{2} + \frac{1}{2}U_{2\mathbf{h}}.$$

If we take $|U_{\mathbf{h}}| = |U_{2\mathbf{h}}| = 0.4$–which still correspond to large magnitudes– the inequality is satisfied by both positive and negative signs of $U_{2\mathbf{h}}$. A set of intensity data collected from a centrosymmetric crystal rather often contains pairs of relections **h** and 2**h** associated with magnitudes of the structure factors that are large but not large enough for the Harker–Kasper inequality. This, and other considerations, led to the question of what is the probability that $U_{2\mathbf{h}}$ is positive, given its magnitude and the magnitude of $U_{\mathbf{h}}$. This conditional probability was investigated in the earliest days of direct phase determination with the aid of the central limit theorem. An expression for this probability–of particular importance, since the reflection 2**h** corresponds to a structure seminvariant–was given by Cochran and Woolfson (1955) in terms of unitary structure factors and, represented in terms of normalized structure factors, is given by

$$p_+(2\mathbf{h}|\mathbf{h}) = \frac{1}{2} + \frac{1}{2}\tanh\left[\frac{\sigma_3}{2\sigma_2^{3/2}}|E_{2\mathbf{h}}|(|E_{\mathbf{h}}|^2 - 1)\right], \tag{8.56}$$

where $p_+(2\mathbf{h}|\mathbf{h})$ is the probability that $E_{2\mathbf{h}}$ is positive, given the magnitudes $|E_{\mathbf{h}}|$ and $|E_{2\mathbf{h}}|$ and $\sigma_n = \sum_{j=1}^{N} Z_j^n$, where Z_j^n is the nth power of the atomic number of the jth atom. Note that for an equal-atom structure $\sigma_3/(2\sigma_2^{3/2})$ reduces to $1/(2\sqrt{N})$.

Equation (8.56) is known as the $\sum_1$ formula and is usually applied in the initial stages of a phase determination in order to obtain as many signs of structure seminvariants as is possible.

Another interesting inequality, for the centrosymmetric case, is based on the sum of two unitary structure factors. We shall follow here the derivation of Woolfson (1961).

$$U_{\mathbf{h}} + U_{\mathbf{k}} = \sum_{j=1}^{N} u_j[\cos(2\pi\mathbf{h}\cdot\mathbf{r}_j) + \cos(2\pi\mathbf{k}\cdot\mathbf{r}_j)]$$
$$= \sum_{j=1}^{N} 2u_j \cos\left[2\pi \times \frac{1}{2}(\mathbf{h}+\mathbf{k})\cdot\mathbf{r}_j\right] \cos\left[2\pi \times \frac{1}{2}(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_j\right]. \tag{8.57}$$

If we define

$$a_j = \sqrt{2u_j}\cos\left[2\pi \times \frac{1}{2}(\mathbf{h}+\mathbf{k})\cdot\mathbf{r}_j\right]$$

and

$$b_j = \sqrt{2u_j}\cos\left[2\pi \times \frac{1}{2}(\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_j\right]$$

for use with Cauch's inequality given in eqn (8.14), we then have

$$\left(\sum_{j=1}^{N} a_j b_j\right)^2 = (U_{\mathbf{h}} + U_{\mathbf{k}})^2,$$

and

$$\sum_{j=1}^{N} a_j^2 = \sum_{j=1}^{N} 2u_j \cos^2\left[2\pi \times \frac{1}{2}(\mathbf{h}+\mathbf{k})\cdot\mathbf{r}_j\right]$$
$$= \sum_{j=1}^{N} u_j\{1 + \cos[2\pi(\mathbf{h}+\mathbf{k})\cdot\mathbf{r}_j]\}$$
$$= \sum_{j=1}^{N} u_j + \sum_{j=1}^{N} u_j \cos[2\pi(\mathbf{h}+\mathbf{k})\cdot\mathbf{r}_j]$$
$$= 1 + U_{\mathbf{h}+\mathbf{k}}.$$

We can show in a similar manner that

$$\sum_{j=1}^{N} b_j^2 = 1 + U_{\mathbf{h}-\mathbf{k}}.$$

If we now substitute these results into Cauchy's inequality, we obtain

$$(U_{\mathbf{h}} + U_{\mathbf{k}})^2 \leq (1 + U_{\mathbf{h}+\mathbf{k}})(1 + U_{\mathbf{h}-\mathbf{k}}). \tag{8.58}$$

If we start with a difference rather than a sum of unitary structure factors and follow the same steps as above, we can obtain the following inequality:

$$(U_{\mathbf{h}} - U_{\mathbf{k}})^2 \leq (1 - U_{\mathbf{h}+\mathbf{k}})(1 - U_{\mathbf{h}-\mathbf{k}}). \tag{8.59}$$

The inequalities in eqn (8.58) and (8.59) can be combined as follows. Let $U_{\mathbf{h}} = s_{\mathbf{h}}|U_{\mathbf{h}}|$, where $s_{\mathbf{h}}$ equals $+1$ or -1 if $U_{\mathbf{h}}$ is positive or negative, respectively. If $s_{\mathbf{h}}s_{\mathbf{k}} = +1$, we have

$$(U_{\mathbf{h}} + U_{\mathbf{k}})^2 = (|U_{\mathbf{h}}| + |U_{\mathbf{k}}|)^2 \leq (1 + s_{\mathbf{h}}s_{\mathbf{k}}U_{\mathbf{h+k}})(1 + s_{\mathbf{h}}s_{\mathbf{k}}U_{\mathbf{h-k}}),$$

and if $s_{\mathbf{h}}s_{\mathbf{k}} = -1$ eqn (8.59) becomes

$$(U_{\mathbf{h}} - U_{\mathbf{k}})^2 = (|U_{\mathbf{h}}| + |U_{\mathbf{k}}|)^2 \leq (1 + s_{\mathbf{h}}s_{\mathbf{k}}U_{\mathbf{h+k}})(1 + s_{\mathbf{h}}s_{\mathbf{k}}U_{\mathbf{h-k}});$$

in either case

$$\begin{aligned}(|U_{\mathbf{h}}| + |U_{\mathbf{k}}|)^2 &\leq (1 + s_{\mathbf{h}}s_{\mathbf{k}}U_{\mathbf{h+k}})(1 + s_{\mathbf{h}}s_{\mathbf{k}}U_{\mathbf{h-k}}) \\ &= (1 + s_{\mathbf{h}}s_{\mathbf{k}}s_{\mathbf{h+k}}|U_{\mathbf{h+k}}|)(1 + s_{\mathbf{h}}s_{\mathbf{k}}s_{\mathbf{h-k}}|U_{\mathbf{h-k}}|). \qquad (8.60)\end{aligned}$$

It can be shown from eqn (8.60) that with structure factors of sufficient magnitude, either or both of the relationships

$$s_{\mathbf{h}}s_{\mathbf{k}}s_{\mathbf{h+k}} = +1 \qquad (8.61)$$

$$s_{\mathbf{h}}s_{\mathbf{k}}s_{\mathbf{h-k}} = +1 \qquad (8.62)$$

is true. Although the inequality in eqn (8.60) may be useful as it stands, its real importance is in the (qualitatively) probable positivity of the triple sign products. This led to derivations of joint probabilities that the triple sign product in eqn (8.61) is positive, conditioned on the magnitudes of the three structure factors involved (Woolfson 1954; Cochran and Woolfson 1955). These derivations, based on the central limit theorem, led to an expression (represented here in terms of normalized structure factors) given by

$$p_+(\mathbf{h}, \mathbf{k}, \mathbf{h}+\mathbf{k}) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{\sigma_3}{\sigma_2^{3/2}}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h+k}}|\right). \qquad (8.63)$$

This equation played a most important role in the early days of direct methods when mainly centrosymmetric structures were studied. We shall mention in the next section some practical approaches to the application of all these formulas to direct phase determination. It may be pointed out that in the centrosymmetric case $E_{\mathbf{h+k}} = E_{-\mathbf{h}-\mathbf{k}}$ and the triple product in eqn (8.63) is a structure invariant. The probability that this structure invariant is positive (that is, its phase is zero) is always greater than that for it being negative (that is, its phase is π). A somewhat similar behavior is also seen in the noncentrosymmetric case, where the probability density function of a three-phase invariant, conditioned on the magnitudes of the structure factors involved, has its maximum at $\Phi_{\mathbf{hk}} = 0$ (see the previous section).

The above probabilistic treatments which evolved from the corresponding Harker–Kasper inequalities, are based on the central limit theorem and presume

an equal-atom structure (or nearly so). Exact joint pdfs, conditioned on the magnitudes of the structure factors involved and allowing introduction of effects of space-group symmetry and arbitrary atomic composition, were derived for the Σ_1 case (Shmueli and Weiss 1985) and the triple product case (Shmueli and Weiss 1986) and were discussed further by these authors (Shmueli and Weiss 1995). It was seen, similarly to test calculations of conditional pdfs of three-phase invariants, that the central-limit-theorem pdfs underestimated the exact ones when an outstandingly heavy atom was present or when the number of atoms in the asymmetric unit was small, and that the exact pdfs agree with the central-limit-theorem pdfs when the assumptions underlying the central limit theorem are satisfied. The exact pdfs are derived as Fourier series representations similarly to those for three-phase invariants.

8.8 Some practical comments on the solution of the phase problem

The foregoing sections of this chapter are by no means a complete description of direct methods and just seek to highlight some important early developments related to this topic. Application of direct methods to the solution of the phase problem for crystals consisting of small and medium-sized molecules (up to a few hundred atoms in the asymmetric unit ...) is more or less routine, insofar as the analysis is cleverly "buried" in software which runs satisfactorily on modern personal computers. The situation with applications to crystals of macromolecules is less simple but recent developments of direct methods seem to be very promising. Many of these developments are reviewed by Giacovazzo (1998) in his comprehensive book and in more recent literature, mainly in *Acta Crystallographica*, Section A.

In spite of the elementary nature of this chapter, it is possible to outline the principles of the procedure of phase determination for small and medium-sized molecules.

1. *Calculation of normalized structure factors*. First, the observed magnitudes of the structure factors are brought to an absolute scale, as outlined in Section 7.2, or by any variant of Wilson's method (see, for example, Main 1975). Then, the squared magnitudes of the normalized structure factors are calculated as

$$|E(\mathbf{h})|^2 = \frac{|F(\mathbf{h})|^2}{\epsilon_{\mathbf{h}} \langle |F^0(\mathbf{h})|^2 \rangle \exp(-2Bs^2)},$$

where $s^2 = \sin^2\theta/\lambda^2$, the superscript "0" means that the average in the denominator is computed from the scattering factors of atoms at rest, and B is the overall displacement parameter obtained from Wilson's scaling method (cf. Section 7.2). The quantity $\epsilon_{\mathbf{h}}$ in the denominator depends on the point-group symmetry (Rogers, 1950).

2. *Resolution of possible space-group ambiguity*, as outlined in Chapter 7. Note that routine program packages use only the basic Wilson statistics for this purpose.

3. *Possible use of seminvariants.* For centric reflections, the Σ_1 formula or related formulas are evaluated and signs are assigned according to the probabilities obtained, for example, from eqn (8.56). These, if any, are the first known phases.
4. *Construction of the $\sum_2$ listing.* The reflections are sorted according to decreasing $|E_{\mathbf{h}}|$, and for each $|E_{\mathbf{h}}|$, all the pairs $|E_{\mathbf{k}}|$ and $|E_{\mathbf{h}-\mathbf{k}}|$ present in the reflection set are selected to form triplets $|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h}-\mathbf{k}}|$. For each such triplet the quantity $\kappa_{\mathbf{hk}} = 2|E_{\mathbf{h}}||E_{\mathbf{k}}||E_{\mathbf{h}-\mathbf{k}}|/\sqrt{N}$ is evaluated, to indicate its efficiency. Reflections with $|E_{\mathbf{h}}|$ below a certain threshold (1.2 to 1.5) are usually omitted from the list.

The $\sum_2$ listing is thus subdivided into sections, each section having a common $|E_{\mathbf{h}}|$ and a number of pairs $|E_{\mathbf{k}}|, |E_{\mathbf{h}-\mathbf{k}}|$ with varying $\mathbf{k}$. The purpose of the process is to find as many phases of these pairs as possible, since then the tangent formula is likely to provide a good approximation to the phase of $E_{\mathbf{h}}$. Probabilistic considerations are indispensable here. Thus, eqn (8.53) refers to a single triplet, and one would like to have a probability density function of $\varphi_{\mathbf{h}}$, conditioned on a knowledge of the magnitudes and phases of all the pairs $\mathbf{k}$ and $\mathbf{h} - \mathbf{k}$ within a section of the $\sum_2$ listing. Assuming that all of the r pairs within a section are independent, the required probability density function was given by Karle and Karle (1966) as

$$p(\varphi_{\mathbf{h}}|...) = \prod_{j=1}^{r} \frac{1}{2\pi I_0(\kappa_{\mathbf{hk}_j})} \exp[\kappa_{\mathbf{hk}_j} \cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}_j} - \varphi_{\mathbf{h}-\mathbf{k}_j})] \tag{8.64}$$

$$= A_{\mathbf{h}} \exp\left\{\sum_{j=1} [\kappa_{\mathbf{hk}_j} \cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}_j} - \varphi_{\mathbf{h}-\mathbf{k}_j})]\right\} \tag{8.65}$$

$$= \frac{1}{2\pi I_0(\alpha)} \exp[\alpha \cos(\varphi_{\mathbf{h}} - \beta_{\mathbf{h}})], \tag{8.66}$$

where

$$\alpha_{\mathbf{h}}^2 = \left[\sum_{j=1}^{r} \kappa_{\mathbf{hk}_j} \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})\right]^2 + \left[\sum_{j=1}^{r} \kappa_{\mathbf{hk}_j} \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})\right]^2 \tag{8.67}$$

and

$$\tan \beta_{\mathbf{h}} = \frac{\sum_{j=1}^{r} \kappa_{\mathbf{hk}_j} \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_{j=1}^{r} \kappa_{\mathbf{hk}_j} \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}. \tag{8.68}$$

The simplification of the argument of the exponential in eqn (8.65) to that in eqn (8.66) was obtained by making use of the relation

$$\sum_j \kappa_{\mathbf{hk}_j} \cos[\varphi_{\mathbf{h}} - (\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})] = \alpha_{\mathbf{h}} \cos(\varphi_{\mathbf{h}} - \beta_{\mathbf{h}}),$$

where $\alpha_{\mathbf{h}}^2$ and $\tan \beta_{\mathbf{h}}$ are given by eqns (8.67) and (8.68), respectively, and the factor $1/[2\pi I_0(\alpha)]$ in eqn (8.66) was obtained by normalizing to unity eqn (8.65), with the exponential simplified as above.

The variance of $\varphi_{\mathbf{h}}$, as obtained from eqn (8.66), is

$$V_{\mathbf{h}} = \frac{\pi^2}{3} + \frac{1}{I_0(\alpha_{\mathbf{h}})} \sum_{n=1}^{\infty} \frac{I_{2n}(\alpha_{\mathbf{h}})}{n^2} - \frac{4}{I_0(\alpha_{\mathbf{h}})} \sum_{n=0}^{\infty} \frac{I_{2n+1}(\alpha_{\mathbf{h}})}{(2n+1)^2} \tag{8.69}$$

and it can be shown that the higher $\alpha_{\mathbf{h}}$ is the lower is the variance and therefore the closer is $\beta_{\mathbf{h}}$ to $\varphi_{\mathbf{h}}$.

At this stage, there is no phase information as yet (apart from a possible few seminvariants), and $\alpha_{\mathbf{h}}$ cannot be computed from eqn (8.67). However, the average of $\alpha_{\mathbf{h}}^2$ can be shown (Germain, Main, and Woolfson 1970) to equal

$$\langle \alpha_{\mathbf{h}}^2 \rangle = \sum_{j=1}^{r} \kappa_{\mathbf{hk}_j}^2 + 2 \sum_{\substack{j_1 \\ j_1<j_2}} \sum_{j_2} \kappa_{\mathbf{hk}_{j_1}} \kappa_{\mathbf{hk}_{j_2}} \frac{I_1(\kappa_{\mathbf{hk}_{j_1}})}{I_0(\kappa_{\mathbf{hk}_{j_1}})} \frac{I_1(\kappa_{\mathbf{hk}_{j_2}})}{I_0(\kappa_{\mathbf{hk}_{j_2}})} \tag{8.70}$$

and this can be evaluated, for all the reflections that were retained, without any phase information. Thus, each section of the $\sum_2$ listing, and the associated $|E_{\mathbf{h}}|$, can be assigned a reliability criterion to be used later on.

5. *Choice of origin-fixing phases.* This most important step is based on the considerations outlined in Section 8.3. Apart from the requirements indicated there the reflections to be chosen for origin specification must satisfy several practical requirements. The corresponding structure factors should have large magnitudes and their $\sum_2$ sections should consist of large numbers of $\mathbf{k}$ and $\mathbf{h} - \mathbf{k}$ pairs, with magnitudes as large as can be found. However, this approach may be appropriate for a judicious choice of such phases, but less so for automatic algorithms. A very popular method, suitable for computation, is the convergence procedure of Germain, Main, and Woolfson (1970). The principle of this method is an iterative elimination of reflections with the lowest $\langle \alpha_{\mathbf{h}}^2 \rangle^{1/2} \equiv \alpha_{\text{rms}}$ and all the triplets in which they are involved, followed by an updating of α_{rms} for all the remaining reflections, until a small set of reflections with very high α_{rms} - that is, involved in many triplets and associated with large $|E|$s–remains, from which the origin-fixing phases are chosen. A detailed description of the convergence method is given in the original paper of Germain, Main, and Woolfson (1970), and this and other approaches to this important step have been reviewed, for example, by Giacovazzo (1980, 1998).

6. *Assignment of symbolic phases.* At this point the set of known phases, known as the *starting set*, contains the origin-fixing phases and possibly some seminvariants obtained from $\sum_1$. It is possible to look for strong triplets containing two known phases and approximately estimate the third, etc., but this process is only seldom successful and more often than not comes to an early end. The remedy is to assign to a small number of the $|E|$s symbols, which in the centrosymmetric case stand for the signs "+" and "−", and in the noncentrosymmetric case stand for numerical values of phases, for example, $\pi/4, 3\pi/4, 5\pi/4, and 7\pi/4$. The symbols are now introduced into the starting set which is the basis of further phase determination.

7. *Expansion of the starting set by use of the tangent formula.* There are many approaches to this task. All of them use the starting set, which now consists of the origin-fixing phases, some possible seminvariants, and the symbols to be expanded into numerical values, Whether this is done at the beginning or at the completion of the tangent expansion, the number of different solutions clearly depends on the number of symbols. If m symbols were assigned there will be 2^m or 4^m solutions in the centrosymmetric and noncentrosymmetric cases, respectively. For example, in the latter case if $m = 3$ there will be $4^3 = 64$ solutions (not necessarily different), which will contain the correct one. The criteria for a correct solution are known as the figures of merit, some of which will be discussed below.
8. *Figures of merit.* Given a set of phased $|E|$s, we can rapidly compute the Fourier synthesis

$$\rho_E(\mathbf{r}) = \sum_{\mathbf{h}} E_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{r})$$

which is related to the electron density, and try to interpret it in terms of the molecular structure. However, in the presence of many solutions, this is very tedious and may also be misleading. The commonly accepted solution is the use of figures of merit, some of which will be given below.

- A criterion which used to be very popular is the "absolute figure of merit" (Germain, Main and Woolfson 1970). This is given by

$$ABSFOM = \frac{\sum_{\mathbf{h}} (\alpha_{\mathbf{h}} - \langle \alpha_{\mathbf{h}}^2 \rangle_{\text{random}}^{1/2})}{\sum_{\mathbf{h}} (\langle \alpha_{\mathbf{h}}^2 \rangle^{1/2} - \langle \alpha_{\mathbf{h}}^2 \rangle_{\text{random}}^{1/2})}, \tag{8.71}$$

  where, for the solution tested, $\alpha_{\mathbf{h}}$ is computed from eqn (8.67), $\langle \alpha_{\mathbf{h}}^2 \rangle^{1/2}$ is obtained from eqn (8.70) and $\langle \alpha_{\mathbf{h}}^2 \rangle_{\text{random}}^{1/2}$ is an expected root-mean-square of $\alpha_{\mathbf{h}}$ for a random phase set, given by $\sum_{\mathbf{k}_j} \kappa_{\mathbf{hk}_j}^2$. Thus, in theory, for a random phase set ABSFOM should be zero, while for a correct phase set it should be unity. This is often a useful criterion.
- Another popular criterion is the residual between the actual and estimated αs:

$$R_\alpha = \frac{\sum_{\mathbf{h}} |\alpha_{\mathbf{h}} - \langle \alpha_{\mathbf{h}}^2 \rangle^{1/2}|}{\sum_{\mathbf{h}} |\langle \alpha_{\mathbf{h}}^2 \rangle^{1/2}}. \tag{8.72}$$

  The correct set of phases is likely to be associated with the smallest value of R_α. It is understood that $\alpha_{\mathbf{h}}$ has been brought to the scale of its expectation value. This is similar to the R criterion of Karle and Karle (1966).
- We also wish to outline a rather powerful figure of merit, based on four-phase or quartet invariants (Schenk 1974; Giacovazzo 2001). Briefly, the quartet invariant is the phase sum

$$\Phi_4 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{m}}$$

  or its cosine, subject to $\mathbf{h} + \mathbf{k} + \mathbf{l} + \mathbf{m} = \mathbf{0}$. The magnitudes of E associated with these diffraction vectors are called *main* terms, while the magnitudes

$|E_{\mathbf{h+k}}|, |E_{\mathbf{h+l}}|, |E_{\mathbf{h+m}}|$, also associated with this quartet, are called *cross* terms. It has been shown that if both the main and cross terms are large, and the quantity

$$B_{\mathbf{hklm}} = \frac{2}{N}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}}E_{\mathbf{m}}|$$

is sufficiently large, then the function $\cos\Phi_4$ is most probably $+1$. On the other hand, for large B, large main terms and small cross terms, the function $\cos\Phi_4$ is most probably -1. This is the basis of the following procedure:

– Search for quartets with large main terms and small cross terms (these are called *negative quartets*).

– Evaluate the quantity

$$NQEST = \frac{\sum\limits_{\mathbf{h,k,l}} B_{\mathbf{h,k,l,m}} \cos\Phi_4}{\sum\limits_{\mathbf{h,k,l}} B_{\mathbf{h,k,l,m}}} \tag{8.73}$$

where the summation extends over the linearly independent vectors **h,k,l** and involves negative quartets. The most negative *NQEST* should correspond to the correct phase set. Note that there are also other formulations of *NQEST* used in practical applications.

9. When the most plausible phase set has been found, a synthesis of phased $|E|$s is computed and the peaks in this electron-density-like map are interpreted in terms of the molecular structure. This stage is aided by an automatic computation of interpeak distances and intervector angles, followed by graphic output, usually available in most direct method packages.

Note that there are more figures of merit in use and it is common practice to combine their outcomes with weights based on the user's/programmer's experience.

This page intentionally left blank

9 Atomic displacement parameters

9.1 Introduction

Our discussions of the structure factor have assumed that the effects of atomic displacements are hidden in the expressions for the scattering factor and hence in the electron density. It is, however, possible to write the positional and displacement parameters explicitly–albeit approximately–and even let them vary so that an optimal agreement is obtained between the magnitudes in the model for the structure factor and the corresponding magnitudes that are obtained from the observed intensity. This process of variation, known as the refinement of structural parameters, will be treated in the next chapter. The present chapter is devoted to an approximate description of the effects of atomic displacement, following closely Trueblood *et al.* (1996), and to some simple models of displacement parameters in terms of the positional parameters, such as those proposed by Schomaker and Trueblood (1968).

The structure factor of a reflection $\mathbf{h}$, taking into account effects of displacements, can be represented by the Fourier transform of the *average* density of scattering matter

$$F(\mathbf{h}) = \int \langle \rho(\mathbf{r}) \rangle \exp(2\pi i \mathbf{h} \cdot \mathbf{r}) \, \mathrm{d}^3\mathbf{r} \equiv \mathcal{F}[\langle \rho(\mathbf{r}) \rangle], \tag{9.1}$$

where the integration extends over the repeating structural unit (or a unit cell); the brackets denote a double averaging over the possible displacements of the atoms from their mean positions (a time average over the atomic vibrations in each unit cell, followed by a space average that consists of projecting all the time averaged cells onto one cell and dividing by the number of cells); $\mathbf{h}$ is a diffraction vector obeying the Laue equations; and $\rho(\mathbf{r})$ is the static density of the motif, consistent with the instantaneous local configuration of the nuclei in a unit cell.

The above general picture can be reduced to what is conventionally used in crystal structure analysis if we make the following assumptions:

1. The average density of matter in eqn (9.1) can be regarded as a superposition of independently averaged atomic densities.
2. For X-rays, the static atomic electron density is assumed to have spherical symmetry. We made this assumption in the derivation of a simplified expression for the atomic scattering factor in Chapter 5.

3. The probability density function for an atomic displacement can be approximated by a univariate or trivariate Gaussian.

Let us now consider these assumptions in some detail. Assumption 1 is equivalent to assuming independently displaced atoms, a fair initial approximation, although not generally valid. The average density of scattering matter at the point $\mathbf{r}$ in a unit cell can then be approximated as

$$\langle \rho(\mathbf{r}) \rangle \simeq \sum_{k=1}^{N} n_k \int \rho_k(\mathbf{r} - \mathbf{r}_k) p_k(\mathbf{r}_k - \mathbf{r}_{k0}) \, \mathrm{d}^3\mathbf{r}_k. \tag{9.2}$$

Here, N is the number of atoms in the unit cell, n_k is the occupancy factor of the kth atom, $\rho_k(\mathbf{r} - \mathbf{r}_k)$ is the static density (the electron density for X-rays or a δ function weighted by the scattering length b_k for neutrons) due to atom k at the point $\mathbf{r}$ when the nucleus of that atom is at $\mathbf{r}_k$, and $p_k(\mathbf{r}_k - \mathbf{r}_{k0})$ is the probability density function corresponding to the probability of having atom k displaced by the vector $\mathbf{r}_k - \mathbf{r}_{k0}$ from its reference position $\mathbf{r}_{k0}$ in an average unit cell, which will be the mean position of the nucleus if ρ_k is sufficiently symmetrical. The above approximation includes the assumption that the (static) atoms are not deformable, by bonding or otherwise, even though at this stage, spherical symmetry of the atomic electron density has not yet been assumed.

If eqn (9.2) is substituted in eqn (9.1) and the order of summation and integration is interchanged, the structure factor becomes

$$F(\mathbf{h}) \simeq \sum_{k=1}^{N} n_k F_k(\mathbf{h}), \tag{9.3}$$

where

$$F_k(\mathbf{h}) = \int \left[\int \rho_k(\mathbf{r} - \mathbf{r}_k) p_k(\mathbf{r}_k - \mathbf{r}_{k0}) \, \mathrm{d}^3\mathbf{r}_k \right] \exp(2\pi i \mathbf{h} \cdot \mathbf{r}) \, \mathrm{d}^3\mathbf{r}. \tag{9.4}$$

If the substitutions $\mathbf{r} = \mathbf{t} + \mathbf{r}_k$ and $\mathbf{r}_k = \mathbf{u} + \mathbf{r}_{k0}$ are made, the integral in eqn (9.4) becomes

$$F_k(\mathbf{h}) \simeq \left\{ \int \left[\int \rho_k(\mathbf{t} - \mathbf{u}) p_k(\mathbf{u}) \, \mathrm{d}^3\mathbf{u} \right] \exp(2\pi i \mathbf{h} \cdot \mathbf{t}) \, \mathrm{d}^3\mathbf{t} \right\} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_{k0}). \tag{9.5}$$

The inner integral in eqn (9.5) has the form of a conventional convolution of the density of atom k with the pdf for the displacement of this atom from its reference or mean position; the outer integral is a Fourier transform of this convolution. This transform is multiplied by an exponential that depends on the reference or mean position $\mathbf{r}_{k0}$ of atom k.

By the convolution theorem, the Fourier transform of a convolution equals the product of Fourier transforms of the functions involved. When this theorem is applied to the outer integral in eqn (9.5), we obtain the conventional approximation for the structure factor of a Bragg reflection,

$$F(\mathbf{h}) \simeq \sum_{k=1}^{N} n_k f_k^0(\mathbf{h}) T_k(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_{k0}). \tag{9.6}$$

If we let $\mathbf{v} = \mathbf{r} - \mathbf{r}_k$ and (as before) $\mathbf{u} = \mathbf{r}_k - \mathbf{r}_{k0}$, then in eqn (9.6)

$$f_k^0(\mathbf{h}) = \int \rho_k(\mathbf{v}) \exp(2\pi i \mathbf{h} \cdot \mathbf{v}) \, \mathrm{d}^3\mathbf{v} \tag{9.7}$$

is the static scattering factor or form factor of atom k (for neutrons this is replaced by the scattering length b_k) and

$$T_k(\mathbf{h}) = \int p_k(\mathbf{u}) \exp(2\pi i \mathbf{h} \cdot \mathbf{u}) \, \mathrm{d}^3\mathbf{u} \tag{9.8}$$

is the Fourier transform of the pdf p_k for the displacement of atom k from its reference or mean position $\mathbf{r}_{k0}$. The dependence of the structure factor on the atomic displacements has thus been isolated. The quantity $T_k(\mathbf{h})$ has been known by the names "atomic Debye–Waller factor" and "atomic temperature factor", although it is obvious that it is not only the temperature that plays a role in giving rise to the displacement. There are no restrictions on the functional form of the pdf. in eqn (9.8) and, in fact, a variety of such forms have been devised (for example, Coppens 1993).

Let us now invoke assumption 2, the spherical symmetry of the atomic electron density. This, when applied to eqn (9.7), reduces it to the following form:

$$f_k^0(|\mathbf{h}|) = 4\pi \int_0^\infty \rho_k(u) \frac{\sin(2\pi|\mathbf{h}|u)}{2\pi|\mathbf{h}|u} u^2 \, \mathrm{d}u, \tag{9.9}$$

as shown in Chapter 5. Equation (9.9) has been computed and extensively tabulated for all the neutral elements and many ions (Maslen, Fox, and O'Keefe 1992). Of course, the spherical-atom approximation is an idealization and disregards such important asphericities as those due to bonding. However, it is very convenient to apply it to the determination of the basic features of the structure, and may serve as a starting point for more refined determinations of atomic positions and for charge density studies. It may be pointed out that the great majority of structural studies are based on the spherical-atom approximation.

We finally invoke assumption 3, the Gaussian shape of the pdf for the atomic displacement. This can actually be shown to follow directly from the assumption of independently displaced atoms, but the proof will be omitted. Interested readers may consult texts on lattice dynamics for a more rigorous treatment of this subject (for example Willis and Pryor 1975; Coppens 1997). In this approximation, the atomic Debye–Waller factor obtained from eqn (9.8) becomes

$$T(\mathbf{h}) = \exp[-2\pi^2 \langle (\mathbf{h} \cdot \mathbf{u})^2 \rangle], \tag{9.10}$$

where the atomic subscript has been omitted. This equation may be derived from the theory of lattice dynamics in the harmonic approximation, which considers

the (always present) contribution of motion to the atomic displacement, but in practice it may also be applied to minor static displacive disorder. Other approximations have been proposed for situations in which the Gaussian form is not adequate (see, for example, Trueblood *et al.* 1996), but these are applied rather rarely.

Let us now consider some common variants of eqn (9.10), which can be rewritten as

$$T(\mathbf{h}) = \exp\left[-2\pi^2 \left\langle \left(\frac{\mathbf{u}\cdot\mathbf{h}}{|\mathbf{h}|}\right)^2 \right\rangle |\mathbf{h}|^2\right]. \tag{9.11}$$

This shows that the exponent is proportional to the mean square projection of the atomic displacement $\mathbf{u}$ on the direction of the diffraction vector $\mathbf{h}$, times the squared magnitude of $\mathbf{h}$. If we denote the projection of $\mathbf{u}$ on the direction of $\mathbf{h}$ by $u_{\mathbf{h}}$ and make use of the relation $|\mathbf{h}| = 2\sin\theta/\lambda$, eqn (9.11) becomes

$$T(\mathbf{h}) = \exp\left[-8\pi^2 \langle u_{\mathbf{h}}^2 \rangle \frac{\sin^2\theta}{\lambda^2}\right]. \tag{9.12}$$

If the atomic displacements are anisotropic, the value of the average in eqn (9.12) depends on the direction of $\mathbf{h}$. This is then the anisotropic Gaussian Debye–Waller factor $T(\mathbf{h})$, discussed in some detail in the next section and in the following chapter. If, however, the atomic displacements are isotropic, the average in eqn (9.12) is determined by the structure alone, but may be different for nonequivalent atoms, and the left-hand side of eqn (9.12) no longer depends on the direction of $\mathbf{h}$ but only on its magnitude. This is then the individual isotropic atomic Gaussian Debye–Waller factor,

$$T(|\mathbf{h}|) = \exp\left[-8\pi^2 \langle u^2 \rangle_j \frac{\sin^2\theta}{\lambda^2}\right], \tag{9.13}$$

where the index j runs over the nonequivalent atoms in the unit cell. The lowest-order approximation to $T(\mathbf{h})$ is the overall Debye–Waller factor. It has the same form as eqn (9.13) except that the atomic subscript is absent: it presumes that all the atoms have the same isotropic mean square displacement $\langle u^2 \rangle$. The whole crystal structure is assigned, in this approximation, a single displacement parameter. We have already encountered this approximation in the discussion of Wilson's scaling method in Chapter 7. The overall isotropic displacement parameter is also used in the initial stages of the refinement of structural parameters.

9.2 Atomic displacement parameters in the Gaussian approximation

In order to make eqn (9.10) and its variants useful, the diffraction and displacement vectors must be referred to some basis and the Debye–Waller factor must be expressible in terms of the appropriate components of these vectors. The basis in question may be either non-Cartesian or Cartesian, and only those bases

which are generally used will be considered here. We shall indicate the conventional notations for them but the tensor notation, as introduced in Appendix B but without the summation convention, will also be used in most derivations.

9.2.1 Non-Cartesian representations

The diffraction vector, in the non-Cartesian case, is referred to the basis of the reciprocal lattice as follows:

$$\begin{aligned}\mathbf{h} &= h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \\ &\equiv h_1\mathbf{a}^1 + h_2\mathbf{a}^2 + h_3\mathbf{a}^3 \\ &= \sum_{i=1}^{3} h_i\mathbf{a}^i. \end{aligned} \tag{9.14}$$

The atomic displacement vector $\mathbf{u}$, when referred to a non-Cartesian basis, has two main representations. The more straightforward one, with the conventional basis ($\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$) and respective dimensionless components (Δx, Δy, Δz) is

$$\begin{aligned}\mathbf{u} &= \Delta x\,\mathbf{a} + \Delta y\,\mathbf{b} + \Delta z\mathbf{ç} \\ &\equiv \Delta x^1\,\mathbf{a}_1 + \Delta x^2\,\mathbf{a}_2 + \Delta x^3\,\mathbf{a}_3 \\ &= \sum_{j=1}^{3} \Delta x^j\,\mathbf{a}_j. \end{aligned} \tag{9.15}$$

The second representation, with a dimensionless and generally nonorthonormal basis ($a^*\mathbf{a}$, $b^*\mathbf{b}$, $c^*\mathbf{c}$), and respective components with dimensions of length ($\Delta\xi$, $\Delta\eta$, $\Delta\zeta$), is

$$\begin{aligned}\mathbf{u} &= \Delta\xi\ a^*\mathbf{a} + \Delta\eta\ b^*\mathbf{b} + \Delta\zeta\ c^*\mathbf{c} \\ &\equiv \Delta\xi^1\ a^1\mathbf{a}_1 + \Delta\xi^2\ a^2\mathbf{a}_2 + \Delta\xi^3\ a^3\mathbf{a}_3 \\ &= \sum_{j=1}^{3} \Delta\xi^j\ a^j\mathbf{a}_j. \end{aligned} \tag{9.16}$$

This representation is due to Hirshfeld and Rabinovich (1966). Only in the orthorhombic, tetragonal and cubic systems are these basis vectors orthonormal, since it is only in these systems that the equalities $a^* = 1/a$, $b^* = 1/b$, and $c^* = 1/c$ are necessarily true.

Let us now rewrite eqn (9.10) as

$$T(\mathbf{h}) = \exp[-2\pi^2\langle(\mathbf{h}\cdot\mathbf{u})^2\rangle] = \exp[-2\pi^2\langle(\mathbf{h}\cdot\mathbf{u})(\mathbf{u}\cdot\mathbf{h})\rangle] \tag{9.17}$$

and evaluate T in the above-mentioned bases. For the basis of the direct lattice, we have from eqns (9.14) and (9.15)

$$\begin{aligned}\mathbf{h}\cdot\mathbf{u} &= \left(\sum_{i=1}^{3} h_i \mathbf{a}^i\right)\cdot\left(\sum_{j=1}^{3}\Delta x^j \mathbf{a}_j\right)\\ &\equiv \sum_{i=1}^{3}\sum_{j=1}^{3} h_i\,\Delta x^j\,\delta_j^i\\ &= \sum_{k=1}^{3} h_k\,\Delta x^k \end{aligned} \tag{9.18}$$

and, similarly,

$$\mathbf{u}\cdot\mathbf{h} = \sum_{l=1}^{3}\Delta x^l\, h_l. \tag{9.19}$$

We have used here the relationship between the direct and reciprocal bases

$$\mathbf{a}^i\cdot\mathbf{a}_j = \delta_j^i = \begin{cases} 1 \text{ if } & i = j \\ 0 \text{ if } & i \neq j \end{cases}. \tag{9.20}$$

If we insert eqns (9.18) and (9.19) into eqn (9.17), we obtain

$$\begin{aligned} T(\mathbf{h}) &= \exp\left(-2\pi^2\sum_{k=1}^{3}\sum_{l=1}^{3} h_k\langle\Delta x^k\,\Delta x^l\rangle h_l\right)\\ &\equiv \exp\left(-\sum_{k=1}^{3}\sum_{l=1}^{3} h_k\beta^{kl}h_l\right), \end{aligned} \tag{9.21}$$

where

$$\beta^{kl} = 2\pi^2\langle\Delta x^k\,\Delta x^l\rangle. \tag{9.22}$$

The quantity β^{kl} is one of the frequently employed forms of the anisotropic displacement parameter. Its determination will be discussed in the next chapter. Let us now consider the basis of Hirshfeld and Rabinovich (1966). We have, from eqns (9.14) and (9.16)

$$\begin{aligned}\mathbf{h}\cdot\mathbf{u} &= \left(\sum_{i=1}^{3} h_i \mathbf{a}^i\right)\cdot\left(\sum_{j=1}^{3}\Delta\xi^j\, a^j\mathbf{a}_j\right)\\ &\equiv \sum_{i=1}^{3}\sum_{j=1}^{3} h_i\,\Delta\xi^j\, a^j\delta_j^i\\ &= \sum_{k=1}^{3} h_k a^k\,\Delta\xi^k \end{aligned} \tag{9.23}$$

and, similarly,

$$\mathbf{u} \cdot \mathbf{h} = \sum_{l=1}^{3} \Delta\xi^l \, a^l h_l. \tag{9.24}$$

If we now substitute eqns (9.23) and (9.24) into eqn (9.17), we obtain

$$\begin{aligned} T(\mathbf{h}) &= \exp\left(-2\pi^2 \sum_{k=1}^{3}\sum_{l=1}^{3} h_k a^k \langle \Delta\xi^k \, \Delta\xi^l \rangle a^l h_l\right) \\ &\equiv \exp\left(-2\pi^2 \sum_{k=1}^{3}\sum_{l=1}^{3} h_k a^k U^{kl} a^l h_l\right), \end{aligned} \tag{9.25}$$

where

$$U^{kl} = \langle \Delta\xi^k \, \Delta\xi^l \rangle = \frac{\beta_{kl}}{2\pi^2 a^j a^l}, \tag{9.26}$$

another well-known form of the atomic displacement parameter. The components U^{kl} have dimension of $(length)^2$ and can be directly associated with the mean square displacements of the atom considered, in the corresponding directions (see, for example, Trueblood *et al.* 1996).

Another form of the anisotropic displacement parameter, which is often used in biomolecular crystallography, is

$$B^{kl} = 8\pi^2 U^{kl} \tag{9.27}$$

and the corresponding expression for $T(\mathbf{h})$ becomes

$$T(\mathbf{h}) = \exp\left(-\frac{1}{4}\sum_{k=1}^{3}\sum_{l=1}^{3} h_k \, a^k B^{kl} a^l \, h_l\right). \tag{9.28}$$

Obviously, the three non-Cartesian representations of the Debye–Waller factor can be interrelated very readily and none has any significant computational advantage over the others. From the viewpoint of physical significance, the best seems to be that in terms of U^{kl}, because of its clear relationship to a mean square displacement.

9.2.2 Cartesian representations

The transformation of crystallographic to Cartesian coordinates is dealt with in A.1. Briefly, if we refer the atomic displacement vector $\mathbf{u}$ to the crystallographic basis ($\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ or $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$) and the Cartesian basis ($\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$) by:

$$\mathbf{u} = \Delta\xi_1^{\mathrm{C}} \, \mathbf{e}_1 + \Delta\xi_2^{\mathrm{C}} \, \mathbf{e}_2 + \Delta\xi_3^{\mathrm{C}} \, \mathbf{e}_3 = \Delta x^1 \, \mathbf{a}_1 + \Delta x^2 \, \mathbf{a}_2 + \Delta x^3 \, \mathbf{a}_3, \tag{9.29}$$

where the superscript C denotes Cartesian components, it is shown in Section A.1that

$$\Delta\xi_i^{\mathrm{C}} = \sum_{j=1}^{3} A_{ij} x^j, \tag{9.30}$$

where

$$A_{ij} = \mathbf{e}_i \cdot \mathbf{a}^j.$$

We can similarly transform a *product* of components of $\mathbf{u}$ from the lattice to the Cartesian basis, and of course an *average* of such a product:

$$\langle \Delta\xi_j^{\mathrm{C}}\, \Delta\xi_l^{\mathrm{C}} \rangle = \sum_{j=1}^{3} \sum_{l=1}^{3} A_{jm} A_{ln} \langle \Delta x^m\, \Delta x^n \rangle. \tag{9.31}$$

If we define

$$U_{jl}^{\mathrm{C}} = \langle \Delta\xi_j^{\mathrm{C}}\, \Delta\xi_l^{\mathrm{C}} \rangle \tag{9.32}$$

as a component of an atomic mean square displacement tensor, with dimensions of [length]^2, referred to a Cartesian basis, and make use of eqn (9.22), we obtain

$$U_{jl}^{\mathrm{C}} = (2\pi^2)^{-1} \sum_{j=1}^{3} \sum_{l=1}^{3} A_{jm} A_{ln} \beta^{mn}. \tag{9.33}$$

This is the expression for the transformation of β to Cartesian components. Components of the matrix A are given in Section A.1 for a specific choice of the Cartesian basis.

The transformation of the components of $\mathbf{U}$ proceeds in a similar manner. We express the atomic displacement vector $\mathbf{u}$ in the two bases of interest by

$$\mathbf{u} = \Delta\xi_1^{\mathrm{C}}\, \mathbf{e}_1 + \Delta\xi_2^{\mathrm{C}}\, \mathbf{e}_2 + \Delta\xi_3^{\mathrm{C}}\, \mathbf{e}_3 = \Delta\xi^1\, a^* \mathbf{a}_1 + \Delta\xi^2\, a^2 \mathbf{a}_2 + \Delta\xi^3\, a^* \mathbf{a}_3, \tag{9.34}$$

and, similarly to the derivation of eqn (9.30), we obtain

$$\Delta\xi_i^{\mathrm{C}} = \sum_{j=1}^{3} C_{ij} x^j, \tag{9.35}$$

where

$$C_{ij} = (\mathbf{e}_i \cdot \mathbf{a}^j) a^j.$$

If we proceed to the transformation of products of coordinates and averages of these products, we have

$$\langle \Delta\xi_j^{\mathrm{C}}\, \Delta\xi_l^{\mathrm{C}} \rangle = \sum_{j=1}^{3} \sum_{l=1}^{3} C_{jm} C_{ln} \langle \Delta\xi^m\, \Delta\xi^n \rangle, \tag{9.36}$$

and finally

$$U_{jl}^{\mathrm{C}} = \sum_{j=1}^{3} \sum_{l=1}^{3} C_{jm} C_{ln} U^{mn}, \tag{9.37}$$

where U_{jl}^{C} and U^{mn} are given by eqns (9.32) and (9.26), respectively.

These Cartesian representations of anisotropic displacement parameters are frequently employed in constrained refinement calculations and in computations related to lattice dynamics. They will also be referred to in the next section.

9.3 A rigid-body model

9.3.1 Introduction

Atomic displacements in molecular crystals have two main sources: external displacement (due to motion or disorder) of the molecule as a whole, and internal vibrations, involving interatomic bonds only, and internal rotations. The amplitude of the internal vibrations is usually very small, as compared to the amplitude of the external displacements, and its effect on the observed diffraction intensities negligible. Internal rotations may be more important in flexible molecules but in many instances the molecules are approximately rigid or consist of rigid fragments, and in such cases the displacement vector of an atom has two main contributions: the displacement of the atom due to rotation of the molecule to which it belongs, $\mathbf{u}_{\rm rot}$, and the displacement of the atom due to the translational displacement of the molecule as a whole, $\mathbf{u}_{\rm trans}$. As we shall see later, $\mathbf{u}_{\rm rot}$ depends on the position of the atom, while $\mathbf{u}_{\rm trans}$ is independent of its position.

Let us assume for convenience that all the quantities are referred to the same Cartesian system, and employ a symbolic matrix notation. Thus, if a vector is denoted by $\mathbf{u}$ (bold upright), we denote a column made up of its components (a 3 by 1 matrix) by $\boldsymbol{u}$ (bold italic). The components of an atomic displacement vector can thus be written as

$$\boldsymbol{u} = \boldsymbol{u}_{\rm rot} + \boldsymbol{u}_{\rm trans}$$

and the (Cartesian) mean square displacement tensor can be written as

$$\begin{aligned}\langle \boldsymbol{u}\boldsymbol{u}^{\rm T}\rangle &= \langle (\boldsymbol{u}_{\rm rot} + \boldsymbol{u}_{\rm trans})(\boldsymbol{u}_{\rm rot} + \boldsymbol{u}_{\rm trans})^{\rm T}\rangle \\ &= \langle \boldsymbol{u}_{\rm rot}\boldsymbol{u}_{\rm rot}^{\rm T}\rangle + \langle \boldsymbol{u}_{\rm rot}\boldsymbol{u}_{\rm trans}^{\rm T}\rangle + \langle \boldsymbol{u}_{\rm trans}\boldsymbol{u}_{\rm rot}^{\rm T}\rangle + \langle \boldsymbol{u}_{\rm trans}\boldsymbol{u}_{\rm trans}^{\rm T}\rangle\end{aligned}$$

In this model, the contributions of the molecular displacement to the mean square displacement tensor of an atom are the following: $\langle \boldsymbol{u}_{\rm rot}\boldsymbol{u}_{\rm rot}^{\rm T}\rangle$ represents the contribution of the molecular rotational or librational displacement, $\langle \boldsymbol{u}_{\rm trans}\boldsymbol{u}_{\rm trans}^{\rm T}\rangle$ represents the contribution of the molecular translational displacement, and $\langle \boldsymbol{u}_{\rm rot}\boldsymbol{u}_{\rm trans}^{\rm T}\rangle$ and $\langle \boldsymbol{u}_{\rm trans}\boldsymbol{u}_{\rm rot}^{\rm T}\rangle$ represent the contribution of screw-type, or helical, displacement of the molecule (from the interaction of rotation and translation)

We shall specify this model in some detail, and find out that (i) it leads to a sharp decrease in the number of structural parameters, and (ii) it allows one to determine approximately the parameters of molecular motion.

9.3.2 Rotational displacement

Suppose an atom at $\mathbf{r} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$, belonging to a rigid molecule, is rotated about an axis parallel to a unit vector $\mathbf{k} = \boldsymbol{\lambda}/|\boldsymbol{\lambda}|$ through an angle $|\boldsymbol{\lambda}|$, where $\boldsymbol{\lambda} = \lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + \lambda_3\mathbf{e}_3$ is the vector of rotation. Note that we revert, for the time being, to the vector notation. The rotational displacement $\mathbf{u}_{\rm rot} = u_{rot,1}\mathbf{e}_1 + u_{rot,2}\mathbf{e}_2 + u_{rot,3}\mathbf{e}_3$, is given by

$$\mathbf{u}_{\rm rot} = C_{\boldsymbol{\lambda}/|\boldsymbol{\lambda}|}(|\boldsymbol{\lambda}|)\mathbf{r} - \mathbf{r},$$

where $C_{\boldsymbol{\lambda}/|\boldsymbol{\lambda}|}(|\boldsymbol{\lambda}|)$ is the finite rotation operator, derived in Section A.3. If we make use of this result, we have

$$\mathbf{u}_{\rm rot} = \frac{\boldsymbol{\lambda}}{|\boldsymbol{\lambda}|}\left(\frac{\boldsymbol{\lambda}}{|\boldsymbol{\lambda}|}\cdot\mathbf{r}\right)[1-\cos(|\boldsymbol{\lambda}|)] + \mathbf{r}[\cos(|\boldsymbol{\lambda}|)-1] + \left(\frac{\boldsymbol{\lambda}}{|\boldsymbol{\lambda}|}\times\mathbf{r}\right)\sin(|\boldsymbol{\lambda}|). \quad (9.38)$$

We shall first assume that the angle of rotation is very small, which is consistent with an ordered crystal structure in which the molecules execute small oscillations about their mean positions or have minor departures of rotational nature from perfect periodicity and symmetry. In this case

$$\cos(|\boldsymbol{\lambda}|) \approx 1 - \frac{|\boldsymbol{\lambda}|^2}{2}$$

and

$$\sin(|\boldsymbol{\lambda}|) \approx |\boldsymbol{\lambda}|,$$

and eqn (9.38) reduces to

$$\mathbf{u}_{\rm rot} \approx \frac{1}{2}\left[\boldsymbol{\lambda}\boldsymbol{\lambda} - |\boldsymbol{\lambda}|^2\mathbf{I}\right]\cdot\mathbf{r} + \boldsymbol{\lambda}\times\mathbf{r}, \quad (9.39)$$

where $\mathbf{I}$ is a unit operator. The first term on the right-hand side of eqn (9.39) is interesting in connection with the correction of bond distances for libration but will now be discarded since the mean square displacement is of interest and only quantities of the order of $|\boldsymbol{\lambda}|^2$ can survive after "squaring" eqn (9.39). Hence

$$\mathbf{u}_{\rm rot} \approx \boldsymbol{\lambda}\times\mathbf{r}. \quad (9.40)$$

Following the examples given in Appendix B, the ith component of $\mathbf{u}_{\rm rot}$ is given by

$$u_{rot,i} \approx (\boldsymbol{\lambda}\times\mathbf{r})_i = e_{ijk}\lambda_j x_k, \quad (9.41)$$

where e_{ijk} equals $+1$, -1, or 0, according to whether ijk is a cyclic permutation of 123, a noncyclic permutation of 123, or any two indices are equal, respectively; summations over j and k are implied in eqn (9.41). If we carry out the summation over k, we obtain

$$u_{rot,i} \approx R_{ij}\lambda_j, \quad \text{where} \quad R_{ij} = e_{ijk}x_k, \quad (9.42)$$

and the antisymmetric matrix $\boldsymbol{R}$ is

$$\begin{pmatrix} 0 & x_3 & -x_2 \\ -x_3 & 0 & x_1 \\ x_2 & -x_1 & 0 \end{pmatrix}$$

Finally, the ilth component of the purely rotational contribution to the mean square displacement tensor of the atom is approximated as

$$\langle u_{rot,i}u_{rot,l}\rangle \approx R_{ij}R_{lm}\langle\lambda_j\lambda_m\rangle \equiv R_{ij}R_{lm}L_{jm}. \quad (9.43)$$

The average $L_{jm} = \langle\lambda_j\lambda_m\rangle$ is the jmth component of the symmetric libration tensor $\boldsymbol{L}$.

9.3.3 Rigid-body model parameters

We can now complete the derivation of the rigid-body model. The ith coordinate of the atomic displacement vector is given by

$$\begin{aligned} u_i &= u_{rot,i} + u_{trans,i} \\ &\approx R_{ij}\lambda_j + u_{trans,i} \end{aligned}$$

and the average product of two such coordinates or, in other words, a component of the (Cartesian) atomic mean square displacement tensor, can be expressed as

$$\begin{aligned} \langle u_i u_l \rangle &\approx \langle (R_{ij}\lambda_j + u_{trans,i})(R_{lm}\lambda_m + u_{trans,l}\rangle \\ &= R_{ij}R_{lm}\langle \lambda_j \lambda_m \rangle + R_{ij}\langle \lambda_j u_{utrans,l}\rangle + R_{lm}\langle u_{trans,i}\lambda_m\rangle + \langle u_{trans,i}u_{trans,l}\rangle \end{aligned} \tag{9.44}$$

The averages in eqn (9.44) contain the parameters of the motion of the rigid molecule, subject to the approximation stated above. Thus, $T_{il} \equiv \langle u_{trans,i}u_{trans,l}\rangle$ is a component of the symmetric *translation* tensor, $L_{jm} \equiv \langle \lambda_j \lambda_m \rangle$ is a component of the symmetric *libration* tensor (as also shown above), and $S_{jl} \equiv \langle \lambda_j u_{trans,l}\rangle$ is a component of the nonsymmetric *screw-motion* tensor. The matrices $\boldsymbol{T}$ and $\boldsymbol{L}$ contribute six independent components each, and the trace of the matrix $\boldsymbol{S}$ is indeterminate, so that only two diagonal components can be determined (Schomaker and Trueblood, 1968). Therefore, there are altogether 20 parameters of motion for approximately rigid molecule or molecular fragment.

In the description by anisotropic atomic parameters each anisotropic atom has six independent displacement parameters. Suppose a rigid molecule contains 15 anisotropically refined (see next chapter) atoms and therefore provides 90 displacement parameters. However, the motion of a rigid molecule can be described, in general, in terms of 20 parameters - a sharp decrease of the number of parameters is indeed offered by the rigid body model. This is, however, an approximation since internal molecular motion is not accounted for.

How are these parameters obtained? One way is to set up, for an N-atom molecule, a system of $6N$ linear equations in 20 unknowns, the components of the tensors $\boldsymbol{T}$, $\boldsymbol{L}$ and $\boldsymbol{S}$, and solve it for these parameters by the least-squares method (see next chapter), while regarding the refined components of the atomic displacemement tensors as observations. Another method is to formulate the structure factor in terms of the rigid-body parameters, rather than in terms of atomic anisotropic displacement tensors, and obtain them from a constrained refinement procedure. This will be discussed in the next chapter.

This page intentionally left blank

10 Refinement of structural parameters

10.1 Introduction

Whether a crystal structure is determined by Patterson techniques or by direct methods, the outcome is of necessity approximate and the approximation is usually poorer than that which would have been obtained if all the experimental data had been correctly taken into account. Indeed, Patterson methods are largely dependent on the strongest reflections, which are mainly confined to low-angle regions, and the whole body of data from high-angle reflections is ignored to a large extent. The resulting structure has a lower resolution than what the data might lead to. The direct methods, on the other hand, exploit a relatively small number of reflections with high E values, coming mainly from the high-angle region. The resolution may thus be high, but errors are introduced as a result of omitting most of the data. The practical implication of all of these shortcomings is a low accuracy of the structural parameters. Since a crystallographic experiment can usually be designed so that the ratio of the number of observations to that of parameters is fairly large, it is possible to let the parameters obtained from the structure determination vary so that an optimal agreement of the calculated and observed magnitudes of the structure factors is obtained. The process of carrying out such a confrontation of a model with experimental data, as a result of which the parameters of the model are improved, is known as *refinement of the model parameters*. Several methods allowing such systematic variation are known in numerical analysis. We shall, however, content ourselves with the description of the method most popular in crystal structure analyses, the *least-squares method*. The next two sections will be devoted to an introductory presentation of the relevant variants of the least-squares method, as a preparation for the actual refinement of structural parameters.

10.2 The linear least-squares method

This method is in principle suitable for the treatment of sets of observations which are linear combinations of *unknown* parameters with *exactly known* coefficients. The observations are accompanied by small, unknown errors, and these are assumed to obey a Gaussian distribution. It will be seen that, under this assumption, the "best" set of parameters is obtained by requiring that the sum of the squares of all the errors accompanying the observations is minimal. Hence the name of the method.

Consider an overdetermined system of linear equations

$$b_r = \sum_{j=1}^{n} A_{rj} p_j, \quad r = 1, 2, \ldots, m, \quad m > n. \tag{10.1}$$

Suppose we have to find the quantities $p_1, \ldots, p_n$, on the basis of eqn (10.1), where the coefficients A_{rj} are exactly known while $b_1, \ldots, b_m$ are accompanied by unknown errors $\epsilon_1, \ldots, \epsilon_m$, respectively. Our set of observations can thus be written as

$$o_r = b_r + \epsilon_r, \quad r = 1, 2, \ldots, m. \tag{10.2}$$

As outlined above, we assume further that the errors $\epsilon_1, \ldots, \epsilon_m$ are independent and are normally distributed with zero mean and estimated standard deviations $\sigma_1, \ldots, \sigma_m$, where

$$\sigma_r^2 = \frac{\sigma^2}{w_r}, \quad r = 1, 2, \ldots, m. \tag{10.3}$$

Here, the quantities $w_1, \ldots, w_m$ are *known* weights and σ^2 is an unknown. Our task is therefore to estimate the values of the parameters $p_1, \ldots, p_n$ and the unknown σ on the basis of the observations $o_1, \ldots, o_m$ as formulated above. We shall employ for this purpose the method of maximum likelihood, as described in greater detail by Linnik (1961). We have

$$o_r = \epsilon_r + \sum_{j=1}^{n} A_{rj} p_j \tag{10.4}$$

where ϵ_r is a random variable with zero mean and an estimated standard deviation given by $\sigma/\sqrt{w_r}$, and the second term on the right-hand side of eqn (10.4)depends on the constants A_{rj} and independent parameters $p_1, \ldots, p_n$. It follows that the average $\langle \epsilon_r \rangle$ vanishes, and hence

$$\langle o_r \rangle = \sum_{j=1}^{n} A_{rj} p_j. \tag{10.5}$$

In view of the assumed normal distribution of the errors ϵ_r, the probability density function for an observation o_r is

$$f_r(o_r) = \frac{w_r^{1/2}}{(2\pi\sigma^2)^{1/2}} \exp\left\{ -\frac{w_r}{2\sigma^2} \left(o_r - \sum_{j=1}^{n} A_{rj} p_j \right)^2 \right\}. \tag{10.6}$$

The likelihood function for the sample of independent observations $o_1, \ldots, o_m$ is then given by the product of the pdfs of the individual observations

$$L(o_1,\ldots,o_m)=\prod_{i=1}^{m} f_r(o_r) \tag{10.7}$$

$$=\frac{\left(\prod_{i=1}^{m} w_r\right)^{1/2}}{(2\pi\sigma^2)^{m/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{r=1}^{m} w_r\left(o_r-\sum_{j=1}^{n} A\_rjp_j\right)^2\right\} \tag{10.8}$$

and the likelihood should attain its maximum when the function or its logarithm is at its maximum (Cramér 1951; Linnik 1961). That is,

$$\begin{aligned}\ln L(o_1,\ldots,o_m) &= -\frac{1}{2}m\ln 2\pi-\frac{1}{2}m\ln\sigma^2+\frac{1}{2}\ln\left(\prod_{i=1}^{m} w_i)\right)\\ &\quad -\frac{1}{2\sigma^2}\sum_{r=1}^{m} w_r\left(o_r-\sum_{j=1}^{n} A_{rj}p_j\right)^2\\ &= \max.\end{aligned} \tag{10.9}$$

We see from eqn (10.9) that for each value of σ^2 the likelihood function attains its maximum for the sample $p_1,\ldots,p_n$–independently of σ^2–when

$$M\equiv\sum_{r=1}^{m} w_r\left(o_r-\sum_{j}^{n} A_{rj}p_j\right)^2=\sum_{r=1}^{m} w_r\epsilon_r^2=\min, \tag{10.10}$$

where use has also been made of eqn (10.4). Equation (10.10) is the fundamental condition from which the least-squares equations can be derived, as shown below.

The above requirement for a minimum with respect to the parameters $p_1,\ldots,p_n$ implies that

$$\frac{\partial M}{\partial p_k}=2\sum_{r=1}^{m} w_r\epsilon_r\frac{\partial\epsilon_r}{\partial p_k}=0,\quad k=1,2,\ldots,n. \tag{10.11}$$

Since, now,

$$\epsilon_r=o_r-\sum_{j=1}^{n} A_{rj}p_j$$

and

$$\frac{\partial\epsilon_r}{\partial p_k}=-A_{rk},$$

eqn (10.11) becomes

$$\sum_{r=1}^{m} w_r \sum_{j=1}^{n} A_{rk} A_{rj} p_j = \sum_{r=1}^{m} w_r A_{rk} o_r, \quad k = 1, 2, \ldots, n \tag{10.12}$$

or, interchanging the summations over observations and parameters,

$$\sum_{j=1}^{n} \left\{ \sum_{r=1}^{m} A_{kr}^{\mathrm{T}} w_r A_{rj} \right\} p_j = \sum_{r=1}^{m} A_{kr}^{\mathrm{T}} w_r o_r, \quad k = 1, 2, \ldots, n, \tag{10.13}$$

where the superscript T denotes the transpose. Equation (10.13) constitutes a system of n linear equations in the n parameters $p_1, \ldots, p_n$ and is solvable provided the matrix of the coefficients of the ps on the left-hand side of eqn (10.13) is nonsingular. These equations are known as the *normal* equations of the least-squares problem. If we introduce a diagonal weight matrix $\boldsymbol{W}$, such that $W_{kk} = w_k, k = 1, 2, \ldots, m$, eqn (10.13) can be rewritten in symbolic matrix notation as

$$\boldsymbol{A}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{A} \boldsymbol{p}' = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{o} \tag{10.14}$$

and the solution vector is then

$$\boldsymbol{p}' = (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{A})^{-1} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{o}. \tag{10.15}$$

In many cases the weights are unknown and are assumed to be the same for each observation (or observational equation). In such cases unit weights are usually assumed, $\boldsymbol{W}$ becomes a unit matrix, and eqns (10.14) and (10.15) simplify to

$$\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{p}' = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{o} \tag{10.16}$$

and

$$\boldsymbol{p}' = (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A})^{-1} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{o}, \tag{10.17}$$

respectively. The solution vector is primed in order to emphasize that this is not the actual exact solution but only a least-squares estimate of such a solution, following from the condition in eqn (10.10).

We shall now discuss, without detailed proof, expressions for the estimated standard deviations of parameters that have been estimated by the least-squares method. Our best estimate of the parameter σ^2 is the weighted sum of the squared differences between the actual observations and those recalculated from the least-squares estimates of the parameters, divided by the number of degrees of freedom. That is

$$\sigma^2 = \frac{\sum_{r=1}^{m} w_r (o_r - o_r')^2}{m - n}, \tag{10.18}$$

where

$$o'_r = \sum_{j=1}^{n} A_{rj} p'_j. \tag{10.19}$$

Finally, the variance–covariance matrix of the set of parameters $\{p_i\}$ is given by the corresponding diagonal elements of the inverse matrix of the matrix of the normal equations, scaled by σ^2 as computed from eqn (10.18):

$$\mathrm{cov}(p'_i, p'_j) = [(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}]_{ij}\sigma^2. \tag{10.20}$$

The estimated standard deviation of the parameter p'_i, recently renamed the *estimated standard uncertainty* in the crystallographic literature, is just the square root of the corresponding diagonal element of the variance–covariance matrix.

10.2.1 An example of a linear least-squares problem

There are several straightforward applications of the linear least-squares method in crystallography. We came across one of them in the discussion of rigid-body tensors in Chapter 9, and will illustrate another one in this section. The determination of unit-cell parameters is carried out by a variety of methods, usually dictated by the structure of the diffraction experiment. In some arrangements, notably the moving-crystal–moving-film methods, the diffraction pattern can be indexed in advance of the determination of the unit cell parameters, and such a determination is then easily carried out with the aid of the linear least-squares method. The squared diffraction vector corresponding to the ith reflection can be written as

$$\begin{aligned}
|\boldsymbol{h}_i|^2 &= \frac{4\sin^2\theta_i}{\lambda^2} \\
&= (h_i\mathbf{a}^* + k_i\mathbf{b}^* + l_i\mathbf{c}^*)\cdot(h_i\mathbf{a}^* + k_i\mathbf{b}^* + l_i\mathbf{c}^*) \\
&= h_i^2 a^{*2} + k_i^2 b^{*2} + l_i^2 c^{*2} \\
&\quad + 2k_i l_i b^* c^* \cos\alpha^* + 2l_i h_i c^* a^* \cos\beta^* + 2h_i k_i a^* b^* \cos\gamma^*.
\end{aligned}$$

The observed quantity, which has to include some unknown error, is θ_i. We can therefore write for the ith reflection

$$o_i \equiv \frac{4\sin^2\theta_i}{\lambda^2}.$$

The reflection indices are exactly known integers, and we can denote the ith row of the matrix of observational equations as

$$\begin{aligned}
A_{i1} &\equiv h_i^2, \\
A_{i2} &\equiv k_i^2, \\
A_{i3} &\equiv l_i^2, \\
A_{i4} &\equiv 2k_i l_i, \\
A_{i5} &\equiv 2l_i h_i, \\
A_{i6} &\equiv 2h_i k_i.
\end{aligned}$$

The parameters to be solved for are thus

$$\begin{aligned}
p_1 &\equiv a^{*2} = \mathbf{a}^* \cdot \mathbf{a}^*, \\
p_2 &\equiv b^{*2} = \mathbf{b}^* \cdot \mathbf{b}^*, \\
p_3 &\equiv c^{*2} = \mathbf{c}^* \cdot \mathbf{c}^*, \\
p_4 &\equiv b^* c^* \cos\alpha^* = \mathbf{b}^* \cdot \mathbf{c}^*, \\
p_5 &\equiv c^* a^* \cos\beta^* = \mathbf{c}^* \cdot \mathbf{a}^*, \\
p_6 &\equiv a^* b^* \cos\gamma^* = \mathbf{a}^* \cdot \mathbf{b}^*.
\end{aligned}$$

We shall assume unit weights for simplicity. The normal equations can therefore be constructed as in eqn (10.16), the solution vector estimated as in eqn (10.17), and the variances of the parameters estimated as in eqns (10.18)–(10.20). The parameters so obtained are, of course, not the unit-cell parameters we wish to determine. However, these parameters can be seen to be the elements of the matrix of the metric tensor of the reciprocal space (see eqn (B.8)) The inverse of this matrix can be readily shown to be the matrix of the metric tensor of the direct space (see eqn (B.7)), and so the required (direct) unit-cell parameters can be obtained.

10.3 The non-linear least-squares method

A more frequently occurring situation–and, unfortunately, more difficult to cope with–is a set of observations which are nonlinear functions of the parameters. If, however, the values of the parameters are approximately known, it is possible to "linearize" the problem and apply the method described in the previous section in an iterative manner until convergence to a satisfactory solution has been reached. We shall explain this procedure in what follows.

Consider an overdetermined system of equations

$$f_r(p_1, \ldots, p_n) = g_{r,\mathrm{obs}} + \epsilon_r, \quad r = 1, \ldots, m \ (m > n), \tag{10.21}$$

where the functions f_r, which model the observations $g_{r,\mathrm{obs}}$, depend on the parameters p_i in a nonlinear manner, and each equation contains a small unknown error ϵ. Even if all the errors were zero, the solution of this system would involve treatment of a coupled system of nonlinear equations, which is a complicated undertaking. However, the errors are there, and their presence dictates a statistical approach akin to that presented in the previous section.

Let us assume that the parameters are approximately known. The right-hand side of eqn (10.21) can thus be written as

$$f_r(p_1^0 + \delta p_1, \ldots, p_n^0 + \delta p_n)$$

or, more concisely, as $f_r(\boldsymbol{p}^0 + \delta\boldsymbol{p})$, where $\boldsymbol{p}^0$ is the array of approximately known parameter values and $\delta\boldsymbol{p}$ is the array of unknown deviations of these approximate values from the "true" values of the parameters.

A widely used procedure is to expand the *model function* $f_r(\boldsymbol{p}^0 + \delta\boldsymbol{p})$ in a Taylor series about the vector of the approximately known parameters and regard the deviations δp_j as the unknowns, the values of which are to be estimated. Such an expansion becomes

$$f_r(\boldsymbol{p}^0 + \delta\boldsymbol{p}) = f_r(\boldsymbol{p}^0) + \sum_{j=1}^{n} \left.\frac{\partial f_r(\boldsymbol{p})}{\partial p_j}\right|_{(p_j = p_j^0)} \delta p_j + \frac{1}{2} \sum_{j=1}^{n}\sum_{k=1}^{n} \left.\frac{\partial^2 f_r(\boldsymbol{p})}{\partial p_j \partial p_k}\right|_{(p_j = p_j^0, p_k = p_k^0)} \delta p_j \delta p_k + \cdots \qquad (10.22)$$

This expansion becomes useful when the terms from the third on can be neglected. This can be done only when the approximation is a good one, that is, if the deviations δp_r are small. In that case, $\delta p_j\, \delta p_k$ can be neglected in comparison with δp_i and eqn (10.22) reduces to two terms only: a constant and a linear term. The starting equation, eqn (10.21), can then be approximated by

$$\sum_{j=1}^{n} \left.\frac{\partial f_r(\boldsymbol{p})}{\partial p_j}\right|_{(p_j = p_j^0)} \delta p_j \cong g_{r,\mathrm{obs}} - f_r(\boldsymbol{p}^0) + \epsilon_r, \quad r = 1, \ldots, m \ (m > n), \quad (10.23)$$

and becomes an overdetermined system of linear equations in the deviations δp_j, with exactly calculable coefficients. The solution of such a system for the deviations can be done by the linear least-squares method, described in the previous section: the substitutions to be done are

$$\sum_{j=1}^{n} \left.\frac{\partial f_r(\boldsymbol{p})}{\partial p_j}\right|_{(p_j = p_j^0)} \delta p_j \equiv \sum_{j=1}^{n} A_{rj} p_j \qquad (10.24)$$

and

$$g_{r,\mathrm{obs}} - f_r(\boldsymbol{p}^0) \equiv b_r. \qquad (10.25)$$

When the deviations are available, they can be added to the initial approximate values of the parameters to create a better approximation, and the procedure is repeated in an iterative manner until some convergence is reached. The nonlinear least-quares method really consists in a series of applications of the linear method. The obvious difference is that while the linear least-squares method yields the values of the parameters, the nonlinear method furnishes approximate deviations of the parameters from their "true" values.

10.4 Conventional refinement

The most relevant application of the method described in the previous section to structure determination is to the refinement of structural parameters. The model function to be considered is the magnitude of the conventional structure factor, $|F(\boldsymbol{h})|$, or its square, $|F(\boldsymbol{h})|^2$. We have, for the rth reflection,

$$\begin{aligned} F(\boldsymbol{h}_{(r)}) &= |F(\boldsymbol{h}_{(r)})| \exp(i\varphi) = |F(\boldsymbol{h}_{(r)})|(\cos\varphi + i\sin\varphi) \\ &\equiv A(\boldsymbol{h}_{(r)}) + iB(\boldsymbol{h}_{(r)}) \end{aligned} \tag{10.26}$$

and hence $A(\boldsymbol{h}_{(r)}) = |F(\boldsymbol{h}_{(r)})| \cos\varphi$ and $B(\boldsymbol{h}_{(r)}) = |F(\boldsymbol{h}_{(r)})| \sin\varphi$. It follows from the above that

$$|F(\boldsymbol{h}_{(r)})| = A(\boldsymbol{h}_{(r)}) \cos\varphi + B(\boldsymbol{h}_{(r)}) \sin\varphi. \tag{10.27}$$

The conventional expression for the structure factor of the rth reflection, in symbolic notation, is given by

$$F(\boldsymbol{h}_{(r)}) = \sum_m a_m f_m^0 \sum_s \exp(-\boldsymbol{h}_{(r)}^{\mathrm{T}} \beta_{ms} \boldsymbol{h}_{(r)}) \exp(2\pi i \boldsymbol{h}_{(r)} \cdot \boldsymbol{x}_{ms}), \tag{10.28}$$

where the outer summation ranges over the atoms within the asymmetric unit; a_m is the occupancy factor of the mth atom; f_m^0 is the scattering factor of the mth atom at rest; and β_{ms} is the anisotropic displacement tensor of the mth atom, transformed by the rotation matrix of the sth space-group operation, that is:

$$\beta_{ms} = \boldsymbol{P}_s \beta_m \boldsymbol{P}_s^{\mathrm{T}}. \tag{10.29}$$

The coordinate vector in eqn (10.28) is given by the space-group transformation:

$$\boldsymbol{x}_{ms} = \boldsymbol{P}_s \boldsymbol{x}_m + \boldsymbol{t}_s \equiv (\boldsymbol{P}_s, \boldsymbol{t}_s) \boldsymbol{x}_m. \tag{10.30}$$

The observed quantity, denoted in eqn (10.23) by $g_{(r),\mathrm{obs}}$, is the observed amplitude $|F_{(r),\mathrm{obs}}|$ of the rth reflection. We therefore need the derivatives of the magnitude of the structure factor with respect to all the parameters, the value of the magnitude of the structure factor at the current approximate parameter vector, and the observed structure amplitude. These derivatives are most conveniently calculated if we use eqn (10.27) with the real and imaginary parts of eqn (10.28). It is convenient to introduce the abbreviations

$$A(\boldsymbol{h}_{(r)}) = \sum_m a_m f_m^0 \sum_s C_{(r)sm}$$

and

$$B(\boldsymbol{h}_{(r)}) = \sum_m a_m f_m^0 \sum_s S_{(r)sm}$$

where

$$C_{(r)sm} = \exp(-h_{(r)sp}h_{(r)sq}\beta_m^{pq})\cos[2\pi(h_{(r)sp}x_m^p + h_{(r)i}t_s^i)], \tag{10.31}$$

and where

$$h_{(r)sp} = h_{(r)i}P_{sp}^i.$$

Here $h_{(r)i}$ is the ith component of the diffraction vector of the rth reflection, and P_{sp}^i is the ipth element of the rotation matrix belonging to the sth space-group operator; the summation convention is employed in eqn (10.31). Analogous equations are obtained for $S_{(r)sm}$, by replacing the cosine with a sine.

The derivatives required for the construction of the least-squares normal equations are

$$\begin{aligned}\frac{\partial|F(\boldsymbol{h}_{(r)})|}{\partial a_k} &= \frac{\partial A(\boldsymbol{h}_{(r)})}{\partial a_k}\cos\varphi + \frac{\partial B(\boldsymbol{h}_{(r)})}{\partial a_k}\sin\varphi \\ &= f_k^0\left[\left(\sum_s C_{(r)sk}\right)\cos\varphi + \left(\sum_s S_{(r)sk}\right)\sin\varphi\right],\end{aligned} \tag{10.32}$$

$$\begin{aligned}\frac{\partial|F(\boldsymbol{h}_{(r)})|}{\partial \beta_k^{pq}} &= \frac{\partial A(\boldsymbol{h}_{(r)})}{\partial \beta_k^{pq}}\cos\varphi + \frac{\partial B(\boldsymbol{h}_{(r)})}{\partial \beta_k^{pq}}\sin\varphi \\ &= f_k^0\left(\sum_s -h_{(r)sp}h_{(r)sq}C_{(r)sk}\right)\cos\varphi \\ &+ f_k^0\left(\sum_s -h_{(r)sp}h_{(r)sq}S_{(r)sk}\right)\sin\varphi,\end{aligned} \tag{10.33}$$

$$\begin{aligned}\frac{\partial|F(\boldsymbol{h}_{(r)})|}{\partial x_k^p} &= \frac{\partial A(\boldsymbol{h}_{(r)})}{\partial x_k^p}\cos\varphi + \frac{\partial B(\mathbf{h}_{(r)})}{\partial x_k^p}\sin\varphi \\ &= f_k^0\left[\left(\sum_s -2\pi h_{(r)sp}S_{(r)sk}\right)\cos\varphi + \left(\sum_s 2\pi h_{(r)sp}C_{(r)sk}\right)\sin\varphi\right].\end{aligned} \tag{10.34}$$

In this approximation each atom is specified, in the general case, by ten parameters: three coordinates (known as positional parameters), six components of the symmetric displacement tensor (often referred to as thermal parameters) and one occupancy parameter. The overwhelming majority of structural studies of small and medium-sized molecules employ these specifications.

The parameter vector for an n-atom asymmetric unit is thus

$$\begin{aligned}\boldsymbol{p}^{\mathrm{T}} = (&a_1, \beta_1^{11}, \beta_1^{22}, \beta_1^{33}, \beta_1^{12}, \beta_1^{23}, \beta_1^{31}, x_1^1, x_1^2, x_1^3, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_n, \beta_n^{11}, \beta_n^{22}, \beta_n^{33}, \beta_n^{12}, \beta_n^{23}, \beta_n^{31}, x_n^1, x_n^2, x_n^3),\end{aligned} \tag{10.35}$$

the linearization proceeds just as outlined in the previous section for the general case, and the normal equations are constructed as prescribed by eqn (10.13). According to eqn (10.12), they can be written as

$$\sum_{j=1}^{n}\left\{\sum_{(r)=1}^{m} w_{(r)}\frac{\partial|F_{(r)}(\boldsymbol{p})|}{\partial p_k}\frac{\partial|F_{(r)}(\boldsymbol{p})|}{\partial p_j}\right\}\delta p_j$$
$$=\sum_{(r)=1}^{m} w_{(r)}\frac{\partial|F_{(r)}(\boldsymbol{p})|}{\partial p_k}(|F_{(r),obs}|-|F_{(r)}(\boldsymbol{p}^0|), \tag{10.36}$$

where $w_{(r)}$ is the weight of the rth reflection, often taken as the reciprocal of the variance of the measured intensity and not uncommonly as unity; and $\partial|F_{(r)}(\boldsymbol{p})|/\partial p_k$ is the partial derivative of the model structure-factor function with respect to the kth parameter, evaluated at the currently known approximate values of the parameters; m is the number of reflections participating in the refinement, n is the number of parameters and the unknowns δp_j are the parameter shifts we are after, in order to add them to the current approximation and improve or *refine* it. After this has been done (of course, by the computer program that takes care of this process) eqns (10.36) are solved again, this time with an improved "current approximation" and the parameter vector is updated until the process is thought to have converged. There are several criteria in use, the most common being the discrepancy factor, or R factor, defined as, for example,

$$R^i=\frac{\sum_{(r)}||F_{(r)}(\boldsymbol{p}^i)|-|F_{(r),\mathrm{obs}}||}{\sum_{(r)}|F_{(r),\mathrm{obs}}|} \tag{10.37}$$

for the ith iteration of the refinement. When all the parameters listed in eqn (10.35) have been included for the nonhydrogen atoms, the hydrogen atoms have been located and included with isotropic displacement parameters, and the structure is not disordered, we expect the R factor for the last iteration to have a value not exceeding 0.05.

10.4.1 Rigid-group constraints

The parameters which conventionally participate in the refinement procedure are (i) three positional parameters for each atom, or the atomic coordinates; (ii) one isotropic or six anisotropic displacement parameters for each atom; and (iii) an occupancy factor for each atom, of particular importance in partially disordered structures. Not infrequently, the asymmetric unit of the crystal consists of one or more approximately rigid groups, in which case the number of structural parameters can be considerably reduced, at this level of approximation. In this subsection we shall consider the incorporation of such rigid-group contraints into the structure-factor equation and the general formulation of the derivatives required for the construction of the least-squares normal equations. A general discussion of constrained refinement procedures can be found in Chapter 8.3 of

Volume C of the *International Tables for Crystallography* (Prince, Finger, and Konnert 1999) and the references quoted therein.

For the purpose of our discussion, let us assume that all the quantities appearing in the structure-factor equation are referred to the same Cartesian system. This can be done as indicated in Appendices A and B (see Sections A.1 and B.4) and is frequently done in crystallographic computing.

If anisotropic displacement of the atoms in the Gaussian approximation is assumed, the structure factor is given by

$$F(H_1^{\mathrm{C}}H_2^{\mathrm{C}}H_3^{\mathrm{C}}) = \sum_{j=1}^{N} a(j) f_j^0 \exp(-2\pi^2 H_k^{\mathrm{C}} U_{km}^{\mathrm{C}}(j) H_m^{\mathrm{C}}) \exp[2\pi i H_p^{\mathrm{C}} x_p^{\mathrm{C}}(j)],$$

where N is the number of fully or partially occupied atomic sites; $a(j)$ is the occupancy factor of the jth atomic site; f_j^0 is the scattering factor of the jth atom at rest (usually in the spherical approximation); $H_1^{\mathrm{C}}, H_2^{\mathrm{C}}, H_3^{\mathrm{C}}$ are the components of the diffraction vector $\boldsymbol{h}$, referred to a Cartesian system; $U_{km}^{\mathrm{C}}(j)$ is the kmth element of the matrix of the displacement tensor of the jth atom, referred to the same Cartesian system, and $x_p^{\mathrm{C}}(j)$ is the pth component of the position vector of the jth atom, referred to the same Cartesian system. We shall assume, for generality,the space group $P1$. Other space groups can of course be included in the expression for the structure factor, as shown in Chapter 5.

If the structure is fully ordered (that is, $a(j) = 1$ for all j), the number of structural parameters in the above formulation of the structure factor, for a modest value $N = 20$, is 180, and these have to be conventionally refined as described above. Let us now assume that the 20 atoms constitute a rigid group of known internal structure but approximate location and orientation. The number of structural parameters describing this group now reduces to 26. As shown in Chapter 9, the number of independent parameters in the translation, libration, and screw-motion tensors of a rigid group is 20 [rather than 120 displacement parameters in eqn (10.38)]. Further, as will be shown below, three parameters define the orientation of the rigid group and three coordinates are required to specify its location in the unit cell.

A general displacement of a point can be described as

$$\mathbf{r}_f = C_{\mathbf{k}}(\alpha)\mathbf{r}_i + \rho, \tag{10.38}$$

where $\mathbf{r}_i$ is the initial position vector of the point, $\mathbf{r}_f$ is the position vector of the point after the displacement, $C_{\mathbf{k}}(\alpha)$ is an operator rotating the point about an axis parallel to a unit vector $\mathbf{k}$ through an angle α, and ρ is a linear displacement vector. A matrix representation of the rotation operator, referred to a Cartesian system, is derived in Appendix A (see eqn (A.23)). It depends on three independent parameters: the angle α and two components of the unit vector $\mathbf{k}$. The third component depends on the other two, since $|\mathbf{k}| = 1$. A computationally more transparent equation for a change of orientation and location is

$$\begin{pmatrix} x_{1f}^{\mathrm{C}} \\ x_{2f}^{\mathrm{C}} \\ x_{3f}^{\mathrm{C}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & -s_1 \\ 0 & s_1 & c_1 \end{pmatrix} \begin{pmatrix} c_2 & 0 & s_2 \\ 0 & 1 & 0 \\ -s_2 & 0 & c_2 \end{pmatrix} \begin{pmatrix} c_3 & -s_3 & 0 \\ s_3 & c_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1i}^{\mathrm{C}} \\ x_{2i}^{\mathrm{C}} \\ x_{3i}^{\mathrm{C}} \end{pmatrix} + \begin{pmatrix} \rho_{1i}^{\mathrm{C}} \\ \rho_{2i}^{\mathrm{C}} \\ \rho_{3i}^{\mathrm{C}} \end{pmatrix}, \quad (10.39)$$

where $c_i = \cos\phi_i$, $s_i = \sin\phi_i$, $i = 1, 2.3$, or symbolically, in matrix notation,

$$\boldsymbol{r}_f^{\mathrm{C}} = \boldsymbol{M}(\phi_1)\boldsymbol{M}(\phi_2)\boldsymbol{M}(\phi_3)\boldsymbol{r}_i^{\mathrm{C}} + \boldsymbol{\rho}^{\mathrm{C}}, \quad (10.40)$$

where ϕ_1, ϕ_2, ϕ_3 are angles of rotation about the Cartesian axes 1,2,3 respectively.

The structure-factor equation that incorporates the displacement and positional aspects of this rigid-body model can therefore be written as

$$F(\boldsymbol{h}^{\mathrm{C}}) = \sum_{j=1}^{N} f_j^0 \exp[-2\pi^2 (\boldsymbol{h}^{\mathrm{C}})^{\mathrm{T}} \boldsymbol{U}_{\mathrm{r.b.}}^{\mathrm{C}}(j) \boldsymbol{h}^{\mathrm{C}}] \exp[2\pi i \boldsymbol{h}^{\mathrm{C}} \boldsymbol{r}_f^{\mathrm{C}}(j)], \quad (10.41)$$

where

$$\boldsymbol{U}_{\mathrm{r.b.}}^{\mathrm{C}}(j) = \boldsymbol{R}(j)\boldsymbol{L}\boldsymbol{R}^{\mathrm{T}}(j) + \boldsymbol{R}(j)\boldsymbol{S} + [\boldsymbol{R}(j)\boldsymbol{S}]^{\mathrm{T}} + \boldsymbol{T}$$

is the contribution of the general displacement of the jth atom in the rigid body to the displacement tensor of this atom (see Chapter 9), and $\boldsymbol{r}_f^{\mathrm{C}}(j)$ is given by eqn (10.40). The derivatives now required for the construction of the normal equations are the derivatives of $|F|$ with respect to the angular parameters ϕ_1, ϕ_2, ϕ_3, the components of the location vector $\boldsymbol{\rho}$ and the 20 independent components of the rigid-body tensors $\boldsymbol{T}$, $\boldsymbol{L}$, and $\boldsymbol{S}$. Their computation is conveniently done with the aid of the chain rule. For example, the contributions of an atom (with its subscript omitted) to these derivatives are

$$\frac{\partial|F|}{\partial\phi_m} = \sum_{j=1}^{3} \frac{\partial|F|}{\partial x_{jf}^{\mathrm{C}}} \frac{\partial x_{jf}^{\mathrm{C}}}{\partial\phi_m}, m = 1, 2, 3, \quad (10.42)$$

$$\frac{\partial|F|}{\partial\rho_n} = \sum_{j=1}^{3} \frac{\partial|F|}{\partial x_{jf}^{\mathrm{C}}} \frac{\partial x_{jf}^{\mathrm{C}}}{\partial\rho_n}, n = 1, 2, 3, \quad (10.43)$$

$$\frac{\partial|F|}{\partial L_{pq}} = \sum_{i=1}^{3} \sum_{j=i}^{3} \frac{\partial|F|}{\partial U_{r.b.,ij}^{\mathrm{C}}} \frac{\partial U_{r.b.,ij}^{\mathrm{C}}}{\partial L_{pq}}, p, q = 1, 2, 3,\ p \geq q, \quad (10.44)$$

and similarly for the derivatives with respect to the components of $\boldsymbol{S}$ and $\boldsymbol{T}$. Note, however, that the displacement tensor in the rigid-body approximation also depends on the atomic positions via the antisymmetric matrix $\boldsymbol{R}$ and therefore the derivatives $\partial|F|/\partial x$ must take this into account.

The introduction of space-group symmetry follows in the usual manner, as in the case of conventional refinement. We can, of course, have one or more rigid group in the asymmetric unit, each having its 26 rigid-body parameters. There

are also structures in which the molecule consists of several rigid fragments, and a number of atoms not belonging to a rigid group. The structure factor can be set up to take this situation into account.

Finally, it should be remembered that the rigid-body model is an approximation and that by assuming it we completely neglect internal motion such as bond stretching and bending. Internal vibrations are almost independent of the temperature whereas the external ones depend on it strongly. Therefore, for low-temperature structures the rigid-body model is worse than for structures solved from diffraction data collected at room temperature.

10.4.2 Introduction of restraints

Let us now return to conventional refinement. The solution of the normal equations, eqns (10.36), for the shifts of the parameters involves the inversion of the normal matrix, which is usually very fast with modern personal computers or workstations for moderately small asymmetric units. However, the quality of the results and their standard uncertainties depend on the ratio of the number of measured intensities to the number of refined parameters. A low ratio may be caused by a very large asymmetric unit (for example in the case of protein crystals) or by a weakly diffracting crystal. In such cases, introduction of known stereochemistry of the crystal into the process of refinement may prove to be very helpful or even necessary. An introductory discussion of this subject was given by Waser (1963); it was extended by Konnert (1976) and other authors, reviewed by Prince *et al.* (1999) and implemented in several crystallographic program packages (for example SHELX97, Sheldrick (1997)). The introduction of a known structural feature, such as an interatomic distance or an interbond angle is known as a *restraint* and will be briefly discussed below. We shall follow the papers of Waser (1963) and Konnert (1976) in an occasionally different notation.

The weighted sum of squared residuals, which after minimization and the linearization described above leads to the normal equations (eqns (10.36)), is given by

$$S_1 = \sum_{(r)} w_{(r)}(F)(|F_{(r),\mathrm{obs}}| - |F_{(r),\mathrm{calc}}|)^2, \tag{10.45}$$

where the sum extends over all the observed reflections, $w_{(r)}(F)$ is the weight assigned to the rth reflection, and $F_{(r),\mathrm{calc}}$ is the calculated structure factor based on a parameter vector $\boldsymbol{p}$ (see eqn (10.35)). Suppose we wish to adjust the positional parameters of certain atoms so that they give the best fit to a set of given interatomic distances. The sum of weighted squares of the corresponding residuals is in this case

$$S_2 = \sum_I \sum_K w_{IK}\{[g_{jl}(x_I^j - x_K^j)(x_I^l - x_K^l)]^{1/2} - d_{IK,\mathrm{fix}}\}^2, \quad I \geq K, \tag{10.46}$$

where each term in the double summation corresponds to the pair of atoms I and K, g_{jl} is the jlth element of the metric tensor of the direct lattice, x_I^j is the jth fractional coordinate of atom I, $d_{IK,\mathrm{fix}}$ is the prescribed distance between

atoms I and K, and the repeated-index summation convention is employed in the calculation of the distance between the two atoms.

The least-squares equations are now constructed by requiring that

$$\sum_{i=1}^{m}\left(\frac{\partial S_1}{\partial p_i}+\frac{\partial S_2}{\partial p_i}\right)=0, \tag{10.47}$$

where p_i is a structural parameter and the summation ranges over all m such parameters. Of course, many of the derivatives of S_2 will be zero if only the restraint given by eqn (10.46) is considered. However, there are many possible restraints that can be imposed, related to atomic positions and displacements. A most useful positional restraint is the requirement that a group of atoms be planar and have a fixed geometry–this is almost invariably applied to the peptide linkage which is the basic building block of the backbone of a polypeptide chain in a protein molecule. An interesting restraint involving atomic anisotropic displacement tensors is the rigid-bond restraint. This requires that the average displacement of atom I toward atom J along an IJ bond be the same as the average displacement of atom J toward atom I along this bond, or, quantitatively,

$$\boldsymbol{k}_{IJ}^{\mathrm{T}}\boldsymbol{U}_I\boldsymbol{k}_{IJ}=\boldsymbol{k}_{JI}^{\mathrm{T}}\boldsymbol{U}_J\boldsymbol{k}_{JI} \tag{10.48}$$

where $\boldsymbol{k}_{IJ}$ is the column of components of a unit vector directed from atom I to atom J, and $\boldsymbol{U}_I$ is the matrix of the displacement tensor of atom I, etc. in the reverse direction.

The actual normal equations, in matrix notation, are then given by

$$\boldsymbol{N}\,\delta\boldsymbol{p}=\boldsymbol{b} \tag{10.49}$$

where the kjth element of $\boldsymbol{N}$ is

$$\begin{aligned}N_{kj}=&\sum_{(r)=1}^{m}w_{(r)}\frac{\partial|F_{(r),\mathrm{calc}}|}{\partial p_k}\frac{\partial|F_{(r),\mathrm{calc}}|}{\partial p_j}\\&+\sum_I\sum_K w_{IK}(d_{IK})\frac{\partial d_{IK,\mathrm{calc}}^2}{\partial p_k}\frac{\partial d_{IK,\mathrm{calc}}^2}{\partial p_j}\\&+\text{terms from other restraints,}\end{aligned} \tag{10.50}$$

δp_j is the shift in the jth parameter, and kth component of $\boldsymbol{b}$ is

$$\begin{aligned}b_k=&\sum_{(r)=1}w_{(r)}(F)\frac{\partial|F_{(r),\mathrm{calc}}|}{\partial p_k}(|F_{(r),\mathrm{obs}}|-|F_{(r),\mathrm{calc}}|)\\&+\sum_I\sum_K w_{(r)}(d_{IK})\frac{\partial d_{IK,\mathrm{calc}}^2}{\partial p_k}(d_{IK,\mathrm{fix}}^2-d_{IK,\mathrm{calc}}^2)\\&+\text{terms from other restraints;}\end{aligned} \tag{10.51}$$

$d_{IK,\mathrm{calc}}$, the calculated interatomic distance, was given explicitly in eqn (10.46).

It should, of course, be remembered that restraints are valuable for refinement of structures which are too large for the data that are available and also for weakly diffracting crystals, with either large or small unit cells. Application of restraints is not a "must" if a full, accurately measured data set is available and the ratio of the number of that of observations to conventional parameters is fairly large.

This page intentionally left blank

Appendix A Some geometrical considerations

A.1 Transformation to a Cartesian system

A Cartesian basis $\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ can be readily related to the bases of the direct and reciprocal lattices by taking, for example, $\mathbf{e}_1$ along $\mathbf{a}$, $\mathbf{e}_3$ along $\mathbf{c}^*$, and $\mathbf{e}_2$ along the vector product $\mathbf{e}_3 \times \mathbf{e}_1$. The resulting unit vectors

$$\mathbf{e}_1 = \frac{\mathbf{a}}{|\mathbf{a}|}, \tag{A.1}$$

$$\mathbf{e}_2 = \mathbf{e}_3 \times \mathbf{e}_1, \tag{A.2}$$

$$\mathbf{e}_3 = \frac{\mathbf{c}^*}{|\mathbf{c}^*|}. \tag{A.3}$$

comprise an appropriate, and commonly employed, orthonormal set of basis vectors. In fact, it is readily seen that $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}, i, j = 1, 2, 3$, as required for a Cartesian basis. Let us now refer the vector $\mathbf{r}$ to each of the basis sets $\mathbf{abc}$ and $\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ and write

$$\mathbf{r} = \xi_1^{\mathrm{C}}\mathbf{e}_1 + \xi_2^{\mathrm{C}}\mathbf{e}_2 + \xi_3^{\mathrm{C}}\mathbf{e}_3 = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}. \tag{A.4}$$

If we take the scalar products of the left-hand and right-hand sides of (A.4) with $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, we obtain three linear equations, or a matrix equation of the form

$$\begin{pmatrix} \xi_1^{\mathrm{C}} \\ \xi_2^{\mathrm{C}} \\ \xi_3^{\mathrm{C}} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 \cdot \mathbf{a} & \mathbf{e}_1 \cdot \mathbf{b} & \mathbf{e}_1 \cdot \mathbf{c} \\ \mathbf{e}_2 \cdot \mathbf{a} & \mathbf{e}_2 \cdot \mathbf{b} & \mathbf{e}_2 \cdot \mathbf{c} \\ \mathbf{e}_3 \cdot \mathbf{a} & \mathbf{e}_3 \cdot \mathbf{b} & \mathbf{e}_3 \cdot \mathbf{c} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{A.5}$$

This is a transformation of the components of $\mathbf{r}$, referred to the basis of the direct lattice, to its Cartesian components. The transformation matrix can be evaluated once the Cartesian basis vectors have been defined (for example as in eqns (A.1)–(A.3)) by making use of the fundamental relationships between the direct and reciprocal basis vectors. We leave it to the reader to show that the explicit form of the transformation matrix appearing in eqn (A.5), for the specific Cartesian basis defined in eqns (A.1)–(A.3), is:

$$\begin{pmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & -c\sin\beta\cos\alpha^* \\ 0 & 0 & 1/c^* \end{pmatrix}. \tag{A.6}$$

Of course, a Cartesian basis associated with the direct and/or reciprocal bases can be chosen in an unlimited number of ways. A more general discussion of the construction of Cartesian bases is given elsewhere (Shmueli, 2001*b*).

When a crystallographic calculation involving diffraction vectors is referred to a Cartesian basis, we have to transform the diffraction indices to the same basis as that to which the fractional coordinates have been transformed as shown above. The required transformation matrix can be obtained in a manner analogous to that outlined above; however, this is not needed, since the row of diffraction indices referred to the Cartesian system can be obtained by postmultiplying a row of indices hkl with a matrix which is the inverse of the matrix given by eqn (A.6). Note that the latter matrix premultiplies the fractional coordinates and transforms them to the Cartesian system. A general proof of this relationship is given in Appendix B (see Section B.4).

A.2 Change of basis and trace invariance

A simple but important example of the effect of a change of basis on the components of a vector was discussed in the previous section of this appendix. We now proceed to determine the effect of a change of basis on the elements of a square matrix, and conclude with a proof of the invariance of the trace of the matrix under such a change. The discussion will be general.

Let $\boldsymbol{P}$ be an n by n matrix representation of an operator, which acts on a coordinate column $\boldsymbol{r}_i$ and let $\boldsymbol{r}_f$ be the coordinate column resulting from this operation. We can write this symbolically as

$$\boldsymbol{r}_f = \boldsymbol{P}\boldsymbol{r}_i. \tag{A.7}$$

This is exemplified by eqn (1.37), which shows an application of a rotational symmetry operator to a lattice vector, resulting in a symmetrically related lattice vector.

Now let $\boldsymbol{T}$ be an n by n nonsingular transformation matrix, which when applied to a column vector, say $\boldsymbol{r}_i$, yields the same column vector referred to a different basis. We shall write this as

$$\boldsymbol{r}'_i = \boldsymbol{T}\boldsymbol{r}_i. \tag{A.8}$$

This is exemplified by eqn (A.5) in the previous section. Since the matrix $\boldsymbol{T}$ is nonsingular, we have

$$\boldsymbol{r}_i = \boldsymbol{T}^{-1}\boldsymbol{r}'_i, \tag{A.9}$$

and upon substituting eqn (A.9) into eqn (A.7) we obtain

$$\mathbf{r}_f = \boldsymbol{P}\boldsymbol{T}^{-1}\boldsymbol{r}'_i. \tag{A.10}$$

If we now apply the transformation $\boldsymbol{T}$ to both sides of eqn (A.10), the operation expressed by eqn (A.7) is now referred to the primed basis:

$$\boldsymbol{r}'_f = \boldsymbol{T}\boldsymbol{P}\boldsymbol{T}^{-1}\boldsymbol{r}'_i. \tag{A.11}$$

The matrix representation of the operator has changed from $\boldsymbol{P}$ to $\boldsymbol{T}\boldsymbol{P}\boldsymbol{T}^{-1}$. This rather well-known result is applied in crystallographic computing to situations

in which calculations are being referred to a basis different from that initially assumed.

An interesting consequence of the above result is the fact that the trace of the matrix representation of an operator is invariant under a change of basis. Let us recall that the trace of a square matrix, say $\boldsymbol{A}$, can be conveniently computed as

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i \sum_j \delta_{ij} A_{ij} = \sum_i A_{ii}. \tag{A.12}$$

We now use this result to compute the trace of the matrix representation of our operator after a change of basis has taken place:

$$\begin{aligned}
\mathrm{Tr}(\boldsymbol{TPT}^{-1}) &= \sum_i \sum_j \delta_{ij} (\boldsymbol{TPT}^{-1})_{ij} \\
&= \sum_i \sum_j \delta_{ij} \sum_k \sum_l T_{ik} P_{kl} T^{-1}_{lj} \\
&= \sum_i \sum_k \sum_l T_{ik} P_{kl} T^{-1}_{li} \\
&= \sum_i \sum_k \sum_l T^{-1}_{li} T_{ik} P_{kl} \\
&= \sum_k \sum_l \delta_{lk} P_{kl} \\
&= \sum_k P_{kk} \\
&\equiv \mathrm{Tr}(\boldsymbol{P}).
\end{aligned} \tag{A.13}$$

The trace of the square-matrix representation of an operator is indeed invariant under a change of basis. This general result was used in the derivation of the permissible rotational symmetries in a three-dimensional lattice.

A.3 The finite rotation operator

We shall now consider an operator on a three-dimensional space which has several important applications in crystallography, and to which the considerations of the previous sections of this appendix can be profitably applied. This operator rotates (rigidly) a point about an axis coinciding with a unit vector $\mathbf{k}$, through an angle α, in the clockwise sense when the observer looks in the direction of $+\mathbf{k}$; this is usually referred to as the positive sense. We shall first derive a representation of this operator in vector form, without specifying the basis to which the vectors are referred, and proceed to represent the components of the vectors in a Cartesian system, which is the most convenient type of system for general rotations and general considerations regarding rotational symmetry. The derivation is indicated in Fig. A.1 and is detailed below.

We assume that all the vectors are referred to an origin, located on the axis of rotation. Let $\mathbf{k}$ be a unit vector along the axis of rotation, $\mathbf{r}$ be the vector to be

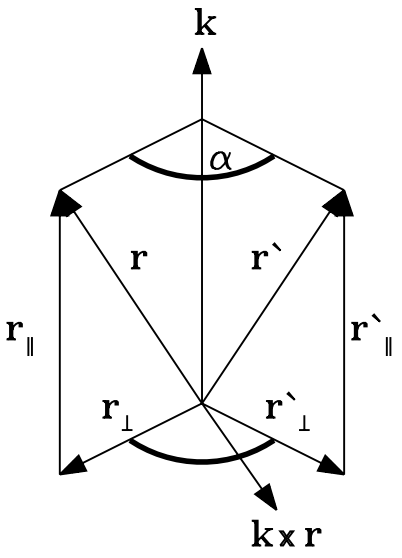


Fig. A.1 Derivation of the rotation operator.

rotated about $\mathbf{k}$, $\mathbf{r}'$ be the vector after its rotation about $\mathbf{k}$, and α be the angle of rotation. We first decompose the unprimed and primed vectors into components parallel and perpendicular to the axis of rotation. That is

$$\mathbf{r} = \mathbf{r}_{\parallel} + \mathbf{r}_{\perp}, \tag{A.14}$$

Since we have assumed a rigid rotation, that is, that the length of the vector does not change, the components of the unprimed and primed vectors parallel to the axis of rotation are also identical, and it is easily seen that they are given by

$$\mathbf{r}_{\parallel} = \mathbf{r}'_{\parallel} = \mathbf{k}(\mathbf{k} \cdot \mathbf{r}). \tag{A.15}$$

We now proceed to express $\mathbf{r}'_{\perp}$ in terms of $\mathbf{r}$, $\mathbf{k}$ and α. For this purpose, we construct an orthogonal system in the plane passing through the origin and perpendicular to the axis of rotation. This plane contains the perpendicular components $\mathbf{r}_{\perp}$ and $\mathbf{r}'_{\perp}$, and also the vector $\mathbf{k} \times \mathbf{r}_{\perp}$. Note, however, that

$$\mathbf{k} \times \mathbf{r}_{\perp} = \mathbf{k} \times \mathbf{r}. \tag{A.16}$$

We now decompose $\mathbf{r}'_\perp$ into components parallel to the mutually perpendicular vectors $\mathbf{r}_\perp$ and $\mathbf{k}\times\mathbf{r}$:

$$\mathbf{r}'_\perp = \left(\mathbf{r}'_\perp \cdot \frac{\mathbf{r}_\perp}{|\mathbf{r}_\perp|}\right)\frac{\mathbf{r}_\perp}{|\mathbf{r}_\perp|} + \left(\mathbf{r}'_\perp \cdot \frac{\mathbf{k}\times\mathbf{r}}{|\mathbf{k}\times\mathbf{r}|}\right)\frac{\mathbf{k}\times\mathbf{r}}{|\mathbf{k}\times\mathbf{r}|}. \tag{A.17}$$

The projections of $\mathbf{r}$ and $\mathbf{r}'$ on the perpendicular plane also have the same length. That is, $|\mathbf{r}_\perp| = |\mathbf{r}'_\perp|$. The first term on the right-hand side of eqn (A.17) thus reduces to $\mathbf{r}_\perp\cos\alpha$. The angle formed by $\mathbf{r}'_\perp$ and $\mathbf{k}\times\mathbf{r}$ is $|\alpha - 90°|$, and we have from eqn (A.16)

$$|\mathbf{k}\times\mathbf{r}| = |\mathbf{k}\times\mathbf{r}_\perp| = |\mathbf{r}_\perp|. \tag{A.18}$$

It follows from eqn (A.15), eqn (A.17), and the above argument that

$$\begin{aligned}\mathbf{r}' &= \mathbf{k}(\mathbf{k}\cdot\mathbf{r}) + \mathbf{r}_\perp\cos\alpha + (\mathbf{k}\times\mathbf{r})\sin\alpha \\ &= \mathbf{k}(\mathbf{k}\cdot\mathbf{r})(1-\cos\alpha) + \mathbf{r}\cos\alpha + (\mathbf{k}\times\mathbf{r})\sin\alpha.\end{aligned} \tag{A.19}$$

The second line of eqn (A.19) is the general expression from which the components of $\mathbf{r}'$ can be computed in *any* system of coordinates, given those of $\mathbf{r}$ and $\mathbf{k}$ referred to the same system, and the angle α. Let us rewrite this result in matrix notation; that is find the matrix which, when applied to an unprimed coordinate column. yields the corresponding primed coordinate column. For simplicity (and for the time being), we choose a Cartesian system of coordinates, and define the various vectors as

$$\mathbf{k} = \sum_{i=1}^{3} k_i\mathbf{e}_i, \tag{A.20}$$

$$\mathbf{r} = \sum_{i=1}^{3} x_i\mathbf{e}_i, \tag{A.21}$$

and

$$\mathbf{r}' = \sum_{i=1}^{3} x'_i\mathbf{e}_i. \tag{A.22}$$

If we rewrite eqn (A.19) in components, as indicated above, we obtain

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} k_1^2 D(\alpha) + \cos\alpha & k_1k_2D(\alpha) - k_3\sin\alpha & k_1k_3D(\alpha) + k_2\sin\alpha \\ k_2k_1D(\alpha) + k_3\sin\alpha & k_2^2D(\alpha) + \cos\alpha & k_2k_3D(\alpha) - k_1\sin\alpha \\ k_3k_1D(\alpha) - k_2\sin\alpha & k_3k_2D(\alpha) + k_1\sin\alpha & k_3^2D(\alpha) + \cos\alpha \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \tag{A.23}$$

where $D(\alpha) = (1-\cos\alpha)$, or, symbolically,

$$\mathbf{x}' = C_{\mathbf{k}}(\alpha)\mathbf{x} \tag{A.24}$$

The trace of this matrix representation

$$\mathrm{Tr}[C_{\mathbf{k}}(\alpha)] = (k_1^2 + k_2^2 + k_3^2)(1-\cos\alpha) + 3\cos\alpha = 1 + 2\cos\alpha \tag{A.25}$$

depends on the angle of rotation alone. This result is the basis of the classification of permissible symmetric rotations presented in the text. The matrix in

eqn (A.23) is the general Cartesian representation of a finite rotation operator on a three-dimensional space. It finds use in several crystallographic applications, such as the Patterson rotation function, (see, for example, Rossmann and Arnold, 1993) and rigid-body motion analysis (Schomaker and Trueblood 1968).

Conversely, if we are given a general rotation matrix, with elements C_{ij}, the angle of rotation may be determined from its trace as

$$\alpha = \cos^{-1}\left[frac(C_{11} + C_{22} + C_{33} - 1)2\right], \tag{A.26}$$

independently of the system of coordinates. When the angle α is different from 180°, the Cartesian components of the unit vector along the axis of rotation can be found from

$$k_1 = \frac{C_{32} - C_{23}}{2\sin\alpha}, \tag{A.27}$$

$$k_2 = \frac{C_{13} - C_{31}}{2\sin\alpha}, \tag{A.28}$$

$$k_3 = \frac{C_{21} - C_{12}}{2\sin\alpha}. \tag{A.29}$$

as can be seen by inspecting the matrix in eqn (A.23).

Appendix B Fundamentals of tensor notation

B.1 Introduction

Tensor notation is of considerable advantage in many areas of crystallographic computing. This appendix is not a substitute for a text on tensor algebra, but is intended to summarize the most important principles and present some relevant examples. It may be noted that among the advantages of tensor notation are conciseness, and ease of transcription of formulas into computer code. We shall adhere to the following conventions:

1. The basis vectors of the direct and reciprocal lattices, discussed in Chapter 1, form mutually reciprocal triplets, and will be denoted as follows:

$$\mathbf{a} = \mathbf{a}_1,\ \mathbf{b} = \mathbf{a}_2,\ \mathbf{c} = \mathbf{a}_3$$
$$\mathbf{a}^* = \mathbf{a}^1,\ \mathbf{b}^* = \mathbf{a}^2,\ \mathbf{c}^* = \mathbf{a}^3.$$

   Tensor algebra associates subscripted quantities with *covariant*, and superscripted quantities with *contravariant* transformation properties. Thus the basis vectors of the direct lattice are represented as covariant quantities and those of the reciprocal lattice as contravariant ones.
2. *The summation convention.* If an index appears twice in an expression, once as a subscript and once as a superscript, a summation over this index is thereby implied and the summation sign is omitted. For example,

$$x^i x^j R_{ij} \quad \text{is understood as} \quad \sum_i \sum_j x^i x^j R_{ij}$$

   by the above definition. Such repeated indices are often called *dummy* indices and, indeed, the replacement of two is by two ks has no effect on the result. The implied summation over repeated indices is often used in practice when the indices are at the same level. This is correct when the coordinate systems are Cartesian, as there is no distinction between covariant and contravariant quantities in Cartesian frames of reference.
3. Components of vectors referred to a covariant basis are written as contravariant quantities and vice versa. For example

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} = x^1\mathbf{a}_1 + x^2\mathbf{a}_2 + x^3\mathbf{a}_3 = x^i\mathbf{a}_i$$

$$\mathbf{h} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* = h_1\mathbf{a}^1 + h_2\mathbf{a}^2 + h_3\mathbf{a}^3 = h_i\mathbf{a}^i.$$

B.2 Transformations

A concept of fundamental importance in tensor algebra is that of the transformation of coordinates. For example, suppose that a position vector $\mathbf{r}$ is referred to two different sets of basis vectors as follows:

$$\mathbf{r} = x^i \mathbf{a}_i \tag{B.1}$$

and

$$\mathbf{r} = x'^k \mathbf{a}'_k. \tag{B.2}$$

Let us now multiply both sides of eqns (B.1) and (B.2) by the vectors $\mathbf{a}^p, p = 1, 2, 3$, that is, the basis vectors that are reciprocal to the set $\{\mathbf{a}_i\}$. We obtain from eqn (B.1)

$$\mathbf{r} \cdot \mathbf{a}^p = x^i \mathbf{a}_i \cdot \mathbf{a}^p = x^i \delta_i^p = x^p,$$

where δ_i^p is the Kronecker delta, which equals 1 when $i = p$ and equals 0 when $i \neq p$. Further, eqn (B.2) yields

$$\mathbf{r} \cdot \mathbf{a}^p = x'^k \mathbf{a}'_k \cdot \mathbf{a}^p \equiv x'^k T_k^p,$$

where $T_k^p = \mathbf{a}'_k \cdot \mathbf{a}^p$. Hence,

$$x^p = x'^k T_k^p \tag{B.3}$$

which can also be written as

$$x^p = T_k^p x'^k. \tag{B.4}$$

A *tensor* is a quantity which transforms as a product of coordinates, and the *rank* or *order* of the tensor is the number of transformations involved. For example, a product of two coordinates transforms from the primed to the unprimed basis as

$$x^p x^q = T_m^p T_n^q x'^m x'^n. \tag{B.5}$$

The same transformation law is valid for a contravariant tensor of order two, the components of which are referred to the primed basis and are to be transformed to the unprimed basis:

$$Q^{pq} = T_m^p T_n^q Q'^{mn}. \tag{B.6}$$

B.3 The scalar product and metric tensors

The scalar product $\mathbf{u} \cdot \mathbf{v}$ is obviously an invariant. However, if it is expressed in terms of the components of the vectors and the bases to which the vectors are referred, several equivalent expressions for the scalar product result. If we assume that the possible bases are the covariant and the corresponding contravariant basis, the following four expressions are obtained:

(1) $\mathbf{u} = u^i \mathbf{a}_i, \quad \mathbf{v} = v^j \mathbf{a}_j$

$$\mathbf{u} \cdot \mathbf{v} = u^i v^j (\mathbf{a}_i \cdot \mathbf{a}_j) \equiv u^i v^j g_{ij}, \tag{B.7}$$

where $g_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j$ is the metric tensor of the direct space.

(2) $\mathbf{u} = u_i \mathbf{a}^i, \quad \mathbf{v} = v_j \mathbf{a}^j$

$$\mathbf{u} \cdot \mathbf{v} = u_i v_j (\mathbf{a}^i \cdot \mathbf{a}^j) \equiv u_i v_j g^{ij}, \tag{B.8}$$

where $g^{ij} = \mathbf{a}^i \cdot \mathbf{a}^j$ is the metric tensor of the reciprocal space.

(3) $\mathbf{u} = u_i \mathbf{a}^i, \quad \mathbf{v} = v^j \mathbf{a}_j$

$$\mathbf{u} \cdot \mathbf{v} = u_i v^j (\mathbf{a}^i \cdot \mathbf{a}_j) \equiv u_i v^j \delta^j_i = u_i v^i. \tag{B.9}$$

(4) $\mathbf{u} = u^i \mathbf{a}_i, \quad \mathbf{v} = v_j \mathbf{a}^j$

$$\mathbf{u} \cdot \mathbf{v} = u^i v_j (\mathbf{a}_i \cdot \mathbf{a}^j) \equiv u^i v_j \delta^i_j = u^i v_i. \tag{B.10}$$

If we compare eqn (B.7) with eqn (B.10), we obtain

$$v_i = v^j g_{ij} = g_{ij} v^j. \tag{B.11}$$

That is, the metric tensor of the direct space transforms the contravariant components of a vector into its covariant components. More pictorially, a multiplication by the metric tensor of the direct space "lowers" one of the indices of the multiplied contravariant (column) vector. Similarly, a comparison of eqn (B.8) with eqn (B.9) leads to

$$v^i = v_j g^{ij} = g^{ij} v_j, \tag{B.12}$$

and to transformation properties inverse to those indicated by eqn (B.11). A comparison of eqn (B.11) with eqn (B.12) clearly shows that the matrices of the direct and reciprocal metric tensors are mutually reciprocal, and so are their determinants. This is consistent with the reciprocity of the volumes of the direct and reciprocal unit cells, already demonstrated in the text.

It can be readily shown that analogous relations exist between the covariant and contravariant basis vectors.

B.4 Examples

The background provided in this Appendix suffices for a variety of highly useful crystallographic applications. Only a few will be mentioned here, by way of an illustration.

1. *Interatomic distance.* Let $\mathbf{r}_1 = x^i_1 \mathbf{a}_i$ and $\mathbf{r}_2 = x^i_2 \mathbf{a}_i$ be the position vectors of atoms 1 and 2, referred to the origin of the unit cell. The interatomic vector is given by $\mathbf{r}_1 - \mathbf{r}_2 = (x^i_1 - x^i_2)\mathbf{a}_i$. If we denote the components of the interatomic vector by $u^i_{12} = x^i_1 - x^i_2$, the squared magnitude of this vector is given by

$$(\mathbf{r}_1 - \mathbf{r}_2) \cdot (\mathbf{r}_1 - \mathbf{r}_2) = u^i_{12} u^j_{12} g_{ij}.$$

This result can be readily extended to the formulation of an intervector angle and dihedral angle.

2. *Magnitude of a diffraction vector.* Let $\mathbf{h} = h_i\mathbf{a}^i$, and recall that $|\mathbf{h}| = 2\sin\theta/\lambda$. We thus have

$$|\mathbf{h}|^2 = \mathbf{h}\cdot\mathbf{h} = h_i h_j g^{ij} = \frac{4\sin^2\theta}{\lambda^2}.$$

This equation is the starting point for the determination of unit-cell parameters from measured Bragg angles, for a given wavelength.

A scalar product, that occurs very frequently in crystallographic expressions is $\mathbf{h}\cdot\mathbf{r}$, where $\mathbf{h}$ is a diffraction vector and $\mathbf{r}$ is the position vector of a point in the direct space. The "natural" frames of reference also lead to the simplest result:

$$\mathbf{h}\cdot\mathbf{r} = (h_i\mathbf{a}^i)\cdot(x^j\mathbf{a}_j) = h_i x^j(\mathbf{a}^i\cdot\mathbf{a}_j) = h_i x^j\delta^i_j = h_i x^i.$$

The number of operations is the same as if the two vectors were referred to a Cartesian system.

3. *The vector product in an oblique system.* Suppose the vectors $\mathbf{u}$ and $\mathbf{v}$ are referred to covariant basis vectors, and let their vector product $\mathbf{u}\times\mathbf{v}$, also be referred to a covariant basis by an appropriate transformation. We have

$$\mathbf{u}\times\mathbf{v} = (u^i\mathbf{a}_i)\times(v^j\mathbf{a}_j) = u^i v^j(\mathbf{a}_i\times\mathbf{a}_j). \tag{B.13}$$

Further, the vector product of the basis vectors is proportional to a basis vector of the opposite variance. In our case,

$$\mathbf{a}_i\times\mathbf{a}_j = V e_{ijk}\mathbf{a}^k, \tag{B.14}$$

where e_{ijk} is an antisymmetric tensor which equals $+1$ if ijk equals 123, 231, or 312, equals -1 if ijk equals 132, 213, or 321, and equals 0 if any two indices are the same. We now have

$$\mathbf{u}\times\mathbf{v} = V e_{ijk}u^i v^j\mathbf{a}^k. \tag{B.15}$$

The transformation of the right-hand side of eqn (B.15) to a covariant basis follows simply, since, in analogy to eqn (B.12), $\mathbf{a}^k = g^{kn}\mathbf{a}_n$. Finally,

$$\mathbf{u}\times\mathbf{v} = V g^{kn}e_{ijk}u^i v^j\mathbf{a}_n. \tag{B.16}$$

4. Suppose we have a Cartesian system, that is, the basis vectors are the mutually orthonormal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. The metric tensor here is $g_{ij} = \mathbf{e}_i\cdot\mathbf{e}_j = \delta_{ij}$. A vector product $\mathbf{u}\times\mathbf{v}$ in this system becomes

$$\begin{aligned}\mathbf{u}\times\mathbf{v} &= (u_i\mathbf{e}_i)\times(v_j\mathbf{e}_j)\\ &= u_i v_j(\mathbf{e}_i\times\mathbf{e}_j)\\ &= e_{ijk}u_i v_j\mathbf{e}_k.\end{aligned}$$

Note that there is no difference between covariant and contravariant quantities when these are referred to a Cartesian system. Let us now recall the expression

for the finite rotation operator, and assume that all the quantities are referred to the same Cartesian system. The original expression reads

$$\mathbf{r}' = \mathbf{k}(\mathbf{k}\cdot\mathbf{r})(1-\cos\theta) + \mathbf{r}\cos\theta + (\mathbf{k}\times\mathbf{r})\sin\theta,$$

where $\mathbf{k}$ is a unit vector along the axis of rotation, and θ is the angle through which $\mathbf{r}$ is rotated to $\mathbf{r}'$ about that axis. If we take $\mathbf{k} = k_i\mathbf{e}_i$, $\mathbf{r} = X_i\mathbf{e}_i$ and $\mathbf{r}' = X_i'\mathbf{e}_i$, we can write

$$X_i' = R_{ij}^{(\mathrm{C})} X_j, \tag{B.17}$$

where

$$R_{ij}^{(\mathrm{C})} = k_i k_j(1-\cos\theta) + \delta_{ij}\cos\theta + e_{ipj}k_p\sin\theta. \tag{B.18}$$

The trace of the matrix of this rotation operator is

$$\begin{aligned}\mathrm{Tr}(\mathbf{r}) &= \delta_{ij}R_{ij}^{(\mathrm{C})}\\ &= \delta_{ij}k_ik_j(1-\cos\theta) + \delta_{ij}\delta_{ij}\cos\theta + \delta_{ij}e_{ipj}k_p\sin\theta\\ &= k_ik_i(1-\cos\theta) + 3\cos\theta\\ &= 1 + 2\cos\theta\end{aligned}$$

since $\delta_{ij}\delta_{ij} = 3$, $\delta_{ij}e_{ipj} = e_{ipi} = 0$, and $k_ik_i = k_1^2 + k_2^2 + k_3^2 = 1$, in agreement with eqn (A.25).

5. Transformations to a specific Cartesian system were dealt with in Appendix A. We wish to treat the problem somewhat more generally. So, let $\mathbf{A}$ be a matrix which transforms by *premultiplication* the column of fractional coordinates $x^1x^2x^3$, referred to the basis of the direct lattice, to a column of coordinates referred to a Cartesian system, say $x_1^{\mathrm{C}}x_2^{\mathrm{C}}x_3^{\mathrm{C}}$, or in index notation

$$x_j^{\mathrm{C}} = A_{jk}x^k, \quad j = 1,2,3.$$

In addition, let $\mathbf{B}$ be a matrix which, when *postmultiplying* a row of diffraction indices $h_1h_2h_3$, referred to the basis of the reciprocal lattice, transforms it to a row of diffraction indices referred to the same Cartesian system, say $H_1^{\mathrm{C}}H_2^{\mathrm{C}}H_3^{\mathrm{C}}$, or in index notation

$$H_n^{\mathrm{C}} = h_mB_{mn}, n = 1,2,3.$$

The scalar product $\mathbf{h}\cdot\mathbf{r}$ is of course invariant under the change of the frame of reference and can be written as

$$\mathbf{h}\cdot\mathbf{r} = h_kx^k = H_j^{\mathrm{C}}x_j^{\mathrm{C}} = h_mB_{mj}A_{jk}x^k.$$

It follows that $B_{mj}A_{jk}$ must be equal to δ_{mk}, or in other words that the matrix $\boldsymbol{B}$ is inverse to the matrix $\boldsymbol{A}$. This result is of course quite general (see Section A.1).

This page intentionally left blank

Appendix C Basic notions from theory of probability

C.1 Introduction

The chapters on structure-factor statistics and direct methods, which are of major importance in the process of structure determination, rely on concepts and ideas from the theory of probability. Although this subject is dealt with extensively in the literature, and there are crystallographic texts in which it is treated (for example, Srinivasan and Parthasarathy 1976; Giacovazzo 1980; Shmueli and Weiss 1995), it was thought that a short appendix devoted to its recapitulation may prove helpful.

We shall first recall and illustrate the concept of a random variable, proceed to the probability density function and its moments, and then recall the notions of joint and conditional probabilities. Some other important concepts such as characteristic functions and cumulants will be briefly explained.

C.2 Random variables

The object underlying probability theory is the *random variable*. There exist a variety of definitions of this object, which vary from colloquial ones, such as *a variable the value of which cannot be predicted in advance of an experiment*, through more precise ones, such as *a variable which can appear with varying degrees of probability*, to rather subtle and intricate definitions. Either one of the definitions stated above will be suitable for our purpose, but the second distinguishes a random variable more clearly from the "ordinary" variable encountered in mathematics. In the latter sense, "a variable" refers to a number which may be any number within a given range; none of the possible values is thought to appear more frequently than the others (that is, with a higher probability). A random variable, on the other hand, is closely associated with the probability of its occurrence. Consider, for example the magnitude of a structure factor, for a centrosymmetric structure. It is seen in most experiments with such structures that small magnitudes appear much more frequently than large ones, with the frequency increasing with decreasing magnitude of $|F|$. As shown in the text, this magnitude is in fact a random variable with a Gaussian frequency function (or probability density function–see below) which is peaked at $|F| = 0$.

Another example of a random variable is the fractional part of the scalar product $\mathbf{h} \cdot \mathbf{r}$, for fixed, known $\mathbf{h}$ and unknown $\mathbf{r}$. The frequency function asso-

ciated with this variable is uniform over the range $(0, 1)$ (see Section 7.1), that is, each of its values appears throughout this range with the same probability.

With the above background in mind, we can proceed to some more formal definitions.

C.3 The probability density function and related quantities

Let X be a real, scalar random variable which can take on a continuum of values. The probability density function (pdf) $p(x)$, is defined so that $p(x)\,\mathrm{d}x$ is the probability to find X in the interval $(x, x + \mathrm{d}x)$. It is essential to distinguish (at least at this stage) between the random variable X and the usual–in the mathematical sense–variable x. The definition of $p(x)$ as a density implies that it must be nonnegative.

The probability that X lies in the finite interval (a, b) is therefore given by

$$\mathrm{Prob}\{a < X \leq b\} = \int_a^b p(x)\,\mathrm{d}x. \tag{C.1}$$

If we assume further that a probability equal to unity means certainty, and that the pdf $p(x)$ exists throughout the range $(-\infty, \infty)$, then the fact that the random variable X must lie *somewhere* in this range can be expressed as

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1. \tag{C.2}$$

A pdf satisfying eqn (C.2) is said to be *normalized to unity*. The cumulative distribution $P(x)$ is defined as the probability that the random variable X does not exceed a given value, say x. According to eqn (C.1), this can be written as

$$P(x) \equiv \mathrm{Prob}\{X|X \leq x\} = \int_{-\infty}^{x} p(y)\,\mathrm{d}y, \tag{C.3}$$

where the symbol $\mathrm{Prob}\{A|B\}$ means "probability for A, subject to the knowledge of B". We shall encounter this symbolism below while discussing conditional pdfs.

An often useful set of statistics associated with a pdf is that of moments. The nth moment of the random variable X is defined as

$$\langle X^n \rangle = \int_{-\infty}^{\infty} x^n p(x)\,\mathrm{d}x \tag{C.4}$$

and is said to exist when the integral in eqn (C.4) converges. The first moment, $\langle X \rangle$, is known as the *mean* or the *average* value of X. The second moment is also used, but a more popular statistic is the *variance*, which is defined by the relation

$$\begin{aligned} \sigma^2 &= \langle (X - \langle X \rangle)^2 \rangle \\ &= \langle X^2 \rangle - 2\langle X \langle X \rangle \rangle + \langle X \rangle^2 \\ &= \langle X^2 \rangle - \langle X \rangle^2. \end{aligned} \tag{C.5}$$

The variance of X is of course a nonnegative quantity.

The probability density function is also often interpreted as a frequency function. Suppose that N occurrences of a random variable, from the set $\{X_i\}$, $i = 1, \ldots, n$, are subdivided into n groups, so that the variable X_1 occurs N_1 times in group 1, X_2 occurs N_2 times in group 2, and so on. The chance of occurrence of the variable X_i is therefore $p_i = N_i/N$, the ith fraction of the total number of occurrences. A higher frequency of occurrence is therefore associated with a higher chance. The concept of a frequency function was utilized in the text in relation to the simulation of distributions and their comparison with the probability density function of $|E|$. Notice that the frequency function is of necessity normalized to unity ($\sum_i p_i = \sum_i (N_i/N) = 1$), while such a normalization was only assumed in the probabilistic definition of the pdf.

Just as the moments are average values of powers of a random variable, the average of any function of a random variable, say $f(X)$, is given by

$$\langle f(X) \rangle = \int_{-\infty}^{\infty} f(x) p(x)\, \mathrm{d}x \tag{C.6}$$

C.4 Joint and conditional pdfs

Crystallographic applications usually require statistics which depend jointly on several random variables. For example, the pdf of the structure factor

$$F = \sum_{j=1}^{N} f_j \exp(i\theta_j)$$

involves N random atomic contributions (each of the θ_js is a random variable). If $X_1, X_2, \ldots, X_n$ are n random variables, then their *joint pdf* is denoted by $p(x_1, x_2, \ldots, x_n)$. In analogy with the meaning of the one-dimensional pdf $p(x)$, the function $p(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1\, \mathrm{d}x_2 ... \mathrm{d}x_n$ is the probability that, simultaneously or jointly, $x_1 < X_1 \leq x_1 + \mathrm{d}x_1, x_2 < X_2 \leq x_2 + \mathrm{d}x_2, \ldots, x_n < X_n \leq x_n + \mathrm{d}x_n$. The normalization condition for the joint pdf analogous to eqn (C.2) is

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1 \cdots x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n = 1 \tag{C.7}$$

and a most important simplification arises when the random variables are independent. In that case, the n-dimensional integral in eqn (C.7) reduces to a product of one-dimensional integrals and, therefore, the independence of the random variables implies that

$$p(x_1, x_2, \ldots, x_n) = p(x_1) p(x_2) \cdots p(x_n). \tag{C.8}$$

The above property is the basis of most crystallographic structure-factor statistics. A useful extension of the concept of the joint pdf arises when some of the random variables have already been determined and can be regarded as known quantities. This leads to the concepts of the *conditional* pdf and *joint conditional*

pdf. We suppose that the set $\{X_i\}$, $i = 1, \ldots, n$, consists of random variables, but assume that $X_k, X_{k+1}, \ldots, X_n$ have somehow been determined. Any such information changes the pdf of the remaining variables. We thus denote the pdf of the random variables $X_1, X_2, \ldots, X_{k-1}$, given the knowledge of the subset $X_k, X_{k+1}, \ldots, X_n$, by

$$p(x_1, x_2, \ldots, x_{k-1} | x_k, x_{k+1}, \ldots, x_n) \tag{C.9}$$

A basic application of the joint conditional pdf is found in the treatment of the fundamentals of direct methods in Chapter 8.

C.5 Characteristic function and cumulants

It is often convenient to introduce the Fourier transform of a pdf, the *characteristic function*. This function is defined as

$$C(\omega) = \int_{-\infty}^{\infty} p(x) \exp(i\omega x)\, \mathrm{d}x, \tag{C.10}$$

and according to eqn (C.6) it equals $\langle \exp(i\omega x) \rangle$, the average of the exponential in the integrand of eqn (C.10). An important consequence of the above is that the pdf $p(x)$ can be found as an inverse Fourier transform of the characteristic function, that is,

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C(\omega) \exp(-i\omega x)\, \mathrm{d}x. \tag{C.11}$$

Indeed, it is sometimes possible to obtain the characteristic function from its definition as an average–without the previous knowledge of the pdf–and thus pave the way to a calculation of the pdf. It follows readily from eqns (C.2) and (C.10) that

$$C(0) = 1. \tag{C.12}$$

It is also possible to show that $|C(\omega)| \leq 1$, for all ω. Indeed, we have

$$\begin{aligned} |C(\omega)| &= \left| \int_{-\infty}^{\infty} p(x) \exp(i\omega x)\, \mathrm{d}x \right| \leq \int_{-\infty}^{\infty} p(x) |\exp(i\omega x)|\, \mathrm{d}x \\ &= \int_{-\infty}^{\infty} p(x)\, \mathrm{d}x \\ &= 1 \end{aligned} \tag{C.13}$$

Finally, we can expand the exponential appearing in eqn (C.10) in a Taylor series,

$$\begin{aligned} C(\omega) &= \int_{-\infty}^{\infty} p(x) \sum_{m=0}^{\infty} \frac{i^m \omega^m x^m}{m!}\, \mathrm{d}x \\ &= \sum_{m=0}^{\infty} \frac{i^m \omega^m}{m!} \int_{-\infty}^{\infty} x^m p(x)\, \mathrm{d}x \\ &= \sum_{m=0}^{\infty} \frac{i^m \omega^m}{m!} \langle X^m \rangle, \end{aligned} \tag{C.14}$$

and the expansion is valid provided all moments exist. The moments of the various orders can be represented as derivatives of the characteristic function at $\omega = 0$:

$$\langle X^n \rangle = (-1)^n \left. \frac{\mathrm{d}^n C(\omega)}{\mathrm{d}\omega^n} \right|_{\omega=0}, \tag{C.15}$$

and the characteristic function is therefore known as a moment-generating function.

Another set of useful quantities, called the *cumulants*, is obtained when the function

$$K(\omega) = \sum_{n=1}^{\infty} \frac{i^n \omega^n}{n!} \kappa_n \tag{C.16}$$

is equated to the logarithm of the characteristic function. The coefficient of $(i\omega)^n/n!$ in eqn (C.16), κ_n, is called the *cumulant of order n*, and the function $K(\omega)$ is known as the cumulant-generating function. Since the nth moment of X is the coefficient of $(i\omega)^n/n!$ in the series expansion (eqn (C.14)) of the characteristic function, it is in principle possible to establish relationships between cumulants and moments, and vice versa. General formulas for carrying out these calculations have been given by Stuart and Ord (1994), and the first few relationships are listed below. For cumulants in terms of moments,

$$\begin{aligned}
\kappa_1 &= \langle X \rangle, \\
\kappa_2 &= \langle X^2 \rangle - \langle X \rangle^2, \\
\kappa_3 &= \langle X^3 \rangle - 3\langle X^2 \rangle \langle X \rangle + 2\langle X \rangle^2, \\
\kappa_4 &= \langle X^4 \rangle - 4\langle X^3 \rangle \langle X \rangle - 3\langle X^2 \rangle^2 + 12\langle X^2 \rangle \langle X \rangle^2 - 6\langle X \rangle^4, ,
\end{aligned} \tag{C.17}$$

and for moments in terms of cumulants,

$$\begin{aligned}
\langle X \rangle &= \kappa_1, \\
\langle X^2 \rangle &= \kappa_2 + \kappa_1^2, \\
\langle X^3 \rangle &= \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3, \\
\langle X^4 \rangle &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4.
\end{aligned} \tag{C.18}$$

This page intentionally left blank

Appendix D The discrete Fourier transform

D.1 Some properties of the discrete Fourier transform

We present here the concept of the discrete Fourier transform and prove those properties which are necessary for a presentation of the basics of the Fast Fourier Transform method. This section is based on Chapter 8 of Weaver (1989).

Given any bounded Nth-order sequence $\{f(k)\} = (f(1), f(2), \ldots, f(N))$, the *discrete Fourier transform pair* is defined as

$$F(j) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-2\pi i k j/N}, \quad j \in [0, N-1], \tag{D.1}$$

$$f(k) = \sum_{j=0}^{N-1} F(j) e^{2\pi i k j/N}, \quad k \in [0, N-1]. \tag{D.2}$$

Equation (D.1) is called the direct discrete Fourier transform, and eqn (D.2) is called the inverse discrete Fourier transform. Each of these transforms maps one Nth-order sequence onto another. In general, both sequences $\{f(k)\}$ and $\{F(j)\}$ will be complex, that is,

$$f(k) = f_{\mathrm{R}}(k) + i f_{\mathrm{I}}(k)$$

and

$$F(j) = F_{\mathrm{R}}(j) + i F_{\mathrm{I}}(j).$$

Equation (D.1) now becomes

$$F(j) = \frac{1}{N} \sum_{k=0}^{N-1} [f_{\mathrm{R}}(k) + i f_{\mathrm{I}}(k)] \left[\cos\left(\frac{2\pi k j}{N}\right) - i \sin\left(\frac{2\pi k j}{N}\right) \right], \quad j \in [0, N-1]$$

or

$$F_{\mathrm{R}}(j) = \frac{1}{N} \sum_{k=0}^{N-1} \left[f_{\mathrm{R}}(k) \cos\left(\frac{2\pi k j}{N}\right) + f_{\mathrm{I}}(k) \sin\left(\frac{2\pi k j}{N}\right) \right], \quad j \in [0, N-1], \tag{D.3}$$

$$F_{\mathrm{I}}(j) = \frac{1}{N} \sum_{k=0}^{N-1} \left[f_{\mathrm{I}}(k) \cos\left(\frac{2\pi k j}{N}\right) - f_{\mathrm{R}}(k) \sin\left(\frac{2\pi k j}{N}\right) \right], \quad j \in [0, N-1].$$

Similarly, the components of eqn (D.2) can be written as

$$f_{\mathrm{R}}(k) = \frac{1}{N}\sum_{k=0}^{N-1}\left[F_{\mathrm{R}}(j)\cos\left(\frac{2\pi kj}{N}\right) - F_{\mathrm{I}}(j)\sin\left(\frac{2\pi kj}{N}\right)\right], \quad j \in [0, N-1],$$
$$f_{\mathrm{I}}(k) = \frac{1}{N}\sum_{k=0}^{N-1}\left[F_{\mathrm{I}}(j)\cos\left(\frac{2\pi kj}{N}\right) + F_{\mathrm{R}}(j)\sin\left(\frac{2\pi kj}{N}\right)\right], \quad j \in [0, N-1]. \tag{D.4}$$

Equations (D.3) and (D.4) are used in actual computations. However, for the purpose of mathematical manipulations it is convenient to abbreviate the exponential term by using

$$W_N = e^{2\pi i/N}$$

and simplify eqns (D.1) and (D.2) to

$$F(j) = \frac{1}{N}\sum_{k=0}^{N-1} f(k)W_N^{-kj}, \quad j \in [0, N-1], \tag{D.5}$$
$$f(k) = \sum_{j=0}^{N-1} F(j)W_N^{kj}, \quad k \in [0, N-1]. \tag{D.6}$$

These concise definitions of the discrete Fourier transform pair will be used in what follows. We shall first examine some important properties of Fourier transforms and then proceed to the basics of the Fast Fourier Transform algorithm. of Cooley and Tukey (1965)

Lemma. If m is any integer not equal to 0 or N, then

$$\sum_{j=0}^{N-1} W_N^{mj} = 0.$$

Proof. We construct the sequence $\{V(j)\}$ whose terms are given by

$$V(j) = \frac{W_N^{mj}}{W_N^m - 1}.$$

It follows from the above that

$$\begin{aligned} V(j+1) - V(j) &= \frac{W_N^{m(j+1)} - W_N^{mj}}{W_N^m - 1} \\ &= \frac{W_N^{mj}(W_N^m - 1)}{W_N^m - 1} \\ &= W_N^{mj}. \end{aligned}$$

We thus have

$$\sum_{j=0}^{N-1} W_N^{mj} = \sum_{j=0}^{N-1} [V(j+1) - V(j) \tag{D.7}$$

$$= V(1) - V(0) + V(2) - V(1) + \cdots + V(N) - V(N-1) \tag{D.8}$$

$$= V(N) - V(0) \tag{D.9}$$

$$= \frac{W_N^{mN} - W_N^{m0}}{W_N^m - 1} \tag{D.10}$$

$$= \frac{e^{2\pi i m} - e^0}{W_N^m - 1} = \frac{1-1}{W_N^m - 1} = 0 \tag{D.11}$$

which was to be shown. We can now prove some theorems which are essential to the understanding of the Fast Fourier Transform.

Theorem 1. The discrete Fourier transform is unique.

Proof. Let us assume that the direct discrete Fourier transform of the sequence $\{(f(k)\}$ is the sequence $\{F(j)\}$, but that the inverse discrete transform of $\{F(j)\}$ is the sequence $\{h(k)\}$. Can we show that $h(k) = f(k)$ for all k? We have

$$\begin{aligned} h(k) &= \sum_{j=0}^{N-1} F(j) W_N^{kj}, \qquad k \in [0, N-1] \\ &= \sum_{j=0}^{N-1} \left[\frac{1}{N} \sum_{l=0}^{N-1} f(l) W_N^{-lj} \right] W_N^{kj}, \qquad k \in [0, N-1] \\ &= \frac{1}{N} \sum_{l=0} f(l) \left[\sum_{j=0}^{N-1} W_N^{(k-l)j} \right], \qquad k \in [0, N-1]. \end{aligned}$$

We have shown above, in the lemma, that the inner summation in brackets in the last line of the above equation vanishes unless $k = l$. In the latter case, we have a sum of N terms, where each equals $W_N^0 = 1$. We therefore have

$$h(k) = \frac{1}{N} \sum_{l=0}^{N-1} f(l) N \delta_{kl} = f(k),$$

for any integer k, which was to be shown.

Theorem 2. The direct and inverse discrete Fourier transforms are both periodic with periodicity N; that is (a) $F(j+N) = F(j)$ and (b) $f(k+N) = f(k)$.

Proof.

(a) From eqn (D.1), we have

$$\begin{aligned} F(j+N) &= \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{-2\pi ik(j+N)/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{-2\pi ikj/N}e^{-2\pi ik} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{-2\pi ikj/N} \\ &= F(j). \end{aligned}$$

(b) From eqn (D.2), we have

$$\begin{aligned} f(k+N) &= \frac{1}{N}\sum_{j=0}^{N-1} F(j)e^{-2\pi i(k+N)j/N} \\ &= \frac{1}{N}\sum_{j=0}^{N-1} F(j)e^{-2\pi ikj/N}e^{-2\pi ij} \\ &= \frac{1}{N}\sum_{j=0}^{N-1} F(k)e^{-2\pi ikj/N} \\ &= f(k). \end{aligned}$$

Theorem 3. If the sequences $\{f(k)\}$ and $\{F(j)\}$ are discrete Fourier transform pairs, then: (a) $F(-j) = F(N-j)$ and (b) $f(-k) = f(N-k)$.

The proof follows similarly to the proof of Theorem 2.
(a) From eqn (D.1), we have

$$\begin{aligned} F(N-j) &= \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{-2\pi ik(N-j)/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{2\pi ikj/N}e^{-2\pi ik} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{-2\pi ik(-j)/N} \\ &= F(-j). \end{aligned}$$

(b) From eqn (D.2), we have

$$\begin{aligned} f(N-k) &= \frac{1}{N}\sum_{j=0}^{N-1} F(j)e^{-2\pi i(N-k)j/N} \\ &= \frac{1}{N}\sum_{j=0}^{N-1} F(j)e^{2\pi ikj/N}e^{-2\pi ij} \\ &= \frac{1}{N}\sum_{j=0}^{N-1} F(k)e^{-2\pi i(-k)j/N} \\ &= f(-k). \end{aligned}$$

D.2 The Fast Fourier Transform

We require N^2 (generally) complex multiplications in order to evaluate a direct discrete Fourier transform of an Nth-order sequence. It was pointed out by Cooley and Tukey (1965) that the computing time (and the number of operations) can be reduced rather drastically if we take N to be a power of 2, and this result is very widely applied. Rather than going through the general derivation, let us make the modest assumption that N is divisible by 2 only. This allows us to split the Nth-order sequence into two equal partial sequences, each of length $N/2$:

$$\{f(k)\} = \{f_1(m)\} \cup \{f_2(m)\},$$

where $k \in [0, N-1]$ and $m \in [0, (1/2)N-1]$. For example, let

$$\begin{aligned} \{f_1(k)\} &= \{f(2k)\}, \quad k = 0, \ldots, \frac{1}{2}N - 1 \\ \{f_2(k)\} &= \{f(2k+1)\}, \quad k = 0, \ldots, \frac{1}{2}N - 1 \end{aligned}$$

and $M = (1/2)N - 1$. The Fourier transforms of the two partial sequences can be written as

$$F_1(j) = \frac{1}{M}\sum_{k=0}^{M-1} f_1(k)W_M^{-kj}, \quad j \in [0, N-1], \tag{D.12}$$

$$F_2(j) = \frac{1}{M}\sum_{k=0}^{M-1} f_2(k)W_M^{-kj}, \quad j \in [0, N-1]. \tag{D.13}$$

By Theorems 2 and 3, both $\{F_1(j)\}$ and $\{F_2(j)\}$ may be considered periodic with periodicity M. We now write the discrete Fourier transform of the Nth-order sequence

$$F(j) = \frac{1}{N}\sum_{k=0}^{N-1} f(k)W_N^{-kj}, \quad j \in [0, N-1], \tag{D.14}$$

and split this equation according to the previous definitions:

$$F(j) = \frac{1}{N}\sum_{k=0}^{M-1} f(2k)W_N^{-2kj} + \frac{1}{N}\sum_{k=0}^{M-1} f(2k+1)W_N^{-(2k+1)j}.$$

However, we see that

$$\begin{aligned} W_N^{-2kj} &= e^{-2\pi i(2k)j/N} = e^{-2\pi ikj/(N/2)} = W_M^{-kj} \\ W_N^{-(2k+1)j} &= e^{-2\pi i(2k+1)j/N} = e^{-2\pi i(2k)j/N}e^{-2\pi ij/N} \\ &= W_M^{-kj}W_N^{-j}. \end{aligned}$$

Using eqns (D.12) and eqn (D.13), we also have

$$\begin{aligned} F(j) &= \frac{1}{N}\sum_{k=0}^{M-1} f_1(k)W_M^{-kj} + \frac{W_N^{-j}}{N}\sum_{k=0}^{M-1} f_2(k)W_M^{-kj} \\ &= \frac{1}{2M}\sum_{k=0}^{M-1} f_1(k)W_M^{-kj} + \frac{W_N^{-j}}{2M}\sum_{k=0}^{M-1} f_2(k)W_M^{-kj} \\ &= \frac{1}{2}[F_1(j) + F_2(j)W_N^{-j}], \qquad j \in [0, N-1]. \end{aligned} \tag{D.15}$$

If j varies from 0 to $M-1$, we have to find $F(j)$ and $F(j+M)$. Now,

$$\begin{aligned} F_1(j+M) &= \frac{1}{M}\sum_{k=0}^{M-1} f_1(k)W_M^{-k(j+M)} \\ &= \frac{1}{M}\sum_{k=0}^{M-1} f_1(k)W_M^{-kj}W_M^{-kM} \\ &= F_1(j) \end{aligned}$$

since $W_M^{-kM} = e^{-2\pi ikM/M} = 1$. Similarly, $F_2(j+M) = F_2(j)$. However,

$$W_N^{-(j+M)} = W_N^{-j}W_N^{-M} = W_N^{-j}e^{-2\pi iM/(2M)} = -W_N^{-j}.$$

We finally have

$$F(j) = \frac{1}{2}[F_1(j) + F_2(j)W_N^{-j}], \qquad j \in [0, M-1]. \tag{D.16}$$

$$F(j+M) = \frac{1}{2}[F_1(j) - F_2(j)W_N^{-j}], \qquad j \in [0, M-1]. \tag{D.17}$$

In order to compute the direct transform $\{F(j)\}$, with $j \in [0, N-1]$, we need the transforms $\{F_1(j)\}$ and $\{F_2(j)\}$, with $j \in [0, M-1]$, and N products of

the second transform with W_N^{-j}, that is, altogether, $N + 2(N/2)^2 = N + N^2/2$ operations. The reduction factor is

$$\frac{N + N^2/2}{N^2} = \frac{1}{N} + \frac{1}{2} \tag{D.18}$$

which tends to 1/2 for large values of N. If the original sequence $\{f(k)\}$ can be split into four equal partial sequences, the reduction factor can be similarly shown to be

$$\frac{2N + 4(N/4)^2}{N^2} = \frac{2}{N} + \frac{1}{4} \tag{D.19}$$

which, for large N, tends to 1/4. If we proceed in this manner and the length N of the original sequence is divisible by 2^p, where p is a positive integer, it can be shown that the total number of operations is

$$pN + 2^p \left(\frac{N}{2^p}\right)^2 = pN + \frac{N^2}{2^p} \tag{D.20}$$

and the reduction factor is

$$\frac{p}{N} + \frac{1}{2^p}. \tag{D.21}$$

If, finally, N is taken as a power of 2, that is, $N = 2^p$, the total number of operations required is now

$$pN + N = N(p + 1) = N(\log_2(N) + 1) \sim N\log_2(N), \tag{D.22}$$

for fairly large N. Of course, N can always be taken as a power of 2 by simply adding the required number of zero terms to the original sequence. Now, this is a tremedous gain (for $N = 2048$, the inverse reduction factor is 186). This is being exploited in many areas of science and engineering.

The above is only the Cooley–Tukey approach, and there are many others (see, for example, Bricogne 2001 and the references quoted there) which promise–with modern computers–further reductions of computing effort.

This page intentionally left blank

References

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.

Amemiya, Y., Arndt, U. W., Buras, B., Chikawa, J., Gerward, L., Langford, I. L., Parrish, W., and de Wolff, P. M. (1999). In *International Tables for Crystallography*, Volume C: *Mathematical, Physical and Chemical Tables*, edited by A. J. C. Wilson and E. Prince, 2nd edition, 612–632. Dordrecht: Kluwer Academic.

Arnold, H. (1983). In *International Tables for Crystallography*, Volume A: *Space-Group Symmetry*, edited by Th. Hahn, pp. 70–79 . Dordrecht: Reidel.

Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M., and Wondratschek, H. (2006). Bilbao Crystallographic Server II: Representations of crystallographic point groups and space groups. *Acta Crystallogr. A* **62**, 115–128.

Beevers, C. A. and Lipson, H. (1936). An improved numerical method of two-dimensional Fourier synthesis for crystals. *Proc. Phys. Soc.* **48**, 772–780.

Bertaut, E. F. (1983). In *International Tables for Crystallography*, edited by Th. Hahn, pp. 50–68. Dordrecht: Reidel.

Beurskens, P. T., Gould, R. O., Bruins Slot, H. J., and Bosman, W. P. (1987). Translation functions for the positionong of a well oriented molecular fragment. *Z. Kristallogr.* **179**, 127–159.

Bricogne, G. (2001). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, 2nd edition, 25–98. Dordrecht: Kluwer Academic.

Buerger, M. J. (1941). *X-ray Crystallography.* New York: Wiley.

Buerger, M. J. (1956). *Elementary Crystallography.* New York: Wiley.

Buerger, M. J. (1959). *Vector Space and Its Applications in Crystal-Structure Investigations.* New York: Wiley.

Buerger, M. J. (1964). *The Precession Method in X-ray Crystallography.* New York: Wiley.

Cochran, W. (1955). Relations between the phases of structure factors. *Acta Crystallogr.* **8**, 473–478.

Cochran, W. and Woolfson, M. M. (1955). The theory of sign relations between structure factors. *Acta Crystallogr.* **8**, 1–12.

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **19**, 297–301.

Coppens, P. (1992). *Synchrotron Radiation Crystallography.* New York: Academic Press.

Coppens, P. (1993). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, 10–22. Dordrecht: Kluwer Academic.

Coppens, P. (1997). *X-ray Charge Density and Chemical Bonding.* Oxford: Oxford University Press.

Coppens, P., Su, Z., and Becker, P. J. (1999). In *International Tables for Crystallography*, Volume C: *Mathematical, Physical and Chemical Tables*, edited by A. J. C. Wilson and E. Prince, 2nd edition, 706–727. Dordrecht: Kluwer Academic.

Cramér, H. (1951). *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press.

Creagh, D. C. (1999). In *International Tables for Crystallography*, Volume C: *Mathematical, Physical and Chemical Tables*, edited by A. J. C. Wilson and E. Prince, 2nd edition, pp. 242–258. Dordrecht: Kluwer Academic.

Crowther, R. A. (1972). In *The Molecular Replacement Method*, edited by M. G. Rossmann. pp. 173–178. New York: Gordon and Breach.

Crowther, R. A. and Blow, D. M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallogr.* **23**, 544–548.

Deas, H. D. and Hamill, C. M. (1957). A note on the geometry of lattice planes. *Acta Crystallogr.* **10**, 541–542.

Ewald, P. P. (1913). Zur Theorie der Interferenzen der Röntgenstrahlen in Kristallen. *Phys. Z.* **14**, 465–472.

Fedorov, E. S. (1891). Symmetry of finite figures. *Proc. S. Peterb. Mineral. Soc. (2)* **28**, 1–146.

Fokkema, D. S. (1983). In *International Tables for Crystallography*, Volume A: *Space-Group Symmetry*, edited by Th. Hahn, pp. xiv–xv. Dordrecht: Reidel.

Friedel, G. (1913). Sur les symétries cristallines que peut révéler la diffraction des rayons Röntgen. *C. R. Acad. Sci. Paris* **157**, 1533–1536.

Friedrich, W., Knipping, P., and Laue, M. (1912). Interferenzerscheinungen bei Roentgenstrahlen. *Proc. Bavarian Acad. Sci.*, pp. 303–322.

Germain, G., Main, P., and Woolfson, M. M. (1970). On the application of phase relationships to complex structures. II. Getting a good start. *Acta Crystallogr. B* **26**, 274–285.

Giacovazzo, C. (1974). A new scheme for seminvariant tables in all space groups. *Acta Crystallogr. A* **30**, 390–395.

Giacovazzo, C. (1980). *Direct Methods in Crystallography.* London: Academic Press.

Giacovazzo, C. (1998). *Direct Phasing in Crystallography, Fundamentals and Applications.* Oxford: Oxford University Press.

Giacovazzo, C. (2001). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, 2nd edition, pp.210–234. Dordrecht: Kluwer Academic.

Goldstein, H. (1956). *Classical Mechanics.* Cambridge, MA: Addison-Wesley.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Tables of Integrals, Series and Products.* New York: Academic Press.

Hahn, Th. (ed.) (1983). *International Tables for Crystallography*, Volume A: *Space-Group Symmetry.* Dordrecht: Reidel.

Hahn, Th. and Klapper, H. (1983). In *International Tables for Crystallography*, Volume A: *Space-Group Symmetry*, edited by Th. Hahn, pp. 746–786. Dordrecht: Reidel.

Hahn, Th. (ed.) (1988). *International Tables for Crystallography: Brief Teachin Edition of* Volume A: *Space-Group Symmetry*, edited by Th. Hahn, 2nd edition. Dordrecht: Reidel.

Hahn, Th. and Wondratschek, H. (1994). *Symmetry of Crystals*: *Introduction to International Tables for Crystallography*, Volume A. Sofia: Heron.

Hahn, Th. (ed.) (1996). *International Tables for Crystallography*, Volume A: *Space-Group Symmetry*, 4th edition. Dordrecht: Kluwer Academic.

Hall, S. R. (1981). Space-group notation with an explicit origin. *Acta Crystallogr. A* **37**, 517–525.

Hall, S. R. (1993). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, Ch. 1.4, Appendix B, pp. 115–126. Dordrecht: Reidel.

Hall, S. R. and Grosse-Kunstleve, R. W. (2001). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, 2nd edition, Appendix 1.4.2, pp.112–119. Dordrecht: Kluwer Academic.

Hamilton, W. C. (1974). In *International Tables for X-ray Crystallography*, Volume IV, edited by J. A. Ibers and W. C. Hamilton, 275–284. Birmingham: Kynoch.

Harker, D. (1936). The application of the three-dimensional Patterson method and the crystal structures of proustite, Ag_3AsS_3, and pyrargyrite, Ag_3SbS_3. *J. Chem. Phys.* **4**, 381–390.

Harker, D. and Kasper, J. (1947). Phases of Fourier coefficients from crystal diffraction data. *J. Chem. Phys.* **15**, 882.

Hauptman, H. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd and R. A. Palmer, pp. 151–197. New York: Plenum.

Hauptman, H. and Karle, J. (1953*a*). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* Pittsburgh: Polycrystal Book Service.

Hauptman, H. and Karle, J. (1953*b*). The probability distribution of the magnitude of a structure factor. II. The noncentrosymmetric crystal. *Acta Crystallogr.* **6**, 136–141.

Hauptman, H. and Karle, J. (1956). Structure invariants and seminvariants for noncentrosymmetric space groups. *Acta Crystallogr.* **9**, 45–55.

Hauptman, H. and Karle, J. (1959). Table 2. Equivalence classes, seminvariant vectors and seminvariant moduli for the centered centrosymmetric space groups, referred to a primitive unit cell. *Acta Crystallogr.* **12**, 93–97.

Helliwell, J. R. (1992). *Macromolecular crystallography with Synchrotron Radiation.* Cambridge: Cambridge University Press.

Henry, N. F. M. and Lonsdale, K. (ed.) (1952). *International Tables for X-ray Crystallography*, Volume I: *Symmetry Groups.* Birmingham: Kynoch.

Henry, N. F. M. and Lonsdale, K. (ed.) (1965). *International Tables for X-ray Crystallography*, Volume I: *Symmetry Groups*, 2nd edition. Birmingham: Kynoch.

Hermann, C. (ed.) (1935). *Internationale Tabellen zur Bestimmung von Kristallstrukturen*, Volume 1. Berlin: Bornträger.

Hilton, H. (1903). *Mathematical Crystallography and the Theory of the Groups of Movements.* Oxford: Clarendon.

Hirshfeld, F. L. and Rabinovich, D. (1966). Rigid-body vibrations in non-orthogonal systems. *Acta Crystallogr.***20**, 146–147.

Hodgkin, D. C., Kamper, J., Lindsey, J., MacKay, M., Pickworth, J., Robertson, J. H., Shoemaker, C. B., White, J. G., Prosen, R. J., and Trueblood, K. N. (1957). The structure of vitamin B_{12}. I. An outline of the crystallographic investigation of vitamin B_{12}. *Proc. R. Soc. London, Ser. A* **242**, 228–263.

James, R. W. (1965). *The Optical Principles of the Diffraction of X-rays.* London: Bell.

Karle, J. and Hauptman, H. (1950). The phases and magnitudes of structure factors. *Acta Crystallogr.* **3**, 181–187.

Karle, J. and Hauptman, H. (1953). The probability distribution of the magnitude of a structure factor. I. The centrosymmetric crystal. *Acta Crystallogr.* **6**, 131–135.

Karle, J. and Karle, I. L. (1966). The symbolic addition procedure for phase determination for centrosymmetric and noncentrosymmetric crystals. *Acta Crystallogr.* **21**, 849–859.

Klug, A. (1958). Joint probability distributions of structure factors and the phase problem. *Acta Crystallogr.* **11**, 515–543.

Koch, E. E. (ed.) (1983). *Handbook of Synchrotron Radiation* (1983). Amsterdam: North-Holland.

Konnert, J. H. (1976). A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. *Acta Crystallogr. A* **32**, 614–617.

Kopsky, V. and Litvin, D. B. (ed.) (2002). *International Tables for Crystallography*, Volume E: *Subperiodic Groups*, Dordrecht: Kluwer Academic.

Laue, M. (1912). Eine quantitative Prüfung der Theorie für die Interferenzerscheinungen bei Röntgenstrahlen. *Proc. Bavarian Acad. Sci.*, pp. 303–322.

Laue, M. (1914). Die Interferenzerscheinungen an Röntgenstrahlen, hervorgerufen durch das Raumgitter der Kristalle. *Jahrb. Radioakt. Elektron.* **11**, 308–345.

Ledermann, W. (1976). *Introduction to Group Theory.* London: Longman.
Linnik, Y. V. (1961). *Method of least squares and principles of the theory of observations.* New York: Pergamon Press.
Lipson, H. and Cochran, W. (1966). *The Determination of Crystal Structures.* London: Bell,
MacGillavry, C. H. (1965). *Symmetry Aspects of M. C. Escher's Periodic Drawings.* Published for the International Union of Crystallography. Utrecht: Oosthoek.
Main, P. (1975). In *Crystallographic Computing Techniques (Prague)*, edited by F. R. Ahmed with co-editors K. Huml and B. Sedláček, pp. 97–105. Copenhagen: Munksgaard.
Main, P. and Rossmann, M. G. (1966). Relationships among structure factors due to identical molecules in different crystallographic environments. *Acta Crystallogr.* **21**, 67–72.
Maslen, E. N. (1999). In *International Tables for Crystallography*, Volume C: *Mathematical, Physical and Chemical Tables*, edited by A. J. C. Wilson and E. Prince, 2nd edition, pp. 593–602. Dordrecht: Kluwer Academic Publishers.
Maslen, E. N., Fox, A. G., and O'Keefe, M. A. (1999). In *International Tables for Crystallography*, Volume C: *Mathematical, Physical and Chemical Tables*, edited by A. J. C. Wilson and E. Prince, 2nd edition, pp.548–589. Dordrecht: Kluwer Academic.
Panofsky, W. K. H. and Philips, M. (1956) *Classical Electricity and Magnetism.* Reading, MA: Addison-Wesley.
Patterson, A. L. (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Phys Rev.* **36**, 372–376.
Prince, E., Finger, L. W., and Konnert, J. H. (1999). In International Tables for Crystallography, Volume C: *Mathematical, Physical and Chemical Tables*, edited by A. J. C. Wilson and E. Prince, 2nd edition, pp. 687–695. Dordrecht: Kluwer Academic.
Rabinovich, D. and Lourie, B. (1987). Use of the polychromatic Laue method for short exposure X-ray diffraction data acquisition. *Acta Crystallogr. A* **43**, 774–780.
Rabinovich, S., Shmueli, U., Stein, Z., Shashua, R., and Weiss, G. H. (1991). Exact random-walk models in intensity statistics. VI. P.d.f.'s of $|E|$ for all plane groups and most space groups. *Acta Crystallogr. A* **47**, 328–335.
Rogers, D. (1950). The probability distribution of X-ray intensities. IV. New methods of determining crystal classes and space groups. *Acta Crystallogr.* **3**, 455–464.
Rossmann, M. G. and Arnold, E. (1993). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, pp.230–263. Dordrecht: Kluwer Academic.
Rossmann, M. G. and Arnold, E. (ed.) (2001*a*). *International Tables for Crystallography*, Volume F: *Crystallography of Biological Macromolecules.* Dordrecht: Kluwer Academic.
Rossmann, M. G. and Arnold, E. (2001*b*). In *International Tables for Crystallography*, Volume B: Reciprocal Space, edited by U. Shmueli, 2nd edition, pp.235–263. Dordrecht: Kluwer Academic.
Rossmann, M. G. and Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 23–31.
Rossmann, M. G., Blow, D. M., Harding, M. M., and Coller, E. (1964). The relative positions of independent molecules within the same asymmetric unit. *Acta Crystallogr.* **17**, 338–342.
Sayre, D. (1952). The squaring method: a new method for phase determination. *Acta Crystallogr.* **5**, 60–65.
Schenk, H. (1974). On the use of negative quartets. *Acta Crystallogr. A* **30**, 477–481.
Schomaker, V. and Trueblood, K. N. (1968). On the rigid-body motion of molecules in crystals. *Acta Crystallogr. B* **24**, 63–76.
Schönflies, A. (1891). *Kristallsysteme und Kristallstruktur.* Leipzig: Teubner.

Schwinger, J. (1949). On the classical radiation of accelerated electrons. *Phys. Rev.* **75**, 1912–1925.
Seitz, F. (1934). A matrix-algebraic development of the crystallographic Groups. I. *Z. Kristall.* **88**, 433–459.
Sheldrick, G. M. (1997). *SHELX97*. University of Göttingen.
Shmueli, U. (1984). Space-group algorithms. I. The space group and its symmetry elements. *Acta Crystallogr. A* **40**, 559–567.
Shmueli, U. (ed.) (2001*a*). *International Tables for Crystallography*, Volume B: *Reciprocal Space*, 2nd edition. Dordrecht: Kluwer Academic Publishers.
Shmueli, U. (ed.) (1993*a*) *International Tables for Crystallography* (1993), Volume B: *Reciprocal Space*. Dordrecht: Reidel.
Shmueli, U. (1993*b*). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, pp. 107–182. Dordrecht: Reidel.
Shmueli, U. (ed.) (2001*a*). *International Tables for Crystallography*, Volume B: *Reciprocal Space*, 2nd edition. Dordrecht: Kluwer Academic Publishers.
Shmueli, U. (2001*b*). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, pp. 2–9. Dordrecht: Kluwer Academic.
Shmueli, U. (2001*c*). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, 2nd edition, pp. 108–111. Dordrecht: Kluwer Academic.
Shmueli, U. and Weiss, G. H. (1985). Exact joint probability distributions for centrosymmetric structure factors. Derivation and application to the Σ_1 relationship in the space group $P\overline{1}$. *Acta Crystallogr. A* **41**, 401–408.
Shmueli, U. and Weiss, G. H. (1986). Exact joint distribution of $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{\mathbf{h}+\mathbf{k}}$ and the probability for the positive sign of the triple product in the space group $P\overline{1}$. *Acta Crystallogr. A***42**, 240–246.
Shmueli, U. and Weiss, G. H. (1995). *Introduction to crystallographic statistics*. Oxford: Oxford University Press.
Shmueli, U. and Wilson, A. J. C. (1981). Effects of space-group symmetry and atomic heterogeneity on intensity statistics, *Acta Crystallogr.* A**37**, 342–353.
Shmueli, U., Rabinovich, S., and Weiss, G. H. (1989). Exact conditional distribution of a three-phase invariant in the space group $P1$. I. Derivation and simplification of the Fourier series. *Acta Crystallogr. A* **45**, 361–367.
Srinivasan, R. and Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon.
Stout, G. H. and Jensen, L. H. (1968). *X-ray Structure Determination: a Practical Guide*. London: Macmillan.
Stuart, A. and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics*. Vol I. *Distribution Theory*, 6th edition, pp.86–95. London: Edward Arnold.
Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U., and Abrahams, S. C. (1996). Atomic displacement parameter nomenclature. *Acta Crystallogr.* A**52**, 770–781.
Vijayan, M. (2001). In *International Tables for Crystallography*, Volume B: *Reciprocal Space*, edited by U. Shmueli, 2nd edition, pp. 264–275. Dordrecht: Kluwer Academic.
Waser, J. (1963). Least-squares refinement with subsidiary conditions. *Acta Crystallogr.* **16**, 1091–1094.
Weaver, H. J. (1989). *Theory of Discrete and Continuous Fourier Analysis*. New York: Wiley.
Willis, B. T. M. and Pryor, A. W. (1975). *Thermal Vibration in Crystallography*. Cambridge: Cambridge University Press.
Wilson, A. J. C. (1942). Determination of absolute from relative intensity data. *Nature (London)***150**, 151–152.
Wilson, A. J. C. (1949). The probability distribution of X-ray intensities. *Acta Crystallogr.* **2**, 318–320.

Wilson, A. J. C. (ed.) (1992). *International Tables for Crystallography*, Volume C: *Mathematical, Physical and Chemical Tables*, Volume C. Dordrecht: Kluwer Academic.

Wilson, A. J. C. and Prince, E. (ed.) (1999). *International Tables for Crystallography*, Volume C: *Mathematical, Physical and Chemical Tables*. 2nd edition. Dordrecht: Kluwer Academic.

Wondratschek, H. and Müller, U. (ed.) (2004). *International Tables for Crystallography*, Volume A1: *Symmetry Relations Between Space Groups*. Dordrecht: Kluwer Academic.

Woolfson, M. M. (1954). The statistical theory of sign relationships. *Acta Crystallogr.* **7**, 61–64.

Woolfson, M. M. (1961). *Direct Methods in Crystallography*. Oxford: Clarendon.

Woolfson, M. M. (1997). *An Introduction to X-ray Crystallography*, 2nd edition. Cambridge: Cambridge University Press.

Wrinch, D. M. (1939). The geometry of discrete vector maps. *Philos. Mag.* **27**, 98–122.

Zachariasen, W. H. (1945). *Theory of X-ray Diffraction in Crystals*. New York: Wiley.

Index

atomic displacement parameters, 207
atomic scattering factor, 128
  assuming spherical symmetry, 129
axial system
  hexagonal, 24
  rhombohedral, 24

Bravais lattices, 45
  derivation of, 46

central limit theorem, 164
change of basis, effect of, 237
coordinates
  Cartesian, 12
  fractional, 5
crystal systems
  and lattice types, 47
  and point groups, 44
  cubic, 25
  hexagonal, 25
  monoclinic, 23
  orthorhombic, 23
  tetragonal, 23
  triclinic, 22
  trigonal, 24
crystallographic point groups
  centrosymmetric, 40
  noncentrosymmetric
    axial, 38
    not axial, 40
cumulative distribution functions of E
  ideal, acentric, 168
  ideal, centric, 168

Debye–Waller factor
  anisotropic
    B tensor, 213
    Cartesian **U** tensor, 214
    U tensor, 213
    β tensor, 212
  isotropic
    individual, 210
    overall, 210
detectors of X-rays
  charge-coupled device (CCD), 102
  imaging plate, 101
  X-ray film, 100

diffracted intensity and interatomic vectors, 143
diffraction conditions, 86
direct lattice, 11
direct methods
  comments on solution of phase problem, 201
  dependence of phase on choice of origin, 178
  example of specification of the origin, 183
  examples of use of seminvariant tables, 186
  figures of merit
    absolute figure of merit, 204
    quartet invariants, 205
    residual, 204
  inequalities
    Harker–Kasper, 186
    Karle–Hauptman, 189
  need for probabilistic considerations, 192
  nonideal joint probability densities, 197
  probability density of three-phase invariant, 192, 195
  Sayre equation, 191
    tangent formula, 191
  structure invariants, 179
  structure seminvariants, 181
  Σ_1 and Σ_2 equations, 198
displacement parameters, 207
  molecular displacement
    rigid-body displacement parameters, 217
    rigid-body model, 215
  underlying assumptions
    Gaussian atomic densities, 208
    spherical atoms, 209
    superposition of independent densities, 207

electron density
  computational approaches
    Beevers–Lipson strips, 133
    Fast Fourier Transform, 133
  Fourier synthesis of, 132
  in terms of structure factors, 132
Eulerian cradle

2θ axis, 107
ϕ axis, 108
χ axis, 108
Ω axis, 108
Ewald sphere of reflection, 90

finite rotation operator
in matrix form, 239
in tensor notation, 245
in vector form, 239
four-circle diffractometer, 107
comments on the experiment, 112
coordinates and angles, 111
diffraction vector representations, 109
orientation matrix, 110
Fourier method in structure-factor statistics, 173
Fourier transform
discrete, 254
fast, 259

group axioms
associativity, 30, 52
closure, 30, 51
identity element, 30, 51
inverse element, 30, 52
group theory, fundamentals, 29
groups
commutativity, 33
composition law, 30
coset decomposition of, 34
cyclic, 33
direct product of, 37
generators, 33
isomorphism of, 34
Lagrange theorem for, 34
multiplication table, 32
order, 33

Hermann–Mauguin notation, 42

interference function, 126
interplanar distance, 8

lattice
centrosymmetry, 16
group of, 31
permissible pairs of axes, 18, 22
permissible rotations, 15
plane, 5
graphical representation, 10
point, 5
vector, 2
Laue conditions
derivation, 87, 89
geometrical interpretation, 88
underlying assumptions, 86
Laue cones, 104

Laue method
calculation of the diffraction pattern, 114
principles of, 114
Laue vs. Bragg conditions, 89
least-squares method
in crystallography, 219
linear, 219
example, 223
expressions for uncertainties, 222
non-linear
application to constrained refinement, 230
application to conventional refinement, 227
application to restrained refinement, 231
nonlinear
formulation of, 225
use of the maximum-likelihood principle, 221
limiting sphere, 90

material unit, 1
metric tensors, direct and reciprocal, 243
Miller indices, 8
moments of E
acentric, 168
centric, 168
monochromatization
crystal monochromator, 96
filter absorption edge, 95
moving-crystal–moving-film methods,
precession method, 106
de Jong–Bouman method, 106
Weisenberg method, 106
mutually reciprocal bases in tensor notation, 241

origin, 1

Patterson function
and interatomic vectors, 144
and overlap of interatomic vectors, 147
and the heavy-atom method, 146
application of Harker peaks, 151
Harker vectors, 148
interunit vectors, 149
intraunit vectors, 148
molecular-replacement method, 157
periodicity of, 145
rotation functions and intraunit vectors, 153
self-vectors, 148
superposition methods, 152
symmetry considerations, 150
translation functions and cross-Patterson vectors, 155

periodicity, 1
point lattice, 1
probabilities
  characteristic function, 250
  conditional pdf, 248
  independence, 249
  joint conditional pdf, 250
  moments, 248
  normalization, 248
  probability density function (pdf), 248
  random variable, 247
  variance, 248
probability density functions of E
  composition- and symmetry-dependent, 170
  ideal, acentric, 166
  ideal, centric, 165

reciprocal lattice, 11
rotating-crystal method
  determination of a period, 105
  principle of, 103
rotation operator, 15
  matrix representation of, 33

scattering of X-rays
  by an arbitrary material unit, 123
  by dispersive scatterers, 121
  by free electrons, 121
  electrodynamical basis of, 120
  principle of, 119
  Thomson scattering cross section, 123
  Thomson scattering formula, 123
space groups, 52
  ambiguities, 164
  change of basis, 62, 75
  classification of, 65
  computational aspects of, 76
  computer readable
    concise symbols, 76
    explicit symbols, 77
  derivation of, 56
  effect of origin shift, 60
  general equivalent positions, 64
  graphical representation of, 68
  location of symmetry elements, 66
  non-symmorphic, 52
  special equivalent positions, 64
  symbols for symmetry elements, 57
  symmetry operators, 51
    glide plane, 55
    screw axis, 54
  symmorphic, 52
  table of explicit symbols, 79
  transformation of coordinates, 62
structure factor
  and electron density, 125
  and space-group symmetry, 130
  effects of space-group symmetry, 136
  for a row of identical material units, 126
  formulation in terms of rigid-body parameters, 230
  of a 3D periodic assembly, 127
  simplified expressions for, 131
  systematic absences
    caused by glide planes, 139
    caused by lattice types, 140
    caused by screw axes, 141
structure-factor statistics, why probability?, 161
subgroup
  invariant, 59
subgroups
  invariant, 36
  proper, 33
  trivial, 33
summation convention, 241
symmetry
  macroscopic, 29
  microscopic, 29
  reflection, 40
  rotation, 13
  rotoinversion, 40
  translation, 1
synchrotron radiation
  an implementation of, 98
  classical electromagetism, 97
  dependence on acceleration, 98

the phase problem of crystallography, 135
trace invariance, 237
trace of a matrix, 14
transformations in tensor algebra, 242
transformations to Cartesian basis, 235

unit cell, 3
  nonprimitive, 18
  primitive, 4, 18

vector lattice, 1
vector product in non-Cartesian system, 244

Wilson's method of scaling intensities, 162

X-ray tube, 91
  and white radiation, 93
  characteristic radiation and electronic transitions, 94
X-rays, 91
  history of spectral brightness, 99

ψ axis, 109