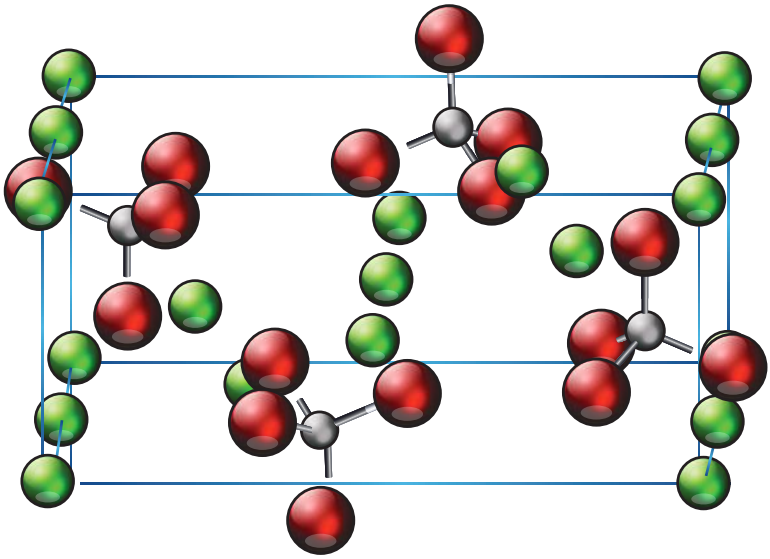
FOURTH EDITION

SOLID STATE CHEMISTRY

AN INTRODUCTION



CRC Press
Taylor & Francis Group

FOURTH EDITION

SOLID STATE CHEMISTRY

AN INTRODUCTION

FOURTH EDITION

SOLID STATE CHEMISTRY

AN INTRODUCTION

Lesley E. Smart • Elaine A. Moore

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20120106

International Standard Book Number-13: 978-1-4398-4792-3 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Dedicated to our children

Sam and Laura,
and
Graham and Rosemary

Contents

Preface to the Fourth Edition

Solid-state chemistry is still a rapidly advancing field, and the past few years have seen many new developments and discoveries in this field. This edition aims, as previously, not only to teach the basic science that underpins the subject, but also to direct the reader to the most modern techniques and to expanding and new areas of research. In order to do this, some of the subjects covered in the previous editions have either been briefly discussed or dropped altogether, to make way for the new. More space has been devoted to electron microscopy and to the techniques that are available with synchrotron radiation; there are new sections on nanomaterials and their synthesis and on computer modeling of solids, periodic mesoporous organosilicas (PMOs), metal organic frameworks (MOFs), metal oxide frameworks, metamaterials, and multiferroics. The section on molecular metals has been dropped and the discussion on conducting polymers has been added at the end of the expanded section on carbon nanoscience. Throughout, new scientific terms and concepts are highlighted in bold typeface where they are first introduced.

We thank our colleagues and families, who have supported us throughout this project, Professor Mark Weller for his very helpful review, and, in particular, Drs. Craig McFarlane and Mike Mortimer and Professor Peter Atkins, who have generously given their time and expertise in explaining and correcting certain points of science.

We wish to acknowledge the use of the Chemical Database Service at Daresbury (Fletcher, D.A., Meeking, R.F., and Parkin, D., The United Kingdom Chemical Database Service, *J. Chem. Inf. Comput. Sci.* **36**, 746–749, 1996), and the Inorganic Crystal Structure Database. (Berghoff, G and Brown, I.D., The Inorganic Crystal Structure Database (ICSD), in *Crystallographic Databases,* F. H. Allen et al., eds. Chester, International Union of Crystallography, 1987.) Also, WebLab Viewer Pro or Discovery Studio Visualiser from Accelerys Inc. was used to create the crystal structures.

Lesley Smart and Elaine Moore,
The Open University, Walton Hall, Milton Keynes, UK

Preface to the Third Edition

Solid-state and materials chemistry is a rapidly moving field, and the aim of this edition has been to bring the text as up to date as possible with new developments. A few changes of emphasis have been made along the way. Single crystal X-ray diffraction has now been removed from Chapter 2 to make way for a wider range of the physical techniques used for characterising solids, and the number of synthetic techniques has been expanded in Chapter 3. Chapter 5 now contains a section on fuel cells and electrochromic materials. In Chapter 6, the section on low-dimensional solids has been replaced with sections on conducting organic polymers, organic superconductors and fullerenes. Chapter 7 now covers mesoporous solids and ALPOs, and Chapter 8 includes a section on photonics. Giant magnetoresistance (GMR) and colossal magnetoresistance (CMR) have been added to Chapter 9, and *p*-wave (triplet) superconductors to Chapter 10. Chapter 11 is new, and it looks at the solid-state chemical aspects of nanoscience. We thank our readers for the positive feedback on the first two editions and for the helpful advice, which has led to this latest version. As ever, we thank our friends in the chemistry department at the OU, who have been such a pleasure to work with over the years and have made enterprises such as this possible.

Lesley Smart and Elaine Moore,
The Open University, Walton Hall, Milton Keynes, UK

Preface to the Second Edition

We were very pleased to be asked to prepare a second edition of this book. When we tried to decide on the changes (apart from updating) to be made, the advice from our editor was 'if it ain't broke, don't fix it'. However, the results of a survey of our users requested about five new subjects but with the provisos that nothing was taken out, that the book did not get much longer and, above all, that it did not increase in price! So what you see here is an attempt to do the impossible, and we hope that we have satisfied some, if not all, of the requests.

The main changes from the first edition are two new chapters—Chapter 2 on X-ray diffraction and Chapter 3 on preparative methods. A short discussion of symmetry elements has been included in Chapter 1. Other additions include an introduction to ALPOs and to clay minerals in Chapter 7 and to ferroelectrics in Chapter 9. We decided that there simply was not enough room to cover the phase rule properly and for that we refer you to the excellent standard physical chemistry texts such as Atkins. We hope that the book now covers most of the basic undergraduate teaching material on solid-state chemistry.

We are indebted to Professor Tony Cheetham for kindling our interest in this subject with his lectures at Oxford University and the beautifully illustrated articles that he and his collaborators have published over the years. Our thanks are also due to Dr. Paul Raithby for commenting on part of the manuscript.

As always, we thank our colleagues at the Open University for all their support and especially the members of the lunch club, who not only keep us sane but also keep us laughing. Finally, thanks go to our families for putting up with us and, in particular, to our children for coping admirably with two increasingly distracted academic mothers—our book is dedicated to them.

Lesley Smart and Elaine Moore,
The Open University, Walton Hall, Milton Keynes, UK

Preface to the First Edition

The idea for this book originated with our involvement in an Open University inorganic chemistry course (S343: Inorganic Chemistry). When the course team met to decide the contents of this course, we felt that solid-state chemistry had become an interesting and important area that must be included. It was also apparent that this area was playing a larger role in the undergraduate syllabus at many universities, due to the exciting new developments in the field.

Despite the growing importance of solid-state chemistry, however, we found that there were few textbooks that tackled solid-state theory from a chemist's rather than a physicist's viewpoint. Of those that did, most, if not all, were aimed at final-year undergraduates and postgraduates. We felt there was a need for a book written from a chemist's viewpoint that was accessible to undergraduates earlier in their degree programme. This book is an attempt to provide such a text.

Because a book of this size could not cover all topics in solid-state chemistry, we have chosen to concentrate on structures and bonding in solids and on the interplay between crystal and electronic structure in determining their properties. Examples of solid-state devices are used throughout the book to show how the choice of a particular solid for a particular device is determined by the properties of that solid.

Chapter 1 is an introduction to crystal structures and the ionic model. It introduces many of the crystal structures that appear in later chapters and discusses the concepts of ionic radii and lattice energies. Ideas such as close-packed structures and tetrahedral and octahedral holes are covered here; these are used later to explain a number of solid-state properties.

Chapter 2 introduces the band theory of solids. The main approach is through the tight binding model, seen as an extension of the molecular orbital theory familiar to chemists. Physicists more often develop the band model through the free electron theory, which is included here for completeness. This chapter also discusses electronic conductivity in solids and, in particular, properties and applications of semiconductors.

Chapter 3 is concerned with solids that are not perfect. The types of defects that occur and the way they are organised in solids form the main subject matter. Defects lead to interesting and exploitable properties and several examples of this appear in this chapter, including photography and solid-state batteries.

The remaining chapters each deal with a property or a special class of solids. Chapter 4 covers low-dimensional solids, the properties of which are not isotropic. Chapter 5 deals with zeolites, an interesting class of compounds used extensively in the industry (as catalysts, for example), the properties of which strongly reflect their structure. Chapter 6 deals with optical properties and Chapter 7 with magnetic properties of solids. Finally, Chapter 8 explores the exciting field of superconductors, in particular the relatively recently discovered high-temperature superconductors.

The approach adopted is deliberately nonmathematical and assumes only the chemical ideas that a first-year undergraduate would have. For example, differential

calculus is used on only one or two pages and nonfamiliarity with this would not hamper an understanding of the rest of the book; topics such as ligand field theory are not assumed.

As this book originated with an Open University text, it is only right that we should acknowledge the help and support of our colleagues on the course team, in particular Dr. David Johnson and Dr. Kiki Warr. We are also grateful to Dr. Joan Mason who read and commented on much of the script and to the anonymous reviewer to whom Chapman & Hall sent the original manuscript and who provided very thorough and useful comments.

The authors have been sustained through the inevitable drudgery of writing by an enthusiasm for this fascinating subject. We hope that some of this transmits itself to the student.

Lesley Smart and Elaine Moore,
The Open University, Walton Hall, Milton Keynes, UK

Authors

Lesley Smart studied chemistry at Southampton University, United Kingdom, and after completing a PhD in Raman spectroscopy, she moved to a lectureship at the (then) Royal University of Malta. After returning to the United Kingdom, she took an SRC Fellowship to Bristol University to work on X-ray crystallography. From 1977 to 2009, she worked at the Open University chemistry department as a lecturer, senior lecturer and Molecular Science Programme director, and since retiring, holds an honorary senior lectureship there.

At the Open University, she was involved in the production of undergraduate courses in inorganic and physical chemistry and health sciences. She was the coordinating editor and an author of *The Molecular World* course, a series of eight books and DVDs copublished with the Royal Society of Chemistry, authoring two of these (2002), *The Third Dimension* and *Separation, Purification and Identification*. Her most recent books are (2007) *Alcohol and Human Health* and (2010) *Concepts in Transition Metal Chemistry*. She has an entry in *Mothers in Science: 64 Ways to Have It All* (downloadable from the Royal Society website). She has served on the Council of the Royal Society of Chemistry and as the chair of their Benevolent Fund.

Her research interests are in the characterisation of the solid state, and she has publications on single-crystal Raman studies, X-ray crystallography, Zintl phases, pigments and heterogeneous catalysis and fuel cells.

Elaine Moore studied chemistry as an undergraduate at Oxford University and then stayed on to complete a DPhil in theoretical chemistry with Peter Atkins. After a two-year postdoctoral position at Southampton, she joined the Open University in 1975 as course assistant, becoming a lecturer in chemistry in 1977, senior lecturer in 1998 and reader in 2004. She has produced OU teaching texts in chemistry for courses at levels 1, 2 and 3 and written texts in astronomy at level 2 and physics at level 3. She is coauthor of *Metal–Ligand Bonding*, and the book *Molecular Modelling and Bonding*, which forms part of the OU level 2 chemistry course, was copublished by the Royal Society of Chemistry as part of *The Molecular World* series. She oversaw the introduction of multimedia into chemistry courses and has designed multimedia material at levels 1, 2 and 3. She is coauthor of *Metals and Life* and of *Concepts in Transition Metal Chemistry,* which are part of a level 3 Open University course in inorganic chemistry and copublished with the Royal Society of Chemistry.

Her research interests are in theoretical chemistry applied to solid-state systems and to NMR spectroscopy. She is author or coauthor of over 50 papers in scientific journals.

Solid-State Chemistry was first produced in 1992. Since then, it has been translated into French, German, Spanish and Japanese and is currently being translated into Arabic.

List of Units, Prefixes, and Constants

BASIC SI UNITS

Physical Quantity (and Symbol)	Name of SI Unit	Symbol for Unit
Length (l)	Metre	m
Mass (m)	Kilogram	kg
Time (t)	Second	s
Electric current (I)	Ampere	A
Thermodynamic temperature (T)	Kelvin	K
Amount of substance (n)	Mole	mol
Luminous intensity (I_v)	Candela	cd

DERIVED SI UNITS

Physical Quantity (and Symbol)	Name of SI Unit	Symbol for SI Derived Unit and Definition of Unit
Frequency (v)	Hertz	Hz ($=s^{-1}$)
Energy (U), enthalpy (H)	Joule	J ($= kg\ m^2\ s^{-2}$)
Force	Newton	N ($= kg\ m\ s^{-2} = J\ m^{-1}$)
Power	Watt	W ($= kg\ m^2\ s^{-3} = J\ s^{-1}$)
Pressure (p)	Pascal	Pa ($= kg\ m^{-1}\ s^{-2} = N\ m^{-2} = J\ m^{-3}$)
Electric charge (Q)	Coulomb	C ($= A\ s$)
Electric potential difference (V)	Volt	V ($= kg\ m^2\ s^{-3}\ A^{-1} = J\ A^{-1}\ s^{-1}$)
Capacitance (c)	Farad	F ($= A^2\ s^4\ kg^{-1}\ m^{-2} = A\ s\ V^{-1} = A^2\ s^2\ J^{-1}$)
Resistance (R)	Ohm	O ($= V\ A^{-1}$)
Conductance (G)	Siemen	S ($= A\ V^{-1}$)
Magnetic flux density (B)	Tesla	T ($= V\ S\ m^{-2} = J\ C^{-1}\ s\ m^{-2}$)

SI PREFIXES

10^{-18}	10^{-15}	10^{-12}	10^{-9}	10^{-6}	10^{-3}	10^{-2}	10^{-1}	10^{3}	10^{6}	10^{9}	10^{12}	10^{15}	10^{18}
alto	femto	pico	nano	micro	milli	centi	deci	kilo	mega	giga	tera	peta	exa
a	f	p	n	μ	m	c	d	k	M	G	T	P	E

FUNDAMENTAL CONSTANTS

Constant	Symbol	Value
Speed of light in a vacuum	c	2.997925×10^{8} m s^{-1}
Charge of a proton and charge of an electron	e	1.602189×10^{-19} C
Avogadro constant	N_A	6.022045×1023 mol^{-1}
Boltzmann constant	k	1.380662×10^{-23} J K^{-1} mo^{l-1}
Gas constant	$R = N_A k$	$8.31441 \times$ J K^{-1}
Faraday constant	$F = N_A e$	9.648456×10^{4} C mol^{-1}
Planck constant	h	6.626176×10^{-34} J s
	$\hbar = \frac{h}{2\pi}$	1.05457×10^{-34} J s
Vacuum permittivity	ε_0	8.854×10^{-12} F m^{-1}
Vacuum permeability	μ_0	$4\pi \times 10^{-7}$ J s^2 C^{-2} m^{-1}
Bohr magneton	μ_B	9.27402×10^{-24} J T^{-1}
Electron g value	g_e	2.00232
Electron mass	m_e	9.1095×10^{-31} kg
Proton mass	m_p	1.6726×10^{-27} kg
Neutron mass	m_n	1.6749×10^{-27} kg

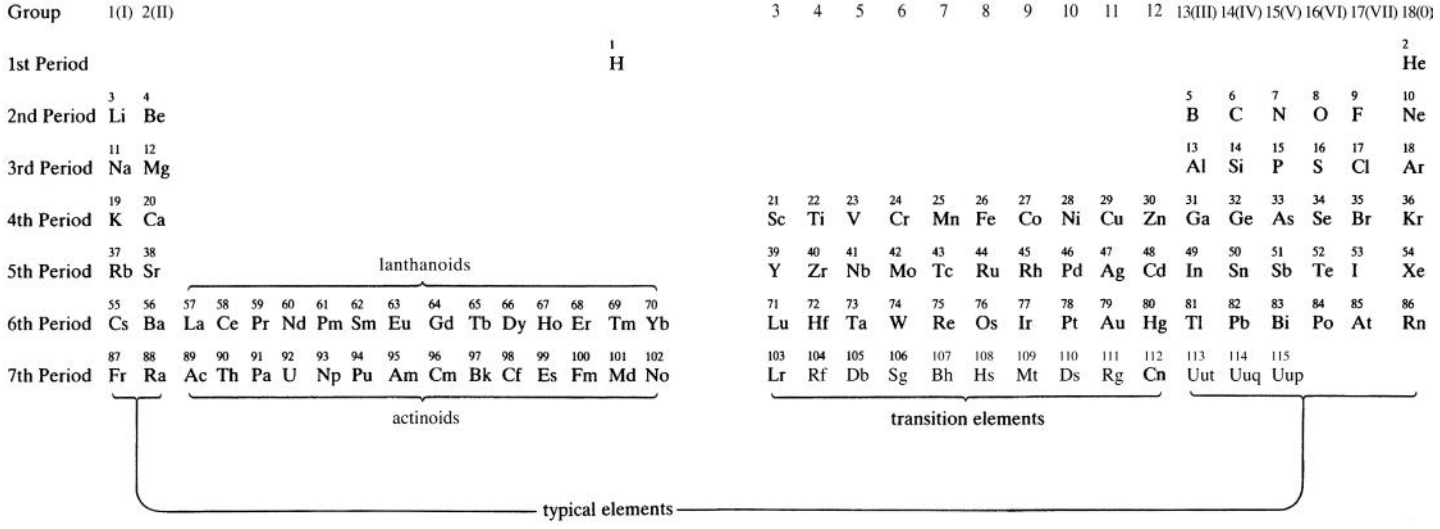
MISCELLANEOUS PHYSICAL QUANTITIES

Name of Physical Quantity	Symbol	SI Unit
Enthalpy	H	J
Entropy	S	J K^{-1}
Gibbs function	G	J
Standard change of molar enthalpy	$\Delta H_m^{\ominus}$	J mol^{-1}
Standard of molar entropy	$\Delta S_m^{\ominus}$	J K^{-1} mol^{-1}
Standard change of molar Gibbs function	$\Delta G_m^{\ominus}$	J mol^{-1}
Wave number	$\sigma = \frac{1}{\lambda}$	cm^{-1}
Atomic number	Z	Dimensionless
Conductivity	σ	S m^{-1}
Molar bond dissociation energy	D_m	J mol^{-1}
Molar mass	$M \left(= \frac{m}{n}\right)$	kg mol^{-1}

THE GREEK ALPHABET

alpha	A	α	nu	N	ν
beta	B	β	xi	Ξ	ξ
gamma	Γ	γ	omicron	O	o
delta	Δ	δ	pi	Π	π
epsilon	E	ε	rho	P	ρ
zeta	Z	ζ	sigma	Σ	σ
eta	H	η	tau	T	τ
theta	Θ	θ	upsilon	Υ	υ
iota	I	ι	phi	Φ	φ
kappa	K	κ	chi	X	χ
lambda	Λ	λ	psi	Ψ	ψ
mu	M	μ	omega	Ω	ω

PERIODIC CLASSIFICATION OF THE ELEMENTS



1 An Introduction to Crystal Structures

In the last few decades, research into solid-state chemistry has expanded very rapidly, fuelled at the start by the search for new and better materials and by the discovery in 1986 of 'high-temperature' ceramic oxide superconductors. We have seen immense strides in the development and understanding of nanotechnology, microporous and mesoporous solids, fuel cells and the giant magnetoresistance effect, to mention but a few areas. It would be impossible to cover all the developments and topics in detail in an introductory text such as this, but we endeavour to give you a flavour of the excitement that some of the research has engendered and perhaps, more importantly, sufficient background with which to understand these developments and those that are yet to come.

Most substances, helium being a notable exception, if cooled sufficiently form a solid phase; the vast majority form one or more **crystalline** phases, where the atoms, molecules or ions pack together to form a regular repeating array. This book is concerned mostly with the structures of metals, ionic solids and extended covalent structures; structures that do not contain discrete molecules as such, but that comprise extended arrays of atoms or ions. We look at the structure of, and bonding in, these solids, how the properties of a solid depend on its structure, and how the properties can be modified by changes to the structure.

1.1 INTRODUCTION

To understand the solid state, we need to have some insight into the structure of simple crystals and the forces that hold them together, so it is here that we start this book. Crystal structures are usually determined by the technique of **X-ray crystallography**. This technique relies on the fact that the distances between the atoms in the crystals are of the same order of magnitude as the wavelength of X-rays (of the order of 1 Å or 100 pm): a crystal thus acts as a three-dimensional diffraction grating to a beam of X-rays. The resulting diffraction pattern can be interpreted to give the internal positions of the atoms in the crystal very precisely, thus defining interatomic distances and angles. (Some of the principles underlying this technique are discussed in Chapter 2, where we review the physical methods available for characterising solids.) Most of the structures discussed in this book would have been determined in this way.

The structures of many inorganic crystal structures can be discussed in terms of the simple packing of spheres, so we will consider this first, before moving on to the more formal classification of crystals.

1.2 CLOSE PACKING

Think for a moment of an atom as a small hard sphere. Figure 1.1 shows two possible arrangements for a layer of such identical atoms. On squeezing the square layer in Figure 1.1a, the spheres would move to the positions in Figure 1.1b so that the layer takes up less space. The layer in Figure 1.1b—(layer A)—is said to be **close-packed**. The hollow spaces between the spheres are of two types, which have been marked with dots and crosses, respectively.

To build up a close-packed structure in three dimensions, we must now add a second layer—layer B. The spheres of the second layer can only sit in half of the hollows of the first layer, and layer B in Figure 1.2 (shown in blue) sits directly over the hollows marked with a cross (although it makes no difference which type we choose). When we come to add a third layer, there are again two possible options as to where it can go. First, it could go directly over layer A, in the unmarked hollows; if we repeated this stacking sequence, we would build up the layers ABABABA... and so on. This is known as **hexagonal close packing (hcp)** (Figure 1.3a). In this structure, the hollows marked with a dot are never occupied by spheres, leaving very small channels through the layers (Figure 1.3b).

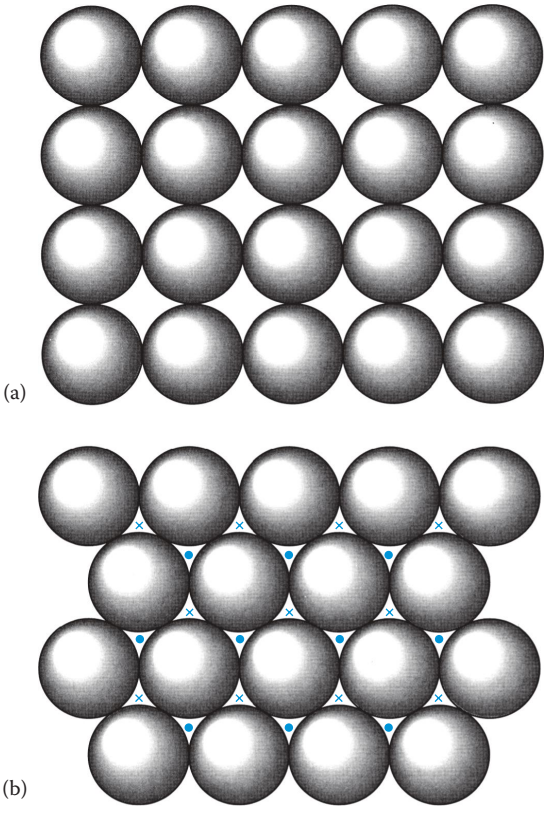


FIGURE 1.1 (a) A square array of spheres. (b) A close-packed layer of spheres.

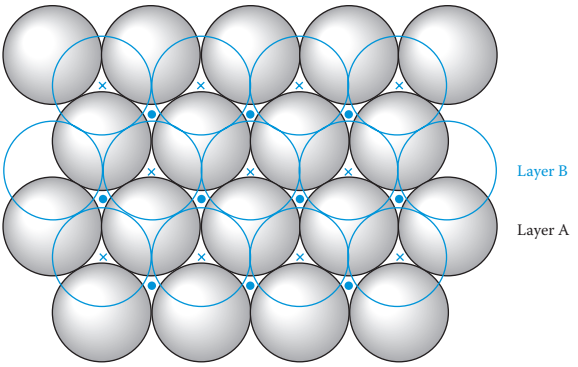


FIGURE 1.2 Two layers of close-packed spheres.

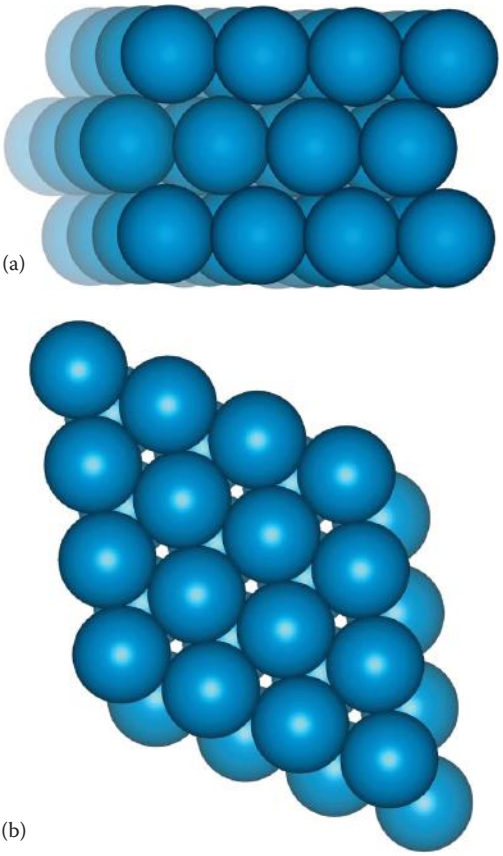


FIGURE 1.3 (a) Three *hcp* layers showing the ABAB... stacking sequence. (b) Three *hcp* layers showing the narrow channels through the layers.

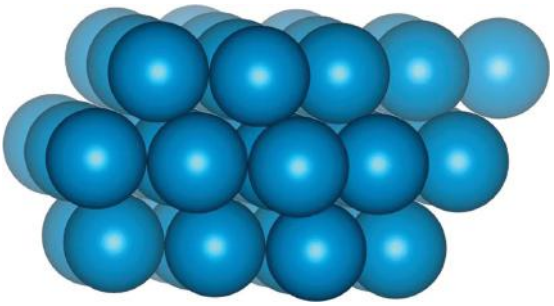


FIGURE 1.4 Three *ccp* layers.

In the second option, the third layer could be placed over those hollows, as in Figure 1.2b, marked with a dot. This third layer, which we now label C, would not be directly over either A or B, and the stacking sequence when repeated would be ABCABCABC... and so on. This is known as **cubic close packing (ccp)** (Figure 1.4). (The names *hexagonal* and *cubic* for these structures arise from the resulting symmetry of the structures—this will be discussed more fully later on.)

Close packing represents the most efficient use of space when packing identical spheres—74% of the total volume is occupied by the spheres: the **packing efficiency** is said to be 74%. Each sphere in the structure is surrounded by twelve equidistant neighbours—six in the same layer, three in the layer above and three in the layer below: the **coordination number** of an atom in a close-packed structure is thus 12.

Another important feature of close-packed structures is the shape and number of the small amounts of space trapped in between the spheres. There are two different types of spaces within a close-packed structure: the first that we will consider is called an **octahedral hole**. Figure 1.5a shows two close-packed layers again, but with the octahedral holes shaded. Each of these holes is surrounded by *six* spheres, three in layer A and three in layer B; the centres of these spheres lie at the corners of an octahedron, hence the name (Figure 1.5b). If there are n spheres in the array, then there are also n octahedral holes.

Similarly, Figure 1.6a shows two close-packed layers, now with the second type of space, **tetrahedral holes**, shaded. Each of these holes is surrounded by four spheres with centres at the vertices of a tetrahedron (Figure 1.6b). If there are n spheres in the array, then there are $2n$ tetrahedral holes.

The octahedral holes in a close-packed structure are much bigger than the tetrahedral holes—they are surrounded by six atoms rather than four. It is a matter of simple geometry to calculate that in a close-packed array of spheres of radius r, the radius of a sphere that will just fit in an octahedral hole is $0.414r$. For a tetrahedral hole, the radius is $0.225r$ (Figure 1.7).

There are, of course, innumerable stacking sequences possible when repeating close-packed layers; however, hexagonal close-packed and cubic close-packed structures are those of maximum simplicity and are those most commonly encountered in the crystal structures of the noble gases and of the metallic elements. Only two other stacking sequences are found in perfect crystals of the elements—an ABAC repeat in La, Pr, Nd and Am; and a nine-layer repeat ABACACBCB in Sm.

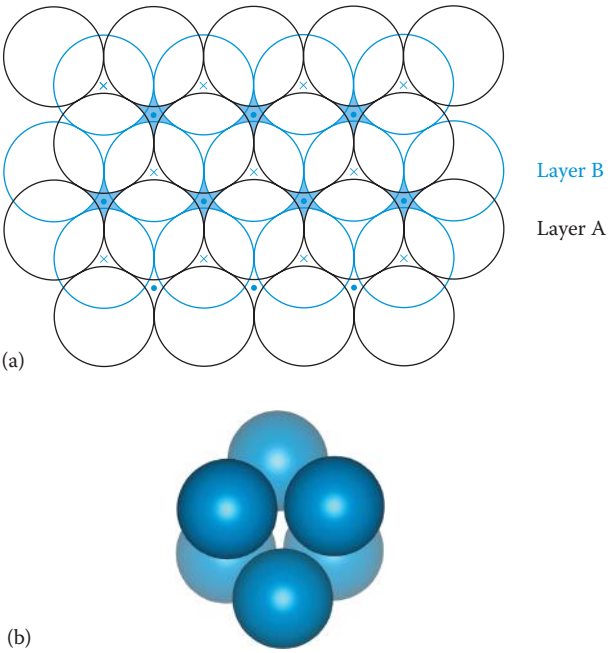


FIGURE 1.5 (a) Two layers of close-packed spheres with the enclosed octahedral holes shaded. (b) A computer representation of an octahedral hole.

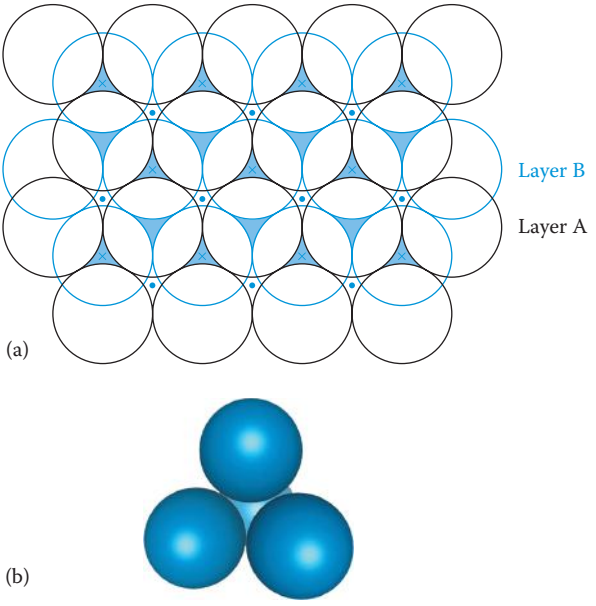


FIGURE 1.6 (a) Two layers of close-packed spheres with the tetrahedral holes shaded. (b) A computer representation of a tetrahedral hole.

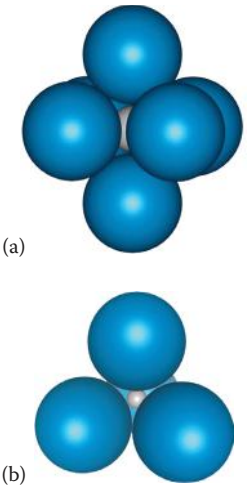


FIGURE 1.7 (a) A sphere of radius 0.414*r* fitting into an octahedral hole. (b) A sphere of radius 0.225*r* fitting into a tetrahedral hole.

1.3 BODY-CENTRED AND PRIMITIVE STRUCTURES

Some metals do not adopt a close-packed structure, but have a slightly less efficient packing method: it is the **body-centred cubic (bcc) structure**, shown in Figure 1.8. (Unlike the previous diagrams, the positions of the atoms are now represented here, and in subsequent diagrams, by small spheres that do not touch each other: this is merely a device to open up the structure and allow it to be seen more clearly. The whole question of atom and ion size is discussed in Section 1.6.4.) In this structure, an atom in the middle of a cube is surrounded by eight identical and equidistant atoms at the corners of the cube, the coordination number has dropped from 12 to 8 and the packing efficiency is now 68%, compared with 74% for close packing.

The simplest of the cubic structures is the **primitive cubic structure**. This is built by placing square layers like the one shown in Figure 1.1a, directly on top of one

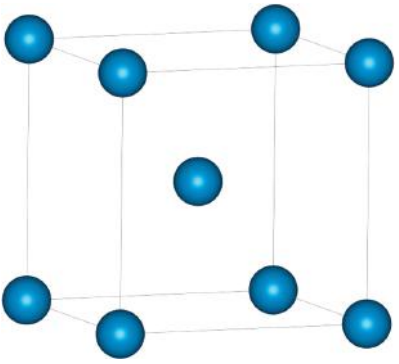


FIGURE 1.8 A body-centred cubic array.

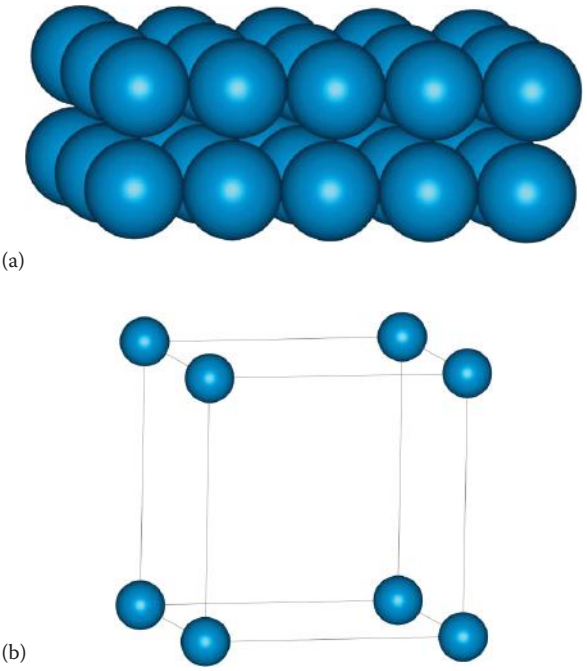


FIGURE 1.9 (a) Two layers of a primitive cubic array. (b) A cube of atoms from this array.

another. Figure 1.9a illustrates this, and we can see in Figure 1.9b that each atom sits at the corner of a cube. The coordination number of an atom in this structure is six.

The majority of metals have one of the three basic structures: *hcp*, *ccp* or *bcc*; polonium alone adopts the primitive structure. The distribution of the packing types among the most stable forms of the metals at 298 K is shown in Figure 1.10. As we noted earlier, very few metals have a mixed *hcp/ccp* structure of a more

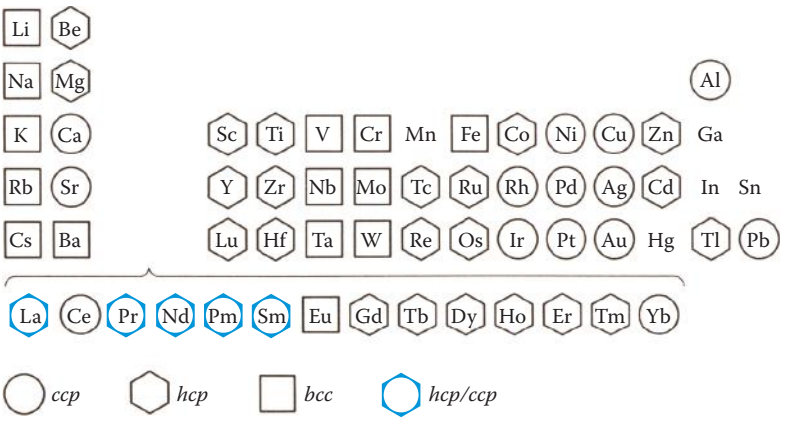


FIGURE 1.10 Occurrence of packing types among the metals.

complex type. The structures of the actinoids tend to be rather complex and are not included.

1.4 SYMMETRY

Before we take the discussion of crystalline structures any further, we will look at the symmetry displayed by chemical species. The concept of symmetry is an extremely useful one when it comes to describing the shapes of both individual molecules and regular repeating structures, as it provides a way of describing similar features in different structures so that they become unifying features. The symmetry of objects in everyday life is something that we tend to take for granted and recognise easily without having to think about it. Take some simple examples illustrated in Figure 1.11. If we imagine a mirror dividing a spoon in half along the plane indicated, then we can see that one half of the spoon is a mirror image or reflection of the other half. Similarly with the paintbrush, only now there are two

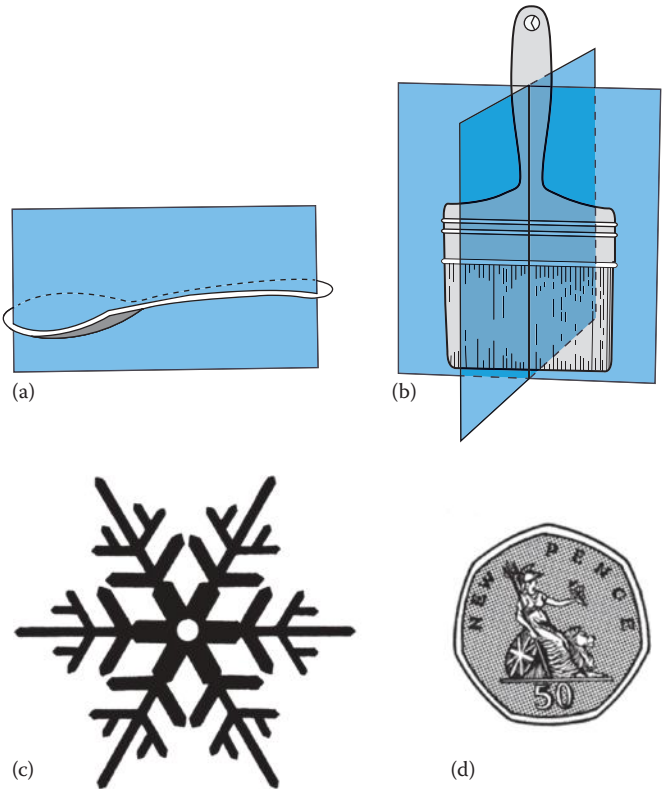


FIGURE 1.11 Common objects displaying symmetry: (a) a spoon, (b) a paintbrush, (c) a snowflake and (d) a British 50p coin.

mirror planes at right angles which divide it. This symmetry feature is called a **plane of symmetry**.

Objects can also possess rotational symmetry. In Figure 1.11c, imagine an axle passing through the centre of the snowflake; in the same way as a wheel rotates about an axle, if the snowflake is rotated through one-sixth of a revolution, then the new position is indistinguishable from the old. Similarly, in Figure 1.11d, rotating the seven-sided UK 50p coin about a central axis by one-seventh of a revolution brings us to the same position as we started (ignoring the pattern on the surface). These axes are known as **axes of symmetry** or **rotational axes**. The symmetry axes and planes possessed by objects are examples of **symmetry elements**.

1.4.1 Symmetry Notation

Two forms of symmetry notation are in common use, and as chemists we will come across both. The **Schoenflies** notation is useful for describing the symmetry of individual molecules and is used by spectroscopists. The **Hermann–Mauguin** notation can be used to describe the symmetry of individual molecules; additionally, it is also used to describe the relationship between different molecules in space—their so-called **space symmetry**—and so it is the form most commonly met in crystallography and in the solid state, although both are used. Here, we give the Schoenflies notation in parentheses after the Hermann–Mauguin notation.

1.4.2 Axes of Symmetry

In the same way that we have previously seen for the snowflake and the 50p coin, molecules can also possess rotational symmetry. Figure 1.12 illustrates this for several molecules.

In Figure 1.12a, the rotational axis is shown as a vertical line through the O atom in OF_2; rotation about this line by 180° in the direction of the arrow produces an identical-looking molecule. The axis of symmetry in this case is a twofold axis because we have to perform the rotation twice to return the molecule to its starting position.

Axes of symmetry are denoted by the symbol n (C_n), where n is the order of the axis. So the rotational axis of the OF_2 molecule is 2 (C_2).

The BF_3 molecule in Figure 1.12b possesses a threefold axis of symmetry (3 (C_3)) because each one-third of a revolution leaves the molecule looking the same, and three turns bring the molecule back to its starting position. In the same way, the XeF_4 molecule in Figure 1.12c has a fourfold axis (4 (C_4)) and four quarter turns are necessary to bring it back to its starting position. Notice that this molecule also has C_2 rotational axes at right angles to the C_4 axis; this highest-order axis is then denoted as the **principal axis**. All linear molecules have an ∞ (C_∞) axis, which is illustrated for the BeF_2 molecule in Figure 1.12d; however small an angle of rotation it always looks identical. The smallest angle of rotation possible is $1/\infty$, and so this principal axis is an infinite-order axis of symmetry.

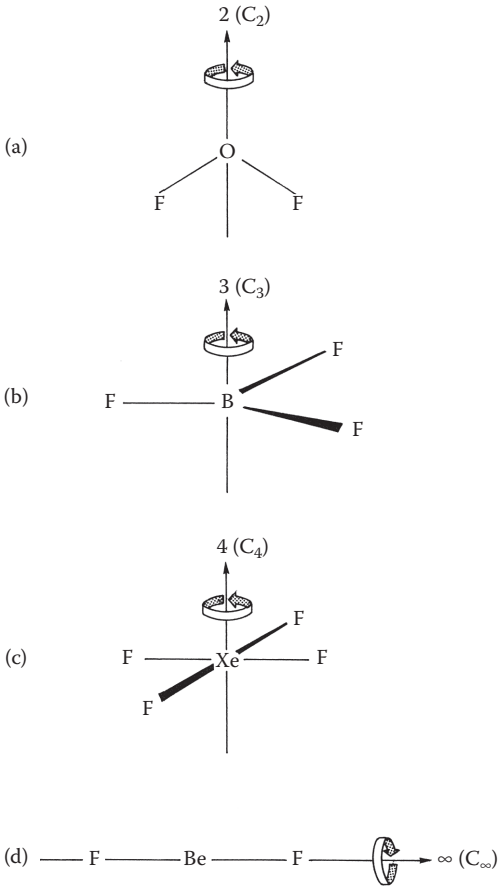


FIGURE 1.12 Axes of symmetry in molecules: (a) twofold axis in OF_2, (b) threefold axis in BF_3, (c) fourfold axis in XeF_4 and (d) ∞-fold axis in BeF_2.

1.4.3 Planes of Symmetry

Mirror planes occur in isolated molecules and in crystals such that everything on one side of the plane is a mirror image of the other. In a structure such a mirror plane is given the symbol *m* (σ). Molecules may possess one or more **planes of symmetry**, and the diagrams in Figure 1.13 illustrate some examples. The planar OF_2 molecule has two planes of symmetry (Figure 1.13a): one is the plane of the molecule and the other is at right angles to the former. For *all* planar molecules, the plane of the molecule is a plane of symmetry. The diagrams for BF_3 (and also planar molecules) only show the planes of symmetry that are perpendicular to the plane of the molecule. A symmetry plane that is parallel to the principal axis is known as *vertical* (σ_v) and the one perpendicular to it is known as *horizontal* (σ_h). A third type of symmetry plane that bisects the angle between 2 twofold rotation axes perpendicular to the principal axis, as shown in Figure 1.13c for XeF_4, is known as *dihedral* (σ_d).

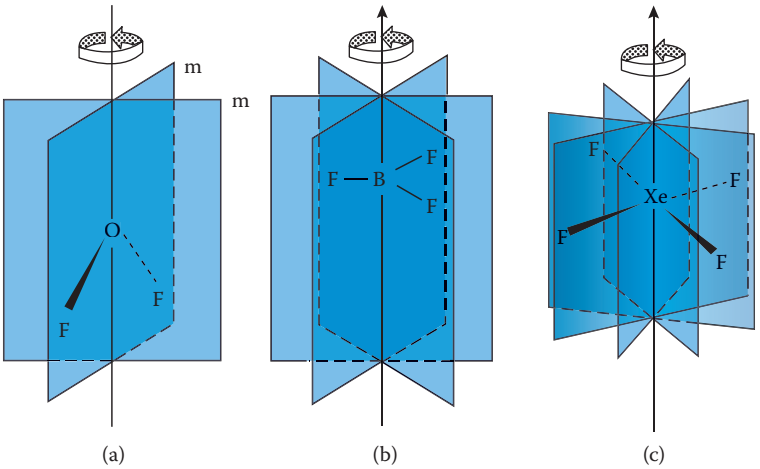


FIGURE 1.13 Planes of symmetry in molecules: (a) planes of symmetry in OF_2, (b) planes of symmetry in BF_3 and (c) planes of symmetry in XeF_4.

1.4.4 Inversion

In this section, we show the third symmetry element called **inversion** through an **inversion centre** or **centre of symmetry**; it is given the symbol $\bar{1}$(i). In this element, you have to imagine a line drawn from any atom in the molecule, through the centre of symmetry and then continued for the same distance the other side; if, for *every* atom, this meets with an identical atom on the other side, then the molecule has a centre of symmetry. Of the molecules in Figure 1.12, XeF_4 and BeF_2 both have a centre of symmetry, while BF_3 and OF_2 do not. The benzene molecule (C_6H_6) possesses a centre of symmetry in the middle of the ring.

The next symmetry element is described differently in the Schoenflies and Hermann–Mauguin notations.

1.4.5 Inversion Axes and Improper Symmetry Axes and the Identity Element

The two following symmetry elements, *inversion axis* and *improper symmetry axis*, give the same result, but are described differently in the two systems. Both descriptions use a combination of the symmetry elements described above. The Hermann–Mauguin **inversion axis** is a combination of rotation and inversion and is given the symbol $\bar{n}$. The symmetry element consists of a rotation by $1/n$ of a revolution about the symmetry axis, followed by inversion through the centre of symmetry. An example of a $\bar{4}$ inversion axis is shown in Figure 1.14 for a tetrahedral molecule such as CF_4. The molecule is shown inside a cube as this makes it easier to see the symmetry elements. Anticlockwise rotation about the axis through 90° takes F_1 to the position shown as a dotted F, and inversion through the centre then takes this atom to the F_3 position.

The equivalent symmetry element in the Schoenflies notation is the **improper axis of symmetry (S_n)** or the **rotation–reflection axis**, which, as its name suggests,

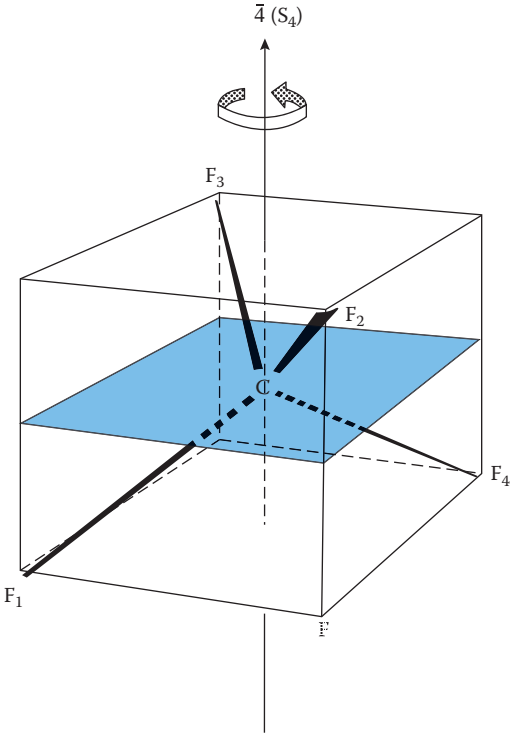


FIGURE 1.14 The $\bar{4}$ (S_4) inversion (improper) axis of symmetry in the tetrahedral CF_4 molecule.

is a combination of rotation and reflection. The symmetry element consists of a rotation by $1/n$ of a revolution about the axis, followed by reflection through a plane at right angles to the axis. Figure 1.14 thus also shows an S_4 axis: F_1 rotates to the dotted position and then reflects to F_2. The equivalent inversion axes and improper symmetry axes for the two systems are shown in Table 1.1.

TABLE 1.1
Equivalent Symmetry Elements in the Schoenflies and Hermann–Mauguin Systems

Schoenflies	Hermann–Mauguin
$S_1 \equiv m$	$\bar{2} \equiv m$
$S_2 \equiv i$	$\bar{1} \equiv i$
S_3	$\bar{6}$
S_4	$\bar{4}$
S_6	$\bar{3}$

The group of five symmetry elements is completed by the so-called **identity element (*E*)**, which is no change—the equivalent of multiplying by 1.

1.4.6 Operations

Whereas the symmetry elements are essentially static properties of the object, actions such as rotating or reflecting a molecule are called **symmetry operations**, and the five symmetry elements have five symmetry operations associated with them. Their notation is distinguished from those of the elements by a caret; thus, $\hat{C}_n$ is the rotation of a molecule around an axis and $\hat{E}$ is the identity operation.

The group of *all* the symmetry operations acting on a single object to describe the repetition of identical parts of the object while one point remains fixed is known as its **point group**. For the determination of the point groups of common molecules, we refer you to the standard texts on the subject (see the Further Reading at the end of the book).

1.4.7 Symmetry in Crystals

The discussion so far has only shown the symmetry elements that belong to individual molecules. However, in the solid state, we are interested in regular arrays of atoms, ions or molecules, and they too are related by these same symmetry elements. Figure 1.15 shows examples (not real) of how molecules could be arranged in

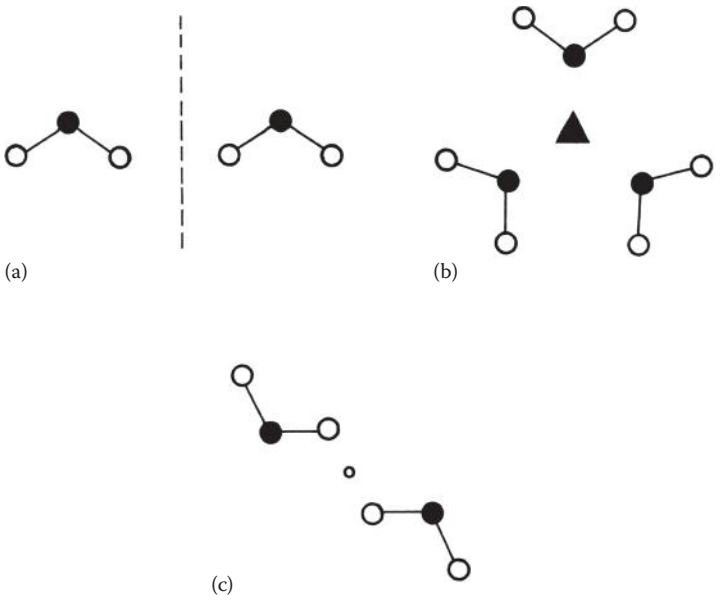


FIGURE 1.15 Symmetry in solids: (a) two OF_2 molecules related by a plane of symmetry, (b) three OF_2 molecules related by a threefold axis of symmetry and (c) two OF_2 molecules related by a centre of inversion.

a crystal. In Figure 1.15a, two OF_2 molecules are related to one another by a plane of symmetry; in Figure 1.15b, three OF_2 molecules are related to one another by a threefold axis of symmetry; and in Figure 1.15c, two OF_2 molecules are related by a centre of inversion. Notice that in Figure 1.15b and 1.15c, the molecules are related in space by a symmetry element that they themselves do not possess; this is said to be their **site symmetry**.

1.5 LATTICES AND UNIT CELLS

Crystals are regular-shaped solid particles with flat shiny faces. In 1664, Robert Hooke first noted that the regularity of their external appearance is a reflection of a high degree of internal order. Crystals of the same substance, however, vary in shape considerably. In 1671, Steno observed that this is not because their internal structure varies, but because some faces develop more than others. The angles between similar faces on different crystals of the same substance are always identical. The constancy of the interfacial angles reflects the internal order within the crystals. Each crystal is derived from a basic 'building block' that repeats over and over again, in all directions, in a perfectly regular way. This building block is known as the **unit cell**.

In order to talk about and compare the many thousands of crystal structures that are known, there has to be a way of defining and categorising the structures. This is achieved by defining the shape, symmetry and also the size of each unit cell and the positions of the atoms within it.

1.5.1 Lattices

The simplest regular array is a line of evenly spaced objects such as that depicted by the commas in Figure 1.16a. There is a dot at the same place in each object: if we now remove the objects leaving the dots, we have a line of equally spaced dots of spacing *a* (Figure 1.16b). The line of dots is called the **lattice**, and by definition each **lattice point** (dot) has *identical surroundings*. This is the only example of a one-dimensional lattice and it can vary only in the spacing *a*. There are five possible two-dimensional lattices, and everyday examples of these can be seen all around in wallpapers and tiling.

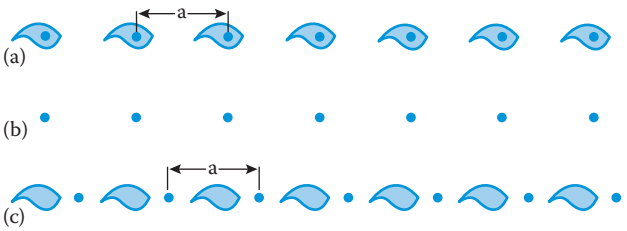


FIGURE 1.16 (a–c) A one-dimensional lattice and the choice of unit cells.

1.5.2 One- and Two-Dimensional Unit Cells

The unit cell for the one-dimensional lattice in Figure 1.16a lies between the two vertical lines. If we took this unit cell and repeated it over and over again, we would reproduce the original array. Notice that it does not matter where in the structure we place the lattice points as long as they each have identical surroundings. In Figure 1.16c, we have moved the lattice points and the unit cell, but repeating this unit cell will still give the same array—we have simply moved the origin of the unit cell. There is never one unique unit cell that is 'correct'; there are always many that could be chosen and the choice depends both on convenience and convention. This is equally true in two and three dimensions.

The unit cells for the two-dimensional lattices are parallelograms with their corners at equivalent positions in the array, that is, *the corners of a unit cell are lattice points*. In Figure 1.17, we show a square array with several different unit cells depicted. All of these, if repeated, would reproduce the array, but it is conventional to choose the smallest cell that fully represents the symmetry of the structure. Both unit cells (1a) and (1b) are of the same size, but clearly (1a) shows that the array is square, and this would be the conventional choice. Figure 1.18 demonstrates the same principles but for a centred rectangular array, where (a) would be the conventional choice because it includes information on the centring; the smaller unit cell (b)

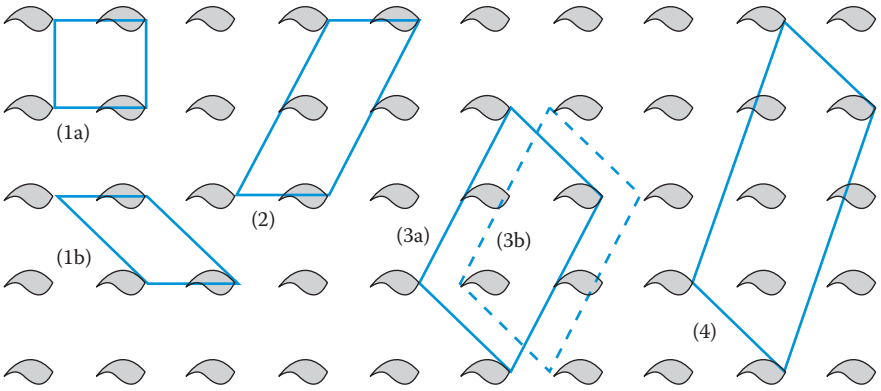


FIGURE 1.17 Choice of unit cell in a square two-dimensional lattice.

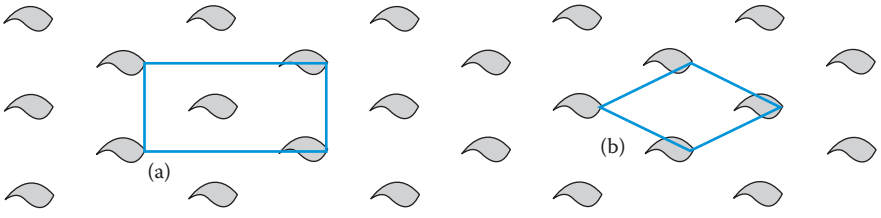


FIGURE 1.18 Choice of unit cell in a centred-rectangular lattice.

loses this information. It is always possible to define a noncentred oblique unit cell, but information about the symmetry of the lattice may be lost by doing so.

Unit cells such as (1a) and (1b) in Figure 1.17 and (b) in Figure 1.18 have a lattice point at each corner. However, they each contain a total of one lattice point because each lattice point is shared by four adjacent unit cells. They are known as **primitive unit cells** and are given the symbol **P**. The unit cell marked (a) in Figure 1.18 contains a total of two lattice points—one from the shared four corners and the other totally enclosed within the cell. This cell is said to be **centred** and is given the symbol **C**.

1.5.3 Translational Symmetry Elements

In Section 1.4, we introduced you to the idea of symmetry, both in individual molecules and for extended arrays of molecules such as those found in crystals. Before going on to discuss three-dimensional lattices and unit cells, it is important to introduce two more symmetry elements; these elements involve translation and are only found in the solid state.

The **glide plane** combines translation with reflection. Figure 1.19 shows an example of this symmetry element. The diagram shows part of a repeating three-dimensional structure projected onto the plane of the page; the circle represents a molecule or ion in the structure and there is a distance a between identical positions in the structure in the x direction (this follows the convention that a unit cell based on the x, y and z axes has unit cell dimensions a, b and c in those respective directions). The + sign next to the circle indicates that the molecule lies above the plane of the page in the z direction. The plane of symmetry is in the xz plane perpendicular to the page and is indicated by the dashed line. The symmetry element consists of reflection through this plane of symmetry, followed by translation. The translation can be either in the x or in the z direction (or along a diagonal) and the translation distance is half of the repeat distance in that direction. In the example illustrated, the translation takes place in the x direction; the repeat distance between identical molecules is therefore a, and the translation is by $a/2$. This symmetry

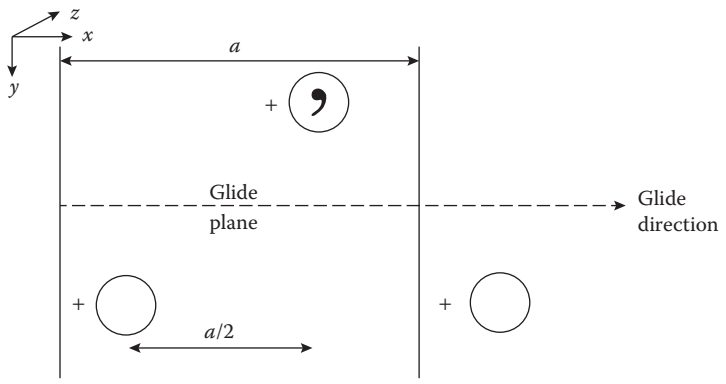


FIGURE 1.19 An *a* glide perpendicular to *b*.

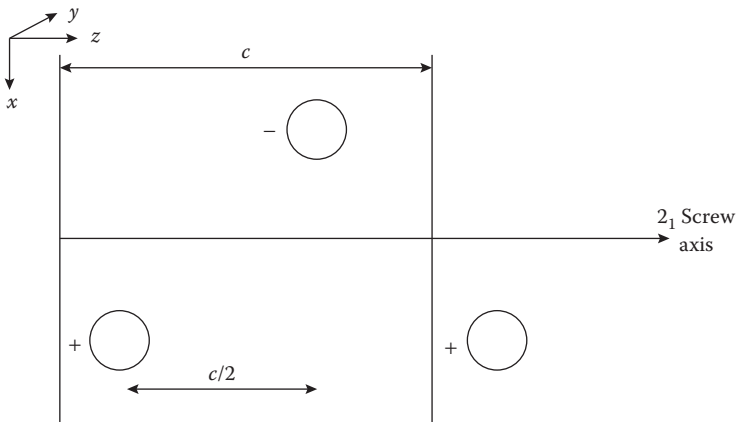


FIGURE 1.20 A 2_1 screw axis along z.

element is called an ***a* glide**. If the translation had been along the z direction, the translation would have been by $c/2$ and is called a ***c* glide**.

You will notice two things about the molecule generated by this symmetry element: first, it still has a + sign against it, because the reflection in the plane leaves the z coordinate the same and, second, it now has a comma on it. Some molecules when

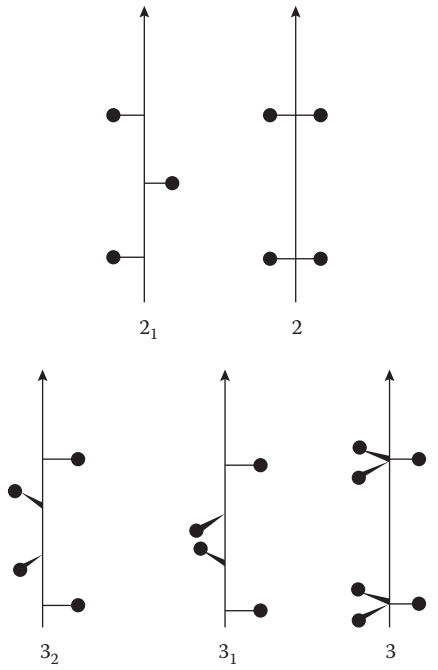


FIGURE 1.21 Comparison of the effects of twofold and threefold rotation axes and screw axes.

reflected through a plane of symmetry are enantiomorphic, which means that they are not superimposable on their mirror image: the presence of the comma indicates that this molecule could be an enantiomorph.

The **screw axis** combines translation with rotation. Screw axes have the general symbol n_i, where n is the rotational order of the axis, that is, twofold, threefold, etc., and the translation distance is given by the ratio i/n. Figure 1.20 illustrates a 2_1 screw axis. In this example, the screw axis lies along the z direction and so the translation must also be in the z direction by $c/2$, where c is the repeat distance in the z direction. Notice that in this case, the molecule starts above the plane of the page (indicated by the + sign), but the effect of a twofold rotation is to take it below the plane of the page (− sign). Figure 1.21 probably illustrates this more clearly and also shows the different effects that the rotational and screw axes of the same order have on a repeating structure. Rotational and screw axes produce objects that are superimposable on the original. All other symmetry elements—glide plane, symmetry plane, inversion centre and inversion axis—produce a mirror image of the original.

1.5.4 Three-Dimensional Lattices and Their Unit Cells

There are seven lattice systems defined by their symmetry: triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal and cubic (Table 1.2).

As we saw in two dimensions, a lattice can always be defined by a primitive unit cell (P) with a lattice point at each corner; the same is true in three dimensions. However, in three dimensions, there are three other types of lattices with unit cells

TABLE 1.2
Seven Lattice Systems

System	Unit Cell	Minimum Symmetry Requirements
Triclinic	$\alpha \neq \beta \neq \gamma \neq 90°$ $a \neq b \neq c$	None
Monoclinic	$\alpha = \gamma = 90°$ $\beta \neq 90°$ $a \neq b \neq c$	One twofold axis or one symmetry plane
Orthorhombic	$\alpha = \beta = \gamma = 90°$ $a \neq b \neq c$	Any combination of three mutually perpendicular twofold axes or planes of symmetry
Tetragonal	$\alpha = \beta = \gamma = 90°$ $a = b \neq c$	One fourfold axis or one fourfold improper axis
Rhombohedral	$\alpha = \beta = \gamma \neq 90°$ $a = b = c$	One threefold axis
Hexagonal	$\alpha = \beta = 90°$ $\gamma = 120°$ $a = b \neq c$	One sixfold axis or one sixfold improper axis
Cubic	$\alpha = \beta = \gamma = 90°$ $a = b = c$	Four threefold axes at 109°28′ to each other

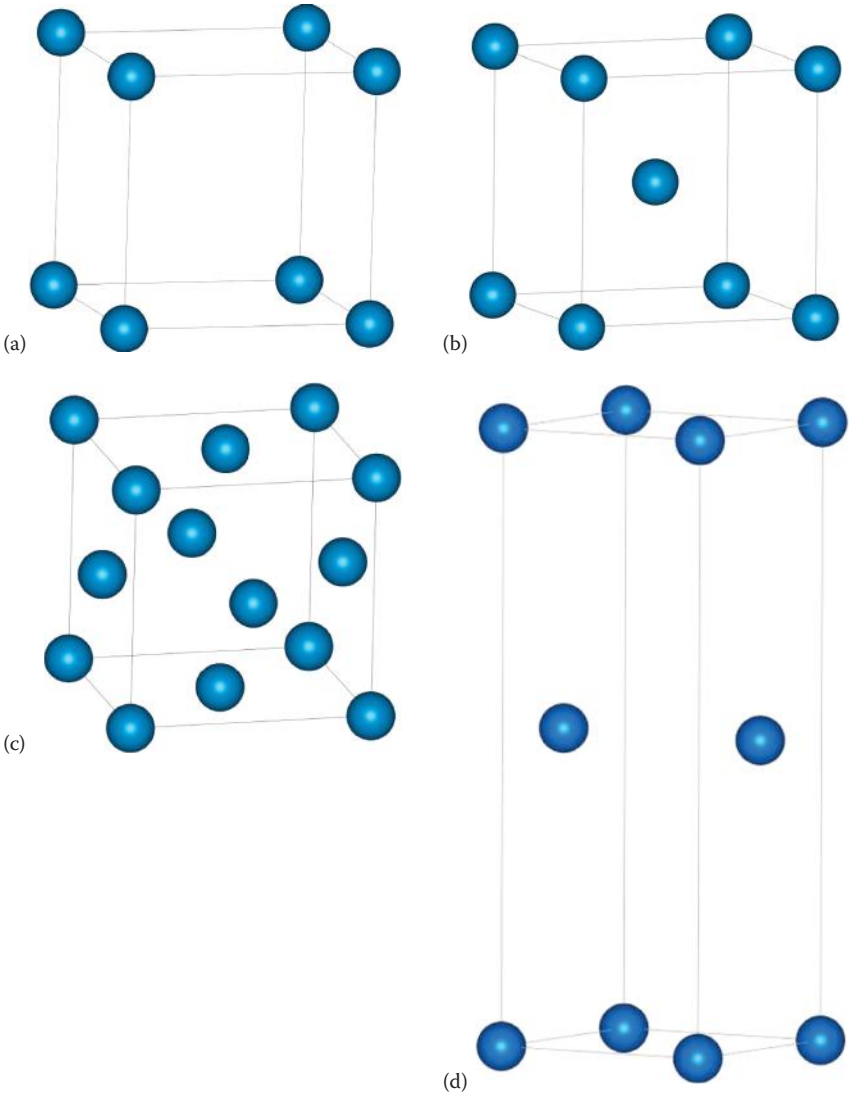


FIGURE 1.22 (a) Primitive (P), (b) body-centred (I), (c) face-centred (F) and (d) face-centred (A, B or C) unit cells.

larger than the primitive ones; all four types are listed next and are illustrated in Figure 1.22.

1. The **primitive** unit cell—symbol **P**—has a lattice point at each corner.
2. The **body-centred** unit cell—symbol **I**—has a lattice point at each corner and one at the centre of the cell.
3. The **face-centred** unit cell—symbol **F**—has a lattice point at each corner and one at the centre of each face.

4. The **face-centred** unit cell—symbol **A**, **B** or **C**—has a lattice point at each corner, and one at the centres of a pair of opposite faces, for example, an A-centred cell has lattice points at the centres of the *bc* faces.

These lattice types are distributed among the lattice systems to give the 14 Bravais lattices listed in Table 1.3. Notice, for instance, that it is not possible to have an A-centred (or B- or C-centred) cubic unit cell; if only two of the six faces are centred, the unit cell necessarily loses its cubic symmetry.

The unit cell of a three-dimensional lattice is a parallelepiped defined by three distances a, b and c, along the x, y and z axes, respectively, and three angles α, β and γ, where α is the angle between the y and z axes; β between x and z; and γ between x and y (Figure 1.23). Because the unit cells are the basic building blocks of the crystals, they are space-filling, that is, they must pack together to fill all space. The unit cells of the seven lattice systems are illustrated in Figure 1.24. The seven lattice systems together with the shapes of their unit cells, as determined by minimum symmetry requirements, are detailed in Table 1.3.

TABLE 1.3
Bravais Lattices

Lattice System	Lattice Types
Cubic	P, I, F
Tetragonal	P, I
Orthorhombic	P, C, I, F
Hexagonal	P
Rhombohedral	R[a]
Monoclinic	P, C
Triclinic	P

[a] The primitive unit cell of the rhombohedral system is normally given the symbol R.

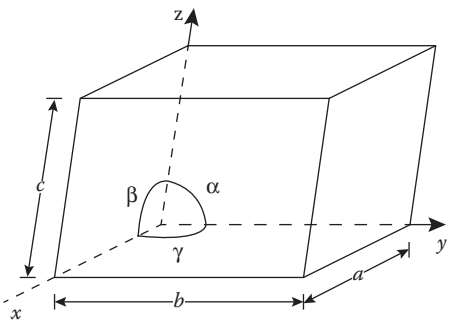
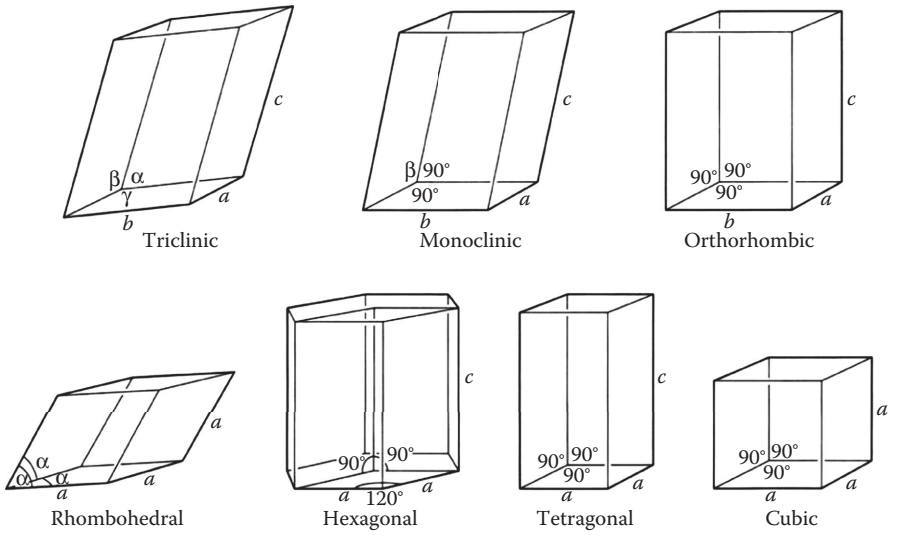


FIGURE 1.23 Definition of axes, unit cell dimensions and angles for a general unit cell.



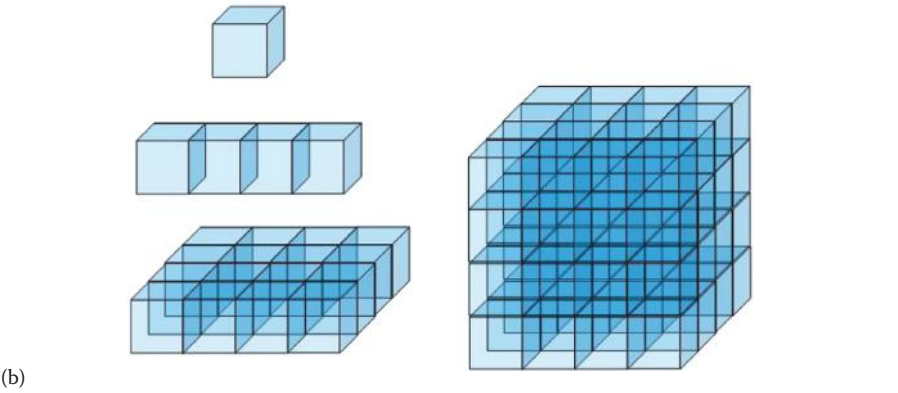


FIGURE 1.24 (a) Unit cells of the seven crystal systems. (b) Assemblies of cubic unit cells in one, two and three dimensions.

The symmetry of a crystal is a point group taken from a point at the centre of a perfect crystal and the set of symmetry operations leave this point fixed while moving each atom of the crystal to the position of an atom of the same kind. Only certain point groups are possible because of the constraint made by the fact that unit cells must be able to stack exactly with no spaces—thus, for instance, only one-, two-, three-, four- and sixfold rotational axes are possible. Combining this with the planes of symmetry and centres of symmetry, we find that there are 32 crystallographic point groups compatible with the translational symmetry in three dimensions. If we combine the 32 crystal point groups with the 14 Bravais lattices, we find that there are 230 three-dimensional **space groups** that crystal structures can adopt, that is, 230 different space-filling patterns. These are all documented in the International Tables for Crystallography (see Bibliography at the end of the book).

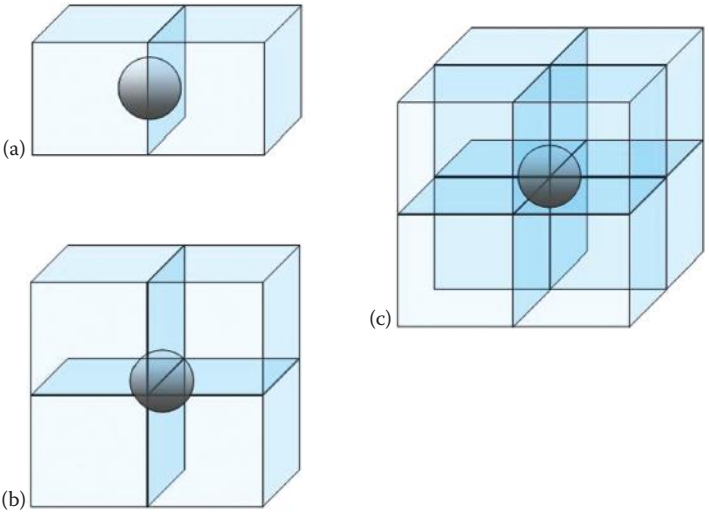


FIGURE 1.25 Unit cells showing a molecule on (a) a face, (b) an edge and (c) a corner.

It is important not to lose sight of the fact that the *lattice points represent equivalent positions* in a crystal structure and *not* atoms. In a real crystal, a lattice point could be occupied by an atom, a complex ion, a molecule or even a group of molecules. The lattice points are used to simplify the repeating patterns within a structure, but they tell us nothing of the chemistry or bonding within the crystal—for that we have to include the atomic positions; this we will do later in the chapter when we look at some real structures.

It is instructive to note how much of a structure these various types of unit cell represent. We noted a difference between the centred and the primitive two-dimensional unit cells, where the centred cell contains two lattice points whereas the primitive cell contains only one. We can work out similar occupancies for the three-dimensional case. The number of unit cells sharing a particular lattice point depends on its site. A corner site is shared by eight unit cells, an edge site by four, a face site by two and a lattice point at the body centre is not shared by any other unit cell (Figure 1.25). Using these figures, we can work out the number of lattice points in each of the four types of cells in Figure 1.22. The results are listed in Table 1.4.

TABLE 1.4
Number of Lattice Points in the Four Types of Unit Cell

Name	Symbol	Number of Lattice Points in Unit Cell
Primitive	P	1
Body-centred	I	2
Face-centred	A or B or C	2
All face-centred	F	4

1.5.5 Miller Indices

The crystal faces, both when they grow and when they are formed by cleavage, tend to be parallel either to the sides of the unit cell or to the planes in the crystals that contain a high density of atoms. It is useful to be able to refer to both the crystal faces and the planes in the crystals in some way—to give them a label—and this is usually done by using **Miller indices**.

First, we will describe how Miller indices are derived for lines in two-dimensional nets and then move on to look at planes in three-dimensional lattices. Figure 1.26 shows a rectangular net with several sets of lines, and a unit cell is marked on each set with the origin of each in the bottom left-hand corner corresponding to the directions of the x and y axes. A set of parallel lines are defined by two indices, h and k, where h and k are the number of parts into which a and b, the unit cell edges, are divided by the lines. Thus, the indices of a line hk are defined so that the line intercepts a at a/h and b at b/k. Start by finding a line next to the one passing through the origin. In the set of lines marked A, the line next to the one passing through the origin leaves a undivided (a/1), but divides b into two (b/2); both intercepts lie on the positive side of the origin, so in this case the indices of the set of lines hk are *12*, called the 'one-two' set. If the set of lines lies parallel to one of the axes, then there is no intercept and the index becomes zero. If the intercepted cell edge lies on the negative side of the origin, then the index is written with a bar on the top, for example, $\bar{2}$, called 'bar-two'. Notice that if we had selected the line on the other side of the origin in A, we would have indexed the lines as the $\bar{1}\,\bar{2}$; there is no difference between the two pairs of indices and always the hk and the $\bar{h}\,\bar{k}$ lines are the same set of lines. Try Question 5 for more examples. Notice also, in Figure 1.26, that the lines with the lower indices are more widely spaced.

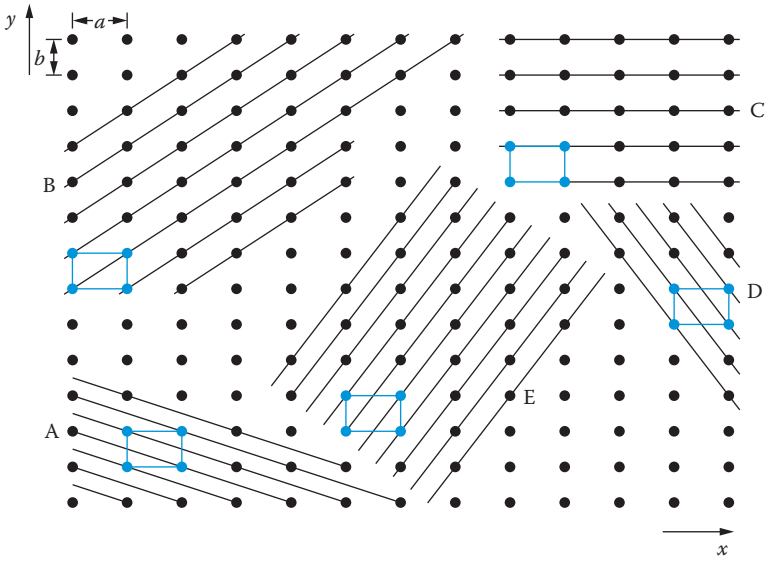


FIGURE 1.26 A rectangular net showing five sets of lines, A–E, with unit cells marked.

The Miller indices for planes in three-dimensional lattices are given by *hkl*, where *l* is now the index for the *z* axis. The principles are exactly the same. Thus, a plane is indexed *hkl* when it makes intercepts *a*/*h*, *b*/*k* and *c*/*l* with the unit cell edges *a*, *b* and *c*, respectively. Figure 1.27 shows some cubic lattices with various planes shaded. The positive directions of the axes are marked and these are orientated to conform with the conventional right-hand rule, as illustrated in Figure 1.28. In Figure 1.27a, the shaded planes lie parallel to *y* and *z*, but leave the unit cell edge *a* undivided;

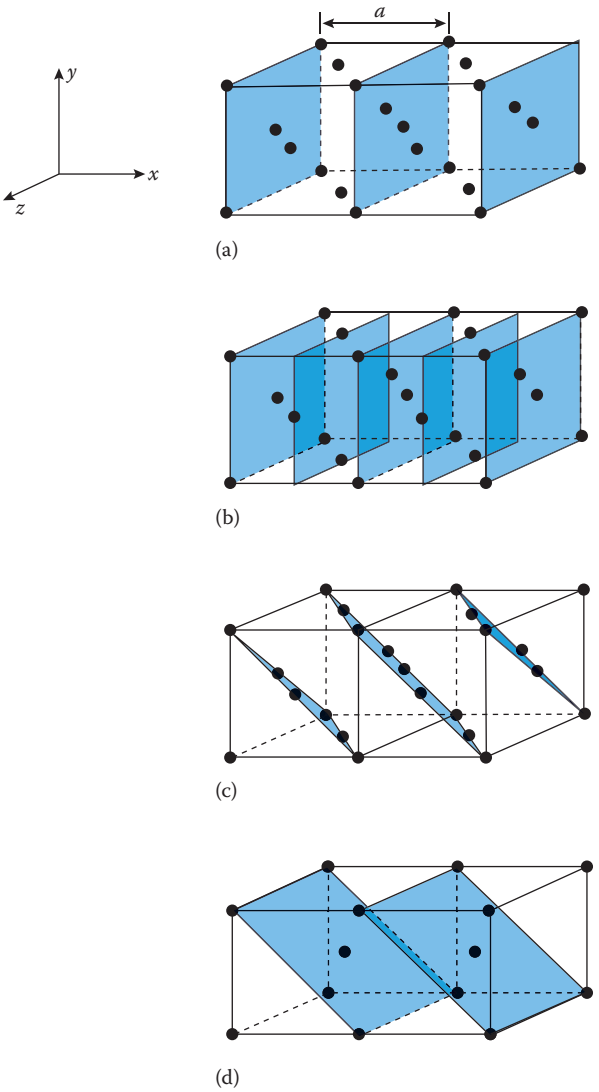


FIGURE 1.27 (a–c) Planes in a face-centred cubic lattice. (d) Planes in a body-centred cubic lattice (two unit cells are shown).

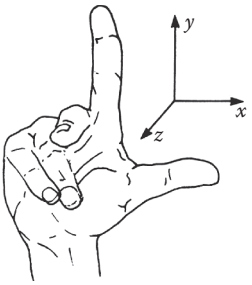


FIGURE 1.28 The right-handed rule for labelling axes.

thus, the Miller indices of these planes are *100*. Again, take note that the *hkl* and $\bar{h}\bar{k}\bar{l}$ planes are the same.

1.5.6 Interplanar Spacings

It is sometimes useful to be able to calculate the perpendicular distance d_{hkl} between parallel planes (Miller indices *hkl*). When the axes are at right angles to one another (orthogonal), the geometry is simple and for an orthorhombic system where $a \neq b \neq c$ and $\alpha = \beta = \gamma = 90°$, this gives:

$$\frac{1}{d_{hkl^2}} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2}$$

Other relationships are summarised in Table 1.5.

1.5.7 Packing Diagrams

Drawing structures in three dimensions is not easy and so crystal structures are often represented by two-dimensional plans or projections of the unit cell contents—in much the same way as an architect makes building plans. These projections are called **packing diagrams** as they are particularly useful in small molecule structures for showing how the molecules pack together in the crystal and thus their intermolecular interactions.

The position of an atom or an ion in a unit cell is described by its **fractional coordinates**; these are simply the coordinates *based on the unit cell axes* (known as the **crystallographic axes**) and are expressed as *fractions of the unit cell lengths*. It has the simplicity of a universal system that enables unit cell positions to be compared from one structure to another regardless of a variation in the unit cell size.

To take a simple example, in a cubic unit cell with $a = 1000$ pm, an atom with an x coordinate of 500 pm has a fractional coordinate of $x/a = 500/1000 = 0.5$ in the x direction. Similarly, in the y and z directions, the fractional coordinates are given by y/b and z/c, respectively.

TABLE 1.5
d-Spacings in Different Crystal Systems

Lattice System	d_{hkl} as a Function of Miller Indices and Lattice Parameters
Cubic	$\frac{1}{d^2} = \frac{h^2 + k^2 + l^2}{a^2}$
Tetragonal	$\frac{1}{d^2} = \frac{h^2 + k^2}{a^2} + \frac{l^2}{c^2}$
Orthorhombic	$\frac{1}{d^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2}$
Hexagonal	$\frac{1}{d^2} = \frac{4}{3}\left(\frac{h^2 + hk + k^2}{a^2}\right) + \frac{l^2}{c^2}$
Monoclinic	$\frac{1}{d^2} = \frac{1}{\sin^2\beta}\left(\frac{h^2}{a^2} + \frac{k^2 \sin^2\beta}{b^2} + \frac{l^2}{c^2} - \frac{2hl\cos\beta}{ac}\right)$

A packing diagram is shown in Figure 1.29 for the body-centred unit cell of Figure 1.8. The projection is shown on the *yx* plane, that is, we are looking at the unit cell straight down the *z* axis. The *x*-fractional coordinate of any atoms/ions lying in the top or bottom face of the unit cell will be 0 or 1 (depending on where you take the origin), and it is conventional for this not to be marked on the diagram. Any *z* coordinate that is not 0 or 1 is marked in a convenient place on the diagram. There is an opportunity to practice constructing these types of diagrams in the questions at the end of the chapter.

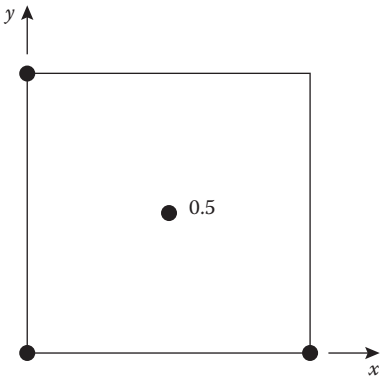


FIGURE 1.29 Packing diagram for a body-centred unit cell.

1.6 CRYSTALLINE SOLIDS

We start this section by looking at the structures of some simple **ionic solids**. Ions tend to be formed by the elements in the groups at the far left and far right of the periodic table. Thus, we expect the metals in Groups 1 (I) and 2 (II) to form cations and the nonmetals of Groups 16 (VI) and 17 (VII) and nitrogen to form anions, because by doing so they are able to achieve a stable noble gas configuration. Cations can also be formed by some of the Group 13 (III) elements, such as aluminium (Al^{3+}), by some of the low oxidation state transition metals and even occasionally by the high atomic number elements in Group 14 (IV), such as tin and lead, giving Sn^{4+} and Pb^{4+}. Each successive ionisation becomes more difficult because the remaining electrons are more strongly bound due to the greater effective nuclear charge, so highly charged ions are rather rare.

An **ionic bond** is formed between two oppositely charged ions because of the electrostatic attraction between them. Ionic bonds are strong, but are also nondirectional; their strength decreases with the increasing separation of the ions. Ionic crystals are therefore composed of infinite arrays of ions that have packed together in such a way as to maximise the coulombic attraction between oppositely charged ions and to minimise the repulsions between ions of the same charge. We expect to find ionic compounds in the halides and oxides of Group 1 and 2 metals and it is with these crystal structures that this section begins.

However, just because it is possible to form a particular ion, it does not mean that this ion will always exist whatever the circumstances. In many structures, we find that the bonding is not purely ionic, but it possesses some degree of **covalency**: the electrons are shared between the two bonding atoms and are not merely transferred from one to the other. This is particularly true for the elements in the centre of the periodic table. This point is taken up in Section 1.6.4, where we discuss the size of ions and the limitations of the concept of ions as hard spheres.

Sections 1.6.5 and 1.6.6 look at the crystalline structures of covalently bonded species. In Section 1.6.5, extended covalent arrays are investigated, such as the structure of **diamond**—one of the forms of elemental carbon—where each atom forms strong covalent bonds to the surrounding atoms, forming an infinite three-dimensional network of localised bonds throughout the crystal. In Section 1.6.6, we look at molecular crystals that are formed from small, individual, covalently bonded molecules. These molecules are held together in the crystal by weak forces known collectively as **van der Waals forces**. These forces arise due to the interactions between the dipole moments in the molecules. Molecules that possess a permanent dipole can interact with one another (**dipole–dipole interaction**) and also with ions (**charge–dipole interaction**). Molecules that do not possess a dipole also interact with each other because ‘transient dipoles’ arise due to the movement of electrons, and these in turn induce dipoles in adjacent molecules. The net result is a weak attractive force known as the **London dispersion force**, which falls off very quickly with distance.

Finally, in this section, we take a very brief look at the structures of some silicates—the compounds that largely form the Earth’s crust.

1.6.1 Ionic Solids with Formula MX

The cubic unit cell of the **caesium chloride structure (CsCl)** is shown in Figure 1.30. It shows a caesium ion (Cs^+) at the centre of the cube, surrounded by eight chloride ions (Cl^-) at the corners. It could equally well have been drawn the other way round with chloride at the centre and caesium at the corners because the structure consists of two interpenetrating primitive cubic arrays. Note the similarity of this unit cell to the body-centred cubic structure adopted by some of the elemental metals such as the Group 1 (alkali) metals. It is important to note, however, that the caesium chloride structure is *not* body-centred cubic. This is because the environment of the caesium at the centre of the cell is not the same as the environment of the chlorides at the corners; a body-centred cell would have the same type of atom at both corners, that is, at (0, 0, 0) etc., *and* at the body centre ($\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$) because they are lattice points. In CsCl, each caesium is surrounded by eight chlorines at the corners of a cube and vice versa, so the coordination number of each type of atom is eight. The unit cell contains one formula unit of CsCl, with each of the eight corner chlorines being shared by eight unit cells. With ionic structures like this, individual molecules are not distinguishable because individual ions are surrounded by ions of the opposite charge.

Caesium is a large ion (ionic radii are discussed in detail in Section 1.6.4), so it is able to coordinate the eight chloride ions around it. Other compounds with large cations that can also accommodate eight anions and crystallise with this structure include CsBr, CsI, TlCl, TlBr, TlI and NH_4Cl.

Common salt, **sodium chloride (NaCl)**, is also known as **rock salt**. It is mined all over the world from underground deposits left by the dried-up remains of ancient seas and has been so highly prized in the past that its possession has been the cause of much conflict, most notably causing the 'salt marches' organised by Gandhi and helping to spark off the French revolution. A unit cell of the sodium chloride structure is illustrated in Figure 1.31. The unit cell is cubic and the structure consists of two interpenetrating face-centred arrays: one of Na^+ and the other of Cl^- ions. Each sodium ion is surrounded by six equidistant chloride ions situated at the corners of an octahedron and, in the same way, each chloride ion is surrounded by six sodium ions: we say that the coordination is 6 : 6.

An alternative way of viewing this structure is to think of it as a cubic close-packed array of chloride ions with sodium ions filling all the octahedral holes. Filling

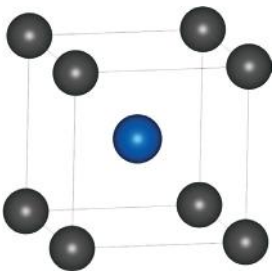


FIGURE 1.30 CsCl unit cell. Key: Cs blue; Cl grey (or vice versa).

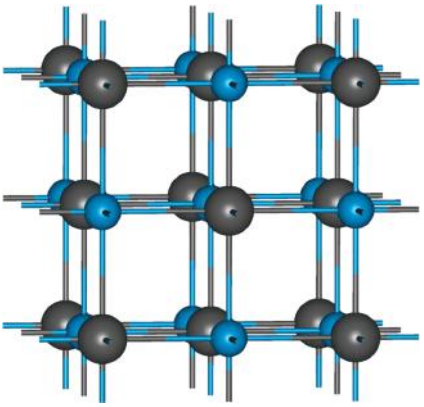


FIGURE 1.31 NaCl unit cell. Key: Na blue; Cl grey (or vice versa).

all the octahedral holes gives a Na : Cl ratio of 1 : 1 with the structure as illustrated in Figure 1.31. The conventional unit cell of a *ccp* array is an F face-centred cube (hence the cubic in *ccp*); the close-packed layers lie at right angles to a cube diagonal (see Figure 1.32). Interpreting simple ionic structures in terms of the close packing of one of the ions with the other ion filling some or all of either the octahedral or the tetrahedral holes, is extremely useful: it makes it particularly easy to see both the coordination geometry around a specific ion and the available spaces within a structure.

The coordination number of sodium (and chlorine) is lower in this structure: as you might expect from its relative position in Group 1, a sodium ion is smaller than a caesium ion, so it is now only possible to pack six chlorides around it and not eight as in caesium chloride.

The sodium chloride unit cell contains four formula units of NaCl. If you find this difficult to see, work it out for yourself by counting up the numbers of ions in the different sites and apply the information given in Table 1.4.

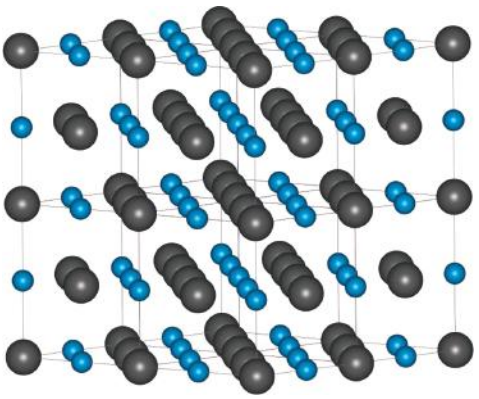


FIGURE 1.32 Close-packed layers in NaCl. Key: Na blue; Cl grey.

TABLE 1.6
Compounds with NaCl (Rock-Salt) Type of Crystal Structure

Most alkali halides, MX, and AgF, AgCl, AgBr
All alkali hydrides, MH
Monoxides, MO, of Mg, Ca, Sr, Ba
Monosulfides, MS, of Mg, Ca, Sr, Ba

Table 1.6 lists some of the compounds that adopt the NaCl structure—there are more than 200 known.

Many of the structures described in this book can be viewed as linked octahedra, where each octahedron consists of a metal atom surrounded by six other atoms situated at the corners of an octahedron (Figure 1.33a and 1.33b). These can also be depicted as viewed from above with contours marked, as in Figure 1.33c. Octahedra can link together through corners, edges and faces, as seen in Figure 1.34. The linking of octahedra by different methods effectively eliminates the atoms because some of the atoms are now shared between them: two MO_6 octahedra linked through a vertex have the formula M_2O_{11}; two MO_6 octahedra linked through an edge have the formula, M_2O_{10}; two MO_6 octahedra linked through a face have the formula, M_2O_9.

The NaCl structure can be described in terms of $NaCl_6$ octahedra sharing edges. An octahedron has 12 edges, and in the NaCl structure each one is shared by two octahedra. This is illustrated in Figure 1.35, which shows an NaCl unit cell with three $NaCl_6$ octahedra shown in outline and one of the resulting tetrahedral spaces depicted by shading.

The **nickel arsenide (NiAs) structure** is the equivalent of the sodium chloride structure in hexagonal close packing. It can be thought of as an *hcp* array of arsenic atoms with nickel atoms occupying the octahedral holes. The geometry of the nickel atoms is thus octahedral. This is not the case for arsenic, however,

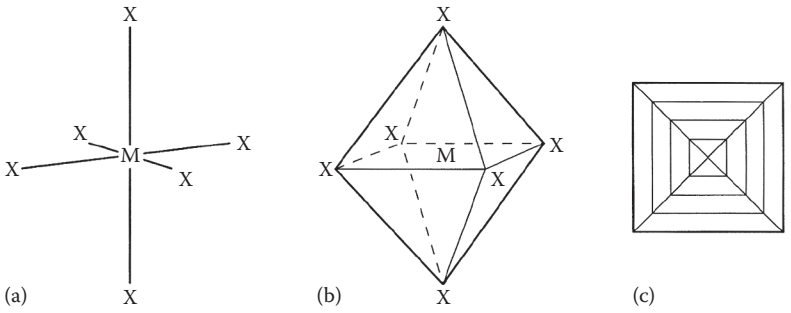


FIGURE 1.33 (a) An $[MX_6]$ octahedron, (b) a solid octahedron and (c) a plan of an octahedron with contours.

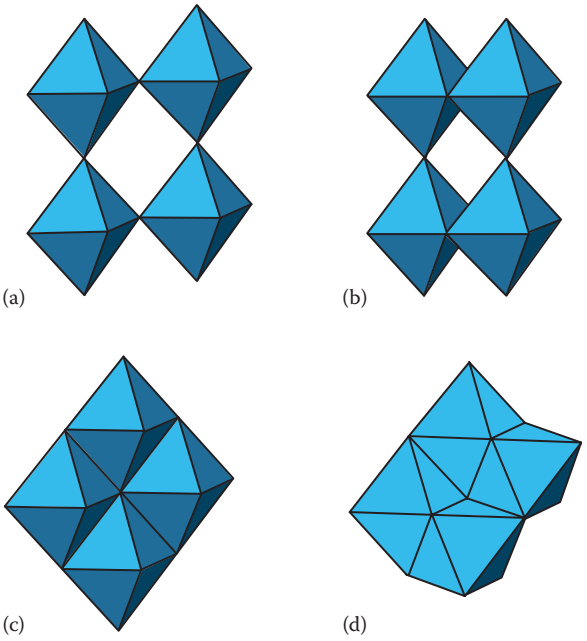


FIGURE 1.34 Conversion of (a) corner-shared MX_6 octahedra to (b) edge-shared octahedra, and (c) edge-shared octahedra to (d) face-shared octahedra.

and each arsenic atom sits at the centre of a **trigonal prism** of six nickel atoms (Figure 1.36).

The unit cells of two structures **zinc blende (or sphalerite) and wurtzite (ZnS)** are shown in Figures 1.37 and 1.38, respectively. They are named after two different naturally occurring mineral forms of zinc sulfide (ZnS). Structures of the same

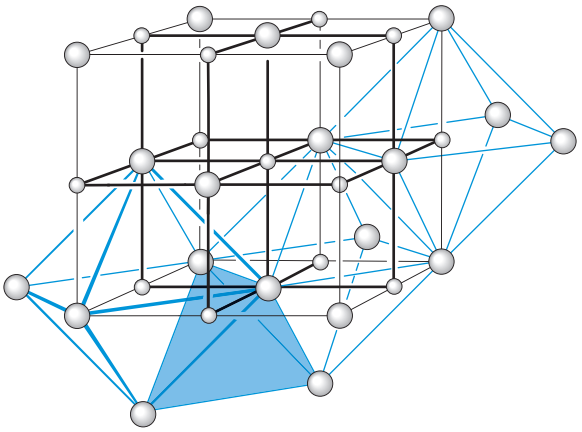
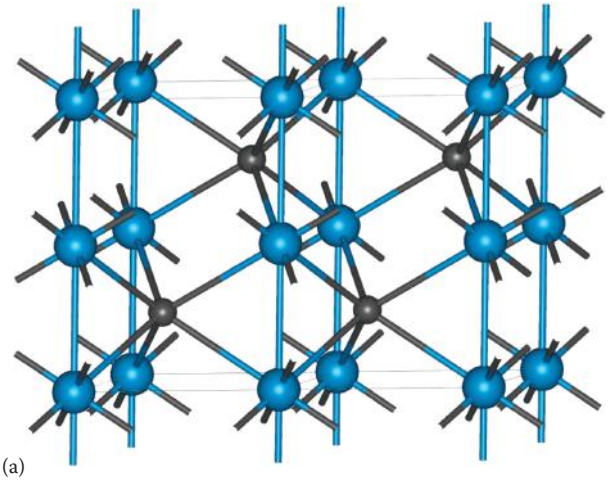


FIGURE 1.35 NaCl structure showing edge-sharing of octahedra and the enclosed tetrahedral space (shaded).



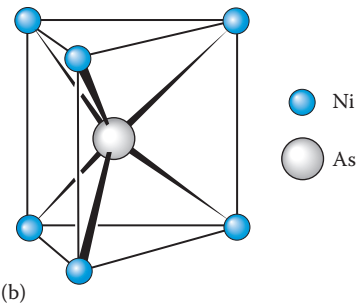


FIGURE 1.36 (a) Unit cell of nickel arsenide, NiAs. (For undistorted *hcp* $c/a = 1.633$, but this ratio is found to vary considerably.) Key: Ni, blue; As, grey. (b) The trigonal prismatic coordination of arsenic in NiAs.

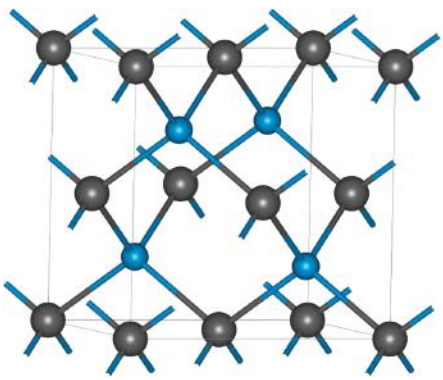


FIGURE 1.37 Crystal structure of zinc blende or sphalerite, ZnS. Key: Zn, blue; S, grey (or vice versa).

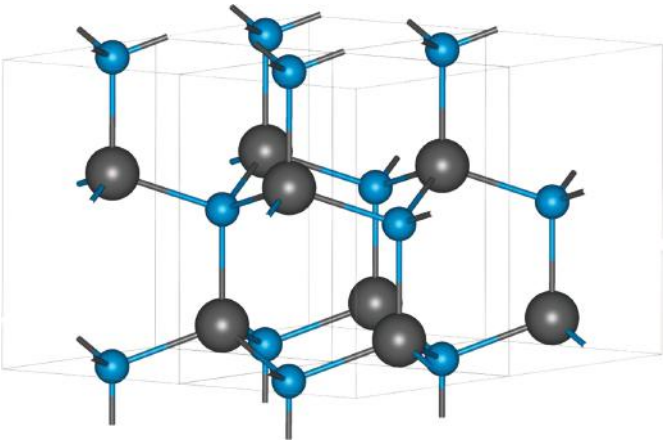


FIGURE 1.38 Crystal structure of wurtzite, ZnS. Key: Zn, blue; S, grey.

element or compound that differ only in their atomic arrangements are termed **polymorphs**. Zinc blende is often contaminated by iron, making it very dark in colour and thus lending it the name of 'Black Jack'.

The zinc blende structure can be thought of as a *ccp* array of sulfide ions with zinc ions occupying every other tetrahedral hole in an ordered manner. Each zinc ion is thus tetrahedrally coordinated by four sulfides and vice versa. Notice that if all the atoms were identical, the structure would be exactly the same as that of diamond (see Section 1.6.5). Notice also that the atomic positions are equivalent, and we could equally well generate the structure by swopping the zinc and sulfur ions. Compounds adopting this structure include the copper halides and the Zn, Cd and Hg sulfides.

The wurtzite structure is composed of an *hcp* array of sulfide ions with alternate tetrahedral holes occupied by zinc ions. Each zinc ion is tetrahedrally coordinated by four sulfide ions and vice versa. Compounds adopting the structure include BeO, ZnO and NH_4F.

Notice how the coordination numbers of the structures we have looked at so far have changed. The coordination number for close packing, where all the atoms are identical, is 12. In the CsCl structure, it is eight; in NaCl, six; and in both the ZnS structures, four. As a general rule, the larger the cation, the more anions it can pack around itself (see Section 1.6.4).

1.6.2 Solids with General Formula MX_2

The **fluorite structure** is named after the mineral form of calcium fluoride (CaF_2) which is found in the United Kingdom in the famous Derbyshire 'Blue John' mines. The structure is illustrated in Figure 1.39. It can be thought of as related to a *ccp* array of calcium ions with fluorides occupying all the tetrahedral holes. There is a problem with this as a description, however, because calcium ions are rather smaller than fluoride ions and so, physically, fluoride ions would not be able to fit into the

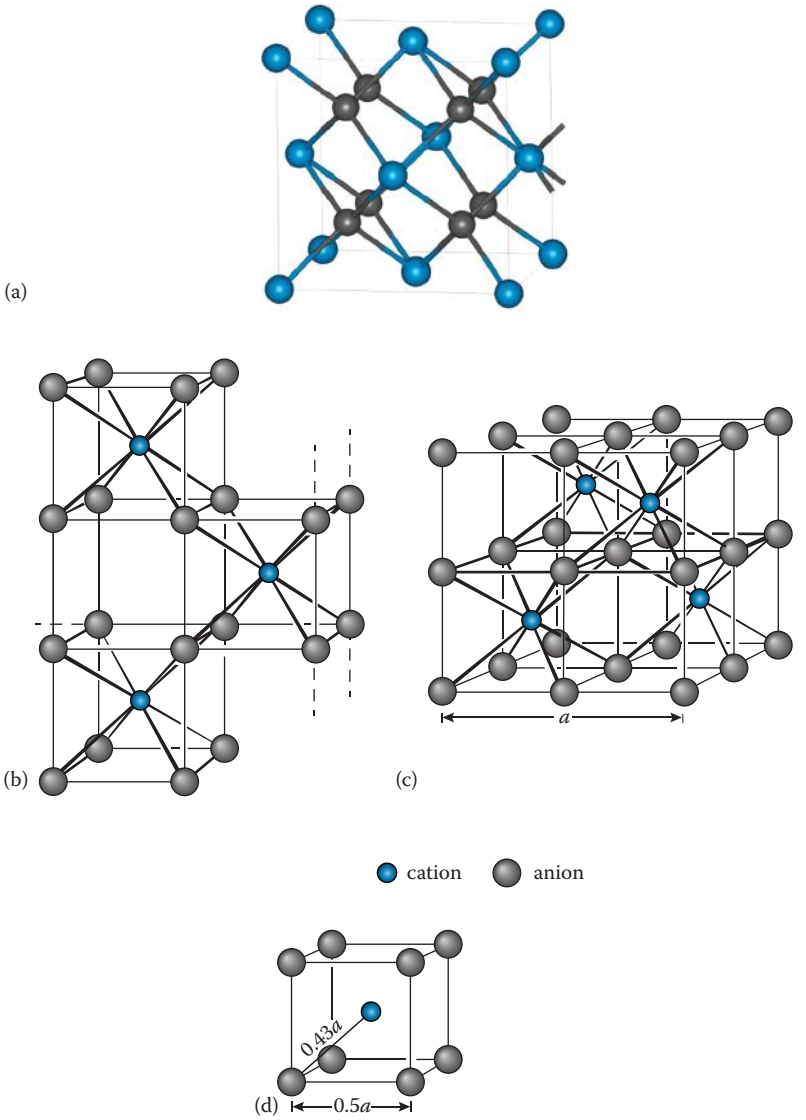


FIGURE 1.39 Crystal structure of fluorite, CaF_2. (a) Computer-generated unit cell as a *ccp* array of cations. Key: Ca, blue; F, grey. (b, c) The same structure redrawn as a primitive cubic array of anions. (d) Relationship of unit cell dimensions to the primitive anion cube (the octant).

tetrahedral holes of a calcium ion array. Nevertheless, it gives an exact description of the *relative* positions of the ions in the structure. The diagram in Figure 1.39a clearly shows the fourfold tetrahedral coordination of the fluoride ions. Notice also that the larger octahedral holes are vacant in this structure—one of them is located at the body centre of the unit cell in Figure 1.39a surrounded by the six (blue) calcium ions.

This becomes a very important feature when we come to look at the movement of ions through defect structures in Chapter 5.

By drawing small cubes with fluoride ions at each corner, as shown in Figure 1.39b, you can see that there is an eightfold cubic coordination of each calcium cation. Indeed, it is possible to move the origin and redraw the unit cell so that this feature can be seen more clearly, as shown in Figure 1.39c. The unit cell is now divided into eight smaller cubes called **octants**, with each alternate octant occupied by a calcium cation.

In the **antifluorite structure**, the positions of the cations and anions are merely reversed, and the description of the structure as cations occupying all the tetrahedral holes in a *ccp* array of anions becomes more physically realistic. In the example with the biggest anion and smallest cation, lithium telluride (Li_2Te), the tellurium ions are approximately close-packed (even though there is a considerable amount of covalent bonding). For the other compounds adopting this structure, such as the oxides and sulfides of alkali metals, M_2O and M_2S, the description accurately shows the relative positions of the atoms, but the anions could not be described as close-packed because they are not touching each other, and the cations are too big to fit in the tetrahedral holes. This means that the anion–anion distance is greater than for close packing.

These are the only structures where 8 : 4 coordination is found. Many of the fast-ion conductors are based on these structures (see Section 5.4).

Both of the **cadmium chloride and cadmium iodide (CdI_2) structures** are based on the close packing of the appropriate anions with half of the octahedral holes occupied by cations. In both structures, the cations occupy all the octahedral holes in every other close-packed anion layer, giving an overall layer structure with 6 : 3 coordination. The cadmium chloride ($CdCl_2$) structure is based on a *ccp* array of chloride ions, whereas the cadmium iodide (CdI_2) structure is based on an *hcp* array of iodide ions. The cadmium iodide structure is shown in Figure 1.40; in (a) we can see that an iodide anion is surrounded by three cadmium cations on one side but by three iodides on the other, that is, it is not completely surrounded by ions of the opposite charge, as we would expect for an ionic structure. This is evidence that the bonding in some of these structures is not entirely ionic, as we have tended to imply so far. This is a point that is taken up again in more detail in Section 1.6.4.

The **rutile structure** is named after a mineral form of titanium dioxide (TiO_2). Rutile has a very high refractive index, scattering most of the visible light incident on it; therefore, it is the most widely used white pigment in paints and plastics. A unit cell is illustrated in Figure 1.41. The unit cell is tetragonal and the structure again demonstrates 6 : 3 coordination but this time it is not based on close packing: each titanium atom is coordinated by six oxygen atoms at the corners of a (slightly distorted) octahedron and each oxygen atom is surrounded by three planar titanium ions that lie at the corners of an (almost) equilateral triangle. It is not geometrically possible for the coordination around Ti to be a perfect octahedron *and* for the coordination around O to be a perfect equilateral triangle.

The structure can be viewed as chains of linked TiO_6 octahedra, where each octahedron shares a pair of opposite edges, and the chains are linked by sharing vertices (Figure 1.41b). Figure 1.41c shows a plan of the unit cell, which is facing down the chains of octahedra so that they are seen in projection.

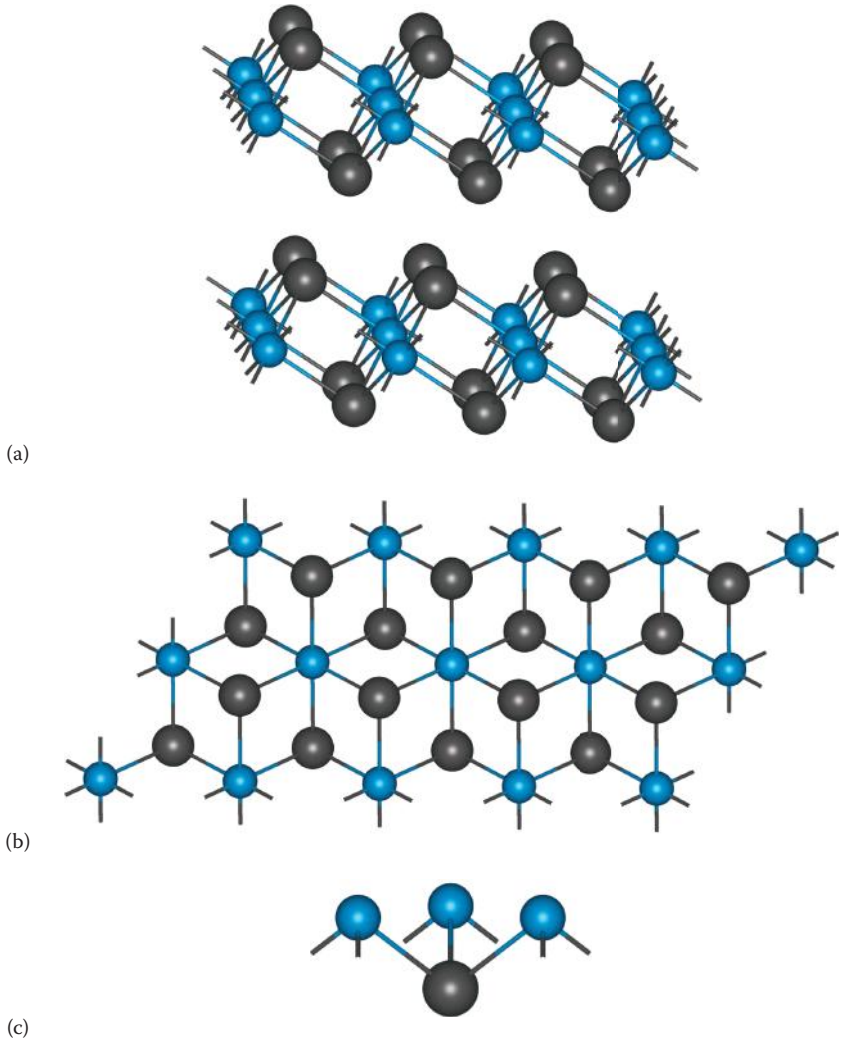


FIGURE 1.40 (a) Crystal structure of cadmium iodide, CdI_2. (b) Structure of the layers in CdI_2 and $CdCl_2$: the halogen atoms lie in planes above and below that of the metal atoms. (c) Coordination around one iodine atom in CdI_2. Key: Cd, blue; I, grey.

Occasionally, the **antirutile structure** is encountered where the metals and the nonmetals have swopped places, such as in titanium nitride (Ti_2N).

The **β-cristobalite structure** is named after a mineral form of silicon dioxide (SiO_2). The silicon atoms are in the same positions as both the zinc and the sulfur ions in zinc blende (or the carbon atoms in diamond, which we look at in Section 1.6.5), and each pair of silicon atoms is bonded to an oxygen atom midway between them. The only metal halide adopting this structure is beryllium

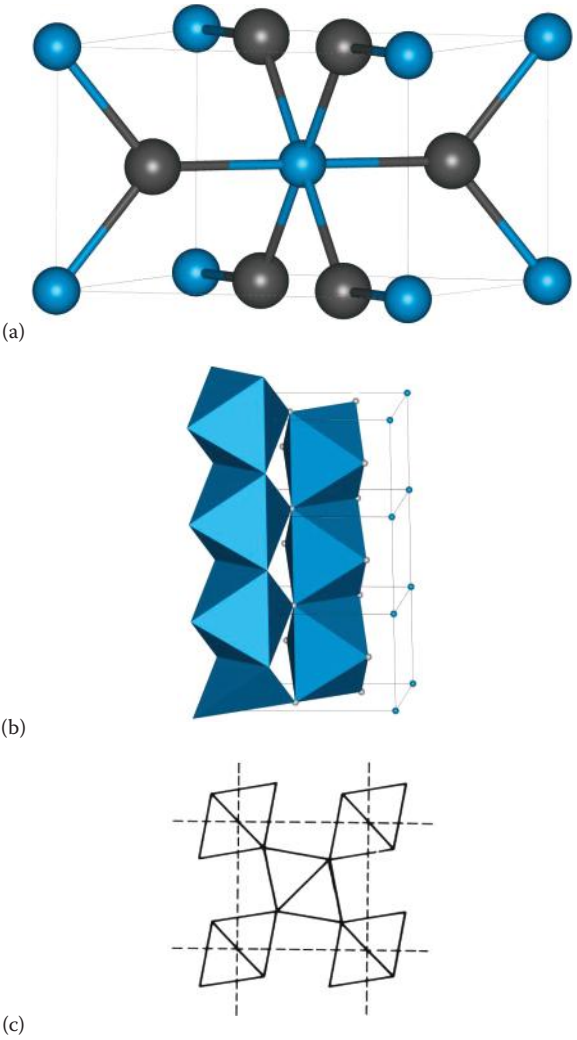


FIGURE 1.41 Crystal structure of rutile, TiO_2. (a) Unit cell, (b) parts of two chains of linked $[TiO_6]$ octahedra and (c) projection of structure on base of unit cell. Key: Ti, blue; O, grey.

fluoride (BeF_2), which is characterised by 4 : 2 coordination, with Be tetrahedrally coordinated by F.

1.6.3 Other Important Crystal Structures

As the valency of the metal increases, the bonding in these simple binary compounds becomes more covalent and the highly symmetrical structures characteristic of the simple ionic compounds occur far less frequently, with molecular and layer structures being common. There are many thousands of inorganic crystal structures;

here, we briefly describe just a few of those that are commonly encountered and those that occur in later chapters of this book.

The **bismuth triiodide structure (BiI_3)** is based on an *hcp* array of iodides with the bismuth atoms occupying one third of the octahedral holes. Alternate pairs of layers have two thirds of the octahedral sites occupied.

The **corundum (α-Al_2O_3)** mineral is the basis for ruby and sapphire gemstones, with their colour depending on the impurities. It is very hard—second only to diamond. This structure can be described as an *hcp* array of oxygen atoms with two thirds of the octahedral holes occupied by aluminium atoms. As we have seen before,

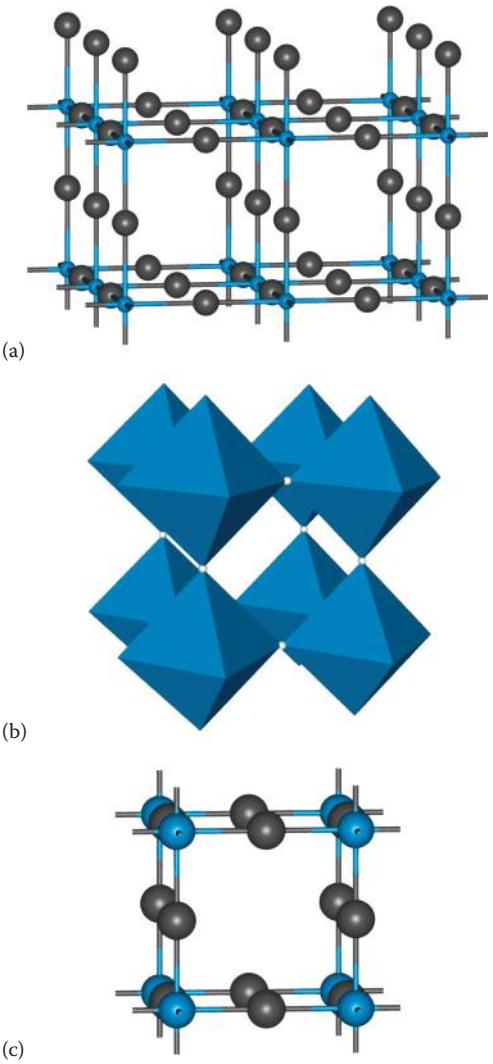


FIGURE 1.42 (a) Part of the ReO_3 structure, (b) ReO_3 structure showing the linking of $[ReO_6]$ octahedra and (c) unit cell. Key: Re, blue; O, grey.

geometrical constraints dictate that octahedral coordination of the aluminium atoms precludes tetrahedral coordination of the oxygen atoms. However, it is suggested that this structure is adopted in preference to other possible ones because the four aluminium atoms surrounding an oxygen atom approximate most closely to a regular tetrahedron. The structure is also adopted by Ti_2O_3, V_2O_3, Cr_2O_3, α-Fe_2O_3, α-Ga_2O_3 and Rh_2O_3.

The **rhenium trioxide structure (ReO_3)** structure (also called the aluminium fluoride structure) is adopted by the fluorides of Al, Sc, Fe, Co, Rh and Pd and also by the oxides WO_3 (at high temperature) and ReO_3 (see Section 5.8.1). The structure consists of ReO_6 octahedra linked together through *each* corner to give a highly symmetrical three-dimensional network with cubic symmetry. Part of the structure is shown in Figure 1.42a, the linking of the octahedra in Figure 1.42b and the unit cell in Figure 1.42c.

There are three important **mixed oxide structures: spinel, perovskite and ilmenite**.

The **spinels** have the general formula AB_2O_4, taking their name from the mineral spinel $MgAl_2O_4$, where, generally, A is a divalent ion (A^{2+}) and B is trivalent (B^{3+}). The structure can be thought of as being based on a cubic close-packed array of oxide ions, with A^{2+} ions occupying tetrahedral holes and B^{3+} ions occupying octahedral holes. A spinel crystal containing $n$$AB_2O_4$ formula units has $8n$ tetrahedral holes and $4n$ octahedral holes, so, accordingly, one eighth of the tetrahedral holes are occupied by the A^{2+} ions and one half of the octahedral holes by the B^{3+} ions. The face-centred unit cell is illustrated in Figure 1.43. The A ions are tetrahedrally coordinated, whereas the B ions occupy octahedral sites. Spinels with this structure include compounds of the formula MAl_2O_4, where M is Mg, Fe, Co, Ni, Mn or Zn.

When compounds of the general formula AB_2O_4 adopt the **inverse-spinel structure**, the formula is better written as $B(AB)O_4$, because this indicates that half of

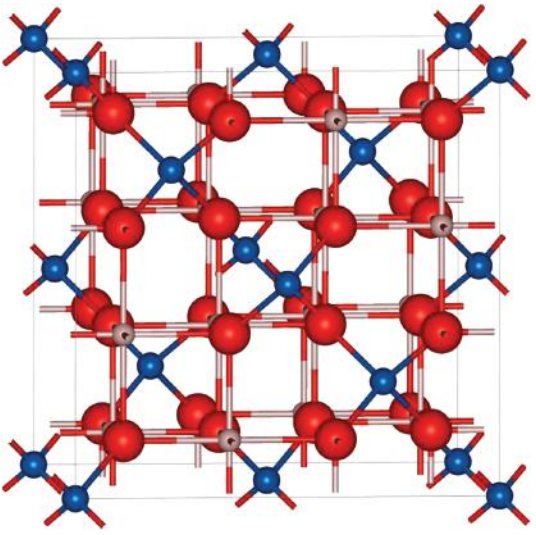


FIGURE 1.43 Spinel structure, $CuAl_2O_4$ (AB_2O_4). Key: Cu, blue; Al, mauve; O, red.

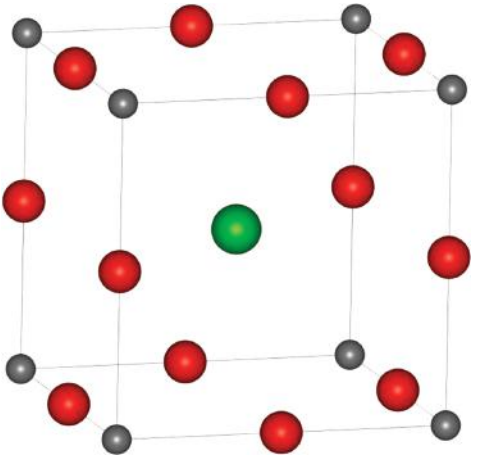


FIGURE 1.44 Perovskite structure of compounds ABX_3, such as $CaTiO_3$. Key: Ca, green; Ti, silver; O, red.

the B^{3+} ions now occupy tetrahedral sites, and the remaining half, together with the A^{2+} ions, occupy the octahedral sites. Examples of inverse spinels include magnetite, Fe_3O_4 (see Section 9.7), $Fe(MgFe)O_4$ and $Fe(ZnFe)O_4$.

The **perovskite structure** is named after the mineral $CaTiO_3$. A unit cell is shown in Figure 1.44: this unit cell is termed the A-type, because if we take the general formula ABX_3 for the perovskites, then in this cell, the A atom is at the centre. The central Ca (A) atom is coordinated to 8 Ti atoms (B) at the corners and to 12 oxygens (X) at the midpoints of the cell edges. The structure can be usefully described in other ways. First, perovskite has the same octahedral framework as ReO_3 based on BX_6 octahedra with an A atom added in at the centre of the cell (see Figure 1.42b). Secondly, it can be thought of as a *ccp* array of A and X atoms, with the B atoms occupying the octahedral holes (compare with the unit cell of NaCl in Figure 1.31 to check this). Compounds adopting this structure include $SrTiO_3$, $SrZrO_3$, $SrHfO_3$, $SrSnO_3$ and $BaSnO_3$. The structures of the high-temperature superconductors are based on this structure (see Chapter 9).

The **ilmenite structure** is adopted by oxides of the formula ABO_3, where A and B are similar in size and their charges add up to a total charge of +6. The structure is named after the mineral $Fe^{II}Ti^{IV}O_3$, and is very similar to the corundum structure described above—an *hcp* array of oxygen atoms—but now there are two different cations present, occupying two thirds of the octahedral holes.

The structures related to close packing are summarised in Table 1.7.

1.6.4 Ionic Radii

We know from quantum mechanics that atoms and ions do not have precisely defined radii. However, from the foregoing discussion of ionic crystal structures, we have seen that ions pack together in an extremely regular fashion in crystals,

TABLE 1.7
Structures Related to Close-Packed Arrangements of Anions

Formula	Cation: Anion Coordination	Type and Number of Holes Occupied	Examples	
			Cubic Close Packing	Hexagonal Close Packing
MX	6:6	All octahedral	Sodium chloride: NaCl, FeO, MnS, TiC	Nickel arsenide: NiAs, FeS, NiS
	4:4	Half tetrahedral—every alternate site occupied	Zinc blende: ZnS, CuCl, γ-AgI	Wurtzite: ZnS, β-AgI
MX_2	8:4	All tetrahedral	Fluorite: CaF_2, ThO_2, ZrO_2, CeO_2	None
	6:3	Half octahedral; alternate layers have fully occupied sites	Cadmium chloride: $CdCl_2$	Cadmium iodide: CdI_2, TiS_2
MX_3	6:2	One-third octahedral; alternate pairs of layers have two thirds of the octahedral sites occupied		Bismuth iodide: BiI_3, $FeCl_3$, $TiCl_3$, VCl_3
M_2X_3	6:4	Two-thirds octahedral		Corundum: α-Al_2O_3, α-Fe_2O_3, V_2O_3, Ti_2O_3, α-Cr_2O_3
ABO_3		Two-thirds octahedral		Ilmenite: $FeTiO_3$
AB_2O_4		One-eighth tetrahedral and one-half octahedral	Spinel: $MgAl_2O_4$ inverse spinel: $MgFe_2O_4$, Fe_3O_4	Olivine: Mg_2SiO_4

and that their atomic positions and thus, their interatomic distances can be measured very accurately. It is a very useful concept, therefore, particularly for those structures based on close packing, to think of ions as hard spheres, each with a particular radius.

If we take a series of alkali metal halides, all with the rock-salt structure, if we replace one metal ion with another, say sodium with potassium, we would expect the metal–halide internuclear distance to change by the same amount each time if the concept of an ion as a hard sphere with a particular radius holds true. Table 1.8 shows the results of this procedure for a range of alkali halides, with the change in internuclear distance on swopping one ion for another, emboldened.

From Table 1.8, we can see that the change in internuclear distance upon changing the ion is not constant, and also that the variation is not great. This provides us with an experimental basis that it is not unreasonable to approximate the ions as having fixed radii, despite the picture being not precisely true. We would expect a model based on hard spheres to be only an approximation, because atoms and ions are squashable entities, they are polarisable and their shape and size will be affected

TABLE 1.8
Interatomic Distances of Some Alkali Halides, r_{M-X} (pm)

	F^-		Cl^-		Br^-		I^-
Li^+	201	**56**	257	**18**	275	**27**	302
	30		**24**		**23**		**21**
Na^+	231	**50**	281	**17**	298	**25**	323
	35		**33**		**31**		**30**
K^+	266	**48**	314	**15**	329	**24**	353
	16		**14**		**14**		**13**
Rb^+	282	**46**	328	**15**	343	**23**	366

by their environment. Nevertheless, it is a useful concept to develop a bit further as it enables us to describe some of the ionic crystal structures in a simple pictorial way.

There have been many suggestions as to how individual ionic radii can be assigned, and the literature contains several different sets of values, each set being named after the person(s) who originated the method of determining the radii. We briefly describe some of these methods before listing the values most commonly used at present. It is most important to remember that we must not mix radii from more than one set of values, because, even though the values vary considerably from one set to another, each set is internally consistent, that is, if we add together two radii from one set of values, we will obtain an approximately correct internuclear distance as determined from the crystal structure.

The internuclear distances can be determined by X-ray crystallography, and in order to obtain values for individual ionic radii from these, the value of one radius needs to be fixed by some method. Originally, in 1920, Landé suggested that in the alkali halide with the largest anion and smallest cation—LiI—we can assume that the iodide ions must be in contact with each other and that the tiny Li^+ ion is inside the octahedral hole. As the Li-I distance is known, it is then a matter of simple geometry to determine the iodide radius. Once the iodide radius is known, and assuming it stays constant, then the radii of the metal cations can be found from the structures of the other metal iodides. Bragg and Goldschmidt later extended the list of ionic radii using similar methods. It is very difficult to come up with a consistent set of values for the ionic radii because, of course, the ions are *not* hard spheres, they are somewhat elastic and their radii are affected by their environment, such as the nature of the oppositely charged ligand and the coordination number. Pauling proposed a theoretical method of calculating the radii from the internuclear distances, and he produced a set of values that are internally consistent and also show the expected trends in the periodic table. Pauling's method was to take a series of alkali halides with isoelectronic cations and anions and assume that they are in contact with each other; if we assume that each radius is inversely proportional to the effective nuclear charge felt by the outer electrons of the ion, the radius for each ion can be calculated from the internuclear distance. Divalent ions undergo additional compression in a lattice and compensation has to be made for this effect in calculating their radii.

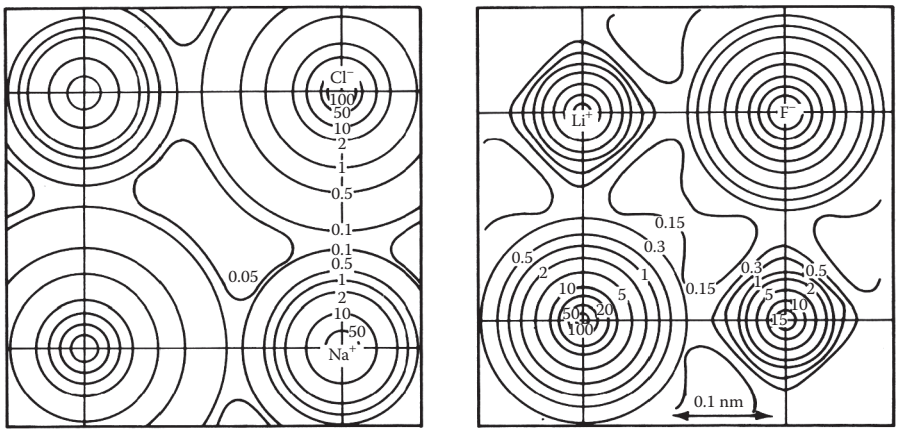


FIGURE 1.45 Electron density maps for (a) NaCl and (b) LiF.

With some refinements, this method gave a consistent set of values that were widely used for many years; they are usually known as **effective ionic radii**.

It is also possible to determine accurate electron density maps for the ionic crystal structures using X-ray crystallography. Such a map is shown for NaCl and LiF in Figure 1.45. The electron density contours fall to a minimum—although, note, not to zero—in between the nuclei, and it has been suggested that this minimum position should be taken as the radius position for each ion. These experimentally determined ionic radii are often called **crystal radii**; the values are somewhat different from the older sets and tend to make the anions smaller and the cations bigger than previously calculated. Shannon and Prewitt have compiled the most comprehensive set of radii, using data from almost a thousand crystal structure determinations and based on the conventional values of 126 and 119 pm for the radii of the O^{2-} and F^- ions, respectively. These values differ by a constant factor of 14 pm from the older traditional values, but it is generally accepted that they correspond more closely to the actual physical sizes of ions in a crystal. A selection of these data is shown in Table 1.9.

Several important trends in the sizes of ions can be noted from the data in Table 1.9:

1. The radii of ions within a group of the periodic table, such as the alkali metals, increase with atomic number, Z: as you go down the group, there are more electrons and the outer ones are further from the nucleus.
2. In a series of isoelectronic cations, such as Na^+, Mg^{2+} and Al^{3+}, the radii decrease rapidly with increasing positive charges. The number of electrons is constant, but the nuclear charge increases pulling the electrons in, and the radii decrease.
3. For pairs of isoelectronic anions, for example, F^- and O^{2-}, the radii increase with increasing charges because the more highly charged ion has a smaller nuclear charge.

TABLE 1.9
Crystal Radii/pm

						H										He
						126(−1)										—
Li	Be											B	C	N	O	F	Ne
90(+1)	59(+2)											41(+3)	30(+4)	132(−3)[a]	126(−2)	119(−1)	—
														30(+3)			
Na	Mg											Al	Si	P	S	Cl	Ar
116(+1)	86(+2)											68(+3)	54(+4)	58(+3)	170(−2)	167(−1)	—
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
152(+1)	114(+2)	89(+3)	100(+2)	93(+2)	87/94(+2)[b]	81/97(+2)[b]	75/92(+2)[b]	79/89(+2)[b]	83(+2)	91(+1)	88(+2)	76(+3)	87(+2)	72(+3)	184(−2)	182(−1)	—
			81(+3)	78(+3)	76(+3)	72/79(+3)[b]	69/79(+3)[b]	69/75(+3)[b]	70/74(+3)[b]	87(+2)			67(+4)				
			75(+4)							68(+3)							
Rb	Sr	Y	Zr							Ag	Cd	In	Sn	Sb	Te	I	Xe
166(+1)	132(+2)	104(+3)	86(+4)							129(+1)	109(+2)	94(+3)	136(+2)[c]	90(+3)	207(−2)	206(−1)	62(+8)
													83(+4)				
Cs	Ba	Lu	Hf							Au	Hg	Tl	Pb	Bi	Po	At	Rn
181(+1)	149(+2)	100(+3)	85(+4)							151(+1)	116(+2)	103(+3)	133(+2)	117(+3)	—	—	—
													92(+4)				
Fr	Ra																
194(+1)	—																

Source: Values taken from Shannon, R.D. (1976) *Acta Crystallographica*, A32, 751, and are for octahedral coordination except where noted. The figures in parentheses indicate the charge on the ion.

[a] Tetrahedral.

[b] High spin/low spin.

[c] Eight coordination.

4. For elements with more than one oxidation state, for example, Ti^{2+} and Ti^{3+}, the radii decrease as the oxidation state increases: in this case, the nuclear charge stays the same but the number of electrons that it acts on decreases.
5. As you move across the periodic table, for a series of similar ions, such as the first row transition metal divalent ions (M^{2+}), there is an overall decrease in radii. This is due to an increase in the nuclear charge across the table because electrons in the same shell do not screen the nucleus from each other very well. A similar effect is seen for the M^{3+} ions of the lanthanoids and this is known as the **lanthanoid contraction**.
6. For transition metals, the spin state affects the ionic radii.
7. The crystal radii increase with an increase in the coordination number—see examples Cu^{+} and Zn^{2+} in Table 1.9. One can think of fewer ligands around the central ion as allowing the counterions to compress the central ion.

The picture of ions as hard spheres works best for fluorides and oxides, both of which are small, fairly uncompressible ions. As the ions get larger, they are more easily compressed—the electron cloud is more easily distorted—and they are said to be more **polarisable**.

When we were discussing particular crystal structures in the previous section, we noted that a larger cation, such as Cs^{+}, was able to pack eight chloride ions around itself, whereas the smaller Na^{+} could accommodate only six. If we continue to think of ions as hard spheres, for a particular structure, as the ratio of the cation and anion radii changes, there will come a point when the cation is so small that it will no longer be in contact with the anions. The lattice would not be stable in this state because the negative charges would be too close together for comfort and we would predict that the structure would change to one of lower coordination, allowing the anions to move further apart. If the ions are hard spheres, using simple geometry, it is possible to quantify the radius ratio (i.e., $r_{cation}/r_{anion} = r_+/r_-$) at which this happens, and this is illustrated for the octahedral case in Figure 1.46.

Taking a plane through the centre of an octahedrally coordinated metal cation, the stable situation is shown in Figure 1.46a, and the limiting case for stability, when the anions are touching each other, is shown in Figure 1.46b. The anion radius (r_-) in Figure 1.46b, is OC, and the cation radius (r_+) is (OA − OC). From the geometry of the right-angled triangle, we can see that $\cos 45° = OC/OA = 0.707$. The radius ratio (r_+/r_-) is given by $(OA - OC)/OC = ((OA/OC) - 1) = (1.414 - 1) = 0.414$. Using similar calculations, it is possible to calculate the limiting ratios for the other geometries: these are summarised in Table 1.10.

On this basis, we would expect to be able to use the ratio of ionic radii to predict possible crystal structures for a given substance. But does it work? Unfortunately, only about 50% of the structures are correctly predicted. This is because the model is too simplistic—ions are *not* hard spheres, but are polarised under the influence of other ions. In larger ions, the valence electrons are further away from the nucleus and also shielded from its influence by the inner core electrons, so the electron cloud is more easily distorted. The ability of an ion to distort an electron cloud—its polarising power—is greater for small ions with high charge, and the distortion of the electron cloud means that the bonding between two such ions becomes more directional in character. This means that

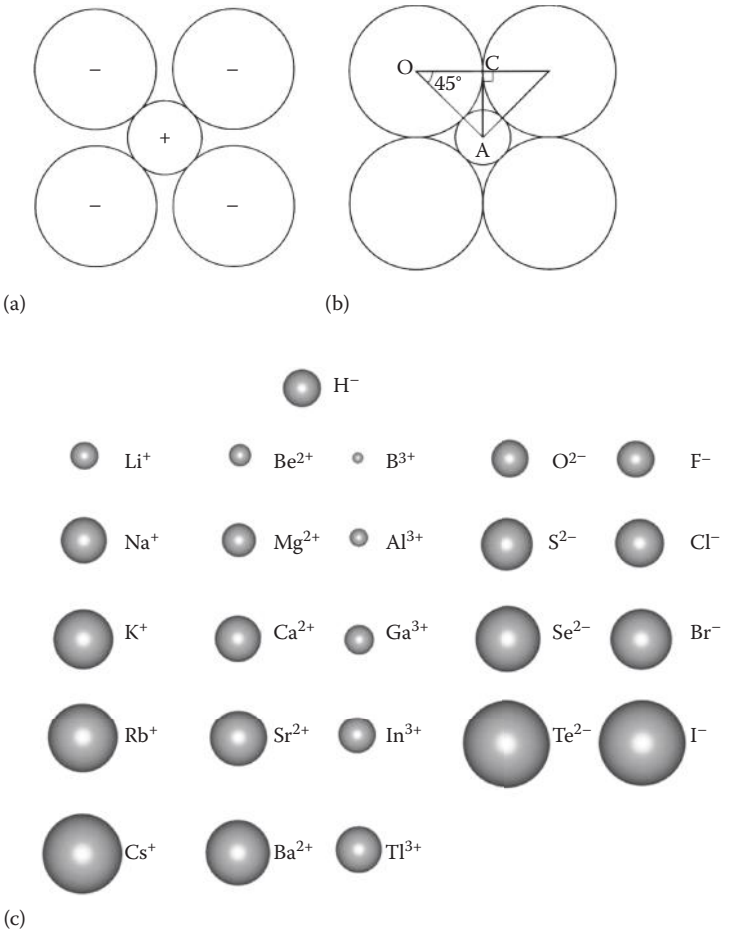


FIGURE 1.46 (a) Anions packed around a cation on a horizontal plane, (b) anion–anion contact on a horizontal plane through an octahedron and (c) relative sizes of the typical element ions.

in these cases the bonding involved is rarely truly ionic but frequently involves at least some degree of covalency. The higher the formal charge on a metal ion, the greater is the proportion of covalent bonding between the metal and its ligands. The higher the degree of covalency, the less likely is the concept of ionic radii and radius ratios to work. It also seems that there is little energy difference between the six-coordinate and eight-coordinate structures and that the six-coordinate structure is usually preferred. In fact, eight-coordinate structures are rarely found, for instance there are *no* eight-coordinate oxides. The preference for the six-coordinate rock-salt structure is thought to be due to the covalent bonding contribution: in this structure, the three orthogonal p orbitals lie in the same direction as the vectors joining the cation to the surrounding six anions, so they are well placed for good overlap of the orbitals necessary for σ bonding to take place. The potential overlap of the p orbitals in the caesium chloride structure is less favourable.

TABLE 1.10
Limiting Radius Ratios for Different Coordination Numbers

Coordination Number	Geometry	Limiting Radius Ratio	Possible Structures
		0.225	
4	Tetrahedral		Wurtzite, zinc blende
		0.414	
6	Octahedral		Rock salt, rutile
		0.732	
8	Cubic		Caesium chloride, fluorite
		1.00	

1.6.5 Extended Covalent Arrays

In the last section, we noted that many 'ionic' compounds in fact possess some degree of covalency in their bonding. As the formal charge on an ion becomes greater, we expect the degree of covalency to increase, and so we would generally expect compounds of elements in the centre of the periodic table to be covalently bonded. Indeed, some of these elements themselves are covalently bonded solids at room temperature. Examples include elements such as Group 3 (III), boron; Group 14 (IV), carbon, silicon and germanium; Group 15 (V), phosphorus and arsenic; and Group 16 (VI), selenium tellurium, form **extended covalent arrays** in their crystal structures.

Take, for instance, one of the forms of carbon—**diamond**. Diamond has a cubic crystal structure with an F-centred lattice (Figure 1.47); the positions of the atomic centres are the same as in the zinc blende structure, with carbon now occupying both the zinc and the sulfur positions. Each carbon is equivalent and is tetrahedrally coordinated to four others, forming a covalently bonded **giant molecule** throughout

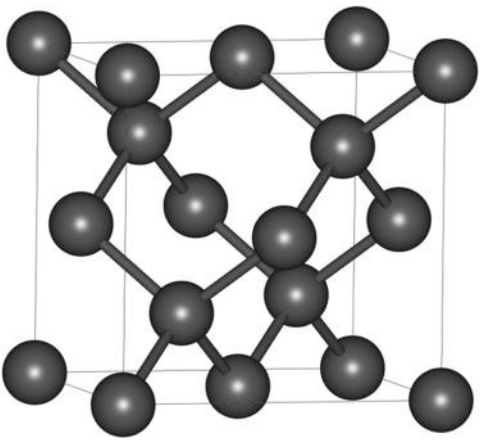


FIGURE 1.47 A unit cell of the diamond structure.

the crystal. The carbon–carbon distances are all identical (154 pm). It is interesting to note how the different types of bonding have affected the coordination number, as here we have identical atoms, but now restricted to four because this is the maximum number of covalent bonds that carbon can form. In the case of a metallic element, such as magnesium forming a crystal, the structure is close-packed with each atom 12-coordinated (bonding in metals is discussed in Chapter 4). The covalent bonds in diamond are strong, and the rigid three-dimensional network of atoms makes diamond the hardest substance known; it also has a high melting temperature (m.t.) (3773 K). Silicon carbide (SiC), known as carborundum, also has this structure with silicon atoms and carbon atoms alternating throughout; it, too, is very hard and is used for polishing and grinding.

Silica (SiO_2) gives us other examples of giant molecular structures. There are two crystalline forms of silica at atmospheric pressure—**quartz** and **cristobalite**. Each of these also exists in low- and high-temperature forms, α and β, respectively. We have already discussed the structure of β-cristobalite in terms of close packing in Section 1.6.2. Quartz is commonly encountered in nature. The structure of β-quartz is illustrated in Figure 1.48, and it consists of SiO_4 tetrahedra linked so that each oxygen atom is shared by two tetrahedra, thereby giving the overall stoichiometry of SiO_2. Notice how, once again, the covalency of each atom dictates the coordination around itself, silicon having four bonds and oxygen two, rather than the larger coordination numbers that are found for metallic and some ionic structures. Quartz is unusual in that the linked tetrahedra form spirals or **helices** throughout the crystal, which are all either left- or right-handed, producing laevorotatory or dextrorotatory crystals, respectively; these are known as **enantiomorphs**.

For our final example in this section, we will look at the structure of another polymorph of carbon. Normal **graphite** is illustrated in Figure 1.49a (there are other more complex graphite structures). The structure of normal graphite consists of two-dimensional layers of carbon atoms joined together in a hexagonal array. Within the layers, each carbon atom is strongly bonded to three others at a distance of 142 pm. This carbon–carbon distance is rather shorter than the one observed in diamond, due

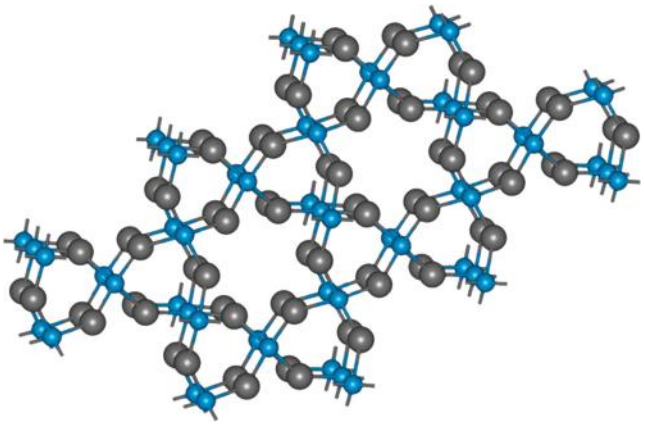


FIGURE 1.48 β-Quartz structure. Key: Si, blue; O, grey.

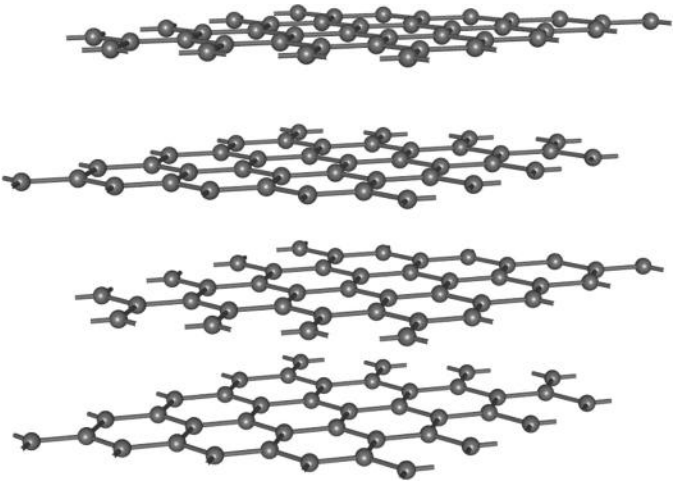


FIGURE 1.49 Crystal structure of graphite.

to the presence of some double bonding (the planar hexagonal configuration of carbon atoms puts some 2p orbitals in a suitable position for π overlap and is thus able to form a delocalised orbital in a similar fashion to the π bonding in benzene, where the carbon–carbon distance is 139.7 pm). The distance *between* the layers is much greater at 340 pm; this is indicative of the weak bonding between the layers due to van der Waals forces. Graphite is a soft grey solid with a high m.t. and low density; its softness is attributed to the weak bonding between the layers, which allows them to slide over one another. The weak bonding and large distance between the layers also explain the low density of graphite. It is also found that graphite crystals shear easily, parallel to the layers. It is a popular myth, however, that this ease of shearing makes graphite a useful lubricant. In fact, its lubricant properties are dependent on an adsorbed layer of nitrogen and when this is lost under extreme conditions such as low pressure or high temperature, then the lubricant properties are also lost. When graphite is used as a lubricant under high vacuum conditions, such as in space, then surface additives have to be incorporated to maintain the low-friction properties. The electrons in the delocalised π orbital are mobile, so graphite electrically conducts in the layers like a two-dimensional metal, but it is a poor conductor perpendicular to the layers.

1.6.6 Bonding in Crystals

Although bonding in solids will be dealt with in detail in Chapter 4, at this point it is convenient to summarise the different types of bonding that we have met so far in crystal structures. In Section 1.2, we considered the structures of metallic crystals held together by *metallic bonding*. In Section 1.6, we looked at structures such as NaCl and CsCl, which have *ionic bonding*, and later saw the influence of *covalent bonding* in the layer structures of $CdCl_2$ and CdI_2. In the graphite structure, we saw covalently bonded layers of carbon atoms held together by weak *van der Waals*

forces, which we will discuss again, together with *hydrogen bonding*, in Section 1.6.8 on molecular crystal structures.

A simple picture of **metallic bonding** is that metals consist of a regular array of metal cations surrounded by a 'sea' of electrons. These electrons occupy the space between the cations, binding them together, but they are able to move under the influence of an external field, thus accounting for the electrical conductivity of metals.

An **ionic bond** forms between two oppositely charged ions due to the electrostatic attraction between them. The attractive force (F) is given by **Coulomb's law**, $F \propto [(q_1 q_2)/r^2]$, where q_1 and q_2 are the charges on the two ions and r is the distance between them. A similar but opposite repulsive force is experienced by two ions of the *same* charge. Ionic bonds are strong and nondirectional; the energy of the interaction is given by *force* × *distance* and is inversely proportional to the separation of the charges, *r.* Ionic forces are effective over large distances compared with other bonding interactions. Ions pack together in regular arrays in ionic crystals, in such a way as to maximise coulombic attraction and minimise repulsions.

In **covalent bonding**, the electrons are shared between two atoms, resulting in a build-up of electron density between the atoms. Covalent bonds are strong and directional.

In a covalent bond, electronegative elements such as oxygen and nitrogen attract an unequal share of the bonding electrons, such that one end of the bond acquires a partial negative charge (δ−) and the other end acquires a partial positive charge (δ+). The separation of the negative and positive charge creates an **electric dipole**, and the molecule can align itself in an electric field. Such molecules are said to be **polar** and possess a permanent **dipole moment**. The partial electric charges on polar molecules can attract one another in a **dipole–dipole interaction**. The dipole–dipole interaction is about 100 times weaker than ionic interactions and falls off quickly with distance, as a function of $1/r^3$.

Polar molecules can also interact with ions in a **charge–dipole interaction**, which is about 10–20 times weaker than ion–ion interactions and which decreases with distance as $1/r^2$.

Even if molecules do not possess a permanent dipole moment, weak forces can exist between them. The movement of the valence electrons creates 'transient dipoles', and these, in turn, induce dipole moments in the adjacent molecules. The transient dipole in one molecule can be attracted to the transient dipole in a neighbouring molecule, and the result is a weak, short-range attractive force, known as the **London dispersion force**. These dispersion forces drop off rapidly with distance, decreasing as a function of $1/r^6$.

The weak nonbonded interactions that occur between molecules are often collectively referred to as **van der Waals forces**.

In one special case, polar interactions are strong enough for them to be exceptionally important in dictating the structure of the solid and liquid phases. Where hydrogen is bonded to a very electronegative element, such as oxygen or fluorine, there is a partial negative charge (δ−) on the electronegative element and an equal and opposite charge (δ+) on the hydrogen. The positively charged $H^{\delta+}$ can also be attracted to the partial negative charge on a neighbouring molecule, forming a weak bond, known as a **hydrogen bond**, O–H-----O, pulling the three atoms almost into a

straight line. A network of alternating weak and strong bonds is built up, for example, as can be seen in water (H_2O) and in hydrogen fluoride (HF). The longer, weaker hydrogen bonds can be thought of as dipole–dipole interactions and are particularly important in biological systems and in any crystals that contain water.

1.6.7 Atomic Radii

An atom in a covalently bonded molecule can be assigned a **covalent radius (r_c)** and also a nonbonded radius, known as the **van der Waals radius**.

Covalent radii are calculated from half the interatomic distance between two singly bonded, like atoms. For diatomic molecules such as F_2, this is no problem, but for other elements, like carbon, which do not have a diatomic molecule, an average value is calculated from a range of compounds that contain a C–C single bond.

The van der Waals radius is defined as a nonbonded distance of closest approach and is calculated from the smallest interatomic distances in crystal structures that are considered to be not bonded to one another. Again, these are average values compiled from many crystal structures. If the sum of the van der Waals radii of two adjacent atoms in a structure is greater than the measured distance between them, then it is supposed that there is some bonding between them. Table 1.11 gives the covalent and the van der Waals radii for the typical elements.

1.6.8 Molecular Structures

Finally, we consider crystal structures that do not contain any extended arrays of atoms—the many crystals that contain small, discrete, covalently bonded molecules that are held together only by weak forces. The example of graphite in the previous section, in a way, forms a bridge between these structures and the structures with infinite three-dimensional arrays.

TABLE 1.11
Single-Bond Covalent Radii and van der Waals Radii (in Parentheses) for the Typical Elements/pm

Group 1	Group 2	Group 3	Group 14	Group 15	Group 16	Group 17	Group 18
							He
							—(140)
Li	Be	B	C	N	O	F	Ne
135	90	80	77 (170)	74 (155)	73 (152)	71 (147)	—(154)
Na	Mg	Al	Si	P	S	Cl	Ar
154	130	125	117 (210)	110 (180)	104 (180)	99 (175)	—(188)
K	Ca	Ga	Ge	As	Se	Br	Kr
200	174	126	122	121 (185)	117 (190)	114 (185)	—(202)
Rb	Sr	In	Sn	Sb	Te	I	Xe
211	192	141	137	141	137 (206)	133 (198)	—(216)
Cs	Ba	Tl	Pb	Bi	Po	At	Rn
225	198	171	175	170	140	—	—

Examples of molecular crystals are found throughout organic, organometallic and inorganic chemistry. The crystals are characterised by low melting and boiling temperatures. We will look at just two examples, carbon dioxide and water (ice), both familiar, small, covalently bonded molecules.

When gaseous carbon dioxide (CO_2) is cooled sufficiently, it forms a molecular crystalline solid, as illustrated in Figure 1.50. Notice that the unit cell contains clearly discernible CO_2 molecules, which are covalently bonded, and are held together in the crystal by weak van der Waals forces.

The structure of one form of ice (crystalline water) is shown in Figure 1.51. Each H_2O molecule is tetrahedrally surrounded by four others. The crystal structure is

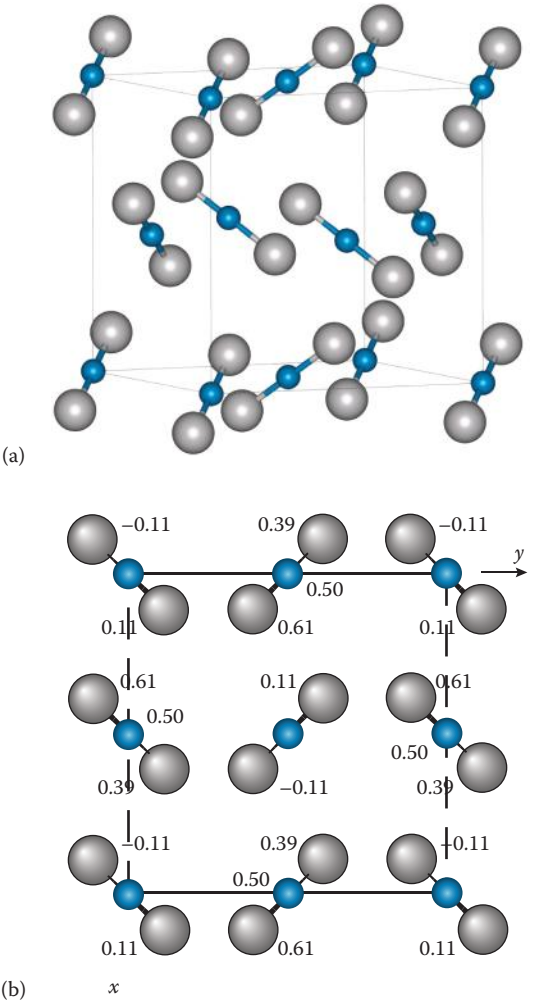


FIGURE 1.50 (a) Crystal structure of CO_2. (b) A packing diagram of the unit cell of CO_2 projected onto the *xy* plane. The heights of the atoms are expressed as fractional coordinates of *c*. Key: C, blue; O, grey.

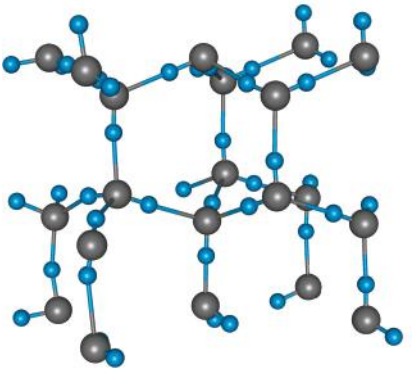


FIGURE 1.51 Crystal structure of ice. Key: H, blue; O, grey.

held together by the hydrogen bonds formed between a hydrogen atom of one water molecule and the oxygen of the adjacent molecule, forming a three-dimensional arrangement throughout the crystal. This open hydrogen-bonded network of water molecules makes ice less dense than water, so that it floats on the surface of water.

A summary of the various types of crystalline solids is given in Table 1.12, relating the type of structure to its physical properties. It is important to realise that this

TABLE 1.12
Classification of Crystal Structures

Type	Structural Unit	Bonding	Characteristics	Examples
Ionic	Cations and anions	Electrostatic, nondirectional	Hard, brittle, crystals of high m.t.; moderate insulators; melts are conducting	Alkali metal halides
Extended covalent array	Atoms	Mainly covalent	Strong hard crystals of high m.t.; insulators	Diamond, silica
Molecular	Molecules	Mainly covalent between atoms in molecules, van der Waals or hydrogen bonding between molecules	Soft crystals of low m.t. and large coefficient of expansion; insulators	Ice, organic compounds
Metallic	Metal atoms	Band model (see Chapter 4)	Single crystals are soft; strength depends on structural defects and grain; good conductors; m.t. vary but tend to be high	Iron, aluminium, sodium

only gives a broad overview and is intended as a guide only: not every crystal will fall exactly into one of these categories.

1.6.9 Silicates

The **silicates** form a large group of crystalline compounds with rather complex but interesting structures. A great part of the Earth's crust is formed from these complex oxides of silicon.

Silicon itself crystallises with the same structure as diamond. Its normal oxide, silica (SiO_2), is polymorphic, and in the previous sections we have discussed the crystal structure of two of its polymorphs—β-cristobalite (Section 1.5.2) and β-quartz (Section 1.5.5). Quartz is one of the most common minerals on the Earth, occurring as sand on the seashore, as a constituent in granite and flint and, in less pure form, as agate and opal. The silicon atom in all these structures is tetrahedrally coordinated.

The silicate structures are most conveniently discussed in terms of the SiO_4^{4-} unit. The SiO_4^{4-} unit has tetrahedral coordination of silicon by oxygen and is represented in these structures by a small tetrahedron, as shown in Figure 1.52a. The silicon–oxygen bonds possess considerable covalent character.

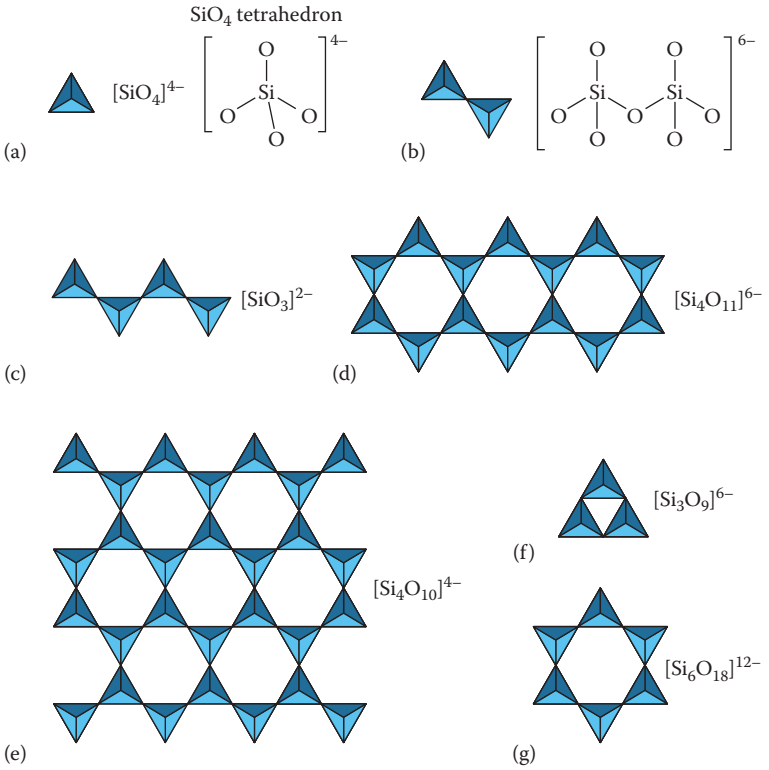


FIGURE 1.52 (a–g) Structural classification of mineral silicates.

Some minerals—**olivines** for instance—contain discrete SiO_4^{4-} tetrahedra (Figure 1.53a): these compounds do not contain Si–O–Si–O–Si–… chains, but there is considerable covalent character in the metal–silicate bonds. These are often described as **orthosilicates** or salts of orthosilicic acid ($Si(OH)_4$) or (H_4SiO_4) which is a very weak acid. The structure of olivine itself ($(Mg,Fe)_2SiO_4$), which can be described as an assembly of SiO_4^{4-} ions and Mg^{2+} (or Fe^{2+}) ions, appears in Table 1.7 because it can be alternatively described as an *hcp* array of oxygens with silicons occupying one eighth of the tetrahedral holes and magnesium ions occupying one half of the octahedral holes.

In most silicates, however, the SiO_4^{4-} tetrahedra are linked by oxygen sharing through a vertex, as illustrated in Figure 1.52b for two linked tetrahedra to give $Si_2O_7^{6-}$. Notice that each terminal oxygen confers a negative charge on the anion and the shared oxygen is neutral. The diagrams of silicate structures showing the silicate frameworks, such as are shown in Figure 1.52, omit these charges because they can be readily calculated. By sharing one or more oxygen atoms through the vertices, the tetrahedra are able to link up to form chains, rings, layers, etc. The negative charges on the silicate framework are balanced by metal cations in the lattice. Some examples are discussed below.

Examples of **discrete SiO_4^{4-} units** are found in: **olivine** (see Table 1.7 and Figure 1.53a), an important constituent of basalt; **garnets**, $M^{II}_3M^{III}_2(SiO_4)_3$ (where M^{II} can be Ca^{2+}, Mg^{2+} or Fe^{2+} and M^{III} can be Al^{3+}, Cr^{3+} or Fe^{3+}), the framework of which is composed of $M^{III}O_6$ octahedra, which are joined to six others through vertex-sharing SiO_4 tetrahedra—the M^{II} ions are coordinated by eight oxygens in dodecahedral interstices; **Ca_2SiO_4**, found in mortars and Portland cement; and **zircon** ($ZrSiO_4$), which has eight-coordinate Zr.

Structures containing **disilicate units**, $Si_2O_7^{6-}$ (Figure 1.52b) are not common but occur in **thortveitite**, the primary source of scandium ($(Sc,Y)_2Si_2O_7$), and **hemimorphite**, a rare zinc silicate ($Zn_4(OH)_2Si_2O_7$).

SiO_4^{4-} units share two corners to form infinite **chains** (see Figure 1.52c). The repeat unit is SiO_3^{2-}. Minerals with this structure are called **pyroxenes**, which together with olivine, form the primary constituents of the Earth's upper mantle, for example, **diopside** ($CaMg(SiO_3)_2$) and **enstatite** ($Mg_2(SiO_3)_2$). The silicate chains lie parallel to one another and are linked together by the cations that lie between them.

In **double chains** alternate tetrahedra share two and three oxygen atoms, respectively, as in Figure 1.52d. This class of minerals are known as the **amphiboles**, usually containing magnesium and/or iron (Figure 1.53b), an example of which is **tremolite** ($Ca_2Mg_5(OH)_2(Si_4O_{11})_2$). Most of the asbestos minerals fall into this class. The repeat unit is $Si_4O_{11}^{6-}$.

In **infinite layers** the tetrahedra all share three oxygen atoms (see Figure 1.52e). The repeat unit is $Si_4O_{10}^{4-}$. Examples are the **mica** group, named for their glittering appearance and used as insulators in high-voltage appliances, of which **biotite** ($K(Mg,Fe)_3(OH)_2Si_3AlO_{10}$) (Figure 1.53c) is a member containing a sandwich of two layers with octahedrally coordinated cations between the layers; and clay minerals such as **kaolinite**, found in china clay ($Al_4(OH)_8Si_4O_{10}$) and **talc** ($Mg_3(OH)_2Si_4O_{10}$).

In **rings** Each SiO_4^{4-} unit shares two corners as in the chains. Figure 1.52f and 1.52g shows three and six tetrahedra linked together; these have the general formula

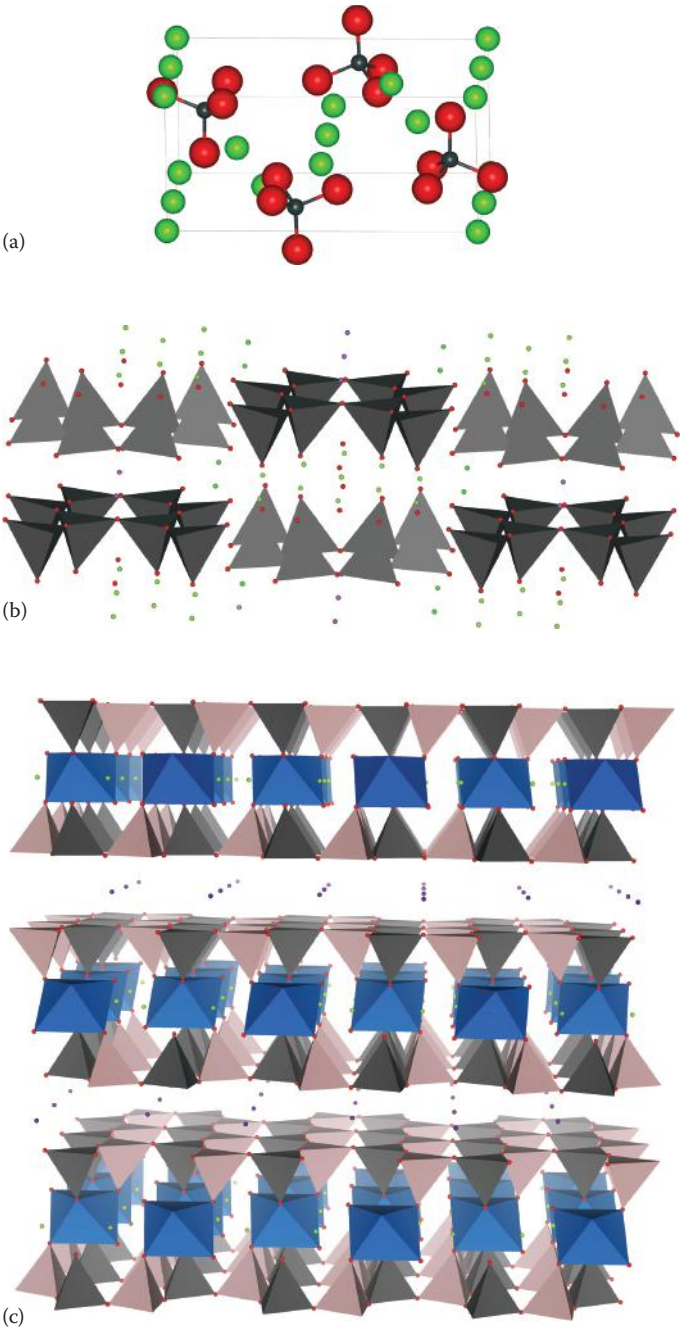


FIGURE 1.53 (a) Unit cell of olivine. Key: Mg and Fe, green; Si, grey; O, red. (b) Structure of an amphibole. Key: Mg, green; Si, grey; O, red; Na, purple. (c) Structure of biotite. Key: Mg, green; Si, grey; O, red; K, purple; Al, pink; Fe, blue.

$(SiO_3)_n^{-2n}$; rings may also be made from four tetrahedra. An example of a six-tetrahedra ring is **beryl** (emerald, $(Be_3Al_2Si_6O_{18})$); here, the rings lie parallel to the metal ions between them. Other examples include the uncommon, desert mineral **dioptase** ($Cu_6Si_6O_{18}.6H_2O$) and **benitoite**, California's very rare state gemstone ($BaTiSi_3O_9$).

If SiO_4^{4-} tetrahedra share all four oxygens, then the **three-dimensional structure** of silica (SiO_2) is produced. However, if some of the silicon atoms are replaced by the similarly sized atoms of the Group III element aluminium (i.e., if SiO_4^{4-} is replaced by AlO_4^{5-}), then other cations must be introduced to balance the charges. Such minerals include the **feldspars** (general formula $M(Al,Si)_4O_8$), the most abundant of the rock-forming minerals; the **zeolites**, which are used as ion exchangers, molecular sieves and catalysts (these are discussed in detail in Chapter 7); and the **ultramarines**, which are coloured silicates manufactured for use as pigments, **lapis lazuli** being a naturally occurring mineral of this type.

As one might expect, there is an approximate correlation between the solid-state structure and the physical properties of a particular silicate. For instance, cement contains discrete SiO_4^{4-} units and is soft and crumbly; asbestos minerals contain double chains of SiO_4^{4-} units and are characteristically fibrous; mica contains infinite layers of SiO_4^{4-} units and the weak bonding between the layers is easily broken and micas show cleavage parallel to the layers; granite contains feldspars, which are based on three-dimensional SiO_4^{4-} frameworks and are very hard.

1.6.10 Quasicrystals

In the classical crystallography that we have described so far, the basic unit of the crystal, the unit cell, repeats with translational symmetry in all three directions to fill all space. The only rotational symmetry axes allowed for the lattices of these crystals are 1-, 2-, 3-, 4- and 6-fold. However, a class of crystals that breaks these rules and that can exhibit 5-, 8-, 10- or 12-fold symmetry has been found—something that was previously thought impossible.

In the 1970s, Roger Penrose developed methods of two-dimensional tiling using pentagons or rhombi that can completely cover a flat surface. An example with two different types of rhombus-shaped tiles (fat and thin) is shown in Figure 1.54. The patterns are ordered in that they can exhibit reflection and fivefold symmetry, but they do not repeat in a translational fashion; the pattern never repeats itself exactly and is thus nonperiodic. An equivalent three-dimensional structure can be made from rhombohedrons.

In 1982, Dan Shechtman observed an electron diffraction pattern from a rapidly cooled Al–Mn alloy that showed a pattern of sharp spots which exhibited 10-fold symmetry—a symmetry that cannot be space-filling. An example of a similar diffraction pattern is shown in Figure 1.55. The sharpness of the peaks shows that the crystal must contain long-range order. Similar to Penrose's two-dimensional tiling, this structure was found to be ordered but aperiodic and the name **quasicrystal** was given to this and similar structures which were subsequently found. In this crystal, the ordered areas of fivefold symmetry cannot fit together to fill all space, so other atomic arrangements fit into the gaps, thereby removing the translational symmetry (Figure 1.55b). Quasicrystals that have 5-, 8-, 10 and 12-fold symmetry have also been found.

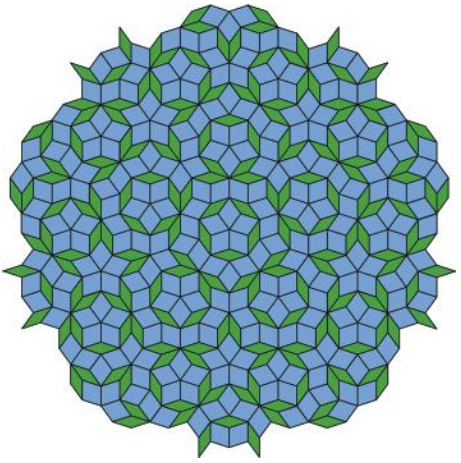


FIGURE 1.54 A Penrose tiling using thick and thin rhombi which exhibits exact fivefold symmetry.

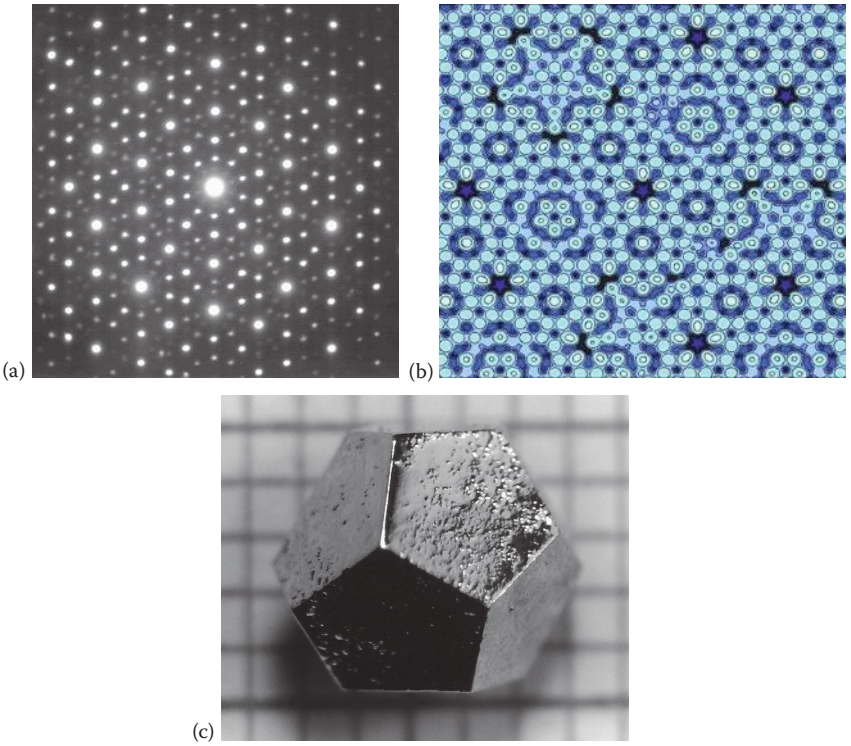


FIGURE 1.55 (a) Electron diffraction pattern of an icosahedral Zn–Mg–Ho quasicrystal. (b) Atomic model of fivefold icosahedral–Al–Pd–Mn quasicrystal surface. (c) Icosahedral Ho–Mg–Zn quasicrystal grown from the ternary melt. The edges are 2.2 mm long. Note the pentagonal faces and dodecahedral morphology.

The original publication of this work in 1984 met with a great deal of scepticism, but Shechtman's brilliant discovery and perseverance in developing this new class of crystals were rewarded with the Nobel Prize in 2011.

Quasicrystals fall into two categories: those with icosahedral symmetry (Figure 1.55c), which have no periodic directions, and the polygonal quasicrystals with only one periodic direction in the direction of 8-, 10- or 12-fold rotational symmetry. Because of the lack of translational symmetry, the structures can no longer be described by Miller indices, but polygonal quasicrystals require five, and icosahedral quasicrystals require six, independent vectors to describe them.

The first quasicrystals were grown from rapidly cooled molten binary alloys, but were found to be metastable. Now, many stable quasicrystals have been grown from ternary alloys often containing aluminium, although others containing rare-earth elements are proving to be of interest.

Quasicrystals show some unusual properties: they are extremely hard and rather brittle and have a very low coefficient of friction; they are poor conductors of electricity, but their conductivity *increases* with increasing temperature—rather surprising in view of their metallic composition (see Chapter 4). They also have poor thermal conductivity and do not exhibit magnetic ordering. Aluminium alloys that develop nano-sized quasicrystals during tempering have been found to form very strong but light materials suitable for surgical instruments. Other potential uses being developed are in surface coatings such as for nonstick frying pans and for hydrogen storage.

1.7 LATTICE ENERGY

The **lattice energy (*L*)** of an ionic crystal is a measure of the strength of the bonding and is defined as the standard enthalpy change when one mole of the solid is formed from the *gaseous ions*, for example, for NaCl, $L = \Delta H_m^{\ominus}$ for the reaction in Equation 1.1 = −787 kJ mol^{-1}; an exothermic process:

$$Na^+(g) + Cl^-(g) = NaCl(s) \tag{1.1}$$

Because of the difficulty in producing gaseous ions, it is not possible to measure lattice energies directly, but they can be determined experimentally using a thermochemical **Born–Haber cycle**, or they can be calculated from electrostatics, as we will see in Section 1.7.2.

1.7.1 Born–Haber Cycle

A Born–Haber cycle is the application of **Hess's law** to the enthalpy of formation of an ionic solid at 298 K. Hess's law states that the enthalpy of a reaction is the same whether the reaction takes place in one or in several steps. A Born–Haber cycle for a metal chloride (MCl) is shown in Figure 1.56; the metal chloride is formed from the constituent elements in their standard state in the equation at the bottom and by the clockwise series of steps above. From Hess's law, the sum of the enthalpy changes

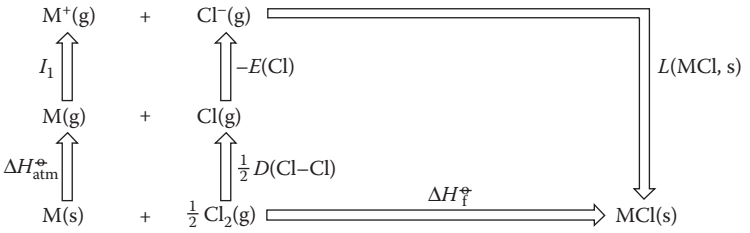


FIGURE 1.56 The Born–Haber cycle for a metal chloride, MCl.

for each step around the cycle can be equated with the standard enthalpy of formation, and we get

$$\Delta H_f^{\ominus}(\mathrm{MCl,s}) = \Delta H_{atm}^{\ominus}(\mathrm{M,s}) + I_1(\mathrm{M}) + \tfrac{1}{2}D(\mathrm{Cl-Cl}) - E(\mathrm{Cl}) + L(\mathrm{MCl,s}). \qquad (1.2)$$

The terms in the Born–Haber cycle are defined in Table 1.13 together with some sample data. By rearranging Equation 1.2, we can write an expression for the lattice energy in terms of the other quantities, which can then be calculated if the values for these are known. Notice that the way in which we have defined lattice energy gives **negative** (exothermic) values; you may find Equation 1.1 written the other way round in some texts, in which case they will quote positive lattice energies. Notice also that electron affinity is defined as the heat *evolved* when an electron is added to an atom; as an enthalpy change refers to the heat *absorbed*, the electron affinity and the enthalpy change for that process have opposite signs.

Cycles such as this can be constructed for other compounds such as oxides, MO, sulfides, MS, higher-valent metal halides, MX_n, etc. The difficulty in these cycles usually comes in the determination of values for the electron affinity (E). In the case of an oxide, it is necessary to know the double electron affinity for oxygen (the negative of the enthalpy change of the following reaction):

TABLE 1.13
Terms in the Born–Haber Cycle

Term	Definition of the Reaction to Which the Term Applies	NaCl/kJ mol^{-1}	AgCl/kJ mol^{-1}
$\Delta H_{atm}^{\ominus}$(M)	Standard enthalpy of atomisation of metal M: M(s) = M(g)	107.8	284.6
I_1(M)	First ionisation energy of metal M: M(g) = M$^+$(g) + e$^-$(g)	494	732
$\frac{1}{2}D$(Cl–Cl)	Half the dissociation energy of Cl_2: $\frac{1}{2}Cl_2$(g) = Cl(g)	122	122
$-E$(Cl)	*Minus* the electron affinity of chlorine is the enthalpy change for: Cl(g) + e$^-$(g) = Cl$^-$(g)	−349	−349
L(MCl,s)	Lattice energy of MCl(s): M$^+$(g) + Cl$^-$(g) = MCl(s)		
$\Delta H_f^{\ominus}$(MCl,s)	Standard enthalpy of formation of MCl(s): M(s) + $\frac{1}{2}Cl_2$(g) = MCl(s)	−411.1	−127.1

$$2e^-(g) + O(g) = O^{2-}(g) \tag{1.3}$$

This can be broken down into two stages:

$$e^-(g) + O(g) = O^-(g) \tag{1.4}$$

and

$$e^-(g) + O^-(g) = O^{2-}(g). \tag{1.5}$$

It is impossible to determine the enthalpy of reaction for Equation 1.5 experimentally, so this value can only be found if the lattice energy is known—a Catch 22 situation! To overcome problems such as this, methods of calculating (rather than measuring) the lattice energy have been devised and they are described in the next section.

1.7.2 Calculating Lattice Energies

For an ionic crystal of known structure, it should be a simple matter to calculate the energy released on bringing together the ions to form the crystal, using the equations of simple electrostatics. The energy of an ion pair, M^+ and X^- (assuming they are point charges), separated by distance r is given by **Coulomb's law**:

$$E = -\frac{e^2}{4\pi\varepsilon_0 r} \tag{1.6}$$

and where the *magnitudes* of the charges on the ions are Z_+ and Z_-:

$$E = -\frac{Z_+ Z_- e^2}{4\pi\varepsilon_0 r} \tag{1.7}$$

(e is the electronic charge, 1.6×10^{-19} C and ε_0 is the permittivity of a vacuum, 8.854×10^{-12} F m^{-1}).

The energy due to coulombic interactions in a crystal is calculated for a particular structure by summing all the ion-pair interactions, thereby producing an infinite series. The series includes terms due to the attraction of the opposite charges on cations and anions and also repulsion terms due to cation–cation and anion–anion interactions. Some of these interactions are shown in Figure 1.57 for the NaCl structure: the Na^+ ion at the centre is surrounded by 6 Cl^- ions as next neighbours at a distance r, then by 12 Na^+ cations at distance $\sqrt{2}r$, then by 8 Cl^- anions at distance $\sqrt{3}r$, followed by a further 6 cations at $2r$ and so on. The coulombic energy of interaction is given by the summation of all these interactions:

$$E_C = -\frac{e^2}{4\pi\varepsilon_0 r}\left(6 - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{2} + \frac{24}{\sqrt{5}} \cdots\right)$$

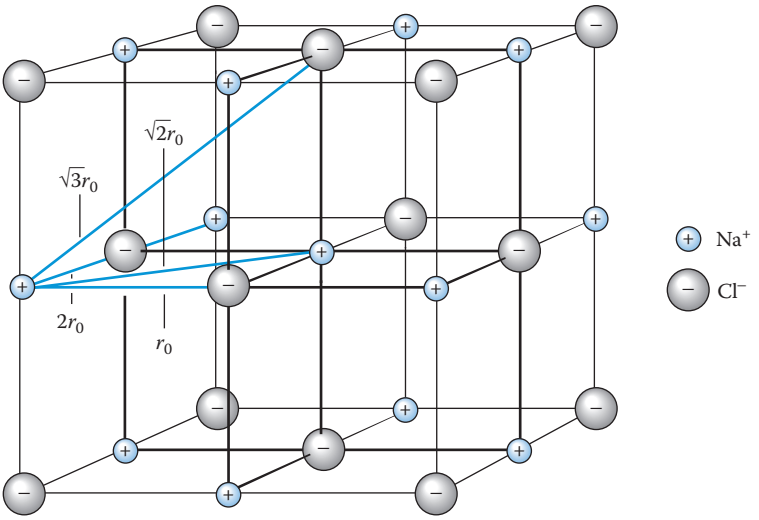


FIGURE 1.57 Sodium chloride structure showing internuclear distances.

or

$$E_C = -\frac{e^2}{4\pi\varepsilon_0 r}\left(\frac{6}{\sqrt{1}} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{\sqrt{4}} + \frac{24}{\sqrt{5}}\cdots\right) \tag{1.8}$$

The term inside the brackets is known as the **Madelung constant** (***A***), in this case, for the NaCl structure. The series is slow to converge; nevertheless, the values of the Madelung constant have been computed, not only for NaCl but also for most of the simple ionic structures. For one mole of NaCl, we can write

$$E_C = -\frac{N_A A e^2}{4\pi\varepsilon_0 r} \tag{1.9}$$

where N_A is the **Avogadro number**, 6.022×10^{23} mol^{-1} (Note that the expression is multiplied by N_A and not by $2N_A$, even though there are N_A cations and N_A anions present; this avoids counting every interaction twice!). The value of the Madelung constant is dependent only on the geometry of the lattice and not on its dimensions; values for various structures are given in Table 1.14.

Ions, of course, are **not** point charges but consist of positively charged nuclei surrounded by electron clouds. At small distances, these electron clouds repel each other, and this too needs to be taken into account when calculating the lattice energy of the crystal. At large distances, the repulsion energy is negligible; however, as the ions approach one another closely, it increases very rapidly. Max Born suggested that the form of this repulsive interaction could be expressed by

$$E_R = \frac{B}{r^n} \tag{1.10}$$

TABLE 1.14

Madelung Constants for Some Common Ionic Lattices

Structure	Madelung Constant, A	Number of Ions in Formula Unit, ν	$\frac{A}{\nu}$	Coordination
Caesium chloride (CsCl)	1.763	2	0.882	8:8
Sodium chloride (NaCl)	1.748	2	0.874	6:6
Fluorite (CaF_2)	2.519	3	0.840	8:4
Zinc blende (ZnS)	1.638	2	0.819	4:4
Wurtzite (ZnS)	1.641	2	0.821	4:4
Corundum (Al_2O_3)	4.172	5	0.835	6:4
Rutile (TiO_2)	2.408	3	0.803	6:3

where B is a constant and ***n*** (known as the **Born exponent**) is large and also a constant.

We can now write an expression for the lattice energy in terms of the energies of the interactions that we have considered:

$$L = E_C + E_R = -\frac{N_A A Z_+ Z_- e^2}{4\pi\varepsilon_0 r} + \frac{B}{r^n} \tag{1.11}$$

The lattice energy will be a minimum when the crystal is at equilibrium, that is, when the internuclear distance is at the equilibrium value of r_0. If we minimise the lattice energy (see Box), we get

$$L = -\frac{N_A A Z_+ Z_- e^2}{4\pi\varepsilon_0 r_0}\left(1 - \frac{1}{n}\right) \tag{1.12}$$

This is known as the **Born–Landé equation**: the values of r_0 and n can be obtained from X-ray crystallography and from compressibility measurements, respectively. The other terms in the equation are well-known constants, and when values for these are substituted, we get

$$L/\text{kJ mol}^{-1} = -\frac{1.389 \times 10^5 A Z_+ Z_-}{r_0\,(\text{pm})}\left(1 - \frac{1}{n}\right) \tag{1.13}$$

If the unit of r_0 is picometre, then the unit of L will be kilojoules per mole.

Pauling showed that the values of n can be approximated with reasonable accuracy for compounds of ions with noble gas configurations, by averaging empirical constants for each ion. The values of these constants are given in Table 1.15. For example, n for rubidium chloride (RbCl) is 9.5 (average of 9 and 10) and for strontium chloride ($SrCl_2$) it is 9.33 (average of 9, 9 and 10).

Notice the dramatic effect that the charge on the ions has on the value of the lattice energy. A structure containing one doubly charged ion has a factor of 2 in the

DERIVATION OF THE BORN–LANDÉ EQUATION

We can minimize the lattice energy function by using the standard mathematical technique of differentiating with respect to r and then equating to zero:

$$\frac{dL}{dr} = \frac{N_A A Z_+ Z_- e^2}{4\pi\varepsilon_0 r^2} = \frac{nB}{r^{n+1}}.$$

But, $\frac{dL}{dr} = 0$ when $r = r_0$.

so $$\frac{nB}{r_0^{n+1}} = \frac{N_A A Z_+ Z_- e^2}{4\pi\varepsilon_0 r_0^2}$$

$$B = \frac{N_A A Z_+ Z_- e^2 r_0^{n+1}}{n4\pi\varepsilon_0 r_0^2} = \frac{N_A A Z_+ Z_- e^2 r_0^{n-1}}{n4\pi\varepsilon_0}$$

But, $$L = -\frac{N_A A Z_+ Z_- e^2}{4\pi\varepsilon_0 r} + \frac{B}{r^n}$$

So, $$L = -\frac{N_A A Z_+ Z_- e^2}{4\pi\varepsilon_0 r_0} + \frac{N_A A Z_+ Z_- e^2 r_0^{n-1}}{4\pi\varepsilon_0 r_0^n}$$

$$L = -\frac{N_A A Z_+ Z_- e^2}{4\pi\varepsilon_0 r_0}\left(1 - \frac{1}{n}\right) \qquad (1.12)$$

Equation ($Z_+Z_- = 2$), whereas one containing *two* doubly charged ions is multiplied by a factor of 4 ($Z_+Z_- = 4$). Structures containing multiply charged ions tend, therefore, to have much larger (numerically) lattice energies.

The Russian chemist A.F. Kapustinskii noted that if the Madelung constants (A) for a number of structures are divided by the number of ions in one formula unit

TABLE 1.15
Constants Used to Calculate *n*

Ion Type	Constant
[He]	5
[Ne]	7
[Ar]	9
[Kr]	10
[Xe]	12

of the structure (v), the resulting values are almost constant (see Table 1.14), varying only between approximately 0.88 and 0.80, and this led to the idea that it would be possible to set up a general lattice energy equation that could be applied to any crystal regardless of its structure. We can now set up a general equation and use the resulting equation to calculate the lattice energy of an unknown structure. First, replace the Madelung constant (A) in the Born–Landé equation (Equation 1.13) with the value from the NaCl structure ($0.874v$) and r_0 by ($r_+ + r_-$), where r_+ and r_- are the cation and anion radii for six-coordination, giving

$$L/\text{kJ mol}^{-1} = -\frac{1.214 \times 10^5 \nu Z_+ Z_-}{r_+ + r_- \,(\text{pm})}\left(1 - \frac{1}{n}\right) \tag{1.14}$$

If n is assigned an average value of 9, we arrive at

$$L/\text{kJ mol}^{-1} = -\frac{1.079 \times 10^5 \nu Z_+ Z_-}{r_+ + r_- \,(\text{pm})} \tag{1.15}$$

Equations 1.14 and 1.15 are known as the **Kapustinskii equations**.

Some lattice energy values that have been calculated by various methods are shown in Table 1.16 for comparison with experimental values that have been computed using a Born–Haber cycle. Remarkably good agreement is achieved, considering all the approximations involved. The largest discrepancies are for the larger polarisable ions, where, of course, the ionic model is not expected to be perfect. The equations can be improved to help with these discrepancies by including the effect of van der Waals forces, zero point energy (the energy due to the vibration of the ions at 0 K) and heat capacity. The net effect of these corrections is only of the order of 10 kJ mol^{-1} and the values thus obtained for the lattice energy are known as **extended-calculation values**.

It is important to note that the good agreement achieved between the Born–Haber and the calculated values for lattice energy do not in any way prove that the ionic model is valid. This is because the equations possess a self-compensating feature in that they use *formal charges* on the ions but take *experimental internuclear distances*. The values of r_0 are the result of *all* the various types of bonding in the crystal, not just of the ionic bonding, so are rather shorter than one would expect for purely ionic bonding.

As we conclude this section, let us briefly revisit ionic radii. Many ionic compounds contain complex or polyatomic ions. Clearly, it is going to be extremely difficult to measure the radii of ions such as ammonium (NH_4^+) or carbonate (CO_3^{2-}), for instance. However, Yatsimirskii has devised a method that determines a value of the radius of a polyatomic ion by applying the Kapustinskii equation to lattice energies determined from thermochemical cycles. Such values are called **thermochemical radii** and Table 1.17 lists some values.

TABLE 1.16
Lattice Energies of Some Alkali and Alkaline Earth Metal Halides at 0 K

		L/kJ mol^{-1}			
Compound	Structure	Born–Haber Cycle[a]	Born–Landé Equation 1.12[b]	Extended Calculation[c]	Kapustinskii Equation 1.15
LiF	NaCl	−1025		−1033	−1033
LiI	NaCl	−756		−738	−729
NaF	NaCl	−910	−904	−906	−918
NaCl	NaCl	−772	−757	−770	−763
NaBr	NaCl	−736	−720	−735	−724
NaI	NaCl	−701	−674	−687	−670
KCl	NaCl	−704	−690	−702	−677
KI	NaCl	−646	−623	−636	−603
CsF	NaCl	−741	−724	−734	−719
CsCl	CsCl	−652	−623	−636	−620
CsI	CsCl	−611	−569	−592	−558
MgF_2	Rutile	−2922	−2883	−2914	−3158
CaF_2	Fluorite	−2597	−2594	−2610	−2779
$CaCl_2$	Deformed rutile	−2226		−2223	−2304

[a] Johnson, D.A., *Some Thermodynamic Aspects of Inorganic Chemistry*, 2nd edn, Cambridge University Press, 1982.

[b] Morris, D.F.C., *Journal of Inorganic and Nuclear Chemistry*, **4**, 8, 1957.

[c] Cubiociotti, D., *Journal of Chemical Physics*, **34**, 2189, 1961; Brackett, T.E. and Brackett, E.B., *Journal of Physical Chemistry*, **69**, 3611, 1965; Jenkins, H.D.B. and Pratt, K.F., *Proceedings of the Royal Society A*, **356**, 115, 1977.

TABLE 1.17
Thermochemical Radii of Polyatomic Ions

Ion	pm	Ion	pm	Ion	pm
NH_4^+	151	ClO_4^-	226	MnO_4^-	215
Me_4N^+	215	CN^-	177	O_2^{2-}	144
PH_4^+	171	CNS^-	199	OH^-	119
$AlCl_4^-$	281	CO_3^{2-}	164	PtF_6^{2-}	282
BF_4^-	218	IO_3^-	108	$PtCl_6^{2-}$	299
BH_4^-	179	N_3^-	181	$PtBr_6^{2-}$	328
BrO_3^-	140	NCO^-	189	PtI_6^{2-}	328
CH_3COO^-	148	NO_2^-	178	SO_4^{2-}	244
ClO_3^-	157	NO_3^-	165	SeO_4^{2-}	235

Source: Taken from Huheey, J.E., *Inorganic Chemistry*, 3rd edn, Harper & Row, New York, 1983.

1.7.3 Calculations Using Thermochemical Cycles and Lattice Energies

It is not yet possible to measure a lattice energy directly, which is why the best experimental values for the alkali halides, as listed in Table 1.16, are derived from a thermochemical cycle. This in itself is not always easy for compounds other than the alkali halides, because as we noted before, not all of the data are necessarily available. Electron affinity values are known from experimental measurements for most of the elements, but when calculating a lattice energy for a sulfide, for example, we need to know the enthalpy change for the reaction in Equation 1.16:

$$2e^-(g) + S(g) = S^{2-}(g) \tag{1.16}$$

which is minus a **double electron affinity** and is not so readily available. This is where lattice energy calculations come into their own, because we can use one of the methods discussed above to calculate a theoretical value of L for the appropriate sulfide and then plug it into the thermochemical cycle to calculate the enthalpy change of Equation 1.16.

Proton affinities can be found in a similar way: proton affinity is defined as the enthalpy change of the reaction shown in Equation 1.17, where a proton is lost by the species, A.

$$AH^+(g) = A(g) + H^+(g) \tag{1.17}$$

Again, this value can be obtained from a suitable thermochemical cycle, provided the lattice energy is known. Take as an example the formation of the ammonium ion, NH_4^+ (g):

$$NH_3(g) + H^+(g) = NH_4^+(g) \tag{1.18}$$

The enthalpy change of the reaction in Equation 1.18 is minus the proton affinity of ammonia, $-P(NH_3,g)$. This could be calculated from the thermochemical cycle shown in Figure 1.58, provided the lattice energy of ammonium chloride is known.

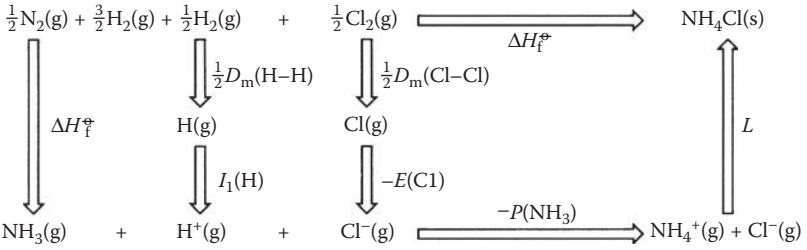


FIGURE 1.58 Thermochemical cycle for the calculation of the proton affinity of ammonia.

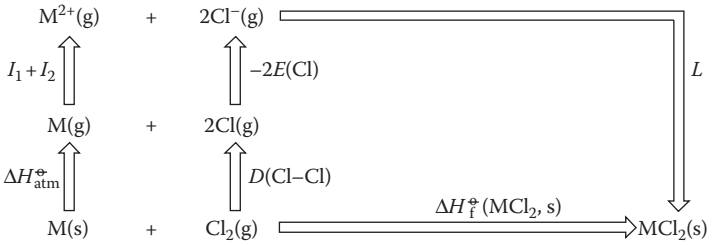


FIGURE 1.59 Born–Haber cycle for a metal dichloride (MCl_2).

Thermochemical cycles can also be used to provide information on the thermodynamic properties of compounds that contain metals in unusual oxidation states, which have not yet been prepared. For instance, we could use arguments to determine whether it is possible to prepare a compound of sodium in a higher oxidation state than normal, $NaCl_2$, for instance. In order to calculate a value for the enthalpy of formation and then $\Delta G_f^{\ominus}(NaCl_2,s)$, we need to set up a Born–Haber cycle for $NaCl_2(s)$ of the type shown in Figure 1.59. The only term in this cycle that is not known is the lattice energy of $NaCl_2$, and for this we make the approximation that it will be the same as that of the isoelectronic compound $MgCl_2$, which *is* known. The summation becomes

$$\Delta H_f^{\ominus}(NaCl_2,s) = \Delta H_{atm}^{\ominus}(Na,s) + I_1(Na) + I_2(Na) + D(Cl-Cl) - 2E(Cl) + L(NaCl_2,s) \quad (1.19)$$

The relevant values are listed in Table 1.18.

We noted earlier that solids containing doubly charged ions tend to have large negative lattice energies, yet the calculation for $NaCl_2$ has shown just the opposite,

TABLE 1.18
Values of the Born–Haber Cycle Terms for $NaCl_2$ and $MgCl_2$/kJ mol^{-1}

	Na	Mg
$\Delta H_{atm}^{\ominus}$	108	148
I_1	494	736
I_2^a	4565	1452
D(Cl–Cl)	244	244
$-2E$(Cl)	−698	−698
$L(MCl_2,s)$	−2523	−2523
$\Delta H_f^{\ominus}(MCl_2,s)$	2190	−641

[a] The second ionisation energy, I_2, refers to the energy change of the reaction: $M^+(g) - e^-(g) = M^{2+}(g)$.

indicating that it is highly unstable. Why? A glance at the figures in Table 1.18 reveals the answer: the second ionisation energy for sodium is huge because this relates to the removal of an inner shell electron, and this far outweighs the larger lattice energy. So, $NaCl_2$ does not exist because the extra stabilisation of the lattice due to a doubly charged ion is not enough to compensate for the large second ionisation energy.

The calculation we have just performed shows that the decomposition reaction of $NaCl_2(s)$ is likely to be highly favourable:

$$NaCl_2(s) = Na(s) + Cl_2(g): \qquad \Delta H_m^{\ominus} = -2190 \text{ kJ mol}^{-1}.$$

But to be sure of the stability of the compound, we need to calculate a value for $\Delta G_m^{\ominus}$. For the analogous reaction of $MgCl_2(s)$,

$$\Delta S_m^{\ominus} = 166.1 \text{ J K}^{-1} \text{ mol}^{-1}$$

and so

$$\Delta G_m^{\ominus} = \Delta H_m^{\ominus} - T\Delta S_m^{\ominus} = -2190 - (298.15 \times 0.166) = -2239 \text{ kJ mol}^{-1}$$

suggesting that $NaCl_2$ would indeed be unstable with respect to decomposition of sodium and chlorine.

Interestingly, it was arguments and calculations of this sort that led Neil Bartlett to the discovery of the first noble gas compound, $XePtF_6$. Bartlett had prepared a new complex, O_2PtF_6, which, by analogy with the diffraction pattern of $KPtF_6$, he formulated as containing the dioxygenyl cation, $[O_2^+][PtF_6^-]$. He realised that the ionisation energies of oxygen and xenon are very similar and that although the radius of the Xe^+ ion is slightly different, because the PtF_6^- anion is very large, the lattice energy of $[Xe^+][PtF_6^-]$ should be very similar to that of the dioxygenyl complex and, therefore, should exist. Accordingly, he mixed xenon and PtF_6 and obtained the orange-yellow solid of xenon hexafluoroplatinate—the first noble gas compound. (Although, in fact, the compound turned out not to have the structure that Bartlett predicted because at room temperature $XePtF_6$ reacts with another molecule of PtF_6 to give a product containing $[XeF]^+[PtF_6]^-$ and $[PtF_5]^-$.)

Several examples of problems involving these types of calculations are included in the questions at the end of the chapter.

1.8 CONCLUSION

In this opening chapter, we have introduced many of the principles and ideas that lie behind a discussion of the crystalline solid state. We have discussed in detail the structure of a number of important ionic crystal structures and shown how they can be linked to a simple view of ions as hard spheres that pack together as closely as possible, but can also be viewed as the linking of octahedra or tetrahedra in various ways. Taking these ideas further, we have investigated the size of these ions in terms

of their radii and then the energy involved in forming a lattice with ionic bonding. We also noted that covalent bonding is present in many structures and that when only covalent bonding is present, we tend to see a rather different type of crystal structure.

QUESTIONS

1. Figure 1.60 shows several molecules. Find and draw all the symmetry elements possessed by each molecule.
2. Does the CF_4 molecule in Figure 1.14 possess a centre of inversion? What other rotation axis is coincident with the $\bar{4}$ (S_4)?
3. How many centred cells are drawn in Figure 1.17?
4. Name the symmetry properties found in these five commonly occurring space groups: $P\bar{1}$, $P2_1/c$, C2/m, Pbca and $F\bar{4}3m$.
5. Index the sets of lines in Figure 1.26 marked B, C, D and E.
6. Index the sets of planes in Figure 1.27b–d.
7. Figure 1.22c shows a unit cell of a face-centred cubic structure. If a single atom is placed at each lattice point, then this becomes the unit cell of the *ccp* (cubic close packed) structure. Find the *100*, *110* and *111* planes and calculate the density of atoms per unit area for each type of plane. (Hint: calculate the area of each plane assuming a cell length *a*. Decide the fractional contribution made by each atom to the plane.)
8. Using Figures 1.30, 1.31 and 1.37 together with the models if necessary, draw unit cell projections for (a) CsCl, (b) NaCl and (c) ZnS (zinc blende).
9. How many formula units (ZnS) are there in the zinc blende unit cell?
10. Draw a projection of a unit cell for both the *hcp* and *ccp* structures, seen perpendicular to the close-packed layers (i.e., assume that the close-packed layer is the *ab* plane, draw in the *x* and *y* coordinates of the atoms in their correct positions and mark the third coordinate *z* as a fraction of the corresponding repeat distance *c*).
11. The unit cell dimension *a* of NaCl is 564 pm. Calculate the density of NaCl in kilograms per cubic metre.

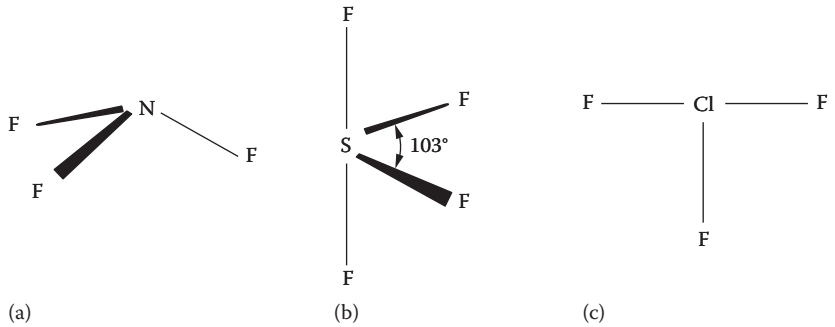


FIGURE 1.60 (a) NF_3, (b) SF_4 and (c) ClF_3.

TABLE 1.19
Values of the Born–Haber Cycle Terms for $CaCl_2$/kJ mol^{-1}

Term	Value
$\Delta H^{\ominus}_{atm}$	178
I_1	590
I_2	1146
D(Cl–Cl)	244
$-2E$(Cl)	−698
$\Delta H^{\ominus}_f(CaCl_2,s)$	−795.8

12. A compound AgX has the same crystal structure as NaCl, a density of 6477 kg m^{-3} and a unit cell dimension of 577.5 pm. Identify X.
13. Estimate a value for the radius of the iodide ion. The distance between the lithium and iodine nuclei in lithium iodide is 300 pm.
14. Calculate a radius for F^- from the data in Table 1.8 for NaI and NaF. Repeat the calculation using RbI and RbF.
15. The diamond structure is shown in Figure 1.47. Find the *100*, *110* and *111* planes and determine the relative atomic densities per unit area (see Question 7.)
16. Use the Born–Haber cycle in Figure 1.59 and the data in Table 1.19 to calculate the lattice energy of solid calcium chloride ($CaCl_2$).
17. Calculate the value of the Madelung constant for the structure in Figure 1.61. All bond lengths are equal and all bond angles are 90°. Assume that there are no ions other than those shown in Figure 1.61 and that the charges on the cation and anion are +1 and −1, respectively.

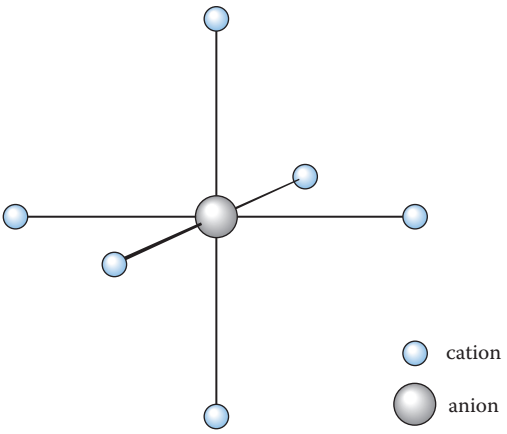


FIGURE 1.61 Structure for Question 17. All bond lengths are equal and bond angles are 90°.

18. Calculate a value for the lattice energy of potassium chloride using Equation 1.15. Compare this with the value that you calculate from the thermodynamic data in Table 1.20.
19. Calculate a value for the electron affinity of sulfur for two electrons. Take iron(II) sulfide (FeS) as a model and devise a suitable cycle. Use the data given in Table 1.21.

TABLE 1.20
Values of the Born–Haber Cycle Terms for KCl/kJ mol^{-1}

Term	Value
$\Delta H^{\ominus}_{atm}$	89.1
I_1	418
$\frac{1}{2}D$(Cl–Cl)	122
$-E$(Cl)	−349
$\Delta H^{\ominus}_f$(KCl,s)	−436.7

TABLE 1.21
Values of the Born–Haber Cycle Terms for FeS/kJ mol^{-1}

Term	Value
$\Delta H^{\ominus}_{atm}$(Fe,s)	416.3
I_1(Fe)	761
I_2(Fe)	1561
$\Delta H^{\ominus}_{atm}$(S,s)	278.8
$\Delta H^{\ominus}_f$(FeS,s)	−100.0
E(S)	200

TABLE 1.22
Values of the Born–Haber Cycle Terms for MgO/kJ mol^{-1}

Term	Value
$\Delta H^{\ominus}_{atm}$(Mg,s)	147.7
I_1(Mg)	736
I_2(Mg)	1452
$\frac{1}{2}D$(O–O)	249
$\Delta H^{\ominus}_f$(MgO,s)	−601.7
E(O)	141

TABLE 1.23
Values of the Born–Haber Cycle Terms for NH_4Cl/kJ mol^{-1}

Term	Value
$\Delta H_f^{\ominus}(NH_3)$	−46.0
$\Delta H_f^{\ominus}(NH_4Cl,s)$	−314.4
$\frac{1}{2}D(H-H)$	218
$\frac{1}{2}D(Cl-Cl)$	122
$I(H)$	1314
$E(Cl)$	349
$r_+(NH_4^+)$	151 pm

TABLE 1.24
Values of the Born–Haber Cycle Terms for NaCl, MgCl, and AlCl/kJ mol^{-1}

Term	Na	Mg	Al
$\Delta H_{atm}^{\ominus}$	108	148	326
I_1	494	736	577
$\frac{1}{2}D(Cl^-Cl)$	122	122	122
$-E(Cl)$	−349	−349	−349
$\Delta H_f^{\ominus}(NaCl,s)$	−411		
$L(MCl,s)$			

20. Calculate a value for the electron affinity of oxygen for two electrons. Take magnesium oxide (MgO) as a model and devise a suitable cycle. Use the data given in Table 1.22.
21. Calculate a value for the proton affinity of ammonia using the cycle in Figure 1.58 and the data in Table 1.23.
22. Compounds of aluminium and magnesium in the lower oxidation states, Al(I) and Mg(I), do not exist under normal conditions. If we make an assumption that the radius of Al^+ or Mg^+ is the same as that of Na^+ (same row of the periodic table), then we can also equate the lattice energies, MCl. Use this information in a Born–Haber cycle to calculate a value of the enthalpy of formation ($\Delta H_f^{\ominus}$) for AlCl(s) and MgCl(s), using the data in Table 1.24.

2 Physical Methods for Characterising Solids

2.1 INTRODUCTION

There is a vast array of physical methods for investigating the structures of solids, each technique with its own strengths and weaknesses—some techniques are able to investigate the local coordination around a particular atom or its electronic properties, whereas others are suited to elucidating the long-range order of the structures. Others investigate just the surface structure. No book could do justice to all the techniques on offer, so we describe just some of the more commonly available techniques and try to show what information can be gleaned from each one and its limitations. We start with X-ray diffraction by powders and single crystals. Single crystal X-ray diffraction is used to determine precisely the atomic positions and therefore, the bond lengths (to a few tens of picometres*) and bond angles of molecules within the unit cell. It gives an overall, average picture of a long-range ordered structure, but is less suited for giving information on the structural positions of defects, dopants and nonstoichiometric regions. It is often very difficult to grow single crystals, but most solids can be made as crystalline powders. Powder X-ray diffraction is probably the most commonly employed technique in solid-state inorganic chemistry and has many uses, from analysis and assessing the phase purity, to determining the structure. Single crystal X-ray diffraction techniques have provided us with the crystal structures which are used as the basis for interpreting powder diffraction data.

2.2 X-RAY DIFFRACTION

2.2.1 Generation of X-Rays

The discovery of X-rays was made by the German physicist Wilhelm Röntgen in 1895, for which he was awarded the first Nobel Prize in Physics in 1901. The benefits of his discovery in terms of medical diagnosis and treatment and in investigating molecular and atomic structures are immeasurable, and yet, Röntgen was a man of such integrity that he refused to make any financial gain out of his discovery, believing that scientific research should be made freely available.

An electrically heated filament, usually tungsten, emits electrons, which are accelerated by a high potential difference (20–50 kV) and allowed to strike a metal

* Different physical techniques commonly use a variety of units, often nanometres, picometres or Ångstroms (Å) (1 Å = 0.1 nm or 100 pm). Although we have tried to use SI units as far as possible throughout this book, you will find that Ångstroms are also used, particularly in this chapter, where it is common in the literature of the subject.

target or an anode that is water-cooled (Figure 2.1a). The anode emits a continuous spectrum of 'white' X-radiation, but superimposed on this are sharp, intense X-ray peaks (K_α, K_β), as shown in Figure 2.1b. The frequencies of the K_α and K_β lines are characteristic of the anode metal; the target metals most commonly used in X-ray crystallographic studies are copper and molybdenum, which have K_α lines at 154.18 pm and 71.07 pm, respectively. These lines occur because the bombarding electrons knock out electrons from the innermost K shell ($n = 1$) and this in turn creates vacancies which are filled by electrons descending from the shells above. The decrease in energy appears as radiation; electrons descending from the L shell ($n = 2$) give the K_α lines, and electrons from the M shell ($n = 3$) give the K_β lines. (These lines are actually very closely spaced doublets—$K_{\alpha1}$, $K_{\alpha2}$ and $K_{\beta1}$, $K_{\beta2}$—which are usually not resolved.) As the atomic number (Z) of the target increases, the lines shift to a shorter wavelength.

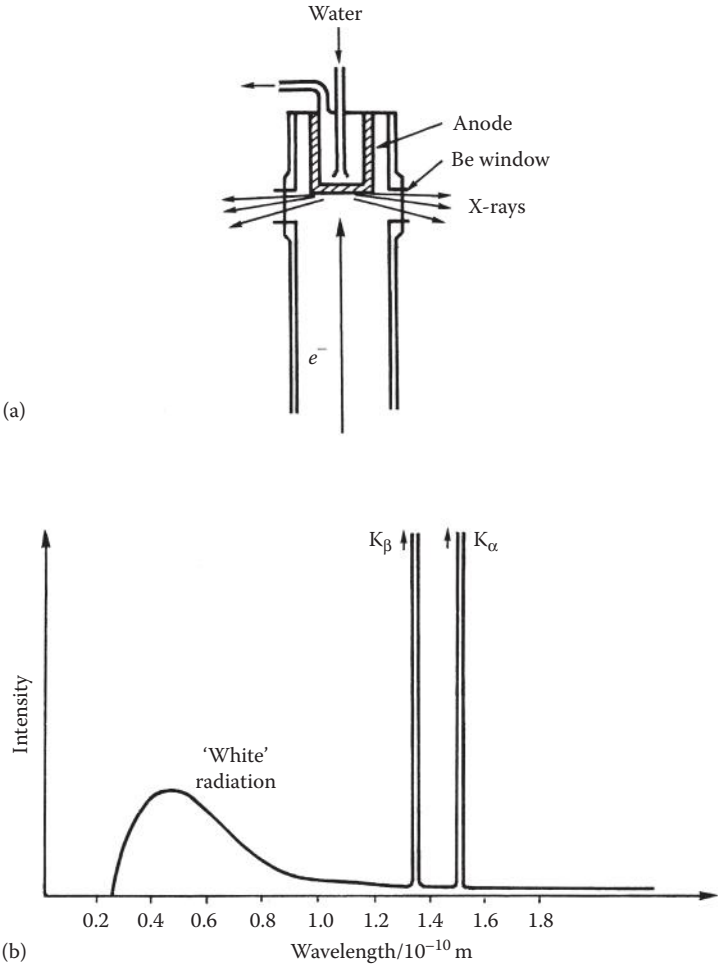


FIGURE 2.1 (a) Section through an X-ray tube. (b) An X-ray emission spectrum.

Normally, in X-ray diffraction, monochromatic radiation (a single wavelength or a very narrow range of wavelengths) is required. Usually, the K_α line is selected and the K_β line is filtered out by using a filter made of a thin metal foil of the element adjacent ($Z - 1$) in the periodic table; thus, nickel effectively filters out the K_β line of copper, and a niobium filter is used for molybdenum. A monochromatic beam of X-rays can also be selected by reflecting the beam from a plane of a single crystal, normally graphite (the reasons why this works will become obvious after you have read the next section).

2.2.2 Diffraction of X-Rays

By 1912, the nature of X-rays—whether they were particles or waves—was still unresolved; a demonstration of X-ray diffraction effects was needed to demonstrate their wave nature. This was eventually achieved by Max von Laue using a crystal of copper sulfate as the diffraction grating, the work that earned him the Nobel Prize in Physics in 1914. Crystalline solids consist of regular arrays of atoms, ions or molecules with interatomic spacings of the order of 100 pm. For diffraction to take place, the wavelength of the incident light has to be of the same order of magnitude as the spacings of the grating. Because of the periodic nature of the internal structure, it is possible for crystals to act as a three-dimensional diffraction grating to light of a suitable wavelength: a Laue photograph is shown in Figure 2.2.

This discovery was immediately noted by W. H. and W. L. Bragg (father and son), and they started experiments on using X-ray crystal diffraction as a means of

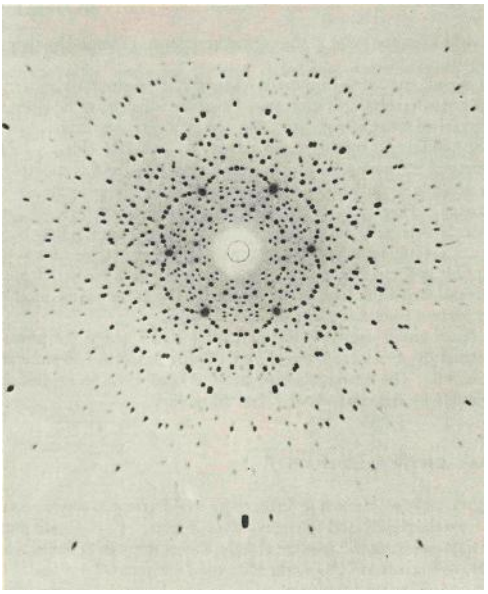


FIGURE 2.2 X-ray diffraction by a crystal of beryl using the Laue method. (From Moore, W.J., *Physical Chemistry*, 5th edn, Longman, 1972. Reprinted with courtesy of Eastman Kodak Company.)

structure determination. In 1913, they first determined the crystal structure of NaCl and then went on to determine many structures, including those of KCl, ZnS, CaF_2, $CaCO_3$ and diamond. W. L. (Lawrence) Bragg noted that X-ray diffraction behaves like a 'reflection' from the planes of the atoms within the crystal and that only at specific orientations of the crystal with respect to the source and detector are X-rays 'reflected' from the planes. It is not like the reflection of light from a mirror, as this requires that the angle of incidence equals the angle of reflection, and this is possible for all angles. With X-ray diffraction, the reflection occurs only when the conditions for constructive interference are fulfilled.

Figure 2.3 illustrates the Bragg condition for the reflection of X-rays by a crystal. The array of black points in the diagram represents a section through a crystal and the lines joining the dots mark a set of parallel planes with Miller indices *hkl* and interplanar spacing d_{hkl}. A parallel beam of monochromatic X-rays ADI is incident on the planes at an angle θ_{hkl}. The ray A is scattered by the atom at B and the ray D is scattered by the atom at F. For the reflected beams to emerge as a single beam of reasonable intensity, they must reinforce or arrive in phase with one another. This is known as constructive interference, and for constructive interference to take place, the path lengths of the interfering beams must differ by an integral number of wavelengths. If BE and BG are drawn at right angles to the beam, the difference in path length between the two beams is given by

$$\text{difference in path length} = EF + FG$$

but

$$EF = FG = d_{hkl} \sin \theta_{hkl}$$

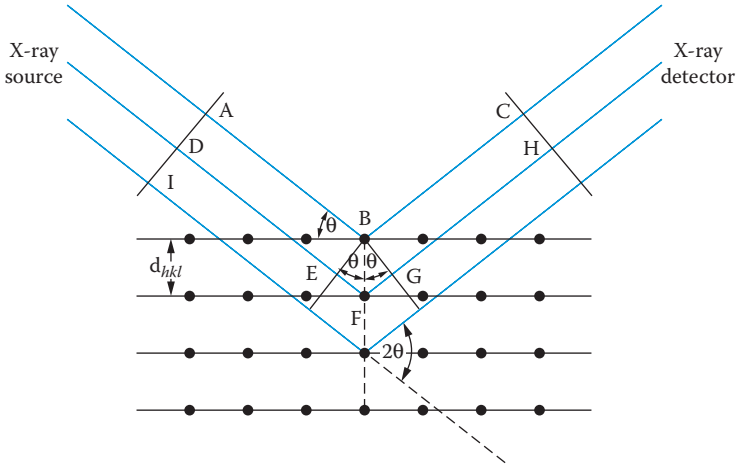


FIGURE 2.3 Bragg reflection from a set of crystal planes with a spacing d_{hkl}.

so

$$\text{difference in path length} = 2d_{hkl} \sin\theta_{hkl} \tag{2.1}$$

This must be equal to an integral number, n, of wavelengths. If the wavelength of the X-rays is λ, then

$$n\lambda = 2d_{hkl} \sin\theta_{hkl} \tag{2.2}$$

This is known as the **Bragg equation**, and it relates the spacing between the crystal planes (d_{hkl}) to the particular Bragg angle (θ_{hkl}) at which reflections from these planes are observed (mostly the subscript hkl is dropped from the Bragg angle θ without any ambiguity as the angle is unique for each set of planes).

When $n = 1$, the reflections are first-order, and when $n = 2$ the reflections are second-order and so on. However, the Bragg equation for a second-order reflection from a set of planes hkl is

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

which can be rewritten as

$$\lambda = 2\frac{d_{hkl}}{2} \sin\theta \tag{2.3}$$

Equation 2.3 represents a first-order reflection from a set of planes with interplanar spacing $d_{hkl}/2$. The set of planes with interplanar spacing $d_{hkl}/2$ has Miller indices $2h$ $2k$ $2l$. So, the second-order reflection from hkl is indistinguishable from the first-order reflection from $2h$ $2k$ $2l$, and the Bragg equation may be written more simply as

$$\lambda = 2d_{hkl} \sin\theta \tag{2.4}$$

2.3 POWDER DIFFRACTION

2.3.1 Powder Diffraction Patterns

A finely ground crystalline powder contains a very large number of small crystals, known as crystallites, which are oriented randomly to one another. If such a sample is placed in the path of a monochromatic X-ray beam, diffraction will occur from planes in those crystallites that happen to be oriented at the correct angle to fulfil the Bragg condition. The diffracted beams make an angle of 2θ with the incident beam. Because the crystallites can lie in all directions while still

maintaining the Bragg condition, the reflections lie on the surface of cones whose *semi*apex angles are equal to the deflection angle 2θ (Figure 2.4a). In the **Debye–Scherrer** photographic method, a strip of film was wrapped around the inside of an X-ray camera (Figure 2.4b) with a hole to allow in the collimated incident beam and a beam-stop to absorb the undiffracted beam. The sample was rotated to bring as many planes as possible into the diffracting condition, and the cones were recorded as arcs on the film. Using the radius of the camera and the distance along the film from the centre, the Bragg angle 2θ and therefore, the d_{hkl} spacing for each reflection can be calculated. Collection of powder diffraction patterns is now almost always performed by automatic diffractometers (Figure 2.5a), using a scintillation or a charge-coupled device (CCD) detector to record the angle and the intensity of the diffracted beams, which are plotted as intensity against 2θ (Figure 2.5b). The resolution obtained using a diffractometer is better than photography as the sample acts like a mirror, helping to refocus the X-ray beam. The data, both position and intensity, are readily measured and stored on a computer for analysis.

The difficulty in the powder method lies in deciding which planes are responsible for each reflection; this is known as 'indexing the reflections', that is, assigning the correct *hkl* index to each reflection. Although this is often possible for simple compounds in high-symmetry systems (as we shall explain in Section 2.4.3), it is extremely difficult to do for many larger and/or less symmetrical systems.

2.3.2 Absences Due to Lattice Centring

First, consider a primitive cubic system. From Equation 2.4, we see that the planes giving rise to the reflection with the smallest Bragg angle will have the

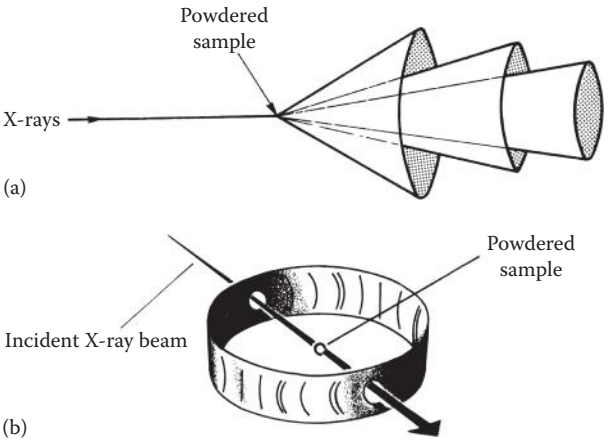


FIGURE 2.4 (a) Cones produced by a powder diffraction experiment. (b) Experimental arrangement for a Debye–Scherrer photograph.

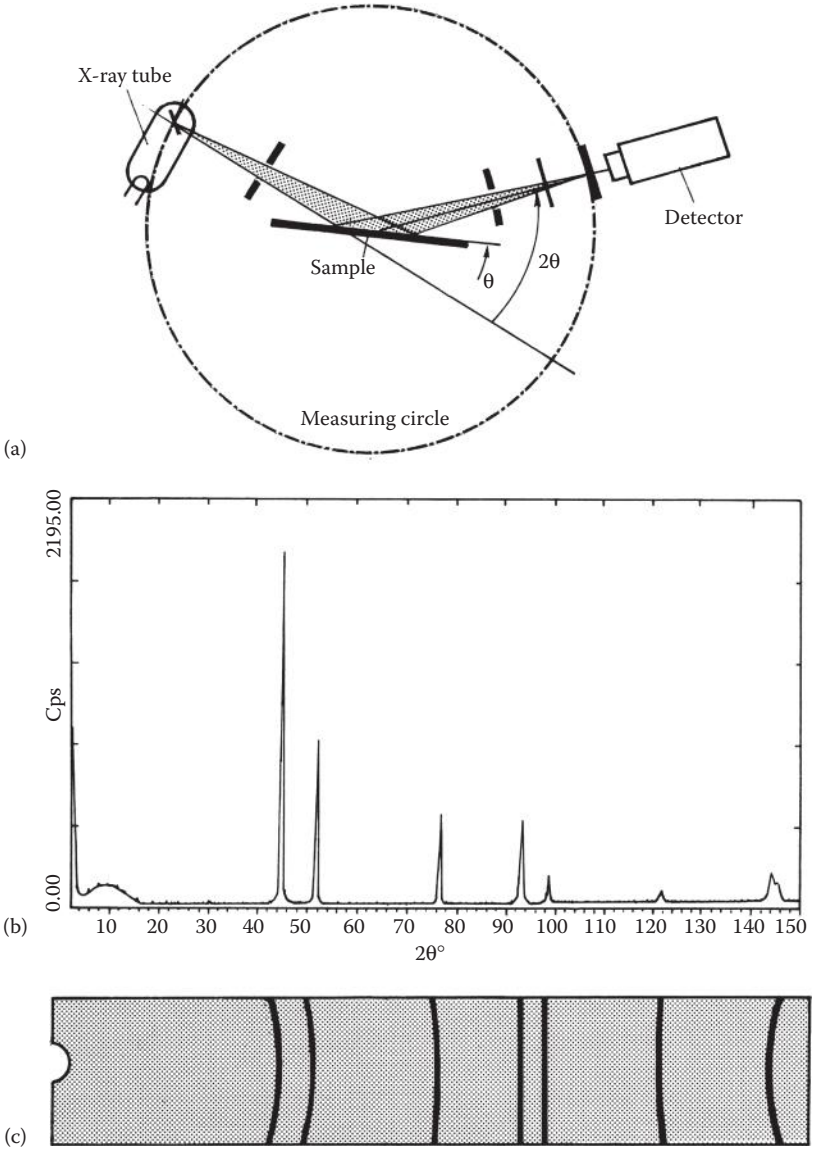


FIGURE 2.5 (a) Diagram of a powder diffractometer. (b) A powder diffraction pattern for Ni powder compared with (c) the Debye–Scherrer photograph of Ni powder.

largest d_{hkl} spacing. In the primitive cubic system, the *100* planes have the largest separation and therefore, give rise to this reflection, and as $a = b = c$ in a cubic system, the *010* and the *001* also reflect at this position. For the cubic system, with a unit cell dimension a, the spacing of the reflecting planes is given by Equation 2.5:

$$d_{hkl} = \frac{a}{\sqrt{(h^2 + k^2 + l^2)}} \tag{2.5}$$

Combining this with the Bragg equation gives

$$\lambda = \frac{2a \sin\theta_{hkl}}{\sqrt{(h^2 + k^2 + l^2)}}$$

and rearranging gives

$$\sin^2\theta_{hkl} = \frac{\lambda^2}{4a^2}(h^2 + k^2 + l^2) \tag{2.6}$$

For the primitive cubic class, all integral values of the indices h, k and l are possible. Table 2.1 shows the values of hkl in order of increasing value of $(h^2 + k^2 + l^2)$ and therefore of increasing sinθ values.

One value in the sequence, 7, is missing: this is because there are no possible integral values for $(h^2 + k^2 + l^2) = 7$. There are also other higher missing values where $(h^2 + k^2 + l^2)$ cannot be an integer (15, 23, 28, etc.), but note that this is only an arithmetical phenomenon and has nothing to do with the structure.

Taking Equation 2.6, if we plot the intensity of the diffraction of the powder pattern of a primitive cubic system against $\sin^2\theta_{hkl}$, we would get six equispaced lines with the 7th, 15th, 23rd, etc., missing. Consequently, it is easy to identify a primitive cubic system and, by inspection, to assign indices to each of the reflections.

The cubic unit cell dimension a can be determined from any of the indexed reflections using Equation 2.6. The experimental error in measuring the Bragg angle is constant for all angles; therefore, to minimise error, either the reflection with the largest Bragg angle is chosen, or, more often, a least-squares refinement to all the data is used.

The pattern of observed lines for the other two cubic crystal systems—body-centred and face-centred—is rather different from that of the primitive system. The differences arise because the centring leads to destructive interference for some reflections and these extra missing reflections are known as **systematic absences**.

Consider the *200* planes, shaded, in the F face-centred cubic unit cells, as shown in Figure 2.6; if a is the cell dimension, they have a spacing $a/2$. Figure 2.7 shows the reflections from four consecutive planes in this structure. The reflection from the *200* planes is exactly out of phase with the *100* reflection. Throughout the crystal,

TABLE 2.1

Values of $(h^2 + k^2 + l^2)$

hkl	*100*	*110*	*111*	*200*	*210*	*211*	*220*	*300 = 221*
$(h^2 + k^2 + l^2)$	1	2	3	4	5	6	8	9

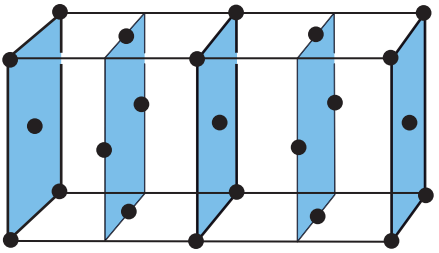


FIGURE 2.6 Two F-centred unit cells showing the *200* planes shaded.

there are equal numbers of the two types of planes with the result that complete destructive interference occurs and no *100* reflection is observed. Examining the reflections from all the planes for the F face-centred system in this way, we find that in order for a reflection to be observed, the indices must be either *all odd* or *all even*. A similar procedure for the body-centred cubic system finds that for reflections to be observed, the *sum of the indices must be even*.

It is possible to characterise the type of Bravais lattice present by the pattern of systematic absences. Although our discussion has centred on cubic crystals, these absences apply to all crystal systems, not only to cubic, and are summarised in Table 2.2. The allowed values of $h^2 + k^2 + l^2$ are listed in Table 2.3 for each of the cubic lattices.

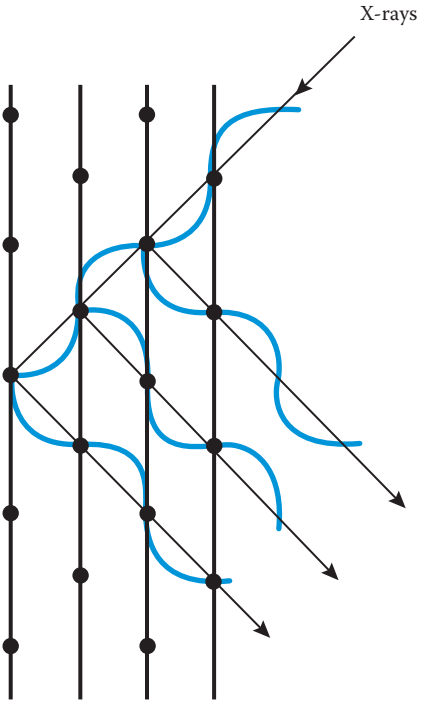


FIGURE 2.7 The *100* reflection from an F-centred cubic lattice.

TABLE 2.2
Systematic Absences due to Translational Symmetry Elements

Symmetry Element		Affected Reflection	Condition for Reflection To Be Present
Primitive lattice	P	*hkl*	None
Body-centred lattice	I	*hkl*	$h + k + l$ = even
Face-centred lattice	A	*hkl*	$k + l$ = even
	B		$h + l$ = even
	C		$h + k$ = even
Face-centred lattice	F		$h\,k\,l$ all odd or all even
2-fold screw, 2_1 along 4-fold screw, 4_2 along 6-fold screw, 6_3 along	*a*	*h*00	h = even
3-fold screw, 3_1, 3_2 along 6-fold screw, 6_2, 6_4 along	*c*	00*l*	l divisible by 3
4-fold screw 4_1, 4_3 along	*a*	*h*00	h divisible by 4
6-fold screw, 6_1, 6_5 along	*c*	00*l*	l divisible by 6
Glide plane perpendicular to	*b*	*h*0*l*	
Translation $\frac{a}{2}$ (*a* glide)			h = even
Translation $\frac{c}{2}$ (*c* glide)			l = even
$\frac{b}{2} + \frac{c}{2}$ (*n* glide)			$h + l$ = even
$\frac{b}{4} + \frac{c}{4}$ (*d* glide)			$h + l$ divisible by 4

Using these pieces of information and Equation 2.6, we can see that if the observed $\sin^2\theta$ values for a pattern are in the ratio 1 : 2 : 3 : 4 : 5 : 6 : 8 : …, then the unit cell is likely to be a primitive cubic unit cell, and the common factor is $\lambda^2/4a^2$.

A face-centred cubic unit cell can also be recognised: if the first two lines have a common factor, A, then dividing all the observed $\sin^2\theta$ values by A gives a series of numbers, 3, 4, 8, 11, 12, 16, …, and A is equal to $\lambda^2/4a^2$.

A body-centred cubic system gives the values of $\sin^2\theta$ in the ratio 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 : … with the values 7 and 15 apparently *not* missing, but now the common factor is $2\lambda^2/4a^2$.

2.3.3 Systematic Absences Due to Screw Axes and Glide Planes

The presence of translational symmetry elements in a crystal structure can be detected because they each lead to a set of systematic absences in the *hkl* reflections. Figures 1.20 and 1.21 show how a twofold screw (2_1) along *z* direction introduces

TABLE 2.3
Allowed Values of ($h^2 + k^2 + l^2$) for Cubic Crystals

Forbidden Numbers	Primitive (P)	Face-Centred (F)	Body-Centred (I)	Corresponding *hkl* Values
	1			*100*
	2		2	*110*
	3	3		*111*
	4	4	4	*200*
	5			*210*
	6		6	*211*
7				–
	8	8	8	*220*
	9			*221, 300*
	10		10	*310*
	11	11		*311*
	12	12	12	*222*
	13			*320*
	14		14	*321*
15				–
	16	16	16	*400*

a plane of atoms exactly halfway between the *001* planes: reflections from these planes will destructively interfere with the reflections from the *001* planes and the *001* reflection will be absent, as will any reflection for which *l* is odd. The effect of a glide plane (Figure 1.19) is to introduce a plane of atoms halfway along the unit cell in the direction of the glide. For an *a* glide perpendicular to *b*, the *10l* reflection will be absent, and, in general, the *h0l* reflections will only be present when *h* is even. Systematic absences are summarised in Table 2.2.

There now exist fairly powerful computer programs for indexing, and powder diffraction patterns can be indexed fairly readily for the high-symmetry crystal classes, such as cubic, tetragonal and hexagonal. For the other systems, the pattern often consists of a large number of overlapping lines, and indexing can be much more difficult or even impossible.

From the cubic unit cell dimension *a*, we can calculate the volume of the unit cell (*V*). If the density (ρ) of the crystals is known, then the mass of the contents of the unit cell (*M*) can also be calculated:

$$\rho = \frac{M}{V} \tag{2.7}$$

From the knowledge of the molecular mass, the number of molecules (*Z*) in the unit cell can be calculated. Examples of these calculations are in the questions at the end of the chapter.

The density of crystals can be determined by preparing a mixture of liquids (in which the crystals are insoluble!) such that the crystals neither float nor sink: the crystals then have the same density as the liquid. The density of the liquid can be determined in the traditional way using a density bottle.

2.3.4 Uses of Powder X-Ray Diffraction

Powder diffraction is difficult to use as a method for determining the structures of anything other than simple high-symmetry crystals because, as the structures become more complex, the number of lines increases so that overlap becomes a serious problem and it is difficult to index and measure the intensities of the reflections. It is usefully used as a fingerprint method for detecting the presence of a known compound or phase in a product. This is made possible by the existence of a huge library of powder diffraction patterns, which is held and regularly updated by the International Centre for Diffraction Data (ICDD). These files are available on CD-ROM. When the powder diffraction pattern of your sample has been measured and both the d_{hkl} spacings and the intensity of the lines have been recorded, these can be matched against the patterns of known compounds in the files. With modern diffractometers, the computer matches the recorded pattern of the sample to the patterns stored in the ICDD files (Figure 2.8).

The identification of compounds using powder diffraction is useful for qualitative analysis, such as mixtures of small crystals in geological samples. It also gives a rough check of the purity of a sample, determining which compounds are present and in what proportions—but note that powder diffraction does not detect amorphous products or impurities of less than about 5%.

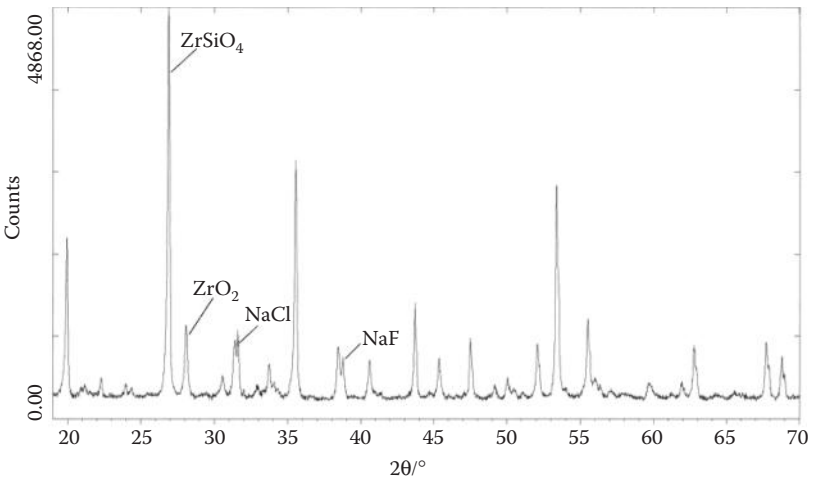


FIGURE 2.8 X-ray diffraction pattern for the preparation of zircon ($ZrSiO_4$) from zirconia (ZrO_2), silica (SiO_2) and sodium halide mineralisers. The peaks show zircon as the main product containing traces of all the starting materials.

Powder diffraction can confirm whether two similar compounds, where one metal substitutes for another for instance, have an isomorphous structure.

As the crystallite size decreases, the width of the diffraction peak increases. To either side of the Bragg angle, the diffracted beam will destructively interfere and we expect to see a sharp peak. However, the destructive interference is the resultant of the summation of all the diffracted beams, and close to the Bragg angle it gets diffracted from very many planes to produce complete destructive interference. In small crystallites, there are not enough planes to produce complete destructive interference, so we see a broadened peak.

The Debye–Scherrer formula enables the thickness of a crystallite to be calculated from the peak widths:

$$T = \frac{C\lambda}{B\cos\theta} = \frac{C\lambda}{\left(B_M^2 - B_S^2\right)^{1/2}\cos\theta} \tag{2.8}$$

where T is the crystallite thickness, λ is the wavelength of the X-rays (T and λ have the same units), θ is the Bragg angle and B is the full-width at half-maximum (FWHM) of the peak (radians) corrected for instrumental broadening. (B_M and B_S are the FWHMs of the sample and of a standard, respectively. A highly crystalline sample with a diffraction peak in a similar position to the sample is chosen, and this gives the measure of the broadening due to instrumental effects.)

This method is particularly useful for plate-like crystals with distinctive shear planes (e.g., the *111*) because measuring the peak width of this reflection gives the thickness of the crystallites perpendicular to these planes.

It is a common feature of solid-state reactions that reaction mixtures become more crystalline upon heating, as is evidenced by the X-ray diffraction pattern becoming sharper.

Powder X-ray diffraction is also a useful method for following the progress of a solid-state reaction and determining mechanisms and also for determining phase diagrams. By collecting an X-ray pattern at regular intervals as the sample is heated on a special stage in the diffractometer, evolving phases can be seen as new lines that appear in the pattern, with a corresponding decrease in the lines due to the starting material(s). Figure 2.9 follows the phase transition of ferrosilicon from the nonstoichiometric α-phase (Fe_xSi_2, $x = 0.77$–0.87) to the stoichiometric β-phase ($FeSi_2$), when it is held at 600°C for a period of time. In Figure 2.10, we see three powder diffraction patterns taken at different temperatures from a solid-state preparation of an Fe-doped zircon from powdered zirconia and silica, according to the equation: $ZrO_2 + SiO_2 = ZrSiO_4$. Sodium halides are used to bring the reaction temperature down, and ferrous sulfate is the source of iron; as the temperature of the reaction mixture is increased, the peaks due to zirconia and silica decrease while those of zircon increase, until at 1060°C, this is the major component. During the reaction, peaks due to intermediates such as Na_2SO_4 and Fe_2O_3 are also seen to evolve.

A careful comparison of the intensities of particular lines using standards, not only enables the different phases to be identified, but also enables the proportions of

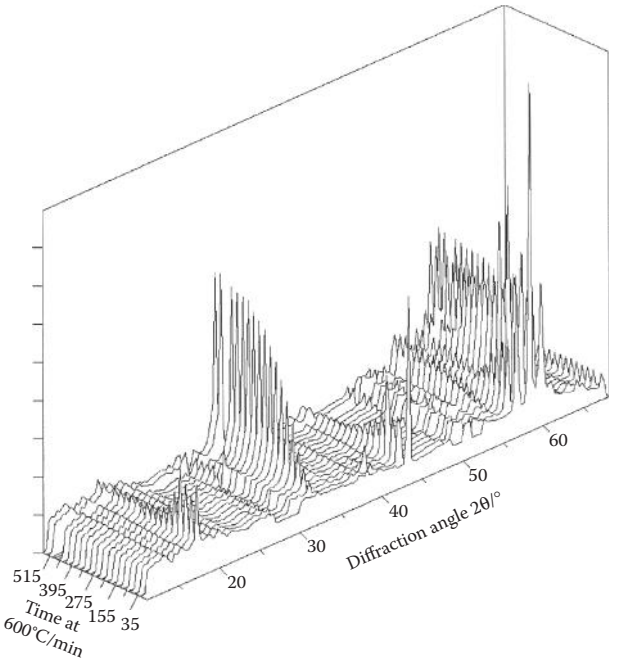


FIGURE 2.9 Powder XRD patterns show the phase changes in ferrosilicon with time, when heated at 600°C. (From Professor Berry, F.J., Open University. With permission.)

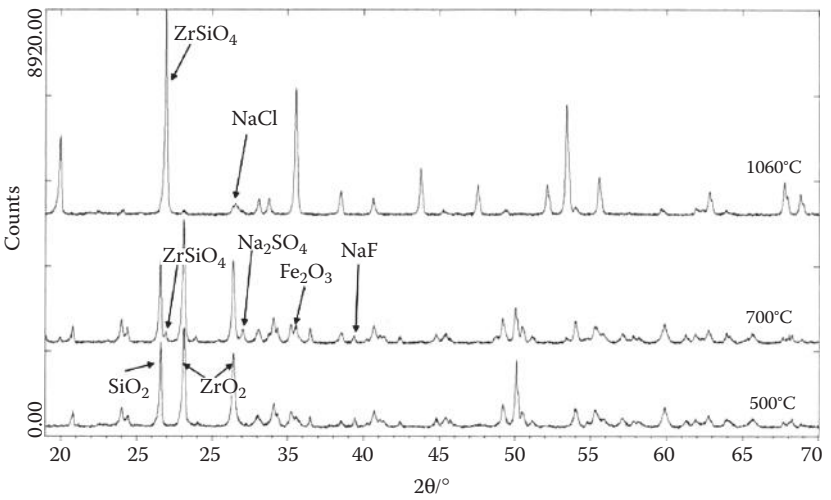


FIGURE 2.10 The phase evolution of iron-doped zircon ($ZrSiO_4$), from zirconia (ZrO_2), silica (SiO_2), ferrous sulfate ($FeSO_4$) and sodium halide mineralisers.

different phases to be determined so that a phase diagram can be constructed. In this way, it is possible to study the samples and phase changes under extreme conditions at high temperatures and pressures.

Rietveld method: In a high-symmetry crystal system, there are very few peaks in the powder pattern, and they are often well resolved and well separated. It is then possible to measure their position and intensity with accuracy and, by the methods we described earlier, index the reflections and solve the structure. For larger and less symmetrical structures, there are far more reflections that overlap considerably, and it becomes impossible to measure the intensities of individual peaks with accuracy.

A method known as **Rietveld analysis** has been developed for solving crystal structures from powder diffraction data. The Rietveld method involves an interpretation of not only the line position but also the line intensities, and because there is so much overlap of the reflections in the powder patterns, the method developed by Rietveld involves analysing the overall line profiles. Rietveld formulated a method of assigning each peak a Gaussian shape and then allowing the Gaussians to overlap so that an overall line profile could be calculated. The method was originally developed for neutron diffraction. In favourable cases, the Rietveld method can be used to solve a structure from its powder diffraction data. One starts by assuming a trial structure, calculating a powder diffraction profile from it and then comparing it with the measured profile. The trial structure can then be gradually modified by changing the atomic positions and refined until a best-fit match with the measured pattern is obtained. The validity of the structure obtained is assessed by an R factor and also by a difference plot of the two patterns, which should be a flat line. The method tends to work best if a good trial structure is already known, for instance if the unknown structure is a slight modification of a known structure, with perhaps one metal exchanged for another (Figure 2.11).

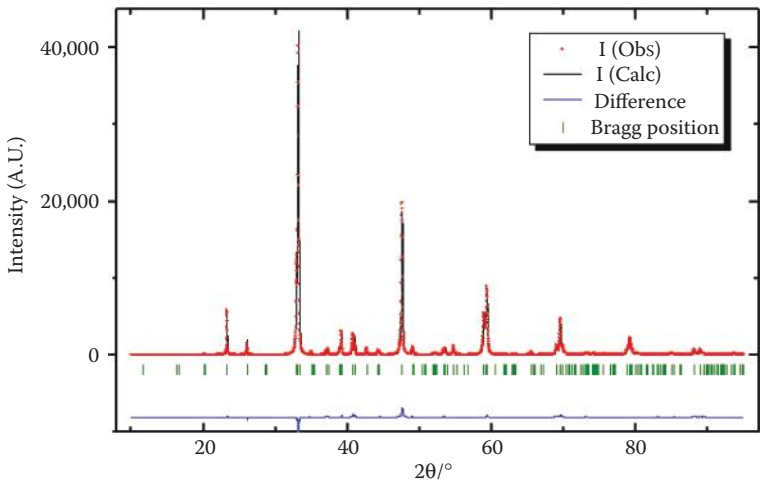


FIGURE 2.11 Rietveld analysis of perovskite with partial substitution of Ti with Ca. (From Reinaudi, L. et al., *Journal of Chemical Society, Dalton Transactions*, 4258–4262, 2000. Reproduced by permission of the Royal Society of Chemistry.)

2.4 SINGLE CRYSTAL X-RAY DIFFRACTION

From a single crystal, it is possible to measure accurately the position and intensity of the *hkl* reflections and, from these data, to determine not only the unit cell dimensions and space group, but also the precise atomic positions. In most cases, this can be done with speed and accuracy, and it is one of the most powerful structural techniques available to a chemist.

2.4.1 Importance of Intensities

So far, we have only discussed the effects of using crystals as three-dimensional diffraction gratings for X-rays. But you might wonder why one goes through all this trouble: if we want to magnify an object to see its structure in more detail, then why not use a lens system as in a microscope or a camera? Here, a lens system focusses the light that is scattered from an object (which if left alone would form a diffraction pattern) and forms an image. Why not use a lens to focus X-rays and avoid all the complications? The problem is that there is no suitable way in which X-rays can be focussed, so the effect of the lens has to be simulated by a mathematical calculation on the information contained in the diffracted beams. Much of this information is contained in the intensity of each beam; however, as always, there is a snag! The recording methods do not record *all* of the information in the beam because they record only the intensity and are insensitive to the *phase*. Intensity is proportional to the *square* of the amplitude of the wave, and the phase information is lost. Unfortunately, it is this information that derives from the atomic positions in a structure. When light is focussed by a lens, this information is retained.

So far, we have seen that if we measure the Bragg angle of the reflections and successfully index them, then we get information on the size of the unit cell and, if it possesses any translational symmetry elements, also on the symmetry. In addition, we have seen that the intensity of each reflection is different and this too can be measured. In early photographic work, the relative intensities of the spots on the film were assessed by eye with reference to a standard, and later a scanning microdensitometer was used. In modern diffractometers, the beam is intercepted by a detector—either a CCD plate or a scintillation counter—and the intensity of each reflection is recorded electronically.

The interaction that takes place between the X-rays and a crystal involves the electrons in the crystal: the more electrons an atom possesses, the more strongly it will scatter the X-rays. The effectiveness of an atom in scattering X-rays is called the **scattering factor** (or **form factor**) and is given the symbol f_0. The scattering factor depends not only on the atomic number, but also on the Bragg angle θ and the wavelength of the X-radiation: as the Bragg angle increases, the scattering power drops off. The decrease in scattering power with angle is due to the finite size of an atom; the electrons are distributed around the nucleus and as θ increases, the X-rays scattered by an electron in one part of the atom become increasingly out of phase with those scattered in a different part of the electron cloud (see Figure 2.14a).

Why are intensities important? A simple example demonstrates this clearly. We know that the heavier an atom is, the better it is at scattering X-rays. On the face of

it then, we might think that the planes containing the heavier atoms will give more intense reflections. While this is true, the overall picture is more complicated than that because there are interactions with the reflected beams from other planes to take account of, and these may produce *destructive interference*. Consider the diffraction patterns produced by NaCl and KCl crystals, both having the same structure (Figure 2.12). The structure can be thought of as two interlocking *ccp* arrays of Na^+ and Cl^- ions. The unit cell shown in Figure 2.12 has close-packed layers of Cl^- ions, which lie parallel to a body diagonal, with indices *111*. Lying exactly halfway in between the *111* layers and parallel to them are close-packed layers of Na^+ ions; this means that a reflection from the Cl^- close-packed layers is exactly out of phase with that from the equivalent Na^+ layers. Because a chloride ion has 18 electrons, it scatters the X-rays more strongly than a sodium ion with 10 electrons, the reflections partially cancel and the intensity of the *111* reflection will be fairly weak. The *222* layers contain the close-packed layers of both Na^+ and Cl^-, and this will be a strong reflection because the reflected waves will now reinforce one another. When we come to look at the equivalent situation in KCl, the reflection from the *111* layers containing K^+ ions is exactly out of phase with the reflection from the Cl^- close-packed layers. However, K^+ and Cl^- are isoelectronic, so their scattering factors for X-rays are virtually identical, and the net effect is that the two reflections cancel each other out and the *111* reflection appears to be absent. Similarly, this means that the first observed reflection in the diffraction pattern from KCl is the *200* reflection, and it would be very easy to make the mistake that this was the *100* reflection from a primitive cubic cell with a unit cell length half that of the real face-centred cell.

The resultant of the waves scattered by all the atoms in the unit cell, in the direction of the *hkl* reflection, is called the **structure factor (F_{hkl})**, and is dependent on both the position of each atom and its scattering factor. It is given by the general expression for *j* atoms in a unit cell:

$$F_{hkl} = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)} \tag{2.9}$$

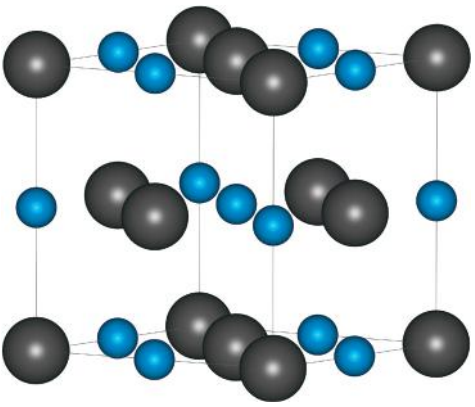


FIGURE 2.12 NaCl unit cell showing the close-packed *111* and *222* planes.

where f_j is the scattering factor of the jth atom and $x_j\ y_j\ z_j$ are its fractional coordinates. Series such as this can also be expressed in terms of sines and cosines, more obviously reflecting the periodic nature of the wave; they are known as **Fourier series**. In a crystal that has a centre of symmetry and n unique atoms in the unit cell (the unique set of atoms is known as the **asymmetric unit**), the expression in Equation 2.9 simplifies to

$$F_{hkl} = 2\sum_n f_n \cos 2\pi(hx_n + ky_n \dotplus lz_n) \tag{2.10}$$

The electron density distribution within a crystal can be expressed in a similar way as a three-dimensional Fourier series:

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx \dotplus ky \dotplus lz)} \tag{2.11}$$

where $\rho(x,y,z)$ is the electron density at a position $x\ y\ z$ in the unit cell and V is the volume of the unit cell. Notice the similarity between the expressions in Equations 2.9 and 2.11. In mathematical terms, the electron density is said to be the **Fourier transform** of the structure factors and vice versa. This relationship means that if the structure factors are known, then it is possible to calculate the electron density distribution in the unit cell and therefore, the atomic positions.

The intensity of the *hkl* reflections (I_{hkl}) is measured as described above and forms the data set for a particular crystal. The intensity of a reflection is proportional to the square of the structure factor:

$$I_{hkl} \propto F_{hkl}^2 \tag{2.12}$$

Taking the square root of the intensity gives a value for the *magnitude* of the structure factor (mathematically, this is known as the *modulus* of the structure factor and is denoted by the vertical bars on either side).

$$\left|F_{hkl}\right| \propto \sqrt{I_{hkl}} \tag{2.13}$$

Because the diffracted intensity is proportional to the square of the modulus of the structure factor ($|F_{hkl}|^2$), the intensities of the $\overline{hkl}$ and $\overline{hkl}$ reflections are equal. This is known as **Friedel's law**, and the *hkl* and $\overline{hkl}$ reflections are known as a **Friedel's pair**. The law only holds true when there is no absorption.

Before this information can be used, the data set has to undergo some routine corrections, a process known as **data reduction**. The **Lorentz correction (*L*)** relates to the geometry of the collection mode; the **polarisation correction (*p*)** allows for the fact that the nonpolarised X-ray beam may become partly polarised on

reflection, and an **absorption correction** is often applied to the data, particularly for inorganic structures, because the heavier atoms absorb some of the X-ray beam, rather than just scatter it. Corrections can also be made for **anomalous dispersion**, which affects the scattering power of an atom when the wavelength of the incident X-ray is close to its absorption edge. These corrections are applied to the scattering factor (f_0) of the atom.

The structure factor (and therefore, the intensity of a reflection) is dependent on both the position of each atom and its scattering factor. The structure factor can be *calculated*, therefore, from the knowledge of the types of atoms and their positions, using equations such as Equation 2.9 or 2.10. It is the great problem of X-ray crystallography that we need to be able to do the reverse of this calculation—we have the measured magnitudes of the structure factors and from them we have to calculate the atomic positions. However, there is the snag, which we mentioned earlier, known as the **phase problem**. When we take the square root of the intensity, we only obtain the modulus of the structure factor, so we only know its magnitude and not its sign. The phase information is unfortunately lost, and we need it in order to calculate the electron density distribution and therefore, the atomic positions.

2.4.2 Solving Single Crystal Structures

It would seem to be an unresolvable problem—to calculate the structure factors, we need the atomic positions, and to find the atomic positions, we need both the amplitude and the phase of the resultant waves, but we only have the amplitude. Fortunately, over the years, many scientists have worked at finding ways around this problem and have been extremely successful, to the extent that for many systems the solving of the structure has become a routine and fast procedure.

Single crystal X-ray diffraction data are nowadays collected using a computer-controlled diffractometer, which measures the Bragg angle θ and the intensity I for each hkl reflection. Modern diffractometers employ a flat-plate detector (CCD), so that all the reflections can be collected and measured simultaneously. A full data set, that is, the intensities of all hkl reflections, which can be thousands of reflections, can be accumulated in hours rather than the days or weeks of earlier times. A full data set is not usually collected because to reduce collection time, crystallographers only collect a unique set of data that are symmetry-related to all the rest—the higher the symmetry of the crystal, the smaller the set needed. It is important, however, to collect all the data, including the Friedel's pairs, if the properties of anomalous dispersion are to be exploited in determining the absolute configuration of a structure.

To summarise what we know about a structure:

- The size and shape of the unit cell are determined, usually from rotation photographs and scanning routines directly on the diffractometer.
- The reflections are indexed and from the systematic absences, the Bravais lattice and the translational symmetry elements of the structure are determined: this information often determines the space group

unequivocally or narrows the possibilities down to a choice of two or three.
- The intensities of the indexed reflections (or a symmetry-related subset) are measured and stored as a data file.
- Correction factors are applied to the raw intensity data.
- The square roots of the corrected data are taken to give a set of observed structure factors. These are known as $\boldsymbol{F}_{\mathbf{obs}}$ or $\boldsymbol{F}_{\mathbf{o}}$.
- In order to calculate the electron density distribution in the unit cell, we need to know not only the magnitudes of the structure factors, but also their phases.

Crystal structures are solved by creating a set of trial phases for the structure factors. There are the two main methods for doing this. The first is known as the **Patterson method**, which relies on the presence of at least one (but not many) heavy atom in the unit cell, so is useful for solving many inorganic molecular structures; the second is called **direct methods**, which is best used for structures where the atoms have similar scattering properties. Direct methods calculate mathematical probabilities for the phase values and hence, an electron density map of the unit cell; theoreticians have produced packages of accessible computer programs for solving and refining structures.

Once the atoms in a structure have been located, a calculated set of structure factors ($\boldsymbol{F}_{\mathbf{calc}}$ or $\boldsymbol{F}_{\mathbf{c}}$) are determined for comparison with the $\boldsymbol{F}_{\mathbf{obs}}$ magnitudes, and the positions of the atoms are refined using **least-squares methods**, for which standard computer programs are available. In practice, atoms vibrate about their equilibrium positions; this is often called thermal motion, although it depends not only on the temperature, but also on the mass of the atom and the strengths of the bonds holding it. The higher the temperature, the larger is the amplitude of the vibration and the electron density is spread out over a larger volume, thus causing the scattering power of the atom to fall off more quickly. Part of the refinement procedure is to allow the electron density of each atom to refine in a sphere around the nucleus. Structure determinations usually quote an adjustable parameter, known as the **isotropic displacement parameter**, $\boldsymbol{B}$ (also called the **isotropic temperature factor**). The electron density of each atom can also be refined within an ellipsoid around its nucleus when an **anisotropic displacement parameter** correction is applied which has six adjustable parameters.

The residual index, or $\boldsymbol{R}$ **factor**, gives a measure of the difference between the observed and calculated structure factors and, therefore, how well the structure has refined. It is defined as

$$R = \frac{\sum |(|F_o| - |F_c|)|}{\sum |F_o|} \tag{2.14}$$

and is used to give a guide to the correctness and precision of a structure. In general, the lower the R value, the better is the structure determination. R values have to be used with caution, as it is not unknown for structures to have a low R value and still be wrong, although fortunately this does not happen often. There are no hard and

fast rules for the expected value of *R*, and interpreting them is very much a matter of experience. For small molecule structures, it is usually taken as a rule of thumb that a correct structure for a reasonable quality data set would refine to well below an *R* of 0.1, and anything above should be viewed with some degree of suspicion. That said, most structures, nowadays, if collected from good-quality crystals on modern diffractometers, would usually refine to below an *R* of 0.05 and often to below an *R* of 0.03.

A good structure determination, as well as having a low *R* value, will also have low standard deviations on both the atomic positions and the bond lengths calculated from these positions. This is probably a more reliable guide to the quality of the refinement.

When a single crystal of a solid can be produced, X-ray diffraction provides an accurate, definitive structure, with bond lengths determined to tenths of a picometre. In recent years, the technique has been transformed from a very slow method reserved only for the most special structures, to a method of almost routine analysis: with modern machines, suites of computer programs and fast computers are used to solve several crystal structures per week.

It is now standard academic practice to deposit final published structures and their data in a crystallographic database, of which there are several: the Cambridge Structural Database, CSD, for small organic and organometallic molecules; the Inorganic Crystal Structure Database, ICSD; CRYSTMET for metals and alloys; the Protein Data Bank, PDB; and the Nucleic Acid Database, NDB. These databases check the deposited structure and its data through their own software to ensure internal consistency. Unfortunately, commercially obtained structures such as those by pharmaceutical companies do not usually appear in these databases.

2.4.3 High-Energy X-Ray Diffraction

The problem of solving the structure of solids that have limited long-range structures has been there for a long time, for instance, in glasses. With the recent upsurge in the preparation of new nanocrystalline materials, there is an increased incentive to be able to solve such structures. A nanoparticle is usually considered to have at least one dimension less than 100 nm. In solids, the limited order on a nanoscale can be due to several reasons: it can be composed of nanoparticles such as buckytubes or sheets; there may be small nanometre-sized ordered areas within larger structures; mesoporous materials have nanometre-size holes (Chapter 7); and nanocomposites are composed of intimate mixtures of nanoparticles. Later sections in this chapter deal with electron microscopy and scanning probe techniques that map the surface-layer materials, but determining the short-range, three-dimensional structure in such materials has presented a real challenge, which has only recently begun to be solved.

Scientists have been turning to X-ray diffraction using synchrotron sources of X-rays or neutron diffraction (see Section 2.5). Synchrotrons produce monochromatic beams of very high energy (and therefore, short wavelength) X-rays

(see Section 2.9 for more detail), which collect over smaller scattering angles and give a higher resolution.

As we have seen, the planes of atoms in a perfectly ordered crystal reflect an X-ray beam giving a sharp spot—the Bragg reflection. In a solid that has only a small volume of ordered structure, the Bragg reflections will be fewer in number, weaker and much less sharp and will be superimposed on a diffuse background from the scattering of the rest of the structure.

The high-energy X-ray diffraction experimental data are collected in much the same way as for powder diffraction, but usually over longer periods of time, and have to be corrected for other factors that contribute to the intensity, such as background scatter, polarisation, absorption and the Compton effect (ionisation of an electron from the sample due to inelastic scattering).

Using a technique called the **atomic pair distribution function (PDF)**, it has proved possible to analyse such patterns and determine the interatomic distances and coordination numbers within the ordered areas. The PDF is based on all the atomic scattering, both sharp and diffuse, and it calculates the probability of finding atomic pairs at a particular distance. As the data do not identify particular atoms, computer models are then fitted to the data to determine the structure. Figure 2.13 shows the results obtained for a molybdenum, sulfur and iodide compound ($Mo_6S_xI_{10-x}$), which forms a nanowire that is not fully crystalline; in Figure 2.13(i) the sharpness of the PDF-refined peaks is shown by comparison with the powder diffraction pattern for crystalline (MoS_2), and in Figure 2.13(ii) the best-fit structure for the nanowire shows that sulfur fills the core and is surrounded by outer iodines.

2.5 NEUTRON DIFFRACTION

Neutron diffraction, the elastic scattering of neutrons from a solid, can also be used for crystallographic structural determinations. The technique was first developed in 1945 at Oak Ridge by Wollan, and he was soon joined by Schull in 1946. It was not until 1994 that Schull received his Nobel Prize for this work (Wollan having died), notable for being the longest gap between doing the work and getting the award! The vast majority of crystal structures published in the literature have been solved using X-ray diffraction; neutrons are much less commonly used because there are very few sources of neutrons available, whereas X-ray diffractometers can be housed in any laboratory. In addition, the flux of a monochromatic source of neutrons is small, and this necessitates the use of large single crystals and long counting times for the experiment, in order to get sufficient intensity. Crystals, typically, have needed to be at least 1 mm in each direction, and it can be extremely difficult if not impossible to grow such large, perfect crystals, and most experiments use powder diffraction, or as we saw in the preceding section, diffraction of nanoparticles. However, new high-energy neutron sources are becoming available, such as the one at Grenoble, and the need for these large single crystals in neutron studies is receding. Neutron diffraction does have different properties that can give certain advantages: it can yield very precise lattice information, determine the position of light atoms, especially hydrogen, and map the structural magnetic data.

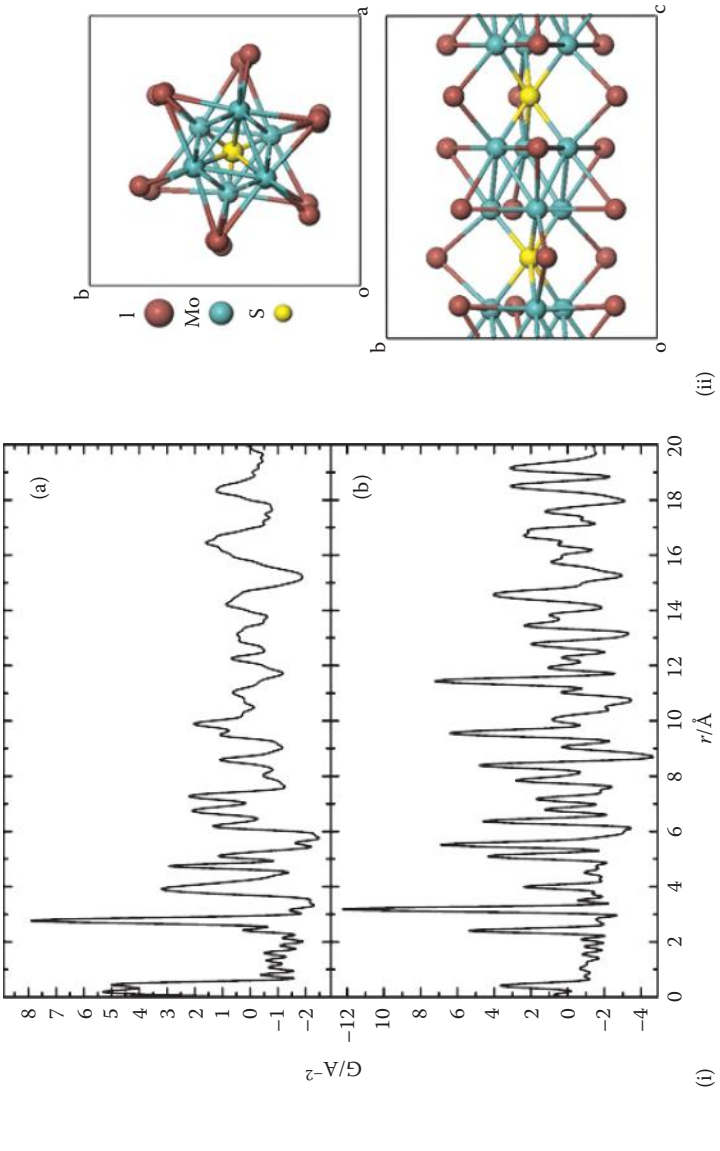


FIGURE 2.13 (i) Experimental PDF for $Mo_6S_xI_{10-x}$ (a) compared with crystalline MoS_2 (b). (ii) The best-fit structure for the $Mo_6S_xI_{10-x}$. (Adapted from Paglia, G., et al., *Chemistry Materials*, **18**, 100, 2006. With permission.)

The de Broglie relationship states that any beam of moving particles will display wave properties according to the formula:

$$\lambda = \frac{h}{p} \tag{2.15}$$

where λ is the wavelength, p is the momentum of the particles ($p = mv$, mass × velocity) and h is Planck's constant. Neutrons are released in atomic fission processes from a uranium target, when they have very high velocities and very small wavelengths. The neutrons generated in a nuclear reactor can be slowed using heavy water so that they have wavelengths of about 0.1 nm (1Å) and are therefore suitable for structural diffraction experiments. The neutrons generated have a spread of wavelengths, and a monochromatic beam is formed using reflection from a plane of a single-crystal monochromator at a fixed angle (according to Bragg's law). Structural studies need a high flux of neutrons, and this usually means that the only appropriate source is a high-flux nuclear reactor such as at Brookhaven and Oak Ridge in the United States and at Grenoble in France.

Alternative spallation sources are also available, such as the Rutherford Laboratory in the United Kingdom, where the neutrons are produced by bombarding metal targets with high-energy protons. The diffraction experiments we have seen so far are set up with X-rays of a single wavelength (λ) so that, in order to collect all the diffracted beams, the Bragg angle θ is varied (Bragg equation: $n\lambda = 2d \sin\theta$). With the spallation source, the whole moderated beam with all its different wavelengths is used at a fixed angle and the diffraction pattern is recorded as the function of the time of flight of the neutrons (if we substitute $v = D/t$, velocity = distance/time, in the de Broglie relationship, we see that the wavelength of the neutrons is proportional to t: $\lambda = (ht/Dm)$). Because this method uses all of the beam, it has the advantage of greater intensity.

The difference between the X-ray and the neutron diffraction techniques lies in the scattering process: X-rays are scattered by the electron cloud around the nucleus, whereas neutrons are scattered by the nucleus. The scattering factors for X-rays therefore increase linearly with the number of electrons in the atom, so that heavy atoms (large atomic number Z) are much more effective at scattering than light atoms (small Z); however, because of the size of the atoms relative to the wavelength of the X-rays, the scattering from different parts of the cloud is not always in phase, so the scattering factor decreases with $\sin\theta/\lambda$ due to the destructive interference (Figure 2.14a). Because the nucleus is very small, neutron scattering factors do not decrease with $\sin\theta/\lambda$, and because nuclei are similar in size, they are all fairly similar in value and that of hydrogen is anomalously large due to the nuclear spin. The fact that the neutron scattering factors are almost invariant with $\sin\theta/\lambda$ means that the intensity of the data does not drop off at high angles of θ as is the case with X-ray patterns, so a neutron powder pattern tends to yield up considerably more data. Neutron scattering factors are also affected randomly by **resonance scattering**, when the neutron is absorbed by the nucleus and released later. This means that neutron scattering factors cannot be predicted

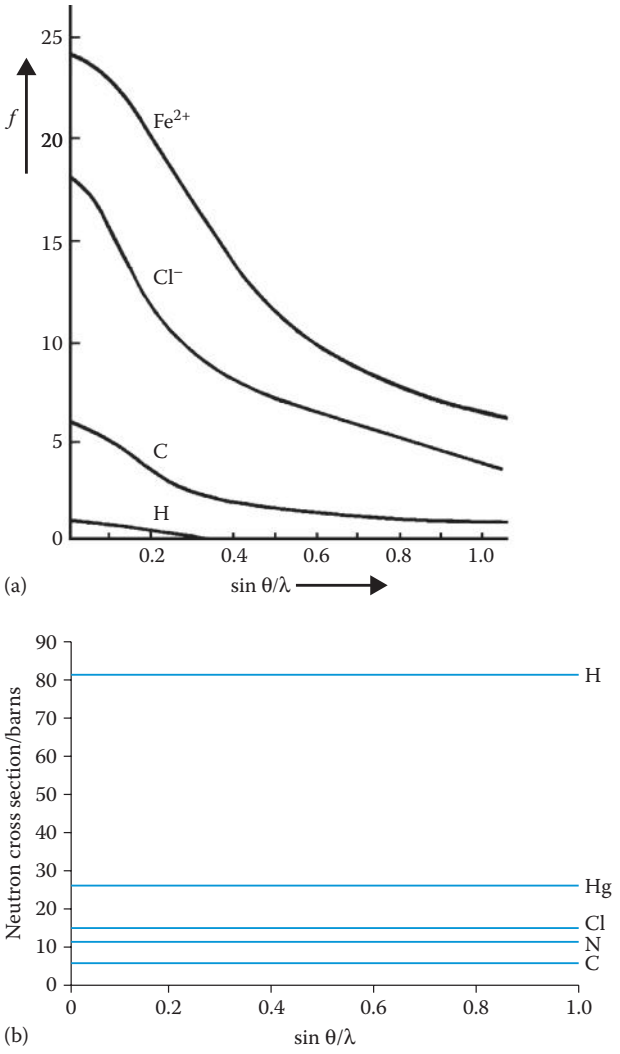


FIGURE 2.14 (a) X-ray scattering factors for hydrogen, carbon, chloride and ferrous ions and (b) the neutron scattering cross sections for several elements, as a function of sin θ/λ.

but have to be determined experimentally, and they vary for different atoms and indeed for different isotopes (Figure 2.14b).

Note that because of the different scattering mechanisms, the bond lengths determined by X-ray and neutron studies will be slightly different. The neutron determination will give the true distance between the nuclei, whereas the X-ray values are distorted by the size of the electron cloud and therefore, are shorter.

The excellent high-angle data obtained from neutron diffraction means that atomic positions are determined very precisely; therefore, structures are solved to

a much higher resolution. The technique of Rietveld profile analysis is used to solve the structures of crystalline powders—in fact, this technique originated with neutron powder diffraction.

2.5.1 Uses of Neutron Diffraction

The fact that neutron scattering factors are similar for all elements means that light atoms scatter neutrons as effectively as heavy atoms and can therefore be located in the crystal structure; for example, the X-ray scattering factors for deuterium and tungsten are 1 and 74, respectively, whereas the equivalent neutron values are 0.667 and 0.486. This property is particularly useful for locating hydrogen atoms in a structure, which can sometimes be difficult to do in an X-ray determination, especially if the hydrogen atoms are in the presence of a heavy metal atom. Accordingly, many neutron studies in the literature have been done with the express purpose of locating hydrogen atoms or of exploring hydrogen bonding.

Neutrons are not absorbed by the crystals, so they are also useful for studying systems containing heavy atoms that absorb X-rays very strongly.

Atoms that are close to each other in the periodic table have very similar X-ray scattering factors and cannot always be distinguished in an X-ray structure determination, for example, oxygen and fluorine, or similar metals in an alloy. A neutron structure determination may be able to identify atoms with similar atomic numbers.

Apart from the scattering of the neutrons by the nuclei, there is additional magnetic scattering of the neutrons from paramagnetic atoms. This arises because a neutron has spin, and therefore possesses a magnetic moment that can interact with the magnetic moment of an atom. The atomic magnetic moment is due to the alignment of the electron spins, and so this interaction, like the scattering of X-rays, falls off with the increasing Bragg angle due to the size of the electron cloud. As you will see in Chapter 8, the magnetic moments of a paramagnetic crystal are arranged randomly; however, in ferromagnetic, ferrimagnetic and antiferromagnetic substances, the atomic magnetic moments are arranged in an ordered fashion. In ferromagnetic substances, the magnetic moments are arranged so that they all point in the same direction, thereby reinforcing one another; in antiferromagnetic substances, the magnetic moments are ordered so that they completely cancel one another out; and in ferrimagnetic substances, the ordering leads to a partial cancellation of magnetic moments. Magnetic scattering of a polarised beam of neutrons from these ordered magnetic moments gives rise to magnetic Bragg peaks. For instance, the structure of NiO, as determined by X-ray diffraction, is the same as that of NaCl. In the neutron study, however, below 120 K, extra peaks appear due to the magnetic interactions; these give a magnetic unit cell that has a cell length *twice* that of the standard cell. This arises because the alternate close-packed layers of Ni atoms have their magnetic moments aligned in opposing directions, giving rise to antiferromagnetic behaviour (Figure 2.15).

Neutron data can yield such accurate lattice constants that the differences for a metal under stress can be mapped. This is used to analyse stresses in aero and car components.

Increasing attention is now being given to determining the structures of nanocrystalline materials.

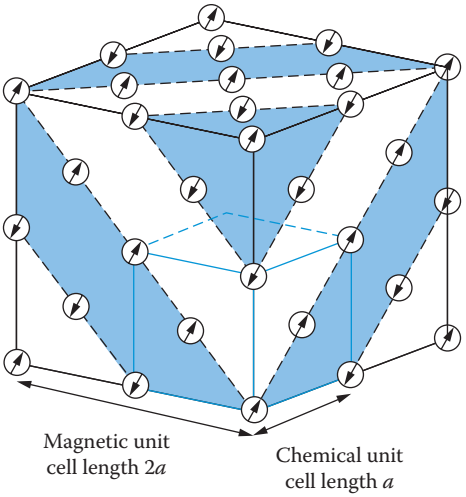


FIGURE 2.15 The magnetic ordering in NiO. Only the Ni planes are shown, and alternate close-packed layers have opposing magnetic moments. Note that the magnetic unit cell length is double that of the normal unit cell.

2.6 ELECTRON MICROSCOPY

Electron microscopy is widely used in the characterisation of solids to study structure, morphology and crystallite size, to examine defects, and to determine the distribution of elements. The basic principle underlying an electron microscope is similar to that of an optical microscope, except that a beam of accelerated electrons is used rather than visible light. Because of the wave particle duality of electrons, they behave like electromagnetic radiation and, at high energies, have very short wavelengths (λ). The electron wavelength imposes a lower limit on the resolution of the microscope. The electron wavelength can be readily calculated:

After acceleration through a voltage V, the kinetic energy of the electron is $\frac{1}{2}mv^2 = eV$. We can write $v = p/m$, where p is the momentum of the electron, so that $p^2/2m = eV$, where e and m are the charge and mass of the electron, respectively. Using the de Broglie relation, $p = h/\lambda$, from Equation 2.15, where h is Planck's constant, we get $h^2/\lambda^2 = eV$. Rearranging, $\lambda = h/\sqrt{2meV}$ and inserting values for h, m and e, we get

$$\lambda = 1.23(V)^{-1/2} \text{ nm} \tag{2.16}$$

(This neglects relativistic corrections, which become significant as the kinetic energy approaches the electron rest mass energy of 511 keV.) For an accelerating potential V of 100 kV, $\lambda = 0.0123$ nm (0.123 Å).

As we will see in the following pages, the lower limit of resolution of even the best electron microscopes is currently well below the actual resolutions achieved, unlike optical microscopes where resolutions close to their limit are obtainable. This reflects the fact that optical lenses are superior to their electromagnetic counterparts.

The electron beam is produced by heating a tungsten filament, a lanthanum hexaboride (LnB_6) crystal, or from a field emission gun (FEG), which uses a cathode of either tungsten or zirconium oxide. The beam is focussed by magnetic coil magnets in a high vacuum (the vacuum prevents interaction of the beam with any extraneous particles) to a fine spot. Detection can be by scintillation counter, film or CCD. Resolution down to subnanometre levels can be achieved in the best instruments.

2.6.1 Scanning Electron Microscopy

In this technique, high-energy monochromatic electrons, generated at a potential of 1–30 kV, are formed into a finely focussed beam, which is rastered across the surface of the sample. Figure 2.16 shows a schematic diagram of the instrument. Electrons are generated at high voltage and are focussed by a condenser lens; the high-angle electrons are excluded by an aperture, and the beam is then further focussed by a second condenser lens. A set of coils sweep the beam across the surface of the sample in a grid, and an objective lens finally focusses the beam onto the sample. The electrons impact on the atoms within an *interaction volume* close to the surface, and the

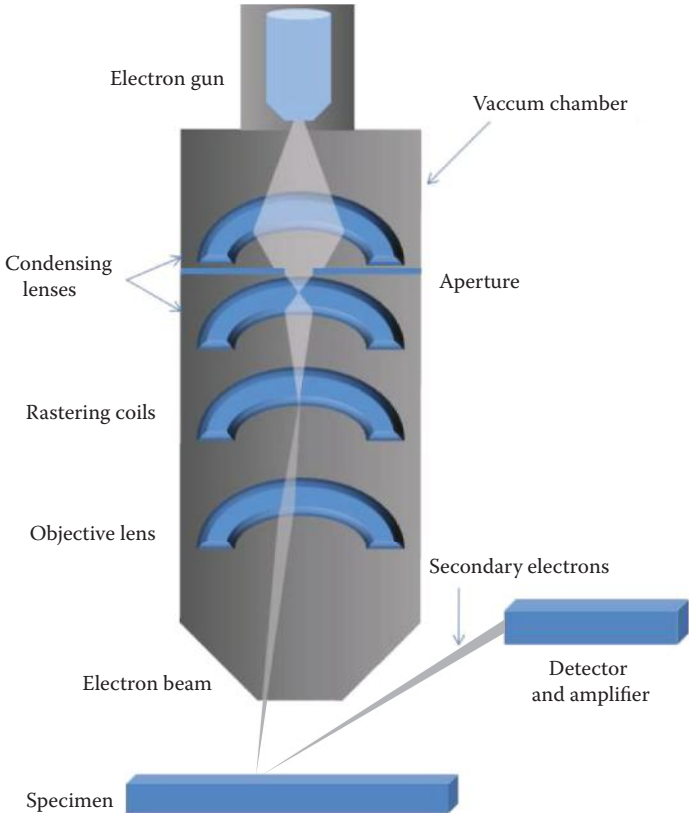


FIGURE 2.16 Schematic diagram of an SEM instrument.

signals produced from these atoms include back-scattered electrons (the reflection of electrons by elastic scattering), emitted secondary electrons from inelastic scattering and characteristic X-rays (due to the production of secondary electrons), among others. An SEM image of the structure is mainly due to the low-energy secondary electrons: those produced deep within the sample will be absorbed, but those closer to the surface will escape and can be collected and detected. These secondary electrons are accelerated towards a scintillation detector, amplified by a photomultiplier and then displayed; the number of electrons determines the brightness of the image.

The back-scattered electrons have higher energy, and when these escape from the surface, they are not deflected into the image detector. However, these electrons can also be collected, and as the higher atomic number elements back-scatter the electrons more strongly, this information, together with the characteristic X-rays emitted from the inelastic scattering, enables both qualitative and quantitative elemental analysis of the surface.

Detection of the different signals can produce maps of the surface topography of samples such as catalysts, minerals and polymers. The maps are useful for determining particle size, crystal morphology, magnetic domains, surface defects (Figure 2.17) elemental analysis and other properties such as electrical conductivity. A large depth of field gives a three-dimensional appearance to the image of the surface. A wide range of magnification can be used, anything from ×10 to ×500,000, and features down to sizes of about 1 nm and less can be resolved in the best instruments.

The samples need to be conducting, at least at the surface, and nonconducting samples may need to be coated with gold, graphite or similar substance; electrical earthing prevents charge building up on the surface.

FIGURE 2.17 SEM showing crystals of $VSbO_4$ growing out of β-Sb_2O_4 following reaction with V_2O_5 (bar = 40 μm). (From Professor Berry, F.J., Open University. With permission.)

The resolution of SEM relies on the size of the focussed beam and on the lens system, and the two have to be balanced carefully to give the best results: the higher the current in the condenser lens, the smaller is the spot and therefore, the higher is the resolution because it is more sensitive to smaller features on the surface; however, this is at the expense of fewer electrons reaching the detector. The technique is not usually good enough to resolve individual atoms, as can the transmission electron microscope, which uses higher energy (and therefore, shorter wavelength) electrons (see Section 2.6.2). SEMs are widely available in laboratories and usually achieve a resolution from 0.5 to <20 nm, depending on the instrument used.

2.6.2 Transmission Electron Microscopy

Transmission electron microscopes (TEMs) provide significantly higher resolution than a conventional microscope using visible light, in addition to SEM, due to the very high energy electrons generated; resolutions of less than 0.1 nm (<1 Å) at magnifications of 50 million and more have been achieved. A TEM produces a direct image of a solid at high magnification from the transmission of electrons through an ultrathin sample, using an electron gun and an assembly of electromagnetic lenses, all under high vacuum. A very thin sample, hundreds of nanometres thick at most, is prepared and subjected to a high-energy, high-intensity beam of electrons, which are usually generated using either a tungsten filament or a lanthanum hexaboride crystal and subjected to a potential of 100–300 kV; after acceleration, the electrons are formed into a beam by two condenser lenses and then focussed onto the sample by an objective lens. Additional complex magnetic lenses (quadrupole, hexapole and octupole) can be incorporated to correct distortions in the beam.

Some of the electrons strike the atoms of the surface and are deflected (the deflected electrons may be elastically or inelastically scattered), whereas those that pass through the sample are detected and form a two-dimensional projection: the emergent beam is expanded by projector lenses onto a detector, such as a fluorescent screen, film or CCD. The instrument can be operated to select either the direct beam (bright-field image) or the diffracted beam (dark-field image).

Most commonly, when the primary beam is collected after passing through the sample, the image is seen as a dark feature on a bright background (**bright-field mode**); areas of the sample where more electrons are absorbed by the higher atomic number elements or thicker parts of the sample are seen as dark, and, effectively, a two-dimensional projection of the sample is obtained (Figure 2.18).

In the **dark-field mode**, however, a detector can be placed *in front* of the sample to collect the coherent back-diffracted electrons, the Bragg reflections. An aperture excludes the primary electron beam, and by tilting the specimen, particular Bragg reflections can be selected, and the image formed is of a dark background with bright areas from the selected reflection(s). This method is particularly useful for studying the position and type of lattice defects (see Chapter 5).

High-resolution instruments use methods of obtaining the phase of the electron beam as well as the amplitude (a technique sometimes called **high-resolution**

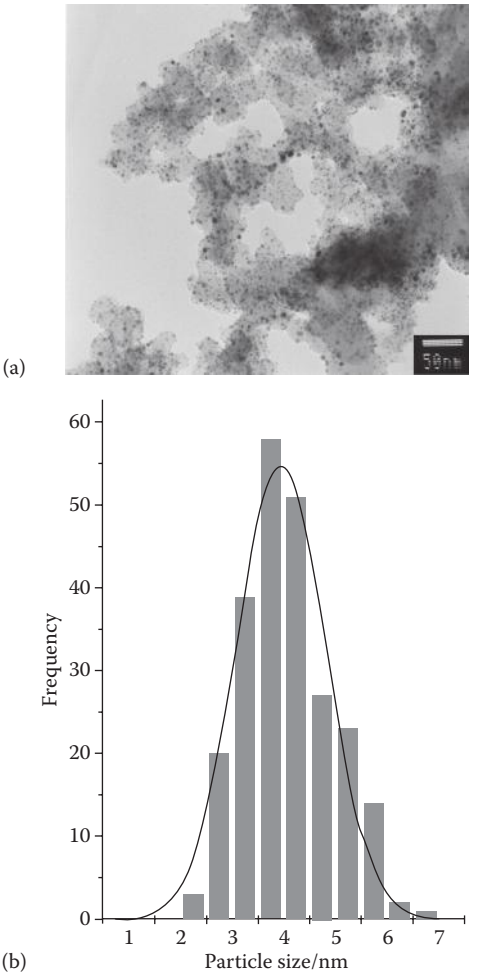


FIGURE 2.18 (a) TEM image of a supported Pt/Cr bimetallic catalyst on C. (b) Analysis of the metal particle sizes of this catalyst.

transmission electron microscopy, **[HRTEM]**). A large aperture allows the incident beam to pass directly through the sample and the scattered beams to pass into a detector. The coherent elastically scattered beams interfere and form an image (the so-called *phase contrast*), which can be interpreted to give information on the atomic position.

Because the electrons pass through the sample, TEM/HRTEM images the bulk structure and can detect crystal defects, such as phase boundaries, shear planes, etc. In the highest-resolution instruments, a resolution of less than 0.1 nm (<1 Å) is achievable, allowing individual atoms and defects to be imaged (Figure 2.19). By combining the views of a crystal taken at different angles, a three-dimensional crystal structure can also be obtained.

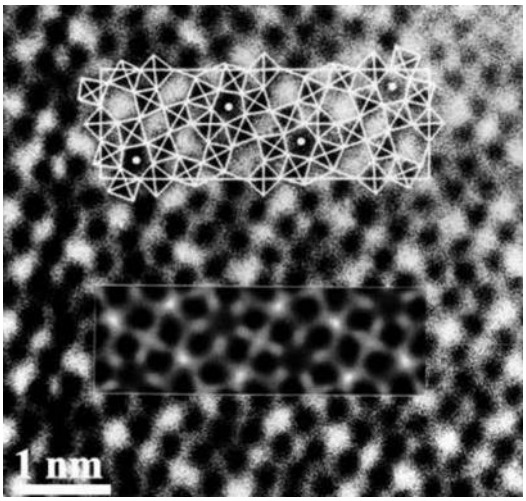


FIGURE 2.19 HRTEM image of $Nb_4W_{13}O_{47}$ tungsten bronze along *011*. The insets show the structural model and a simulation. (From Dr. Frank Krumeich. With permission.)

The resolution limits of TEM are dictated by aberrations in the lens system used, which affect the focussing of the beam. There are three main problems: **astigmatism**, **chromatic aberration (C_c)** and **spherical aberration (C_s)**. Astigmatism is due to the cross section of the beam being elliptical rather than circular. Chromatic aberration occurs because electrons are not quite monochromatic but have a range of energies and, therefore, wavelengths. Spherical aberration is shown diagrammatically in Figure 2.20, and it occurs because, as the electrons pass through the condensing lenses, the outer electrons of the beam, which are diffracted through larger angles, are brought to a focus at a different point on the axis from those in the centre, which are diffracted through small angles, creating a focal disc rather than a

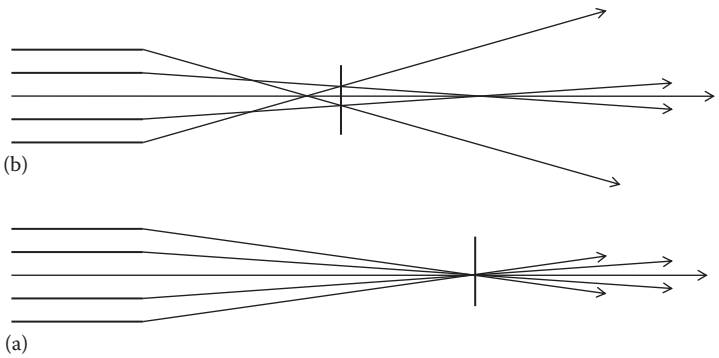


FIGURE 2.20 Diagram showing (a) electrons condensed to a single focal point and (b) electrons focussed to a disc by a lens with spherical aberration.

spot. Modern advances now enable fine corrections to be made for aberrations (see Section 2.6.4) such that a resolution below 0.05 nm (<0.5 Å) has been achieved at the time of writing (2010).

2.6.3 Energy Dispersive X-Ray Analysis

As we saw in Section 2.2.1, an electron beam incident on an element gives rise to the emission of characteristic X-rays. When an electron in an inner shell is excited and ejected from the shell, an electron hole is created; an electron from a higher-energy shell drops down to fill the hole, releasing the excess energy in the form of an X-ray. In the electron microscope, the electrons that are scattered inelastically, the secondary electrons, cause the different elements present to emit characteristic X-rays. These can be separated by a silicon–lithium detector placed within the objective lens system, and each signal can be collected, amplified and corrected for absorption and other effects, to give both qualitative and quantitative analysis of the elements present (for elements of atomic number greater than 11), a technique known as **energy dispersive analysis of X-rays (EDX or EDAX)** (Figure 2.21).

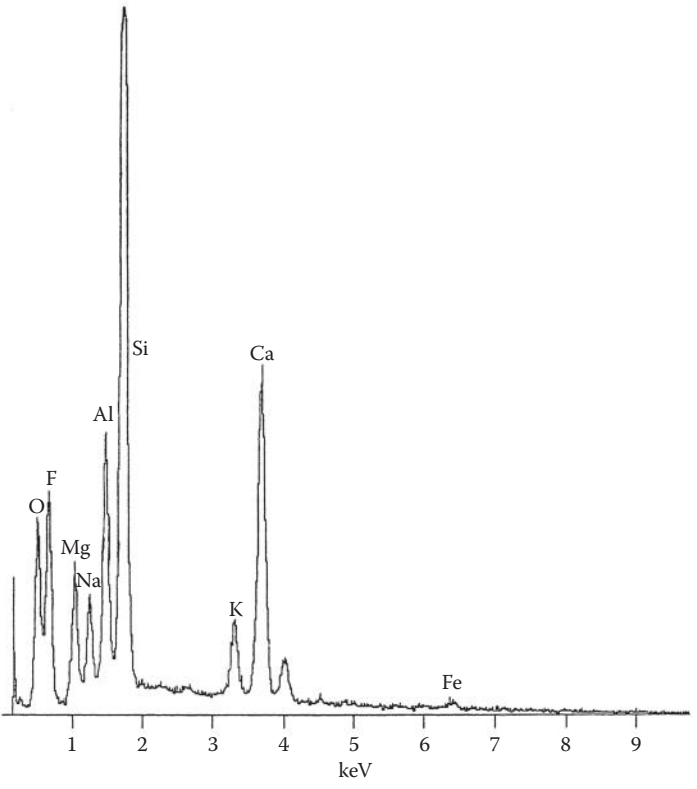


FIGURE 2.21 EDAX analysis of a glaze.

Newer detectors, such as the silicon drift detector (SDD), can offer advantages over the usual Si(Li) variety as they collect over a larger angle, leading to shorter collection times and a better count rate, and also they do not need to be cooled with liquid nitrogen.

2.6.4 Scanning Transmission Electron Microscopy

A **scanning transmission electron microscope (STEM)** combines the scanning ability of SEM with the high resolution achieved in TEM. To achieve this, the electron beam has to be finely focussed down to a very small spot (<1–10 nm), which is then rastered across the sample. The transmitted electrons are collected to provide a high-resolution image of the inner structure and surface of the sample. As in TEM, the secondary and elastically scattered electrons, as well as the characteristic X-ray spectrum, can be collected to give the chemical structure and identification data. As it is a transmission technique, the sample has to be prepared very carefully and has to be ultrathin. Because these instruments are designed to give very high resolution images, they have to be operated under very careful conditions, which are thermally stable, vibration free and radiation free—definitely no banging of doors!

The high-angle scattered electrons in STEM also give rise to a very useful form of imaging, **high-angle annular dark-field imaging (HAADF)** or **Z-contrast imaging**. The Bragg equation, $n\lambda = 2d \sin \theta$, tells us that the high-energy, short-wavelength electrons used in these instruments will give very small Bragg angles for the coherently scattered electrons. However, high-angle scattered electrons are also produced, which have been deflected by close interaction with the nuclei; in this case, each atom acts as an independent scatterer and there is no interference between these electrons—they are incoherent. The intensity of each incoherent scattering can simply be summed over all the atoms and is very sensitive to atomic number. These high-angle electrons can be collected by an annular dark-field detector, which collects electrons that pass through an annulus (ring) around the primary beam, allowing the low-angle Bragg scattered electrons to pass through (Figure 2.22a). The intensity of each position in the image thus obtained is representative of the average atomic number of the column of atoms in that position in the sample, giving detailed information on the position of different types of atoms in the sample (Figure 2.22b and c).

2.6.5 Electron Energy Loss Spectroscopy

The schematic diagram of a STEM instrument, shown in Figure 2.22a, indicates another form of analysis of the transmitted beam of electrons through a thin sample, that of **electron energy loss spectroscopy (EELs)**. In a primary beam, an electron subjected to a 300 kV potential will have a kinetic energy of 300 keV; during inelastic collisions, some of this energy may be transferred to other electrons. The kinetic energy distribution of the transmitted electrons can be measured and the characteristic losses observed and interpreted. Like EDX, this technique often adds another string to the bow of an electron microscopist.

The inelastically scattered incident electrons lose energy during collisions due to interactions with the inner shell, valence, and conduction electrons and also due to the excitation of **phonons** (crystal lattice vibrations). Thus, the electron beam can excite **plasmons** (quantised oscillations of the electron gas in a conducting solid), electronic transitions between the bands in the solid (see Chapter 4) and inner-shell ionisations.

The ionisation of inner-shell electrons is particularly useful for identifying different elements in a solid. During ionisation, a critical threshold energy (E_c) must be transferred from the incident electron to eject another electron from its energy level. The incident electron then travels on through the spectrometer to the detector with a reduced kinetic energy, and the difference between its initial and final energies is

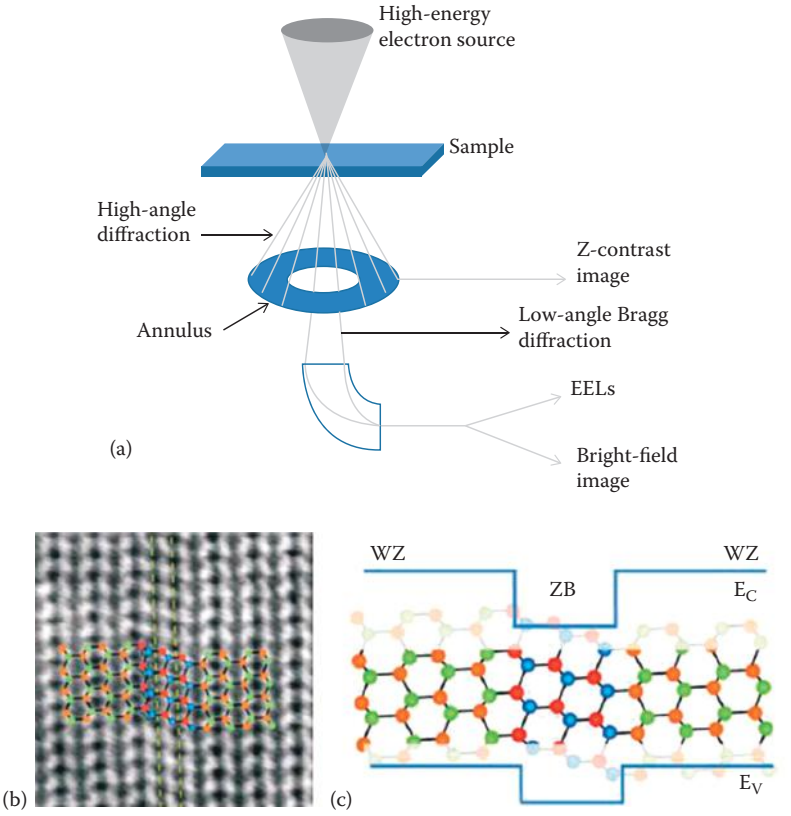


FIGURE 2.22 (a) Schematic diagram of high-angle annular dark-field imaging. (b) Aberration corrected HAADF HRSTEM of a GaAs wurtzite structure with a zinc-blende structure quantum well (Ga, red, As, blue, Ga, orange, As, green). (c) Model of the wurtzite/zinc-blende/wurtzite structures with a schematic band diagram. (Parts (b) and (c) reprinted with permission from Spirkoska, D. et al., *Physics Review B*, **80**, 245325, 2009. Copyright 2009 by the American Physical Society.)
Note: A quantum well is a thin layer of material of differing composition from its surroundings, designed to have a narrower band gap; see Chapters 7 and 10.

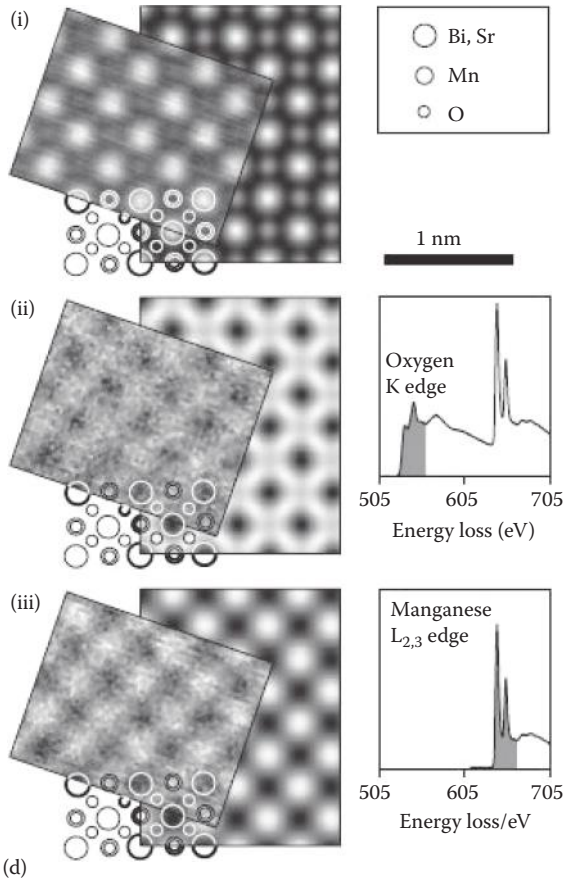


FIGURE 2.22 (Continued) (d) Comparison between experiments and simulations of $Bi_{0.5}Sr_{0.5}MnO_3$ oriented along the 001 zone axis. (i) *Z*-contrast, (ii) oxygen K edge and (iii) manganese $L_{2,3}$ edge STEM images showing atomic structure. (Part (d) reprinted with permission from Bosman, M. et al., *Physical Review Letter*, **99**, 086102, 2007. Copyright 2007 by the American Physical Society.)

measured. More than the critical energy may be given to the ejected electron, $E > E_c$; the intensity of the absorption decreases with increasing energy loss, giving a sharp discontinuity in the spectrum, which corresponds to the characteristic energy of the inner shell and is thus characteristic of the element; this is known as the **absorption edge**. The fine structure of the absorption edges gives information about band structure, which is useful for the accurate measurement of band gaps in semiconductors. Measurement of the absorption intensities gives data on the concentration of the elements. The critical ionisation energy is sensitive to the chemical environment of the atom, so information about the bonding in the structure may be obtained from EELs. The most recent experiments have produced impressive atomic-scale images of solid structures of a barium strontium manganese oxide material showing colossal magnetoresistance (Figure 2.22d; see Chapter 8).

In the order of increasing energy loss, the processes of inelastic electron scattering taking place are of the order: phonons 0.02 eV; intraband transitions 5–25 eV; plasmons 5–25 eV; and ionisation 10–1000 eV. Thus, the spectra naturally divide into two regions with a *low-loss region* with energy losses due to plasmons and interband transitions up to ~50 eV and a *high-loss region* above 50 eV, due to ionisation from the inner shells.

The technique is particularly useful for the detection of light elements.

2.6.6 superSTEM

In order to improve the resolution of STEM, it is necessary not only to have very high energy electrons and an ultrafine spot, but also to get rid of all the problems of instability and aberrations in the beam, and this was the aim of building the **superSTEM** instruments. The first superSTEM instruments in Europe were built at the facility in Daresbury, United Kingdom. This site provides a stable geological basis of sandstone, and the instruments occupy their own building with separate foundations, separate electrical supply and carefully controlled temperature and air-conditioning, and they are shielded from external sources of radiation and sound—everything that might cause a problem.

The ability to correct spherical aberration (C_s) (Figure 2.19) in a beam has progressed immensely in recent years, thanks to the work of Krivanek, Dellby and Brown at Cambridge University, United Kingdom. The approach has been to introduce negative spherical aberration, which then cancels out the positive aberration. They have done this using a series of complex magnetic lenses (quadrupole and octupole) along the direction of travel of the beam, so that all rays focus at a common point. The final specifications of the instrument are impressive, with a current stability of <1 ppm and an improved resolution of about a factor of 3 to less than 0.1 nm (<1 Å) (Figure 2.22).

2.7 SCANNING PROBE MICROSCOPY

The characterisation of solid surfaces is often carried out using a range of techniques known collectively as **scanning probe microscopy (SPM)**. These techniques produce direct images of surfaces down to atomic level features. The common feature of all the techniques is the monitoring of the interaction of a very fine tip with the surface while it is moved across the surface in a very fine grid; information about the surface is gathered from how the interaction changes with position. The pioneer scanning probe technique was **scanning tunnelling microscopy (STM)**, which enables conducting and semiconducting surfaces to be imaged at atomic resolution. It was invented in the early 1980s by Gerd Binnig and Heinrich Rohrer at IBM in Switzerland, for which they were awarded the 1986 Nobel Prize in Physics. Looking for ways to study nonconducting surfaces, Binnig, Quate and Gerber went on to develop the **atomic force microscope (AFM)** several years later. Since then, a plethora of other microscopy techniques, which probe particular properties of surfaces at the atomic level, have been developed (with far too many acronyms to remember); for instance, near-field scanning optical microscopy (NSOM) uses a subwavelength-sized

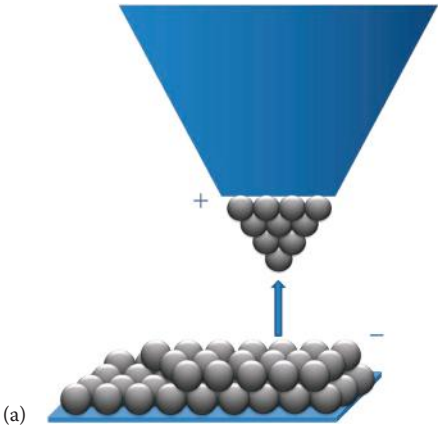
aperture rather than a lens to direct the light onto the sample and a spatial resolution of 20 nm and less can now be achieved; photoscanning microscopy (PSTM) tunnels photons using an optical tip; magnetic force microscopy (MFM) investigates magnetic structure; electric force microscopy (EFM) investigates electronic properties; others enable the mapping of surface friction, capacitance, adhesion and Young's modulus. SPMs can also be used to investigate adsorbed molecules, to manipulate and enable the movement of molecules and to obtain the spectra of single molecules on the surface. More recently, STM resolution has been increased to such an extent that the movement of molecules and atoms on the surface can be tracked, with exciting prospects for the study of diffusion, catalysis and chemical reactions.

2.7.1 Scanning Tunnelling Microscopy

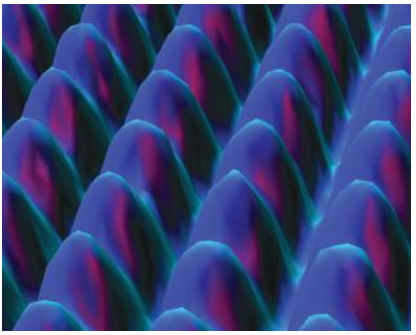
In STM, a sharp metal wire tip is brought sufficiently close (<1 nm) to the surface of a conducting or semiconducting solid sample, such that electrons can move through the vacuum between the two. When a small voltage is applied to the tip, creating a potential difference, electrons from the sample surface are attracted to the positive tip (Figure 2.23a). However, an energy barrier prevents the electrons from leaving the sample. The energy needed to eject an electron from the highest occupied level in the solid, the Fermi level, out into a surrounding vacuum, is known as the **work function** (ϕ) (see Chapter 4). However, if the tip and surface are sufficiently close together, quantum mechanical effects come into play, their electron-wave functions can overlap and there is a finite probability that the electron can be found on the other side of the barrier. This is known as **quantum mechanical tunnelling** and produces a very small tunnelling current in the range of picoamperes to nanoamperes. The magnitude of the current is very sensitive to the size of the gap, changing by a factor of 10 when the distance changes by 0.1 nm.

The metal tip is scanned backwards and forwards across the solid, point-by-point and line-by-line (rastering), and the steep variation of the tunnelling current with distance gives an image of the atoms on the surface (Figure 2.23b and c). The image is usually formed by keeping a constant tunnelling current and measuring the distance, thereby creating contours of constant density of states on the surface (see Chapter 4). The distance from the tip to the sample is typically 0.4–0.7 nm, and this very fine control of the movement of the tip is controlled by the expansion and contraction of piezoelectric crystals (see Chapter 8) using feedback electronics to adjust the height. PZT, lead zirconium titanate ($PbZr_{1-x}Ti_xO_3$) piezoelectric crystals are typically used, moving by about 1 nm with 1 mV. The values of the adjustments are recorded, defining a grid of values, which is then displayed as a greyscale image, which can be artificially coloured by height, curvature, etc., to enhance particular features of interest. A faster scan of flatter surfaces can be achieved by maintaining a constant height, giving an image of the current changes and therefore, a charge density map of the surface.

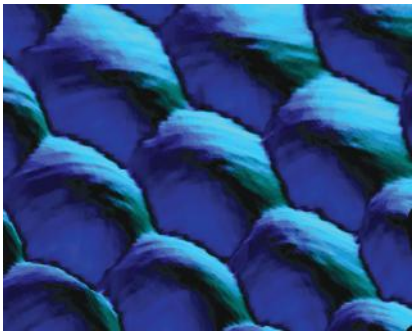
By changing the sign of the potential, the tunnelling direction reverses and thus, STM can map either occupied or unoccupied density of states. Chemical information can be obtained by altering the voltage so that electrons from different types of energy levels tunnel across. The map thus shows features due to both the topography



(a)



(b)



(c)

FIGURE 2.23 (a) Diagram of a positively charged STM tip over a surface. (b) An Ni (*110*) surface showing the rectangular surface unit cell. (c) The Pt (*111*) face clearly showing the close-packed array of atoms. (Images (a) and (b) from IBM Corporation. With permission.)

and the electronic structure, down to the level of showing the positions of individual atoms or defects. As an STM image maps either the density of states or the charge density at the surface, the images have to be manipulated and modelled to finally yield the structure, contours and atomic positions at the surface.

The image resolution relies on the diameter of the scanning metal tip, so it is vital to be able to produce sharp tips with just a few atoms at the end; surprisingly, even cutting a wire at an angle with scissors can achieve this. The most commonly used tips are platinum–iridium, which are produced by mechanical shearing, and tungsten produced by electrochemical etching. Most recently, carbon nanotube tips have been introduced. To get good results, as well as working with very sharp tips, it is also important to work on very clean, stable surfaces in a vibration-free environment.

STM can also be used to manipulate the atoms on a surface—this and nano-lithography will be discussed in Chapters 3 and 10.

The atomic force microscope (AFM) is now widely available in laboratories, and it is to this technique that we turn next.

2.8 ATOMIC FORCE MICROSCOPY

Atomic force microscopy (AFM) is based on the detection of very small (of the order of nanonewtons) forces between a sharp tip (of the order of nanometres) and the atoms on a surface and has the advantage that it can be used on nonconducting surfaces and under ambient conditions. The tip is scanned across the surface at subnanometre distances, and the deflections due to attraction or repulsion by the underlying atoms are detected. The technique produces atomic-scale maps of the surface, and its use has been vital to the understanding of nanotechnology (Chapter 10). It can used in many modes to determine the different forces acting on the surfaces of solids, such as van der Waals, chemical bonding, magnetic, capillary, friction and so on. If the arm is brought very close to a surface, initially (~1–10 nm) a small attractive force is caused by van der Waals forces, but when it approaches even closer, short-range coulombic repulsion forces between the electron clouds of the atoms on the tip and on the surface begin to dominate and the arm is deflected upwards (Figure 2.24a). The resulting interatomic force curve is shown in Figure 2.24b. For small deflections of the arm about the equilibrium position, the curve is almost linear and **Hooke's law**, which tells us that a movement (d) is directly proportional to the force (F) experienced, can be used to determine the repulsive force: $F = -kd$, where k is the force constant or spring constant (stiffness) of the arm. The deflection of the arm is detected by bouncing a laser beam from the arm to a detector consisting of a photodiode array.

The tip and its supporting cantilever arm are typically made from silicon or silicon nitride (Si_3N_4) (Figure 2.24c). Instead of scanning the surface at a constant height, which runs the risk of collision and damage to the tip, a constant force between the tip and the surface is maintained by a feedback loop. The fine control of the movement is through piezoelectric PZT crystals, which allow a subnanometre adjustment in all three directions, thus enabling a three-dimensional image of the surface to be created.

An AFM can be operated in two main ways, static and dynamic. In the static **contact mode**, the tip is dragged across the surface so close as to be in the region of repulsive forces, thereby preventing the tip from being attracted down onto the surface (an Si surface captured this way can be seen in Figure 2.24d); to avoid damage to the tip, it must be kept away from the surface, so thinner more flexible cantilevers are used in this mode. The deflection of the tip is measured, and a feedback loop maintains a constant force as the tip is moved over the surface. By measuring the twist of the cantilever as well as the deflection, the changes in the friction of the surface can also be mapped.

In the **noncontact** dynamic mode, a stiff cantilever is oscillated just above its resonance frequency of (usually) a few nanometres, and it is kept further away from the surface where the attractive forces dominate. The oscillation frequency is modified by changes in the surface as the tip is scanned across. Once again, a feedback loop maintains either a constant oscillation frequency or a constant amplitude. By measuring the deflection at each point of a grid, a three-dimensional map of the surface can be generated. This mode is used for softer surfaces as it reduces the drag on the tip and usually operates under ultrahigh vacuum.

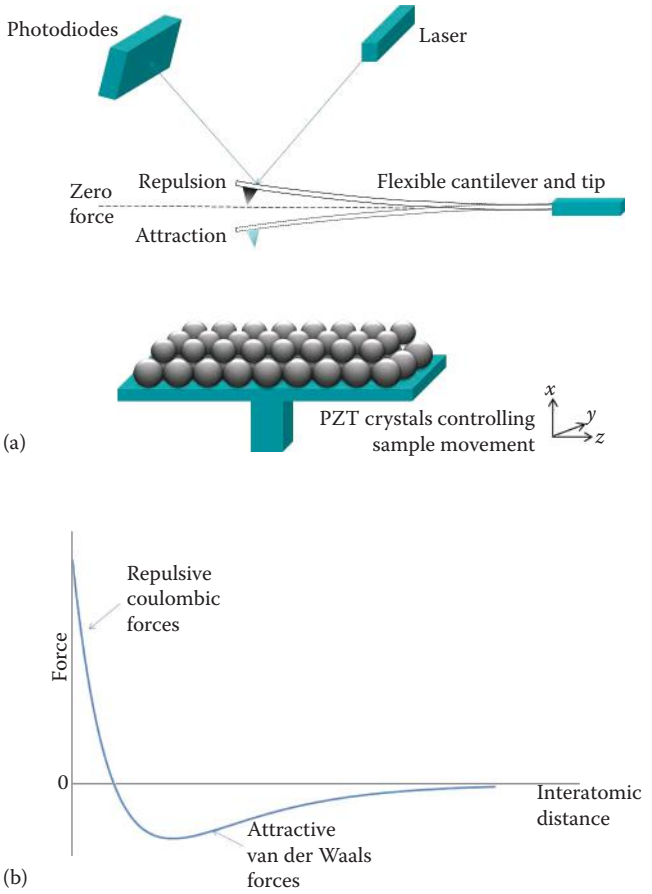
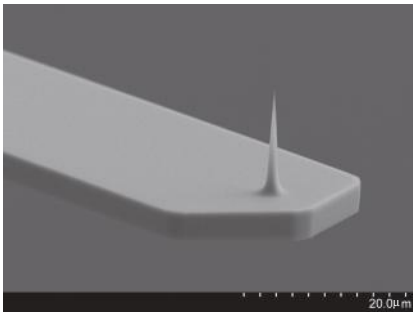


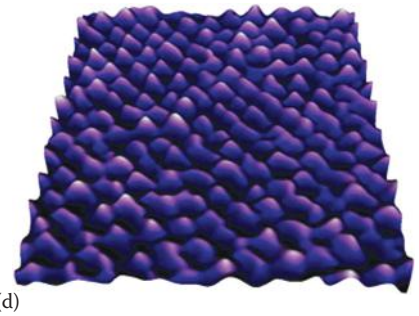
FIGURE 2.24 (a) Schematic diagram of an AFM. (b) an interatomic force distance curve.

One of the problems in investigating surfaces under ambient conditions is the contamination of the surface and, in particular, the surface liquid (mainly water), in which the tip can get stuck. The **tapping** dynamic mode was developed to overcome this, and here the cantilever is oscillated slightly below its resonant frequency and closer to the surface so that it is alternately attracted and repelled. Stiff cantilevers tend to be used in this mode because the attraction by any surface liquid layer can cause the tip to stick. By operating in this way, less damage is caused to both the tip and the surface, and with it, even adsorbed single molecules can be scanned.

Recent developments in AFM have achieved remarkable results. The atoms on the surface of a Sn and Pb alloy deposited on a Si surface have been chemically identified by very accurately measuring the dependence of the force on the distance between the tip and individual atoms and then using modelling to compare with the previously measured fingerprints for each different component. Every atom in the chemical



(c)



(d)

FIGURE 2.24 (Continued) (c) a single crystal silicon AFM tip and cantilever. (From Team Nanotech GmbH, Germany. With permission.) (d) An atomic lattice resolution image of mica imaged in contact mode: scan field: 6.6 × 6.6 nm; *Z* range: 0–0.4 nm. (From JPK Instruments AG, Germany. With permission.)

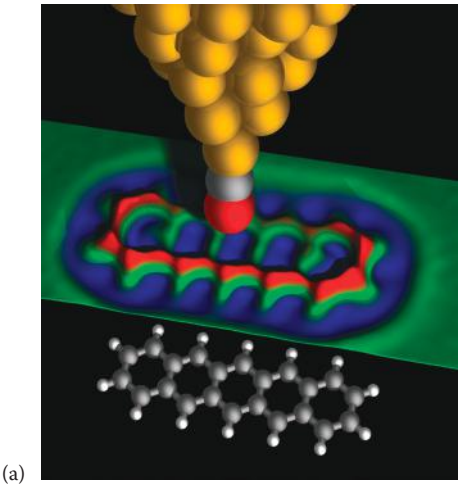
structure of a pentacene molecule (five fused benzene rings) has also been resolved using noncontact AFM. This amazing resolution has been achieved by creating a probe that has a single CO molecule attached to the end and using a stiff cantilever with a resonance amplitude down to 0.02 nm (Figure 2.25a and b). Images of water on a surface have been obtained using tips treated with a hydrophobic compound.

Further applications of AFM in nanoscience can be seen in Chapters 3 and 10.

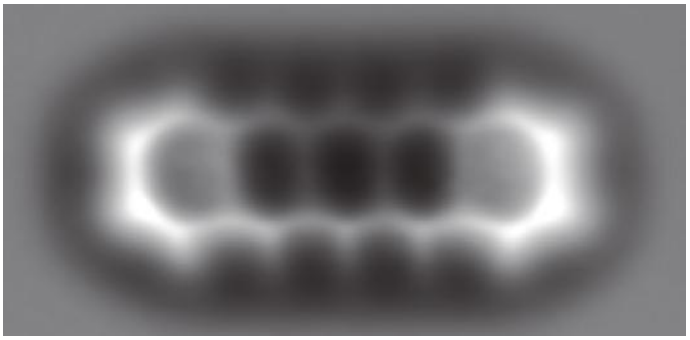
2.9 X-RAY ABSORPTION SPECTROSCOPY

2.9.1 Extended X-Ray Absorption Fine Structure

Extended X-ray absorption fine structure (EXAFS) is used for the determination of short-range order in noncrystalline materials, such as gases, liquids, amorphous powders and nanocrystalline materials, where conventional diffraction techniques are of little use. In high-energy accelerators, electrons are injected into an electron storage ring, where they are captured and accelerated around this circular path by a series of magnets. When the electrons are accelerated to kinetic energies above the mega-electron volt range, they travel very close to the speed of light and emit the so-called **synchrotron radiation** (Figure 2.26). There are only about 50 synchrotron sources in the world of which only three, the so-called 'third generation', are the most powerful: ESRF at Grenoble, France, operates at 6 GeV; APS at Argonne, United States, at 7 GeV; and SPring-8 in Japan, at 8 GeV; these huge storage rings have diameters of approximately 270, 350 and 460 m, respectively. The new UK facility, DIAMOND, has a ring of 180 m in diameter and operates at 3 GeV. The X-radiation generated can be in a range from soft X-rays to hard X-rays (~300 eV–300 keV, 40–0.04 Å). Unlike X-radiation from a conventional generator, the synchrotron radiation is of uniform intensity across a broadband of wavelengths and several orders of magnitude (10^4–10^6) higher in intensity (Figure 2.27). The shortest X-ray wavelengths emerge as almost fully collimated, polarised beams.



(a)



(b)

FIGURE 2.25 (a) Diagram of the AFM tip scanning pentacene. (b) AFM image of pentacene. (From IBM Research, Zurich. With permission.)

In an EXAFS experiment, the X-radiation is absorbed by a bound electron in a core shell (the K, L or M shell) and is ejected as a photoelectron. If you measure the absorption coefficient of the sample as a function of the X-ray frequency, a sharp rise, or **absorption edge**, is observed at the shell threshold energy (Figure 2.28) when the electron is ionised. Each element has its own characteristic shell energies, and this makes it possible to study one type of atom in the presence of many others, by **tuning** the X-ray energy to a particular absorption edge. The appropriate frequency X-radiation from the continuous synchrotron radiation is selected by using the Bragg reflection from a single plane of a carefully cut crystal such as Si *220*; often two crystals are used, as is shown in the schematic diagram of a double-crystal monochromator in Figure 2.29. By changing the Bragg angle of reflection, the frequency of the X-rays selected may be changed, and thus the absorption edges of a wide range of elements can be studied.

The waves of the ejected photoelectron from the shell can be thought of as a spherical wave emanating from the nucleus of the absorbing atom; this encounters

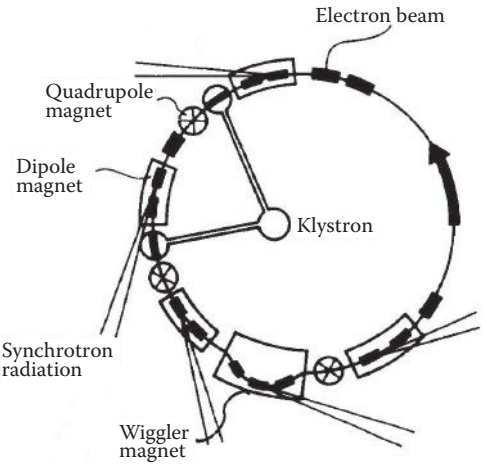


FIGURE 2.26 Storage ring.

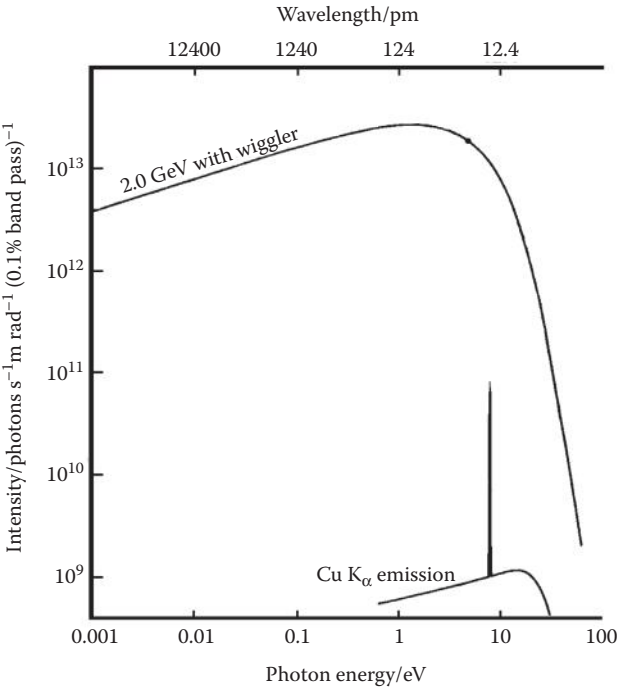


FIGURE 2.27 Synchrotron radiation.

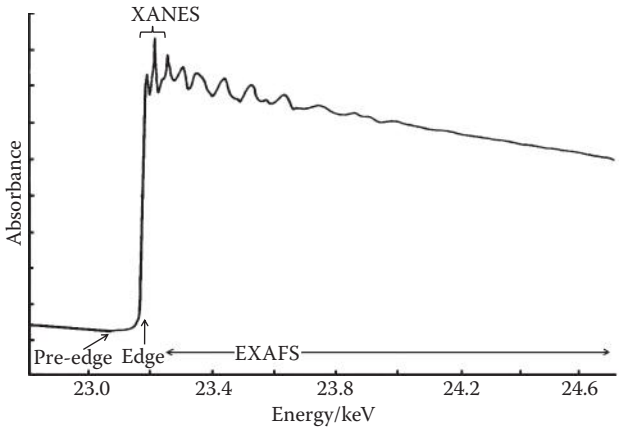


FIGURE 2.28 Rh absorption edge showing the pre-edge XANES and EXAFS regions.

the neighbouring atoms and is partially scattered by them, producing a phase shift (Figure 2.30). Depending on the phase shift experienced by the electron, the reflected waves can then interfere constructively or destructively with the outgoing wave, producing a net interference pattern at the nucleus of the original atom. Absorption by the original atom is now modified, and the effect is seen as sinusoidal oscillations or **fine structure** superimposed on the absorption edge (Figure 2.28), extending out to several hundred electron volts after the edge. The extent to which the outgoing wave is reflected by a neighbouring atom, and thus the intensity of the reflected wave, is

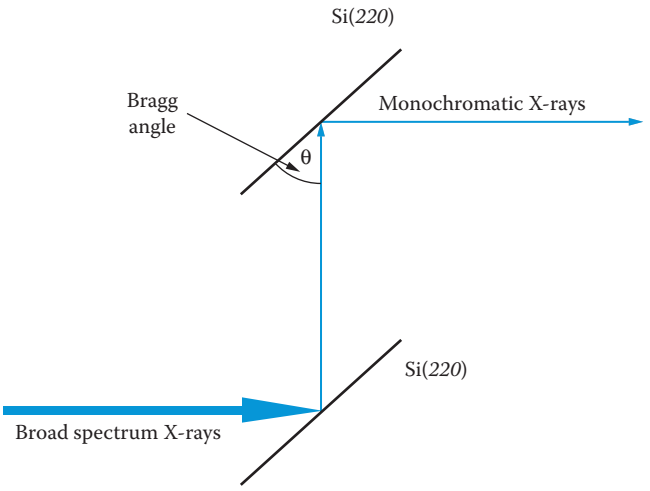


FIGURE 2.29 Bragg reflections from a double-crystal monochromator. From the Bragg equation, $n\lambda = 2d \sin \theta$, d for the planes of the crystal stays constant, so changing the angle changes the wavelength of the X-rays reflected. Two crystals are used to make the exit beam parallel to the entrance beam. Curved crystals focus the X-rays.

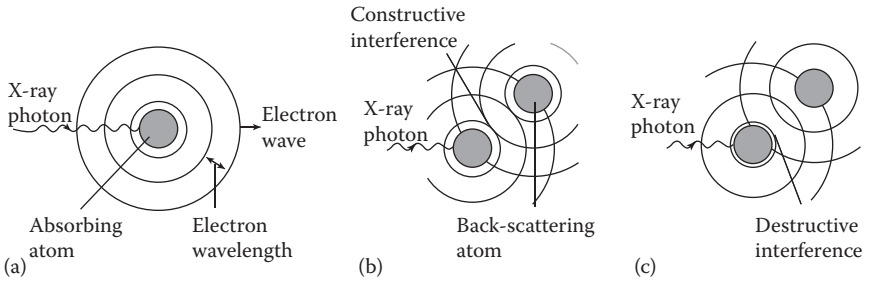


FIGURE 2.30 The EXAFS process: (a) the photoelectron is ejected by X-ray absorption, (b) the outgoing photoelectron wave (solid line) is backscattered constructively by the surrounding atoms (dashed line) and (c) destructive interference between the outgoing and the backscattered wave.

partly dependent on the scattering factor of that atom. The interference pattern making up the EXAFS therefore depends on the *number* and the *type* of neighbouring atoms and their *distance* from the absorbing atom.

The EXAFS function is obtained from the X-ray absorption spectrum by subtracting the absorption due to the free atom. A Fourier transform of the EXAFS data gives a **radial distribution function**, which shows the distribution of the neighbouring atoms as a function of the internuclear distance from the absorbing atom. The absorbing atom is surrounded by shells of neighbouring atoms, known as **coordination shells**. Finally, the radial distribution function is fitted to a series of trial structural models until a structure that best fits the data is obtained, and the data are refined as a series of coordination shells surrounding the absorbing atom. The final structure will refine the number and types of atoms and their distance from the absorbing atom. It is difficult to differentiate atoms of similar atomic number, and it is important to note that EXAFS only gives the data on distance—there is no angular information. Depending on the quality of the data obtained, in favourable cases several coordination shells out to about 6 Å (0.6 nm) can be refined, with an accuracy of about 1 pm. The determination of the number of nearest neighbour atoms is less precise.

In the example shown in Figure 2.31, a clay (a layered double hydroxide, LDH) was intercalated with a transition metal complex, $(NH_4)_2MnBr_4$. The EXAFS data in Figure 2.31a show the Mn K-edge EXAFS of the pure complex and we see one coordination sphere of four Br atoms at a distance of 0.249 nm (2.49 Å), corresponding well to the tetrahedral coordination found in the X-ray crystal structure. However, after intercalation, the complex reacts with the layers in the clay, and the coordination changes to distorted octahedral, where Mn is now surrounded by four O atoms at a distance of 0.192 nm (1.92 Å) and two Br atoms at a distance of 0.225 nm (2.25 Å).

In Figure 2.9 we saw that powder X-ray diffraction can be used to follow phase changes over time with heating; Figure 2.32 shows the corresponding iron K-edge EXAFS analysis for the same ferrosilicon sample. As the α-ferrosilicon changes into the β-phase, a shell of eight silicon atoms at about 2.34 Å is found to surround

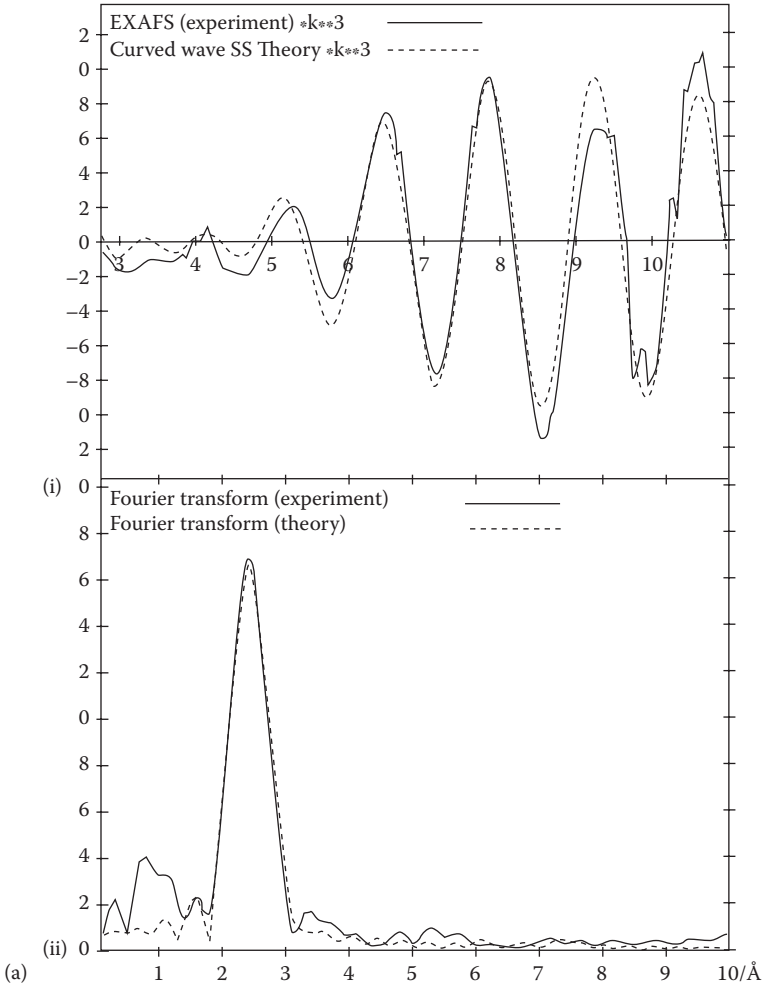


FIGURE 2.31 EXAFS data for (a) $(NH_4^+)_2MnBr_4^{2-}$ and (b) the intercalation of $MnBr_4^{2-}$ in a layered double hydroxide clay: (i) extracted EXAFS data and (ii) the radial distribution function, solid line experimental, dotted line calculated.

Fe in both forms, but the β-form is found to have only two Fe atoms coordinated to Fe 0.298 nm (2.98 Å), compared with 3.4 Fe atoms at 0.268 nm (2.68 Å) in the α-phase.

X-rays are very penetrating, so EXAFS, like X-ray crystallography, examines the structure of the bulk of a solid. It has the disadvantage that it only provides information on interatomic distances, but it has the considerable advantage that it is not confined to crystalline samples and can be used on amorphous solids, glasses and liquids. Additionally, by using different absorption edges, it can investigate the coordination around more than one type of atom in the sample.

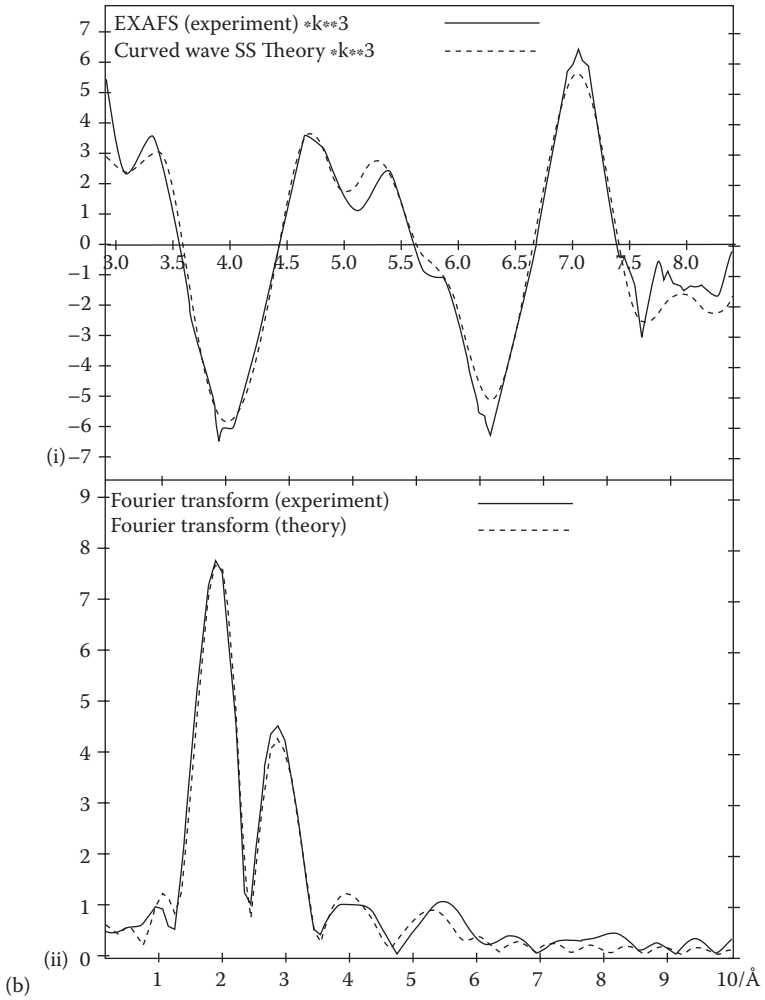


FIGURE 2.31 (Continued)

2.9.2 X-Ray Absorption Near-Edge Structure and Near-Edge X-Ray Absorption Fine Structure

X-ray absorption near-edge structure (XANES) and near-edge X-ray absorption fine structure (NEXAFS) are two techniques that are in principle synonymous and refer to the regions of an X-ray absorption spectrum up to about 150 eV on either side of an absorption edge; the spectrum contains the pre-edge, the position of the edge itself and the XANES fine structure just after the edge (Figure 2.28). The acronym NEXAFS is now used almost exclusively to refer to the spectra of the light elements up to C, N and O. The fine structure close to either side of the edge (Figure 2.28) is very dependent on the nature of the *chemical environment*

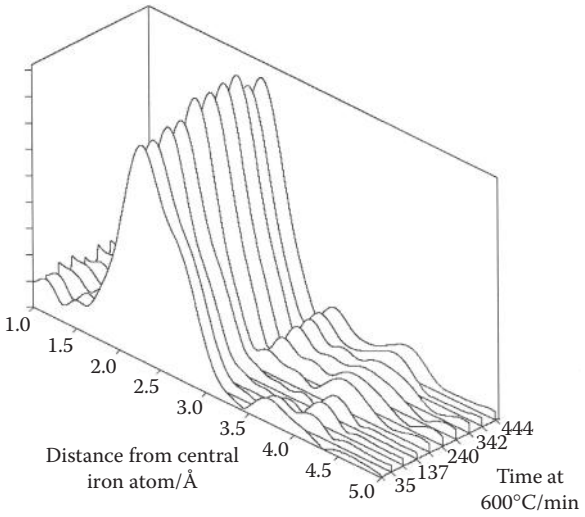


FIGURE 2.32 EXAFS patterns as a function of time, showing the phase evolution of β-ferrosilicon from the α-form when heated at 600°C. (From Professor Berry, F.J., Open University. With permission.)

of the absorbing atom, and the position of the edge itself depends on the *oxidation state*.

The precise position of the absorption edge shifts to higher energy with the increasing chemical oxidation state of the absorbing atom. The example shown in Figure 2.33a shows XANES spectra for manganese in different oxidation states, and these clearly show the edge shift to higher energy as the oxidation state increases from $2 \rightarrow 3 \rightarrow 4 \rightarrow 7$ (the position of the pure metal Mn^0 is marked with a vertical line). An edge may shift by up to 5 eV with each increment in the oxidation number. The shifts and heights of the edge can be used to determine the oxidation state of the absorbing metal. It is often used as a fingerprinting method, using the spectra of known standards, and is thus useful for determining the ratio of the oxidation states in mixed oxides, composites or impure samples.

The fine structure in the pre-edge region is due to both the oxidation state of the metal and the coordination chemistry surrounding it: notice the dramatic changes in the pre-edge region in Figure 2.33 as both the oxidation state and the coordination around Mn change. The peaks here, before complete ionisation occurs at the absorption edge, are due to the excitation of the electrons into the vacant energy levels of the metal. From ligand-field theory, we know just how much the pattern of molecular orbital energy levels changes when we move from octahedral coordination to distorted octahedral, and this is again very different from tetrahedral coordination. The number of electrons in the highest occupied molecular orbitals varies with the oxidation state of the metal—for instance, Mn(II) is d^5 and Mn(IV) is d^3. Both these factors therefore lead to the observation of a different pattern of absorption for each compound.

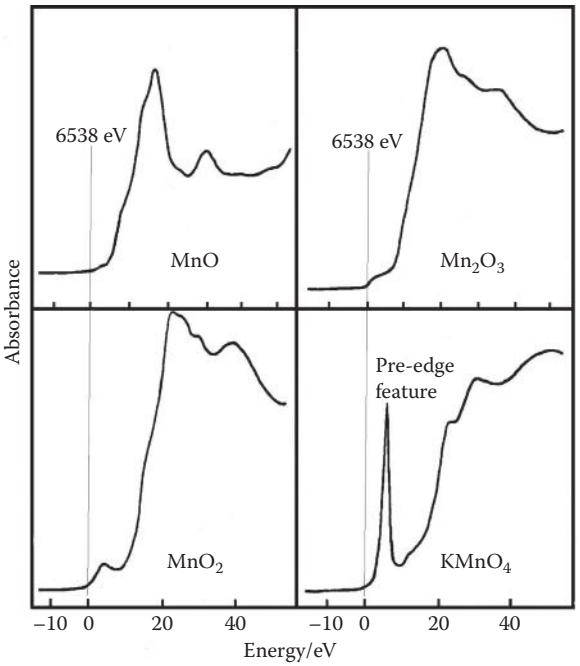


FIGURE 2.33 Mn K-edge XANES data for various manganese oxides. The vertical dashed line is the position of the metal edge. The position of the edge increases with the oxidation state. (From Dr. N. Greaves. With permission.)

The changes in the shape of the immediate postedge structure are due to the multiple-scattering events (when the photoelectron is scattered sequentially by more than one atom), which affect these low kinetic energy photoelectrons. Multiple scattering must also be dependent on atomic position and thus on the coordination and bonding around the absorbing atom.

XANES is not straightforward to interpret, but has some advantages over EXAFS. It is a stronger signal, so can be carried out at a lower concentration or sample size, thereby determining the nature of contaminants or defects; it is not very temperature-sensitive, so it can be recorded at room temperature. In addition, it can provide information on the bonding and molecular orbitals present or on the band structure and density of the available electronic states (see Chapter 4). It can be used to follow the evolution of species during a reaction and to follow the kinetics.

2.10 SOLID-STATE NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

A nuclear magnetic resonance (NMR) nucleus must have a nuclear spin (I) and therefore a nuclear magnetic moment. The size of the magnetic moment depends both on I, which can take values of 0, 1/2, 1, 3/2, 2, 5/2, … and on its magnetogyric ratio

(γ). In the presence of a magnetic field, Zeeman splitting occurs and $2I + 1$ nondegenerate spin states are formed, between which $2I$ allowed transitions can take place between adjacent energy levels. A radio frequency signal at the NMR frequency causes transitions between the spin states to give the NMR spectrum.

For many years, solid-state NMR spectra gave broad resonances that were difficult to interpret. A magnetic dipole interacts with all the other dipoles in its immediate environment. In solution NMR spectroscopy of nuclei with spin $I = 1/2$, the dipolar interactions and anisotropic effects are averaged out by the constant molecular motion, but this is not the case in the solid state, and thus the NMR spectra of their solids tend to be very broad and uninterpretable. In addition, because of the lack of movement in a solid, the spin–lattice relaxation times can be long, giving low signal-to-noise ratio.

Methods have been developed to give more sharply resolved spectra, such that **magnetic dipolar interactions** can be removed by the application of a high-power decoupling field at the resonance frequency. The chemical shielding of a nucleus is not uniform but varies with the orientation of the molecule to the field. In a polycrystalline solid, this gives a range of values, an effect known as the **chemical shielding anisotropy**, which broadens the resonance. In magic angle spinning (MAS), a polycrystalline sample is formed into a cylinder and is then placed in a magnetic field inclined at an angle β to the field. If the sample is then rotated around its cylindrical axis, a time averaging of the anisotropic interactions takes place. As long as the spinning is fast enough, each individual contribution to the resultant NMR signal depends on the term $(3\cos^2\beta - 1)$. If this expression is set equal to zero, it can be solved to find the angle—the so-called magic angle—and this angle can be manually set. The term becomes zero when $3\cos^2\beta = 1$, or $\cos\beta = (1/3)^{1/2}$, that is, $\beta = 54°44'$. As long as the spinning is fast enough, the contributions from all crystallites in the sample are the same and just a single line is observed.

Spinning the sample in this way improves the resolution in the chemical shift of the spectra, and this sometimes lends its name to the technique as **magic angle spinning spectroscopy (MAS NMR)**. The spinning speed has to be greater than the frequency spread of the signal; if it is less, as may be the case for very broad bands, then a set of so-called 'spinning side-bands' are observed, and care is needed in assigning the central resonance. Spinning speeds of up to 70 kHz are now being achieved. The technique for assigning the central resonance is to collect the spectra at two different spinning speeds, when the spinning side bands move but the central resonance stays fixed. MAS NMR is sometimes used as an umbrella term to imply the application of any or all of these techniques in obtaining a solid-state NMR spectrum.

Some isotopes have long spin–lattice relaxation times in solids, which give rise to poor signal-to-noise ratios. In order to increase the signal from these isotopes, the sensitivity can be improved by using a technique known as **cross polarisation**, where a complex pulse sequence transfers polarisation from an abundant spin nucleus to the dilute spin, thereby enhancing the intensity of its signal, improving the signal-to-noise ratios.

Advances in superconducting magnets, which can now provide up to ~20 T, and improved techniques for eliminating the broadening are now giving much higher resolution spectra and rendering the technique more accessible, although it is still

very much the field of the specialist. The improved resolution can allow *J* couplings to be measured, and this has led to the introduction of two-dimensional NMR methods. In some sophisticated experiments, internuclear distances can be measured.

The ability to use density functional theory (DFT) methods (see Chapter 4) to calculate chemical shifts and quadrupolar coupling constants now also means that correlations from structure to parameter can be made.

High-resolution spectra can be measured for many different NMR isotopes. Solid-state NMR has proved very successful in elucidating zeolite structures. Zeolites are three-dimensional framework silicate structures where many of the silicon sites are occupied by aluminium (Chapter 6). Because Al and Si are next to each other in the periodic table, they have similar X-ray atomic scattering factors and consequently are virtually indistinguishable on the basis of X-ray crystallographic data. It is possible to build up a picture of the overall shape of the framework with accurate atomic positions, but it is difficult to decide which atom is Si and which is Al.

^{29}Si has a nuclear spin $I = \frac{1}{2}$, thus giving sharp spectral lines with no quadrupole broadening or asymmetry; the sensitivity is quite high, and ^{29}Si has a natural abundance of 4.7%. In the late 1970s, Lippmaa and Engelhardt carried out pioneering work using solid-state NMR on zeolites. They showed that up to five peaks could be observed for the ^{29}Si spectra of various zeolites and that these corresponded to the five different Si environments that can exist. Each Si is coordinated by four oxygen atoms, but each oxygen can then be attached either to an Si or to an Al atom, giving the five possibilities: $Si(OA1)_4$, $Si(OA1)_3(OSi)$, $Si(OA1)_2(OSi)_2$, $Si(OAl)(OSi)_3$ and $Si(OSi)_4$. They also showed, most importantly, that characteristic ranges of these shifts could be assigned to each coordination type—the more Al linkages, the more positive is the shift. These ranges could then be used in further structural investigations of other zeolites (Figure 2.34a). A solid-state NMR spectrum of the zeolite analcite is shown in Figure 2.34b. Analcite has all five possible environments. Even with this information, it is still an extremely complicated procedure to decide where each linkage occurs in the structure.

Most of the NMR nuclei do not have $I = \frac{1}{2}$, but are quadrupolar nuclei, with a nuclear spin of $I > \frac{1}{2}$ and a nuclear charge distribution that is not spherically symmetric—the distorted spheres of charge may be prolate (elongated) or oblate (flattened) spheroids. These nuclei present much more of a challenge in terms of signal broadening as the quadrupole interacts strongly with local electric field *gradients* caused by any nonsymmetrical local chemical coordination around the nucleus. This provides an efficient spin–lattice relaxation mechanism for the magnetic energy levels, thus giving rise to relatively short spin relaxation times in the solid state. The size of the coupling (ω_Q) depends on both the particular nucleus and its site symmetry and both first-order ($\propto \omega_Q$) and second-order ($\propto \omega_Q^2$) interactions affect the magnetic energy levels (Figure 2.35). In a highly symmetrical environment, such as octahedral or tetrahedral, the lines will be much sharper. The first-order interaction can mostly be suppressed by spinning at the magic angle. Methods of eliminating the broadening due to the second-order interaction have now been found, but they are not straightforward. The **double rotation method (DOR)** spins the sample simultaneously at two angles—the magic angle of 54.74° and 30.56°—to suppress both interactions. This method uses a 'rotor-in-a-rotor'

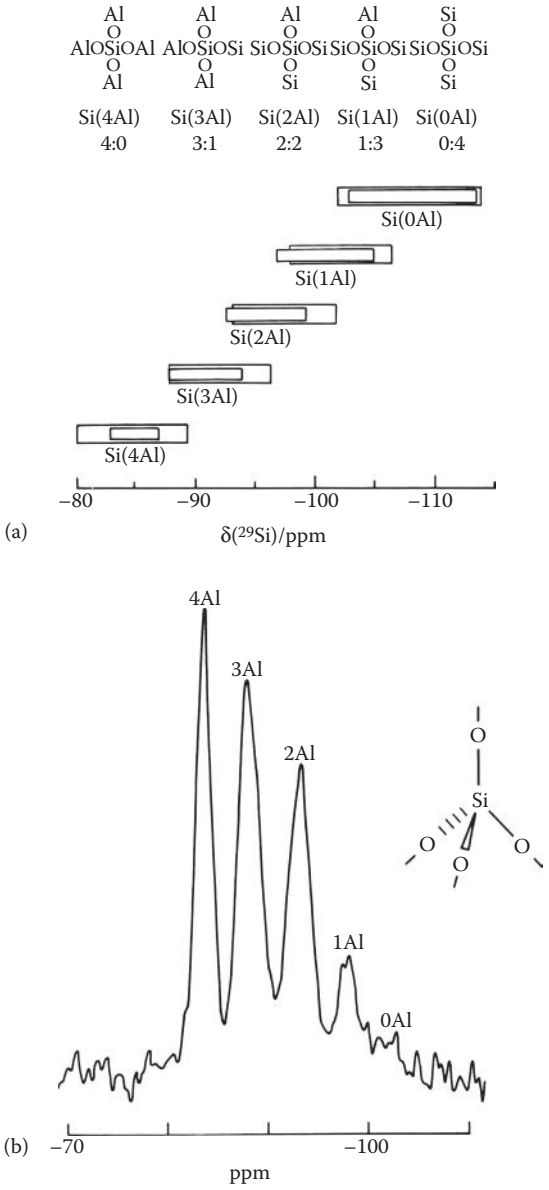


FIGURE 2.34 (a) The five possible local environments of a silicon atom together with their characteristic chemical shift ranges. The inner boxes represent the ^{29}Si shift ranges suggested in the earlier literature. The outer boxes represent the extended ^{29}Si shift ranges, which are more unusual. (b) ^{29}Si solid-state NMR spectrum at 79.6 MHz of analcite, showing five absorptions characteristic of the five possible permutations of Si and Al atoms attached at the corners of the SiO_4 tetrahedron as indicated.

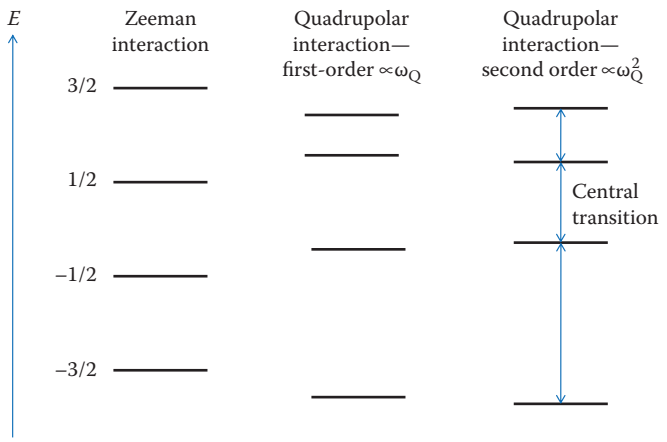


FIGURE 2.35 Diagram showing the splitting of spin states when $I = 3/2$.

design, which is technologically quite difficult. **Dynamic angle spinning (DAS)** spins sequentially at two different angles. More recently, technically simpler techniques have been introduced, known as **multiple-quantum magic angle spinning (MQMAS)**, first applied in 1995, and **satellite transition magic angle spinning (STMAS)** introduced in 2000.

^{27}Al has 100% natural abundance and a nuclear spin of $I = 5/2$, resulting in a strong resonance, which is broadened and rendered asymmetric by the second-order quadrupolar effects. However, interpreting the ^{27}Al solid-state NMR spectrum of minerals and zeolites is now achievable; for instance, it has been used to determine the three aluminium positions in kyanite (Al_2SiO_5). Diagnostically, it can also be used to distinguish different types of aluminium coordination: octahedrally coordinated $[Al(H_2O)_6]^{3+}$, which is frequently trapped as a cation in the pores of zeolites and gives a peak at about 0 ppm ($[Al(H_2O)_6]^{3+}$(aq) is used as the reference); tetrahedral $Al(OSi)_4$, which gives rise to a single resonance with characteristic Al chemical shift values for individual zeolites in the range 50–65 ppm; and $AlCl_4^-$, which may be present as a residue from the preparative process, has a resonance at about 100 ppm.

2.11 THERMAL ANALYSIS

Thermal analysis methods investigate the properties of solids as a function of a change in temperature. They are useful for investigating phase changes, decomposition and loss of water or oxygen and for constructing phase diagrams.

2.11.1 Differential Thermal Analysis

A phase change produces either an absorption or an evolution of heat. The sample is placed in one chamber, and a solid that does not change phase over the temperature range of the experiment is placed in another chamber. Both chambers are

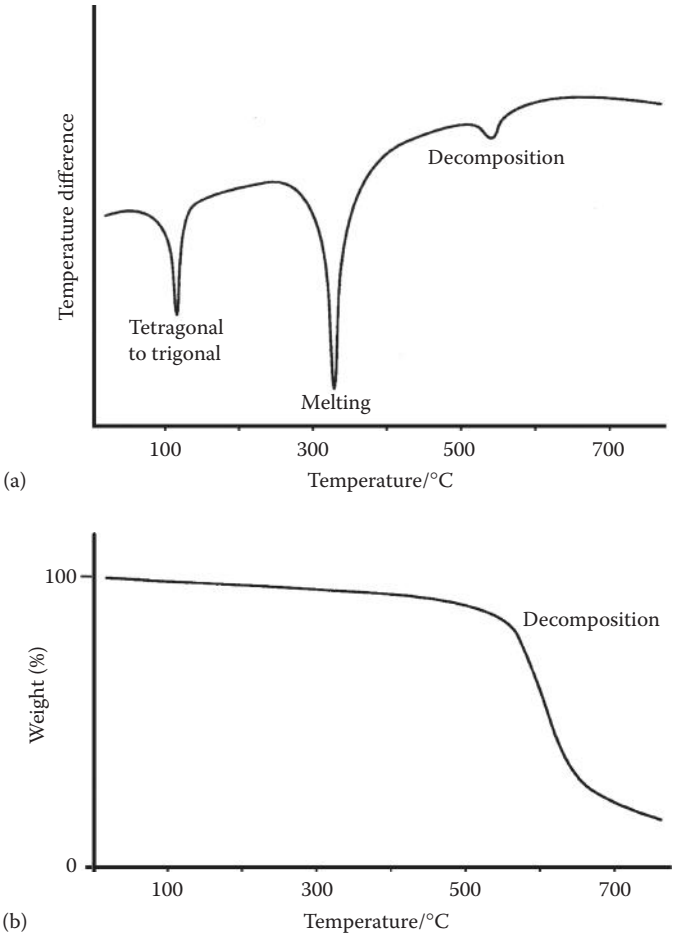


FIGURE 2.36 (a) The DTA and (b) the TGA trace for KNO_3.

heated at a controlled uniform rate in a furnace, and the difference in temperature between the two is monitored and recorded against time. Any reaction in the sample will show as a peak in the plot of differential temperature; exothermic reactions give rise to an increase in temperature, and endothermic reactions, a decrease in temperature, so the peaks appear in opposite directions. Figure 2.36a shows three exotherms in the differential thermal analysis (DTA) of KNO_3, due to (i) a phase change from tetragonal to trigonal at 129°C, (ii) melting at 334°C and (iii) decomposition above 550°C.

2.11.2 Thermogravimetric Analysis

In thermogravimetric analysis (TGA), the weight of a sample is monitored as a function of time as the temperature is increased at a controlled uniform rate. The loss of water due to crystallisation, or volatiles such as oxygen, shows up as a weight loss, as

does decomposition. Oxidation or adsorption of gas shows up as a weight gain. The TGA plot for KNO_3 in Figure 2.36b shows a small weight loss up to about 550°C, probably due to the loss of adsorbed water, followed by a dramatic weight loss when the sample decomposes.

2.11.3 Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) measures the amount of heat released by a sample as the temperature is increased or decreased at a controlled uniform rate. DSC can investigate chemical reactions and measure the heat of reaction for phase changes (Figure 2.37).

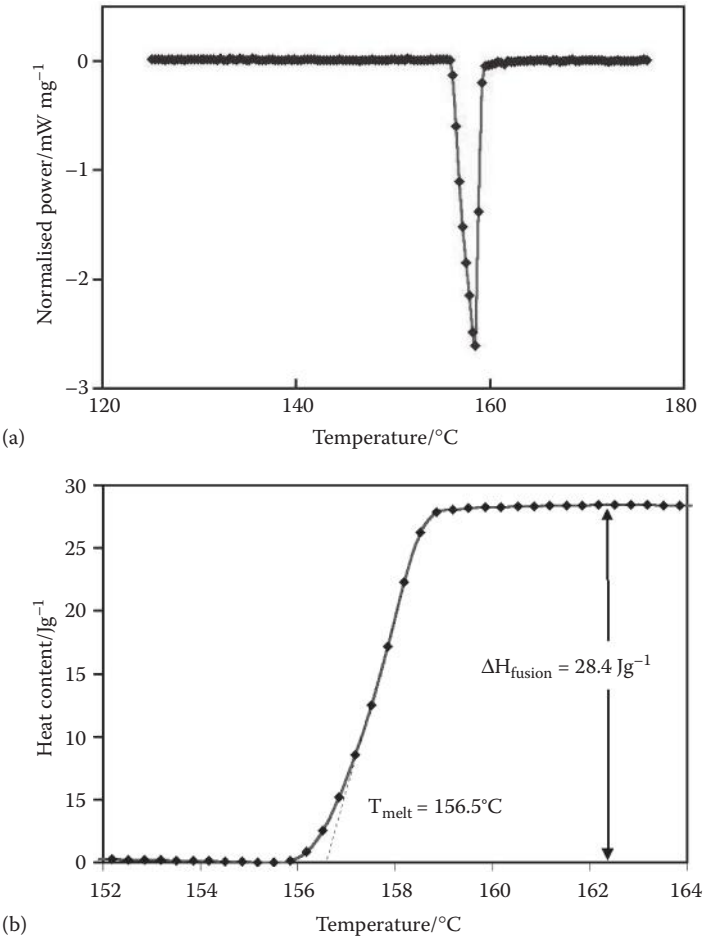


FIGURE 2.37 (a) DSC trace for the melting of indium metal and (b) integration of the power data to give the heat of fusion for In. (From Dr. Albert Sacco Jr., Northeastern University, MA.)

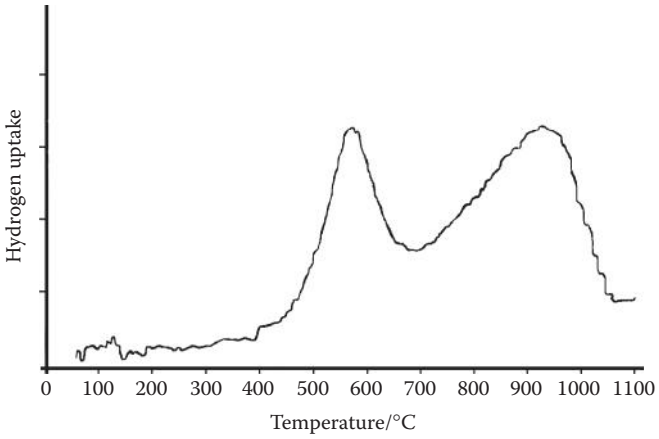


FIGURE 2.38 TPR trace for α-Fe_2O_3.

2.12 TEMPERATURE-PROGRAMMED REDUCTION

Temperature-programmed reduction (TPR) measures the reaction of hydrogen with a sample at various temperatures. The results are interpreted in terms of the different species present in the sample and their amenability to reduction, thereby giving information on the presence of different oxidation states or the effect of a dopant in a lattice. It is useful for measuring the temperature necessary for the complete reduction of a catalyst and is commonly used to investigate the interaction of a metal catalyst with its support or of the effect of a promoter on a metal catalyst.

The sample is heated over time in a furnace under a flowing gas mixture, typically 10% H_2 in N_2. Hydrogen has a high electrical conductivity, so a decrease in hydrogen concentration is marked by a decrease in the conductivity of the gas mixture; this change is measured by a thermal conductivity cell (katharometer) and is recorded against either time or temperature.

The peaks in a TPR profile (Figure 2.38) measure the temperature at which various reduction steps take place. The amount of hydrogen used can be determined from the area under the peak. In the example shown, we see two peaks, indicating that the reduction of Fe_2O_3 takes place in two stages, first to Fe_3O_4 and then to metallic Fe.

Other temperature-programmed techniques include temperature-programmed oxidation (TPO) and temperature-programmed sulfidation (TPS), for investigating oxidation and sulfidation behaviour, and temperature-programmed desorption (TPD) (also called thermal desorption spectroscopy, TDS), which analyses gases desorbed from the surface of a solid or a catalyst upon heating.

2.13 OTHER TECHNIQUES

In an introductory general text such as this, it is impossible to do justice to the plethora of physical techniques that are currently available to the solid-state chemist.

We have therefore concentrated on those that are most widely available either in the laboratory or in national facilities. For other techniques, you will need to refer to more specialised texts. We have not covered vibrational spectroscopy (IR and Raman) as good treatments of this subject are to be found in most undergraduate physical chemistry textbooks. Our coverage is of necessity fairly brief, and much more detail is to be found in the texts listed in the Further Reading section at the end of the book.

QUESTIONS

1. What are the spacings of the *100*, *110* and *111* planes in a cubic crystal system of unit cell dimension a? In what sequence would you expect to find these reflections in a powder diffraction photograph?
2. What is the sequence of the following reflections in a primitive cubic crystal: *220*, *300* and *211*?
3. Nickel crystallises in a cubic crystal system. The first reflection in the powder pattern of nickel is the *111*. What is the Bravais lattice?
4. The $\sin^2\theta$ values for Cs_2TeBr_6 are listed in Table 2.4 for the observed reflections. To which cubic class does it belong? Calculate its unit cell length assuming Cu-Kα radiation of wavelength 154.2 pm.
5. X-ray powder data for NaCl are listed in Table 2.5. Determine the Bravais lattice, assuming that it is cubic.
6. Use the data given in Question 5 to calculate the length for an NaCl unit cell.
7. If the unit cell length of NaCl is $a = 563.1$ pm and the density of NaCl is measured to be 2.17×10^3 kg m^{-3}, calculate Z, the number of formula units in the unit cell. (The atomic masses of sodium and chlorine are 22.99 and 35.45, respectively.)
8. A powder diffraction pattern establishes that silver crystallises in a face-centred cubic unit cell. The *111* reflection is observed at $\theta = 19.1°$ using Cu-Kα radiation. Determine the unit cell length, a.
   If the density of silver is 10.5×10^3 kg m^{-3} and $Z = 4$, calculate the value of the Avogadro constant. (The atomic mass of silver is 107.9.)

TABLE 2.4
$\sin^2\theta$ Values for Cs_2TeBr_6

$\sin^2\theta$
0.0149
0.0199
0.0399
0.0547
0.0597
0.0799
0.0947

TABLE 2.5
θ Values for NaCl

θ_{hkl}
13°41′
15°51′
22°44′
26°56′
28°14′
33°7′
36°32′
37°39′
42°0′
45°13′
50°36′
53°54′
55°2′
59°45′

9. Calcium oxide crystallises with a face-centred cubic lattice, $a = 481$ pm and a density $\rho = 3.35 \times 10^3$ kg m^{-3}. Calculate a value for Z. (The atomic masses of Ca and O are 40.08 and 15.999, respectively.)
10. Thorium diselenide ($ThSe_2$) crystallises in the orthorhombic system with $a = 442.0$ pm, $b = 761.0$ pm, $c = 906.4$ pm and a density $\rho = 8.5 \times 10^3$ kg m^{-3}: calculate Z. (The atomic masses of Th and Se are 232.03 and 78.96, respectively.)
11. Cu crystallises with a cubic close-packed structure. The Bragg angles of the first two reflections in the powder pattern collected using Cu-Kα radiation are 21.6° and 25.15°. Calculate the unit cell length a, and estimate the radius of the Cu atom.
12. Arrange the following atoms in order of their ability to scatter X-rays: Na, Co, Cd, H, Tl, Pt, Cl, F and O.
13. In a cubic crystal, we observe the *111* and *222* reflections, but not *001*. What is the Bravais lattice?
14. Interpret the EXAFS radial distribution function for $Co(CO)_4$ shown in Figure 2.39.
15. Figure 2.40 shows the ^{79}Si MAS NMR spectrum of the zeolite faujasite. Use this figure to determine which Si environments are most likely to be present.
16. Figure 2.41 shows the ^{29}Si spectrum of the same sample of faujasite as in Figure 2.40, but after treatment with $SiCl_4$ and washing with water. What has happened to it?
17. A sample of faujasite was treated with $SiCl_4$ and four ^{27}Al MAS NMR spectra were taken at various stages afterwards (Figure 2.42). Describe carefully what has happened during the process.

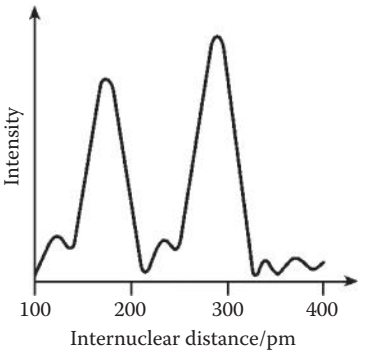


FIGURE 2.39 Co edge EXAFS radial distribution function for $Co(CO)_4$.

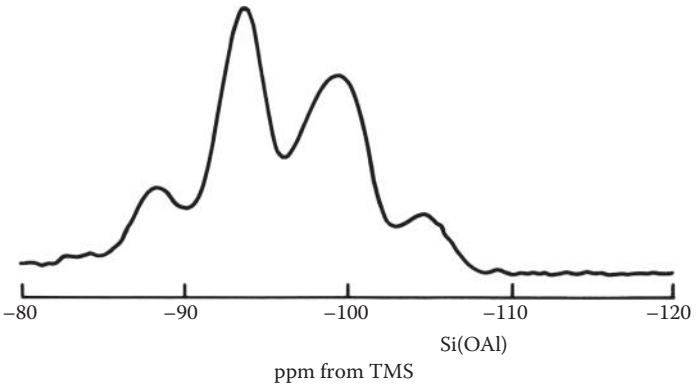


FIGURE 2.40 ^{29}Si MAS NMR spectrum at 79.6 MHz, of faujasite with Si/Al = 2.61 (zeolite-Y).

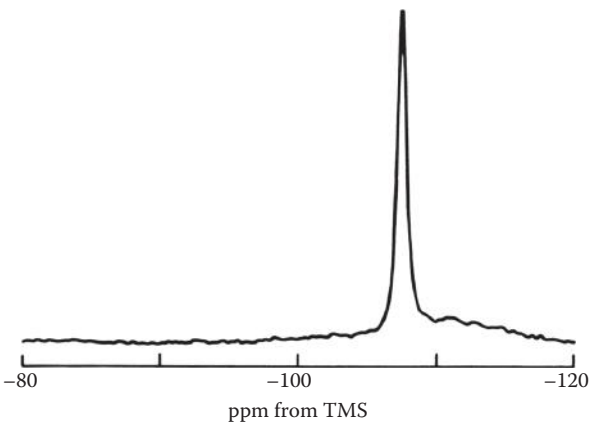


FIGURE 2.41 ^{29}Si MAS NMR spectrum at 79.6 MHz, of faujasite with Si/Al = 2.61 after successive dealumination with $SiCl_4$ and washings.

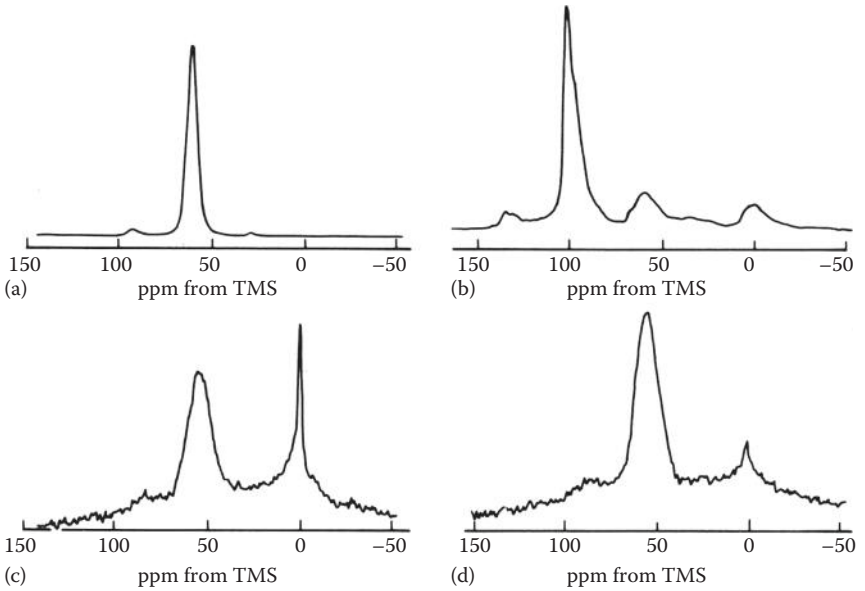


FIGURE 2.42 ^{27}Al MAS NMR spectra at 104.2 MHz obtained on faujasite samples at various stages of the $SiCl_4$ dealumination procedure. (a) Starting faujasite sample. (b) Intact sample after reaction with $SiCl_4$ before washing. (c) Sample (b) after washing with distilled water. (d) After several washings.

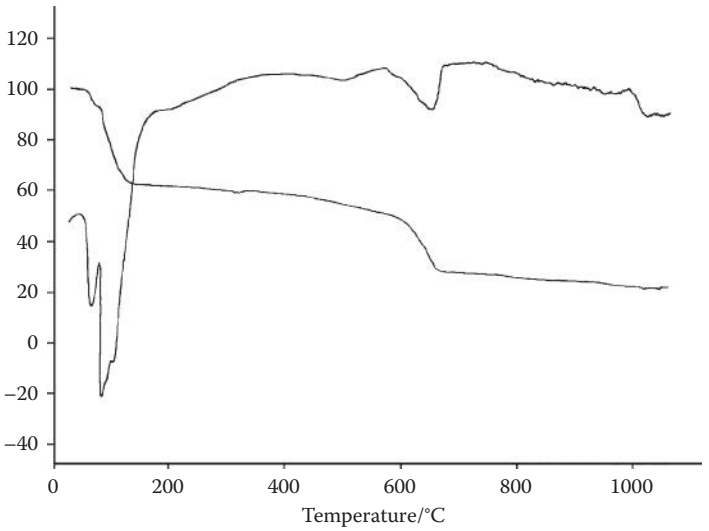


FIGURE 2.43 DTA and TGA traces for ferrous sulfate heptahydrate.

18. A TGA trace for 25 mg of a hydrate of manganese oxalate ($MnC_2O_4 \cdot xH_2O$) showed a weight loss of 20 mg at 100°C. What was the composition of the hydrate? A further weight loss occurs at 250°C, but then a weight *gain* at 900°C. What processes might be taking place?
19. Figure 2.43 shows the DTA and TGA traces for ferrous sulfate heptahydrate. Describe what processes are taking place.
20. What is the standard deviation of a bond length of 2.620(1) Å?

3 Synthesis of Solids

3.1 INTRODUCTION

Interest in the properties of solids and in the development of new materials has given rise to the development of a huge variety of methods for preparing them. The method chosen for any solid will depend not only on the composition of the solid but also on the form it is required in for its proposed use. For example, silica glass for fibre optics needs to be much freer of impurities than silica glass used for manufacturing laboratory equipment. Some methods may be particularly useful for producing solids in forms that are not stable under normal conditions, for example, the synthesis of diamond employs high pressures. Other methods may be chosen because they favour the formation of unusual oxidation states, for example, the preparation of chromium dioxide by the hydrothermal method, or because they promote the production of fine powders or, by contrast, large single crystals. For industrial use, a method that does not employ high temperatures could be favoured because of the ensuing energy savings.

In the preparation of solids, usually, care has to be taken to use stoichiometric quantities, pure starting materials and to ensure that the reaction has been completed, because it is usually not possible to purify a solid once it has formed.

We do not have space here to discuss all the ingenious syntheses that have been employed over the past few years, so we shall concentrate on those that are commonly used, with a few examples of the techniques used for solids with particularly interesting properties. The preparation of organic solid-state compounds and polymers is not covered as, generally, it involves organic synthesis techniques, which is a whole field in itself, and is covered in many organic textbooks.

It is difficult to impose a logical order on such a diverse subject. The chapter starts by considering the most basic and most commonly used method of preparing solids, **the ceramic method**: this grand title disguises the fact that it simply means grinding up the reactant solids and heating them hard until they react! We then go on to look at refinements of this method and ways of improving the uniformity of the reaction and reducing the reaction temperature. The following sections on microwave heating and combustion synthesis show alternative methods of inducing solid-state reactions. Later sections concentrate on less well-known methods of preparing inorganic solids, such as using high pressures and gas-phase reactions. We also consider some methods used for the production of particularly pure solids, which are important in the semiconductor industry, for the preparation of single crystals and for the preparation of nanostructured solids.

3.2 HIGH-TEMPERATURE CERAMIC METHODS

3.2.1 Direct Heating of Solids

Ceramic method: The simplest and most common way of preparing solids is the ceramic method, which consists of heating together two nonvolatile solids, which react to form the required product. This method is widely used both industrially and in the laboratory and can be used to synthesise a whole range of materials, such as mixed metal oxides, sulfides, nitrides, aluminosilicates and many others—the first high-temperature superconductors were made by a ceramic method. To take a simple example, we can consider the formation of zircon ($ZrSiO_4$), which is used in the ceramics industry as the basis of high-temperature pigments to colour the glazes on bathroom porcelain. It is made by the direct reaction of zirconia (ZrO_2) and silica (SiO_2) at 1300°C:

$$ZrO_2(s) + SiO_2(s) = ZrSiO_4(s).$$

The procedure is to take accurately weighed, stoichiometric amounts of the binary oxides, using the purest of solids, grind them in a pestle and mortar to give a uniform small particle size and then heat in a furnace for several hours in an alumina crucible (Figure 3.1).

Despite its widespread use, the simple ceramic method has several disadvantages. High temperatures are generally required, typically between 500°C and 2000°C, and this requires a large input of energy. This is because the coordination numbers in these ionic binary compounds are high, varying from 4 to 12, depending on the size and charge of the ion, and it takes a lot of energy to overcome the lattice energy so that a cation can leave its position in the lattice and diffuse to a different site. In addition, the phase or compound desired may be unstable or decompose at such high temperatures. Reactions such as this can be very slow; increasing the temperature speeds up the reaction as it increases the diffusion rate of the ions. In general, the solids are not raised to their melting temperatures, so the reaction occurs in the solid state. Solid-state reactions can only take place at the interface of two solids, and once the surface layer has reacted, the reaction continues as the reactants diffuse from the bulk to the interface. Raising the temperature enables the reaction at the interface and the diffusion through the solid to be faster than at room temperatures; a rule of thumb suggests that a temperature of about two-thirds of the melting temperature of the solids gives a reasonable reaction time. Even so, diffusion is often the limiting step. Because of this, it is important that the starting materials are ground to give a small particle size and are very well mixed to maximise the surface contact area and minimise the distance that the reactants have to diffuse.

To achieve a homogeneous mix of small particles, it is necessary to be very thorough in grinding the reactants. The number of crystallite faces in direct contact with one another can also be improved by pelletising the mixed powders in a hydraulic press. Commonly, the reaction mixture is removed during the heating process and is reground to bring fresh surfaces in contact, thereby speeding up

(a)

(b)

(c)

FIGURE 3.1 The basic apparatus for the ceramic method: (a) pestles and mortars for fine grinding; (b) a selection of porcelain, alumina and platinum crucibles; and (c) a furnace.

the reaction. Nevertheless, the reaction time is measured in hours; the preparation of the ternary oxide $CuFe_2O_4$ from CuO and Fe_2O_3, for example, takes 23 h. The product is often not homogeneous in composition because as the reaction proceeds, a layer of the ternary oxide is produced at the interface of the two crystals, so ions now need to diffuse through this before they react. It is usual to take the initial product, grind it again and reheat several times before a phase-pure product is obtained. Usually trial and error has to be used to find out the best reaction conditions, with samples tested by powder X-ray diffraction to determine the phase purity.

Solid-state reactions up to 2000°C are usually carried out in furnaces that use resistance heating: the resistance of a metal element results in electrical energy being converted to heat, as in an electric fire. This is a common method of heating up to 2000°C; however, above this temperature, other methods have to be employed. Higher temperatures in samples can be achieved by directing an electric arc at the reaction mixture (to 3000°C), and for very high temperatures, a carbon dioxide laser (with output in the infrared) can give temperatures up to 4000°C.

Containers must be used for the reactions that can withstand both high temperatures and are also sufficiently inert not to react themselves; suitable crucibles are commonly made of silica (to 1400°C), alumina (to 1900°C), zirconia (to 2000°C) or magnesia (to 2400°C), but metals such as platinum (1700°C) and tantalum and graphite linings are used for some reactions.

If any of the reactant oxides are nonstoichiometric, volatile, hygroscopic or air-sensitive, then it may be possible to use carbonates instead. If this simple method of heating in the open atmosphere is no longer appropriate, a sealed tube method will be needed.

Sealed tube methods: Evacuated tubes are used when the products or reactants are sensitive to air or water or are volatile. An example of the use of this method is the preparation of samarium sulfide. In this case, sulfur has a low boiling temperature (444°C) and an evacuated tube is necessary to prevent it from boiling off and being lost from the reaction vessel.

The preparation of samarium sulfide (SmS) is of interest because it contains samarium (a lanthanide element) in an unusual oxidation state of +2 rather than the more common state of +3. Samarium metal in powder form and powdered sulfur are mixed together in stoichiometric proportions and heated to around 730°C in an evacuated silica tube. (Depending on the temperature of the reaction, Pyrex or silica is the common choice for these reaction tubes, as both are fairly inert, and can be sealed onto a Pyrex vacuum system for easy handling.) The product from the initial heating is then homogenised and heated again, this time to around 2000°C in a tantalum tube (sealed by welding) by passing an electric current through the tube, the resistance of the tantalum providing the heating.

A high pressure of oxygen may be used to prepare an oxide of high-oxidation state or hydrogen for a reduction reaction. An inert atmosphere may be used to prevent oxidation or decomposition.

The pressures obtained in sealed reaction tubes can be very high, and it is not unknown for tubes to explode, however carefully they are made; it is thus very important to seal them carefully and to take safety precautions such as surrounding the tube with a protective metal container and using safety screens.

Special atmospheres: The preparation of some compounds must be carried out under a special atmosphere, but not necessarily at high pressures: an inert gas such as argon may be used to prevent oxidation to a higher-oxidation state; an atmosphere of oxygen can be used to encourage the formation of high-oxidation states; or, conversely, a hydrogen atmosphere can be used to produce a low-oxidation state. Experiments of this type are usually carried out with the reactant solids in a small boat placed in a tube in a horizontal-tube furnace. The gas is passed over the reactants for a time to expel all the air from the apparatus, and then it flows over the reactants during the heating and cooling cycles, exiting through a bubbler to maintain a positive pressure and prevent the ingress of air by back diffusion (Figure 3.2).

To obtain more homogeneous products and better reaction rates, methods involving particles smaller than those that can be obtained by grinding have been introduced, and these methods are described in the next section.

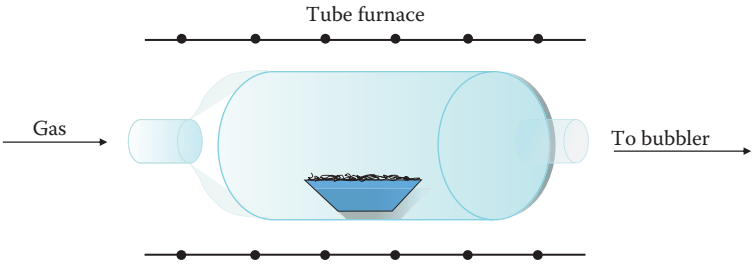


FIGURE 3.2 Reacting solids under a special gas atmosphere.

3.2.2 Reducing the Particle Size and Lowering the Temperature

In a polycrystalline mixture of solid reactants, we might expect the particle size to be of the order of 10 μm; even careful and persistent grinding will only reduce the particle size to around 0.1 μm. Diffusion during a ceramic reaction therefore takes place across anywhere between 100 and 10,000 unit cells. Various ingenious methods, some physical and some chemical, have been pioneered to bring the components of the reaction either into more intimate contact or into contact at an atomic level, thereby reducing this diffusion path; in doing so, the reactions can often take place at lower temperatures.

In **spray drying** the reactants are dissolved in a suitable solvent and sprayed as fine droplets into a hot chamber. The solvent evaporates, leaving a mixture of the solids as a fine powder, which can then be heated to obtain the product. This method can also be used to control the shape and uniformity of the powder particles.

In **freeze drying** the reactants are dissolved in a suitable solvent and are frozen to liquid nitrogen temperatures (77 K). The solvent is then removed by pumping to leave a fine reactive powder.

At a simple level, **precursors** such as nitrates and carbonates can be used as starting materials instead of oxides: they decompose to the oxides upon heating at relatively low temperatures, losing gaseous species and leaving behind fine, more reactive powders. An even more intimate mixture of starting materials can be made by the **co-precipitation** of solids. A stoichiometric mixture of soluble salts of the metal ions is dissolved and then precipitated as hydroxides, citrates, oxalates or formates. This mixture is filtered, dried, ground and then heated to give the final product. It is necessary to be very careful not to introduce impurities during grinding.

The **precursor method** achieves mixing at the atomic level by forming a solid compound, the precursor, in which the metals of the desired compound are present in the correct stoichiometry. For example, if an oxide MM'_2O_4 is required, a mixed salt of an oxyacid, such as an acetate containing M and M′ in the ratio of 1 : 2, is formed. The precursor is then heated to decompose it to the required product. Homogeneous products are formed at relatively low temperatures. A disadvantage is that it is not always possible to find a suitable precursor, but the preparation of barium titanate gives a good illustration of this method.

Barium titanate ($BaTiO_3$) is a ferroelectric material (Chapter 8) that is widely used in capacitors because of its high dielectric constant. It was initially prepared by heating barium carbonate and titanium dioxide at high temperature

$$BaCO_3(s) + TiO_2(s) = BaTiO_3(s) + CO_2(g).$$

However, for modern electronic circuits, it is important to have a product of a controlled grain size and the precursor method is one way to achieve this. (Another way is the sol–gel method, which has also been applied to this material.) The precursor used is an oxalate. The first step in the preparation is to prepare an oxo-oxalate of titanium. Excess oxalic acid solution is added to titanium butoxide, which initially hydrolyses to give a precipitate and then redissolves in the excess oxalic acid

$$Ti(OBu)_4(aq) + 4H_2O(l) = Ti(OH)_4(s) + 4BuOH(aq)$$

$$Ti(OH)_4(s) + (COO)_2^{2-}(aq) = TiO(COO)_2(aq) + 2OH^-(aq) + H_2O(l).$$

Barium chloride solution is then added and barium titanyl oxalate precipitates

$$Ba^{2+}(aq) + (COO)_2^{2-}(aq) + TiO(COO)_2(aq) = Ba[TiO((COO)_2)_2](s).$$

This precipitate contains barium and titanium in the correct ratio and is easily decomposed by heat, to give the oxide. The temperature used for this final heating is 650°C

$$Ba[TiO((COO)_2)_2](s) = BaTiO_3(s) + 2CO_2(g) + 2CO(g).$$

The decomposition of oxalates is also involved in the preparation of ferrites MFe_2O_4, which are important as magnetic materials (Chapter 8).

A precursor can be used in the preparation of promoted bimetallic catalysts. The noble metal of a supported catalyst, for instance, Pt on carbon, is allowed to adsorb hydrogen and then to react with a solution of an organometallic complex of a second metal. The organometallic complex reacts with, and is reduced by, the hydrogen adsorbed in the Pt, to give a metal and a gaseous by-product. The reaction occurs only at the active catalyst sites where the hydrogen has been adsorbed and gives a homogeneous bimetallic product with a very close interface between the two metals.

The products of the precursor method are usually crystalline solids, often containing small particles of large surface area. For applications, such as catalysis and barium titanate capacitors, this is an advantage.

The precipitation methods always have the disadvantage that the stoichiometry of the precipitate(s) may not be exact if one or more ions are left in the solution. The **sol–gel method** overcomes this, as the reactants never precipitate out. First, a

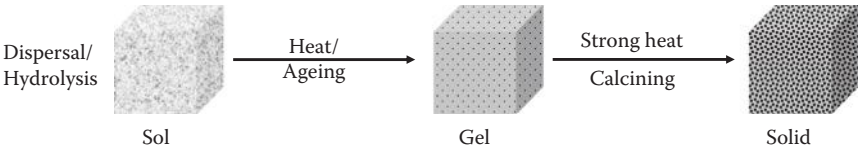


FIGURE 3.3 The steps in the sol–gel synthesis route.

concentrated solution or colloidal suspension of the reactants, the 'sol', is prepared, which is then concentrated or matured to form the 'gel'. This homogeneous gel is then heat-treated to form the product. The main steps in the sol–gel process are outlined in Figure 3.3.

The first investigation of a sol–gel process for synthesis was made in the mid-nineteenth century. This early investigation studied the preparation of silica glass from a sol made by hydrolysing an alkoxide of silicon. Unfortunately, in order to prevent the product cracking and forming a fine powder, an ageing period of a year or more was required. The sol–gel method was further developed in the 1950s and 1960s after it was realised that colloids, which contain small particles (1–1000 nm in diameter), can be highly chemically homogeneous and also that the particles have a very large surface area for reaction.

A sol is a colloidal suspension of particles in a liquid; for the materials being discussed here, these particles will typically be 1–100 nm in diameter. A gel is a semirigid solid in which the solvent is contained in a framework of material that is either colloidal (essentially a concentrated sol) or polymeric.

To prepare solids using the sol–gel method, a sol of reactants is first prepared in a suitable liquid. Sol preparation can be either simply the dispersal of an insoluble solid or the addition of a precursor that reacts with the solvent to form a colloidal product. A typical example of the first is the dispersal of oxides or hydroxides in water with the pH adjusted so that the solid particles remain in suspension rather than precipitate out. A typical example of the second method is the addition of metal alkoxides to water; the alkoxides are hydrolysed, giving the oxide as a colloidal product.

Subsequently, the sol is either treated or simply left to form a gel over time by dehydrating and/or polymerising. To obtain the final product, the gel is heated. This heating serves several purposes—it removes the solvent, it decomposes anions such as alkoxides or carbonates to give oxides, it allows rearrangement of the structure of the solid and it allows crystallisation to occur. Both the time and the temperature needed for the reaction in sol–gel processes can be reduced from those of the direct ceramic method; in favourable cases, the time from days to hours and the temperature by several hundred degrees.

Cracking in the gel can be avoided if it is dried under supercritical conditions, that is, when the solvent is above its critical temperature. The solvent then stays as a gas throughout the drying process, leaving behind a so-called aerogel, which can be fired to give a porous monolith.

Several examples given below illustrate the sol–gel method; two of the examples have been chosen because they have interesting properties and uses, and are

discussed later in this book. Many other materials have also been prepared by the sol–gel method, and other sol–gel preparations have also been employed for the materials chosen; therefore, these examples should be taken as illustrative and not as the main uses of the method.

Lithium niobate ($LiNbO_3$) is a ferroelectric material used as an optical switch (see Chapter 7). Preparation by the simple ceramic method leads to problems in obtaining the correct stoichiometry, and a mixture of phases often results. Several sol–gel preparations have been described, their advantage being the lower temperature required for the preparation and the greater homogeneity of the product. One such preparation starts with lithium ethoxide ($LiOC_2H_5$ (or LiOEt)) and niobium ethoxide ($Nb_2(OEt)_{10}$). Each metal ethoxide was dissolved in absolute ethanol and the two solutions were mixed. The addition of water leads to partial hydrolysis, giving hydroxy-alkoxides, for example:

$$Nb_2(OEt)_{10} + 2H_2O = 2Nb(OEt)_4(OH) + 2EtOH.$$

The hydroxy-alkoxides condense to form a polymeric gel with metal–oxygen–metal links. Lithium niobate is then formed when the gel is heated. Heating removes any remaining ethanol and also any water formed during the condensation. The remaining ethyl groups are pyrolysed (i.e., oxidised to carbon dioxide and water), leaving the oxides.

Tin dioxide (SnO_2) is an oxygen-deficient, *n*-type semiconductor (see Chapter 4) whose conductivity is increased by the addition of dopants such as Sb^{3+} ions. One use of **doped tin dioxide** is to form a transparent electrode. For this, it is deposited onto glass and must have a large surface area with a controlled distribution of dopants. Sol–gel methods have proved suitable for the preparation of such coatings because gels can be sliced finely. In one reported preparation of titanium-doped tin dioxide, the sol was prepared by adding titanium butoxide ($Ti(OC_4H_9)_4$ ($Ti(OBu)_4$)) to a solution of tin dichloride dihydrate ($SnCl_2{\cdot}2H_2O$) in absolute ethanol. Tin(II) salts are easily hydrolysed to give hydroxy complexes and can be oxidised to tin(IV) by oxygen in the air. The sol was left in an open container for 5 days, during which time it formed a gel. The gel was dried by heating at 60°C under reduced pressure for 2 h. Finally, it was heated to around 330°C for 10 min to produce a layer of tin dioxide doped with Ti^{4+} ions.

Coatings for optical fibres are important because the core of an optical fibre carries the light signal, and the cladding on the outside of the core, with a lower refractive index, traps the light in the centre by total internal reflection (see Chapter 7). Optical fibres need to be extremely pure, completely free of impurities such as transition metal ions; the conventional methods of preparing silica glasses are inadequate. The fibres are made by using a modified form of **chemical vapour deposition (CVD)** (Section 3.7) and the coatings and core are drawn out together. However, coatings can also be added around the silica core as the fibre is drawn through a sol. As well as forming the outer lower refractive-index layer, coatings can be added to give protection and added strength and to provide sensor properties.

An interesting application of the sol–gel method is in manufacturing **biosensors**. A biosensor uses a substance of biological origin, often an enzyme, to detect the presence of a molecule. To be of use as a sensor, the biological substance must be in a form that can be easily carried around and placed in a solution, in a stream of gas or on a patient's skin. Biological materials such as enzymes are potentially extremely useful as sensors because they can be very specific; for example, where a chemical test may be positive for any oxidising gas, an enzyme may only give a positive test for oxygen. However, enzymes are easily denatured by removing them from their aqueous environment or by altering the pH of their environment. One way around this problem is to trap the enzyme with its microenvironment in a silica gel. In one such preparation, a sol of silica was first prepared by acid hydrolysis of tetramethyl orthosilicate (TMOS; $Si(OMe)_4$). In a typical sol–gel synthesis, this compound would be dissolved in methanol; however, in this case, methanol was avoided because alcohols denature proteins, so the solvent used was water. The sol was then buffered to be at a suitable pH for the enzyme and the enzyme was added. As the sol formed a gel by condensing, the enzyme was trapped in the cavities in the gel surrounded by an aqueous medium of the correct pH. A urea biosensor for urine analysis can be formed by trapping urease between layers of sol–gel with an indicator in between other layers. The hydrolysis of urea by urease changes the pH, which in turn changes the colour of the indicator.

In this last case, the product was left as a gel, but if the product is to be a glass, a powder or a crystalline material, a high-temperature calcining step is required after the formation of the gel. Thus, the sol–gel method improves the homogeneity of the product, but reaction times are still long (note the 5-day gelling time in the tin dioxide synthesis) and high temperatures (1000°C for silica) are still needed.

The use of microwave ovens rather than conventional heating leads to speeding up of the reaction times in favourable cases.

3.3 MICROWAVE SYNTHESIS

We are familiar with the use of microwave radiation in cooking food, where it increases the speed of reaction. Recently, this method has been used for the synthesis of solid-state materials such as mixed oxides. The first solid-state reaction experiments were performed in modified domestic ovens, which are still used, but more specialised (and expensive!) ovens have also been developed to give more control over the conditions. We shall briefly consider how microwaves heat solids and liquids because this gives us an insight into which reactions will be good candidates for this method.

In a liquid or solid, the molecules or ions are not free to rotate, so the heating is *not* the result of the absorption of microwaves by molecules undergoing rotational transitions as they would in the gas phase. In a solid or liquid, the alternating electric field of the microwave radiation can act in two ways. If charged particles are present that can move freely through the solid or liquid, then these will move under the influence of the field, producing an oscillating electric current. Resistance to their movement causes energy to be transferred to the surroundings as heat. This is **conduction**

heating. If there are no particles that can move freely but there are molecules or units with dipole moments, then the electric field acts to align the dipole moments. It is this effect that produces **dielectric heating**, and when it acts on water molecules in food, it is generally responsible for the heating/cooking in domestic microwave ovens. The electric field of microwave radiation, like that of all electromagnetic radiations, oscillates at the frequency of the radiation. The electric dipoles in the solid do not change their alignment instantaneously, but with a characteristic time (τ). If the oscillating electric field changes its direction slowly so that the time between changes is much greater than τ, then the dipoles can follow the changes. A small amount of energy is transferred to the surroundings as heat each time the dipole realigns, but this is only a small heating effect. If the electric field of the radiation oscillates very rapidly, then the dipoles cannot respond fast enough and do not realign. The frequency of microwave radiation is such that the electric field changes sign at a speed that is the same order of magnitude as τ. Under these conditions, the dipole realignment lags slightly behind the change of electric field and the solid absorbs microwave radiation. This absorbed energy is converted to heat. The quantities governing this process are the dielectric constant (see Chapter 8), which determines the extent of dipole alignment, and the dielectric loss, which governs how efficiently the absorbed radiation is converted to heat.

To use microwave heating in solid-state synthesis, at least one component of the reaction mixture must absorb microwave radiation. The speed of the reaction process is then increased by increasing both the rate of the solid-state reaction and the rate of diffusion, which, as we mentioned earlier, is often the rate-limiting step.

3.3.1 High-Temperature Superconductor $YBa_2Cu_3O_{7-x}$

High-temperature superconductors (see Chapter 9) were first prepared using a conventional ceramic method; in this example, by baking together yttrium oxide, copper(II) oxide and barium carbonate. The synthesis, however, takes 24 h to complete. A microwave method that can prepare the superconductor in under 2 h has been reported. A stoichiometric mixture of copper(II) oxide (CuO), barium nitrate ($Ba(NO_3)_2$) and yttrium oxide (Y_2O_3) was placed in a microwave oven that had been modified to allow safe removal of the nitrogen oxides formed during the reaction. The mixture was treated with 500 W of microwave radiation for 5 min, then reground and exposed to microwave radiation at 130–500 W for 15 min. Finally, the mixture was ground again and exposed to microwave radiation for a further 25 min. The microwaves in this example couple to the copper(II) oxide.

This is one of a range of oxides, mostly nonstoichiometric, that are efficiently heated by microwave radiation. For example, a 5–6 g sample of CuO exposed to 500 W of microwave radiation for half a minute attained a temperature of 800°C. Other oxides that strongly absorb microwave radiation include ZnO, V_2O_5, MnO_2, PbO_2, Co_2O_3, Fe_3O_4, NiO and WO_3. Carbon, $SnCl_2$ and $ZnCl_2$ are also strong absorbers. By contrast, a 5–6 g sample of calcium oxide (CaO) exposed to 500 W of microwave radiation only attained a temperature of 83°C after half an hour. Other oxides that do not absorb strongly include TiO_2, CeO_2, Fe_2O_3, Pb_3O_4, SnO, Al_2O_3 and La_2O_3.

3.3.2 Batteries and Fuel Cells

The spinel structure $Li_4Ti_5O_{12}$ can be used as an anode in lithium-ion batteries and can be made by microwaving Li_2CO_3 and TiO_2. Microwave synthesis is being used to develop new electrode materials (see Chapter 5). One promising high-density cathode material for a solid oxide fuel cell is the perovskite $La_{1-x},Sr_xCo_{1-y}Fe_yO_3$, which can be made in the microwave in 40 min at 800 W from La_2O_3, $SrCO_3$, Co_2O_3 and Fe_2O_3.

3.3.3 Graphene

When graphene sheets (see Chapter 10) are mixed with graphite oxide in an atmosphere of hydrogen and microwaved for only 1 min, the heat generated exfoliates the graphite oxide to give high-quality graphene sheets.

Microwave synthesis has many potential advantages for synthesis: it is fast, clean and much more economical than conventional furnace methods. Also, there can be fewer problems with cracking because the heating is throughout the sample, and not absorbed from outside, and reactions can take place at lower temperatures than in the ceramic method. It is sometimes possible to synthesise metastable states or products with an unusual morphology. However, the disadvantage is that it does not solve the problem of chemical inhomogeneity, which still needs to be maximised carefully; it is also somewhat of a 'black art' and experiments are difficult to design, but may be trial and error.

Increased homogeneity can be obtained with the hydrothermal method discussed in Section 3.5, which also operates at lower temperatures than conventional methods.

3.4 COMBUSTION SYNTHESIS

Combustion synthesis, also known as **self-propagating high-temperature synthesis**, has been developed as an alternative route to the ceramic method, which depends on the diffusion of ions through the reactants, and thus for a uniform product, requires repeated heating and grinding for long periods. Combustion synthesis uses highly exothermic ($\Delta H \sim -170$ kJ mol^{-1}) and even explosive redox reactions to maintain a self-propagating, high reaction temperature and has been used to prepare many refractory materials, including borides, nitrides, oxides, silicides, intermetallics and ceramics. In the solid state, the reactants are mixed together, formed into a cylindrical pellet and then ignited at a single point on the surface (laser, electric arc, heating coil) at high temperature. Once ignited, the reaction propagates as a hot combustion wave, the **synthesis wave**, through the pellet, and the reaction must lose less heat than it generates or it will quench; high temperatures (1700°C–3700°C) are maintained during the fast reactions. Figure 3.4 shows the synthesis wave moving through a mixture of MgO, Fe, Fe_2O_3 and $NaClO_4$ (the images were captured by a thermal-imaging camera).

Ignition is sometimes initiated by heating the whole volume evenly, and self-ignition can sometimes be achieved by ball milling. The method is now used commercially in Russia, Spain and Korea because it is fast, economical and gives

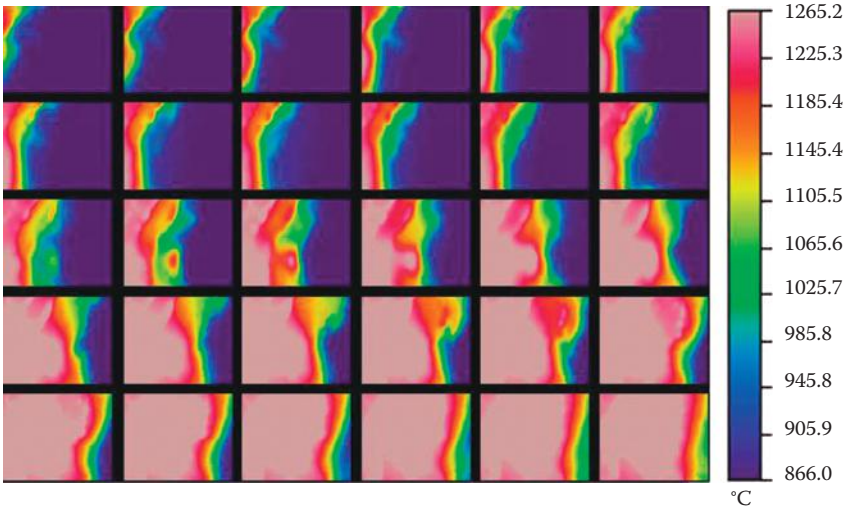


FIGURE 3.4 Thermal images of the synthesis wave moving through a pellet of MgO, Fe, Fe_2O_3 and $NaClO_4$. Each image is of dimension 3 × 2 mm. Images were captured at 0.06 s intervals. The first image is top left and the last is bottom right. (From Professor Ivan Parkin, University College, London. With permission.)

high-purity products. Thermal ignition of a mixture of BaO_2, Fe_2O_3, Cr_2O_3 and Fe produces a ferrite (Chapter 8), $BaFe_{12-x}Cr_xO_{19}$; in the presence of a magnetic field, the reaction proceeds faster and with a higher-temperature propagation wave; the properties of the barium hexaferrite are also improved, demonstrating greater coercivity and remanence. Examples of other useful products of these reactions are

- Hydrides, for example, MgH_2 for hydrogen storage
- Borides, for example, TiB_2 for abrasives and cutting tools
- Carbides, for example, SiC and TiC for abrasives and cutting tools
- Nitrides, for example, Si_3N_4 for high strength, heat-resistant ceramics
- Oxides, for example, cuprates for high-temperature superconductors
- Silicides, for example, $MoSi_2$ for high-temperature heating elements
- Composites, for example, the thermite reaction, $3Fe_3O_4 + 8Al = 4Al_2O_3 + 9Fe$ has been used to coat the inside of steel pipes with an inner layer of Fe and a surface layer of Al_2O_3, using a centrifugal process.

More unexpectedly perhaps, combustion synthesis can also take place in solutions, provided there is a sufficient quantity of a highly exothermic oxidiser present in the solution as a fuel. Various fuels have been used, such as urea or different hydrazides, and they not only provide the carbon and hydrogen necessary for heat production, but also form a complex with the metal salts present, thus ensuring a very homogeneous mixture of reactants. The reactions are initiated by microwave heating and produce flaming or smouldering mixtures with temperatures from about 700°C to 1500°C.

The most commonly used fuel in solution combustion synthesis is urea, which is cheap and easily available; it has been used mostly to produce alumina products, and the list is long and diverse: nanoparticles, abrasives, supported metal catalysts—for example Pt/alumina, doped-alumina anode materials, phosphors and pigments.

3.5 HIGH-PRESSURE METHODS

High pressures are employed directly and indirectly in various synthetic methods:

- High solvent pressures in an autoclave—the hydrothermal method
- Using a high pressure of a reactive gas
- Directly applied hydrostatic pressure on solids
- Detonation
- Ultrasound.

3.5.1 Hydrothermal Methods

The hydrothermal method increases homogeneity and lowers the operating temperatures and is also used to grow single crystals, which are needed in some solid-state applications.

The original **hydrothermal method** involves heating the reactants in a closed vessel, an **autoclave**, with water (Figure 3.5). An autoclave is usually constructed from thick stainless steel to withstand the high pressures and is fitted with safety valves; it may be lined with nonreactive materials, such as the noble metals, quartz or Teflon. The autoclave is heated, the pressure increases and the water remains liquid above its normal boiling temperature of 100°C, so-called 'superheated water'. These conditions, in which the pressure is raised above atmospheric pressure and the temperature is raised above the boiling temperature of water but not to as high a temperature as used in the methods described previously, are known as hydrothermal conditions. Hydrothermal conditions exist in nature, and numerous minerals including naturally occurring zeolites and gemstones, are formed by this process. The term has been extended to other systems with moderately raised pressures and temperatures lower than those typically used in ceramic and sol–gel syntheses. The lower temperatures used are one of the advantages of this method. Others include the preparation of compounds in unusual oxidation states or phases, which are stabilised by the raised temperature and pressure. It is used industrially to prepare large crystals of quartz and synthetic gemstones. It is useful in metal oxide systems, where the oxides are not soluble in water at atmospheric pressure but dissolve in the superheated water of the hydrothermal setup. Where even these temperatures and pressures are insufficient to dissolve the starting materials, alkali or metal salts as mineralisers can be added, whose anions form complexes with the solid and render it soluble. The following examples have been chosen to illustrate these advantages and also the different variations of the method.

The first industrial process for using hydrothermal methods of synthesis was the production of quartz crystals for use as oscillators in radios. **Quartz (SiO_2)** can be

FIGURE 3.5 An autoclave.

used for generating a high-frequency alternating current via the piezoelectric effect (Chapter 8). The hydrothermal growth of quartz crystals employs a temperature gradient. In this variant of hydrothermal processing, the reactant dissolves at a higher temperature, is transported up the reaction vessel by convection, and crystallises out with the aid of seed crystals at a lower temperature in the second cooler chamber at the top of the autoclave. A schematic diagram of the apparatus is shown in Figure 3.6. One end of the autoclave is charged with silica and an alkaline solution. The autoclave is heated to typically between 300°C and 400°C, generating pressures of up to 1500×10^5 Pa; the silica dissolves in the solution and is transported along the autoclave to the cooler part (10°C–25°C cooler), where it crystallises out. The alkaline solution returns to the hotter region, where it can dissolve more silica. In this particular case, the autoclave, which is made of steel, can act as the reaction vessel. Because of the corrosive nature of the superheated solution and the possibility of contamination of the product with material from the autoclave walls, the autoclave is usually lined with an inert substance such as Teflon or the reaction is performed in a sealed ampoule in the autoclave.

Zeolites are a class of crystalline aluminosilicates (see Chapter 6) based on rigid anionic frameworks with well-defined channels and cavities. These cavities contain exchangeable metal cations and can also hold removable and replaceable guest molecules such as water. The primary building units of zeolites are $[SiO_4]^{4-}$ and $[A1O_4]^{5-}$ tetrahedra linked together by corner sharing. Silicon–oxygen tetrahedra are

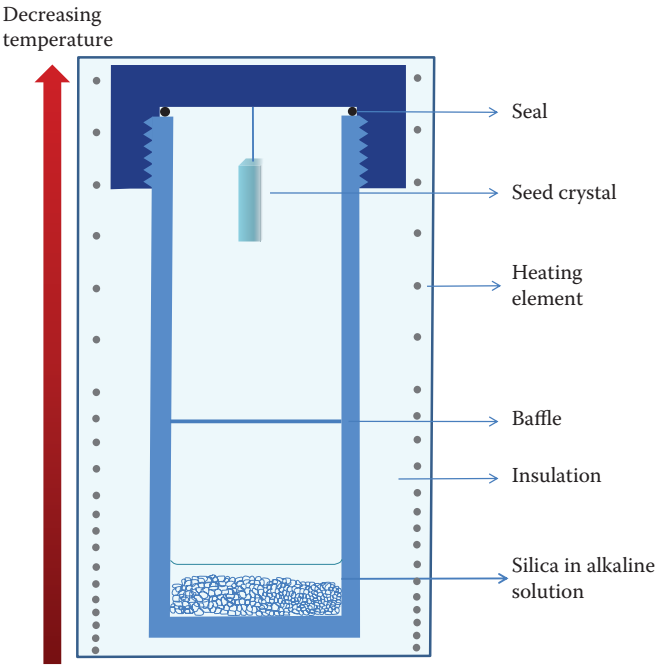


FIGURE 3.6 Growth of quartz crystals in an autoclave under hydrothermal conditions.

electrically neutral when connected together in a three-dimensional network as in quartz. The substitution of Si(IV) by Al(III) in such a structure, however, creates an electrical imbalance, and to preserve overall electrical neutrality, each $[AlO_4]^{5-}$ tetrahedron needs a balancing positive charge, which is provided by exchangeable cations held electrostatically within the zeolite. It is possible for the tetrahedra to link by sharing two, three or all four corners, forming a huge variety of different structures.

Naturally-occurring zeolite minerals are formed through a hydrothermal process geochemically, and in the 1940s and 1950s it was shown that they could be synthesised in the laboratory by simulating these hydrothermal conditions. A general method of preparing zeolites involves mixing an alkali, aluminium hydroxide and silica sol, or an alkali, a soluble aluminate and silica sol. The silica and aluminate condense to form a gel as in the sol–gel method, but the gel is then heated in a closed vessel at temperatures close to 100°C. Under these conditions, zeolites rather than other aluminosilicate phases crystallise out. For example, in the synthesis of a typical zeolite, zeolite A ($Na_{12}[(AlO_2)_{12}(SiO_2)_{12}]\cdot 27H_2O$), hydrated alumina ($Al_2O_3\cdot 3H_2O$) is dissolved in a concentrated sodium hydroxide solution. The cooled solution is then mixed with a solution of sodium metasilicate ($Na_2SiO_3\cdot 9H_2O$) and a thick white gel forms. The gel is then placed in a closed Teflon bottle and heated to about 90°C over 6 h. The reaction time can be reduced by using microwave radiation for this final step: for example, zeolite A has been made by heating for 45 min using pulses of 300 W of microwave radiation.

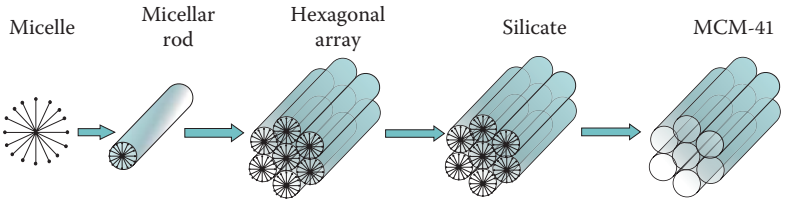


FIGURE 3.7 The formation of mesoporous MCM-41 by liquid crystal templating.

Changes in the form of the alumina, in the pH of the solution, in the type of base used (organic or inorganic) and in the proportions of alkali, aluminium compound and silica, lead to the production of different zeolites.

Microporous materials are defined as having pore diameters of less than 2 nm, and **macroporous** materials are defined as having diameters greater than 50 nm, so the **mesoporous** solids fall in between. Structures such as zeolites (see Chapter 6) can be formed around a single molecule template with pore sizes between 0.2 and 2 nm. In the zeolite synthesis, large cations such as tetramethylammonium (NMe_4^+) and tetrapropylammonium ($N(C_3H_7)_4^+$) are used as a **template** around which the aluminosilicate framework crystallises with large cavities to accommodate the ion. On subsequent heating, the cation is pyrolysed, but the structure retains the cavities.

Mesoporous zeolitic structures with pore sizes of 2–10 nm can be formed by a method known as **liquid crystal templating (LCT)**. The combination of a suitable cationic surfactant with silicate anions, forms arrays of rod-like surfactant micelles (Figure 3.7) surrounded by a polymeric siliceous framework. On calcination, the mesoporous structure is formed. If short-chain alkyltrimethylammonium ions ($R(CH_3)_3N^+$) are used as the surfactant, the alkyl chain length determines whether microporous or mesoporous materials are formed. For alkyl chain lengths of C_6 or shorter, microporous zeolites such as ZSM-5 are formed, but for longer chains, mesoporous materials are isolated.

A different self-assembly approach uses nonionic surfactants such as water soluble polymers—often polyethylene oxide (PEO)—as the hydrophilic group. The self-assembly takes place by hydrogen bonding or weak van der Waals forces to form mesoporous silicates. By carefully adjusting the temperature, pH, and reaction rates, mesoporous structures of other inorganic solids have also been prepared, such as metal phosphates, oxides, sulfides and selenides and even mesoporous metals such as Pt.

The preparation of **yttrium aluminium garnet ($Y_3Al_5O_{12}$; YAG)** illustrates a variation of the hydrothermal method used if each of the starting materials each has a solubility very different from one another. In this case, yttrium oxide (Y_2O_3) is placed in a cooler section of the autoclave and aluminium oxide as sapphire in a hotter section to increase its solubility (Figure 3.8). YAG forms where the two zones meet

$$3Y_2O_3 + 5Al_2O_3 = 2Y_3Al_5O_{12}.$$

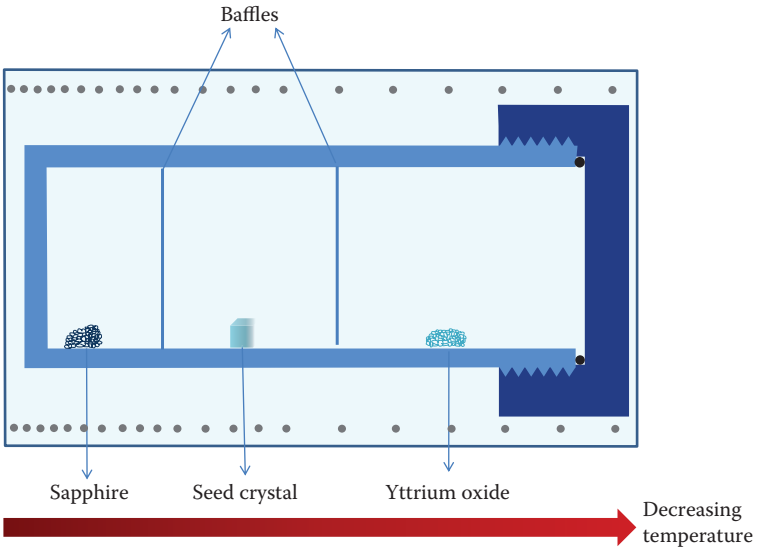


FIGURE 3.8 The hydrothermal arrangement for the synthesis of YAG starting from two reactants of different solubilities.

3.5.2 High-Pressure Gases

This method is usually used for preparing metal oxides and fluorides in less stable high-oxidation states. For instance, the perovskite $SrFeO_3$ containing Fe(IV) can be made from the reaction of $Sr_2Fe_2O_5$ and oxygen at 34 MPa. Fluorine at high pressures (>400 MPa) has also been used in the synthesis of ternary fluorides containing high-oxidation states, for example, Cs_2CuF_6 containing Cu(IV).

3.5.3 Hydrostatic Pressures

The application of high pressures and temperatures can induce reactions and phase changes that are not possible under ambient conditions. Applying very high pressures tends to decrease volume and thus improve the packing efficiency; as a consequence, coordination numbers tend to increase, so, for instance, Si can be transformed from the four-coordinate diamond structure to the six-coordinate white-tin structure, and NaCl (six-coordinate) can be changed to the CsCl (eight-coordinate) structure.

Various apparatus designs are used for achieving exceptionally high pressures. Originally, a **piston and cylinder** arrangement allowed synthesis at pressures up to 5 GPa (50,000 atm) and 1500°C, but the **belt apparatus** using two opposed tungsten carbide cylinders can reach 15 GPa and 2000°C and is used for making synthetic diamonds. Also used for manufacturing synthetic diamonds is the more efficient split-sphere or BARS apparatus, which consists of an inner capsule inside a cube. The capsule is pressed by both inner and outer anvils and heated by pressurised oil. Diamond anvils have been used to reach pressures of 20 GPa. Diamond anvils can be

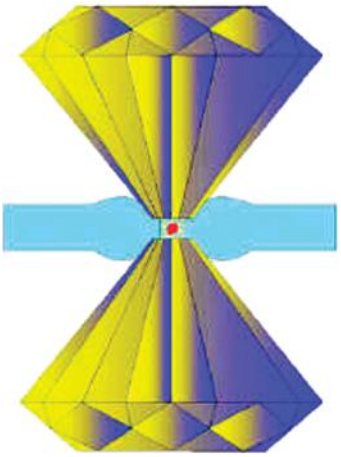


FIGURE 3.9 A high-pressure diamond anvil cell. (From Cornell High Energy Synchrotron Source. With permission.)

arranged as two directly opposed, four tetrahedrally opposed or six in an octahedral configuration (Figure 3.9); however, although they achieve greater pressures than other methods, they have the disadvantage that only milligram quantities of material can be processed, and so are more useful for investigating phase transitions than for synthesis.

The drive to produce **synthetic diamonds** arose during the Second World War; they were urgently needed for the diamond-tipped tools used for manufacturing military hardware, and it was feared that the South African supply of natural stones might dry up. GEC started a research programme to mimic the geothermal conditions that produce the natural stones. In 1955, they eventually succeeded in growing crystals up to 1 mm in length by dissolving graphite in a molten metal, such as nickel, cobalt or tantalum, in a pyrophyllite vessel and subjecting the graphite to pressures of 6 GPa and temperatures of ~2000°C until diamond crystals were formed. The molten metal acts as a catalyst, bringing down the working temperature and pressure. By refining the process, gemstone-quality stones were eventually made in 1970. Synthetic diamond can also be made by CVD methods (see Section 3.6), by detonation of carbon-containing explosives and by high-energy ultrasound (see Section 3.9).

Diamond has the highest thermal conductivity known, about five times that of copper, hence the stones feel cold to the touch and are nicknamed ‘ice’. Increasing the ratio of ^{12}C to ^{13}C in synthetic diamond has been found to improve the thermal conductivity by up to 50%. The isotopic enhancement takes place by producing a CVD diamond film (see Section 3.6) from ^{12}C-enriched methane. This is crushed and is used to make synthetic diamond in the usual way.

High-pressure synthesis has been used to make many high-temperature **cuprate superconductors**, as this facilitates the formation of the two-dimensional Cu–O sheets necessary for this type of superconductivity (see Chapter 9). For instance,

high pressure increases the density of $SrCuO_2$ by 8%, and the Cu–O coordination changes from chains to two-dimensional CuO_2 sheets.

3.5.4 Using Ultrasound

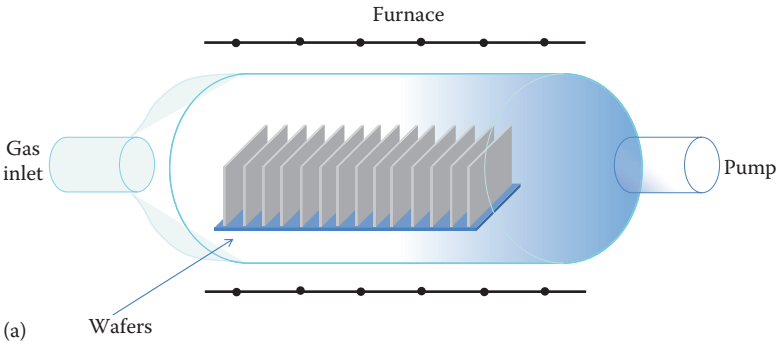
When very high intensity sound waves—ultrasound—travel through a liquid, a phenomenon known as **cavitation** occurs. The travelling sound wave causes high-pressure areas (compression) of the liquid, which is followed by low pressure and sudden expansion (rarefaction) and the formation of tiny bubbles, which expand to an unsustainable size and then collapse. The expansion and collapse of the bubbles create very localised hot spots, which reach instantaneous pressures of more than 100 MPa (1000 atm) and temperatures of up to 5000°C. The consequent collisions and fast cooling can cause chemical reactions and decomposition to occur. For instance, when $Fe(CO)_5$ dissolved in decane under argon is subjected to ultrasound, decomposition to amorphous iron occurs, but in pentane, $Fe_3(CO)_{12}$ is formed. When a suspension of graphite in an organic solvent (at room temperature and atmospheric pressure) is subjected to ultrasound, micron-sized (0.001 mm) diamonds result in about 10% yield.

3.5.5 Detonation

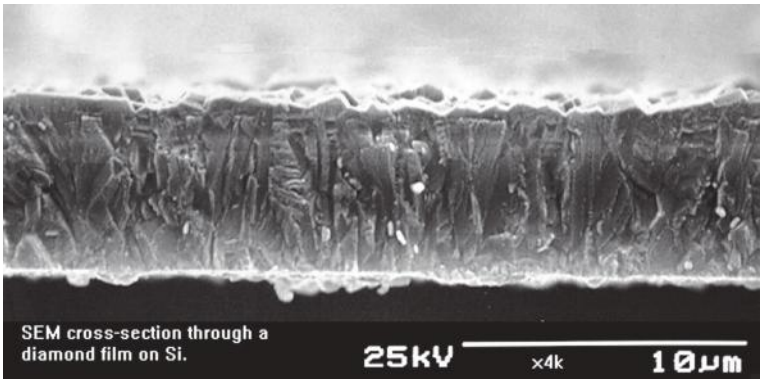
Detonation methods usually prepare nano-sized particles (see Section 3.9). In this method, a suitable starting material is prepared, mixed with oxidising agents and explosives such as TNT or RDX and detonated in a suitably strong containing vessel; the resulting materials have to be treated to remove impurities left from the explosives, fragments from the enclosing vessel walls and unwanted by-products. This method has been used for preparing nano-sized synthetic diamond, graphite, boron nitride, titania, and alumina, among others.

3.6 CHEMICAL VAPOUR DEPOSITION

So far, the preparative methods we have looked at have involved reactants that have mostly been in the solid state or in solution. In CVD, powders and microcrystalline compounds are prepared from reactants in the vapour phase and can, if required, be deposited on a substrate (such as a thin sheet of metal or ceramic, called a **wafer**) to form single-crystal films for devices (see Section 3.7.1). First, the volatile starting materials are heated to form vapours; these are then mixed at a suitable temperature and transported to the substrate for deposition using a carrier gas; finally, the solid product crystallises out. The whole vessel may be heated, which tends to deposit the product all over the walls of the vessel. More commonly, the energy to initiate the reaction is supplied to the substrate. Figure 3.10 shows a schematic setup for CVD. The conditions under which the reactions are carried out can be varied: pressures can be anything from atmospheric pressure to ultrahigh vacuum; ultrasound can be used to generate an aerosol to help transport; a plasma can be created to speed up reaction rates.



(a)



(b)

FIGURE 3.10 (a) A chemical vapour deposition reactor. (b) A cross section of a 100 µm–thick CVD diamond film grown by DC arc jet. The columnar nature of the growth is clearly evident, as is the increase in film quality and grain size with growth time. (From Paul Day and M. Ashfold, Bristol University. With permission.)

Typical starting materials are hydrides, halides and organometallic compounds, as these compounds tend to be volatile; if an organometallic precursor is used, the method is often referred to as **metal organic CVD (MOCVD)**.

The advantages of this method are that reaction temperatures are relatively low, the stoichiometry is easily controlled and dopants can also be incorporated.

3.6.1 Preparation of Semiconductors

The CVD method has been widely used for preparing the III–V semiconductors and silicon, for devices. Various reactions to prepare GaAs, involving chlorides and hydrides, have been tried; for example, the reaction of $AsCl_3$ with Ga in the presence of hydrogen:

$$AsCl_3 + Ga + \tfrac{3}{2}H_2 = GaAs + 3HCl\,.$$

Silicon for semiconductor devices can be prepared using this method from the decomposition of silane (SiH_4):

$$SiH_4 = Si + 2H_2.$$

The silane can be either pure, or mixed with nitrogen or hydrogen at low pressures of about 100 Pa, and temperatures of around 600°C are used.

3.6.2 Diamond Films

Diamond has many useful properties. As well as being the hardest substance known, it also has the best thermal conductivity (at room temperature), is an electrical insulator and has the highest transparency in the IR region of the spectrum of any known substance. Diamond films are made by a CVD process at lower pressures of ~27 kPa and temperatures of around 800°C. Carbon-containing gas mixtures, such as methane (or acetylene), and hydrogen, can be broken down in various ways using microwaves, laser or a hot filament; the carbon atoms and carbon-containing radical species are then deposited on a carefully cleaned and oriented substrate.

Early experiments managed to deposit monocrystalline layers onto a seed diamond and thus build up larger diamonds, layer by layer. Microcrystalline diamond layers have also been deposited onto silicon and metal surfaces. Such layers have been used for nonscratch optical coatings and for coating knives and scalpels so that they retain their sharpness. Additionally, such layers have the potential to be used as a hard, wear-resistant coating for moving parts or as an electrically insulating but heating-conducting layer.

3.6.3 Optical Fibres

Optical fibres now carry most of the world's communication traffic. In order to transmit light over distances of kilometres without loss of intensity, optical fibres have to be manufactured to very high specifications (see Chapter 7). They need to be free of impurities such as transition metal ions, and they also must have a varying refractive index across the fibre, with an outer sheath of lower refractive glass and an inner highly refractive core. This is so that light travelling down the centre cannot escape, but is totally internally reflected back into the fibre. Conventional methods of preparing silica glasses are inadequate and a modified form of CVD (MCVD) is usually employed.

The process uses volatile silicon tetrachloride which can be easily purified, for example, by fractional distillation. Additives such as germanium tetrachloride are incorporated into the centre of the fibre to give a higher refractive index. The process starts by making a 'preform'. A silica tube, which forms the outer low refractive layer, known as the cladding of the fibre, is heated to about 1600°C by a flame passing up and down the outside of the tube as it is rotated (Figure 3.11). $SiCl_4$ and O_2 pass down the tube and react together to form SiO_2, which deposits evenly along the

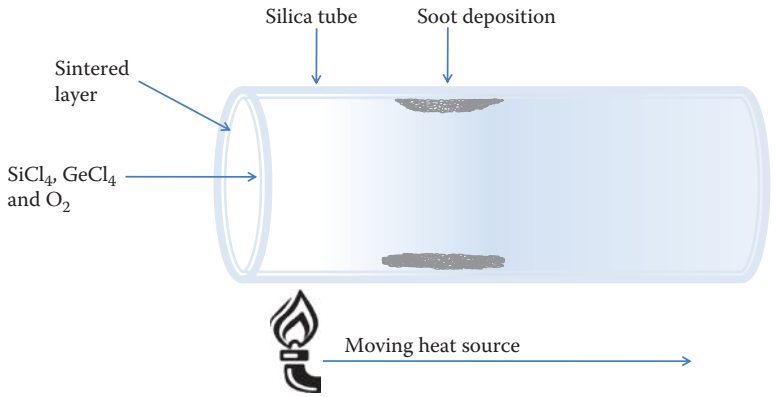


FIGURE 3.11 A diagram for the MCVD method of making optical fibres.

tube wall as soot. Small amounts of $GeCl_4$ and other chemicals are gradually added into the process to deposit as oxides on the wall in the inner layers to increase the refractive index. The flame burner is then used to melt the soot deposit onto the glass preform. After cooling, the cylindrical preform is heated in a drawing tower to about 1900°C until it melts sufficiently to fall under gravity into a fine thread, which is slowly pulled and wound onto a spool. One preform can produce 2 km of fibre.

3.6.4 Lithium Niobate

The sol–gel method for preparing lithium niobate used lithium and niobium alkoxides. Alkoxides are often used in CVD methods, but unfortunately for the preparation of lithium niobate, lithium alkoxides are much less volatile than niobium alkoxides and to get the two metals deposited together, it is better to use compounds of similar volatility. One way around this problem is to use a more volatile compound of lithium. One reported synthesis uses a β-diketonate of lithium in which lithium is coordinated to 2,2,6,6-tetramethylheptan-3,5-dione ($Me_3CCOCH_2COCMe_3$) (Figure 3.12).

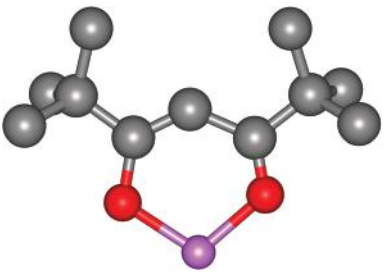


FIGURE 3.12 β-Diketonate of lithium from 2,2,6,6-tetramethylheptan-3,5-dione (lithium, purple; oxygen, red; carbon, grey).

The lithium compound was heated at about 250°C and niobium pentamethoxide at 200°C in a stream of argon containing oxygen. Lithium niobate was deposited on the reaction vessel, which was heated to 450°C.

In this example, the volatile precursor compounds were heated to obtain the product. Other energy sources are also used, notably electromagnetic radiation. An example of vapour-phase deposition involving photodecomposition is given in the next section on vapour-phase epitaxy.

3.7 PREPARING SINGLE CRYSTALS

Single crystals that are of high purity and are defect free may be needed for many applications, none more so than in the electronics and semiconductor industry. Various methods have been developed for preparing different crystalline forms, such as large crystals, films, etc. A few of the important methods are described below.

3.7.1 Epitaxy Methods

CVD methods are now used to make high-purity thin films in electronics where the deposited layers must have the correct crystallographic orientation. In epitaxial growth, a precursor is decomposed in the gas phase and a single crystal is built up layer by layer on a substrate, adopting the same crystal structure as the substrate. This is an important method for semiconductor applications where single crystals of a controlled composition are needed; semiconductors such as gallium nitride, gallium arsenide and indium phosphide are made this way. Gallium arsenide (GaAs) has been prepared by several **vapour-phase epitaxial (VPE) methods**.

In one method for the preparation of *gallium arsenide*, arsenic(III) chloride ($AsCl_3$) (boiling temperature 103°C) is used to transport gallium vapour to the reaction site where gallium arsenide is deposited in layers. The reaction involved is

$$2Ga(g) + 2AsCl_3(g) = 2GaAs(s) + 3Cl_2(g)$$

An alternative to gallium vapour is an organometallic compound of gallium. One preparation reacts trimethyl gallium ($Ga(CH_3)_3$) with the highly volatile and toxic arsine (AsH_3)

$$Ga(CH_3)_3(g) + AsH_3(g) = GaAs(s) + 3CH_4(g).$$

Mercury telluride (HgTe) was first made by vapour-phase epitaxy in 1984. This preparation illustrates the use of ultraviolet radiation as the energy source for decomposition. Diethyltellurium ($(C_2H_5)_2Te$) vapour in a stream of hydrogen carrier gas is passed over heated mercury where it picks up the mercury vapour. As the gas passes over the substrate, it is subjected to ultraviolet illumination and mercury telluride deposits. The substrate is heated to 200°C, which is about 200°C lower than the temperature needed for thermal decomposition. The use of the lower temperature

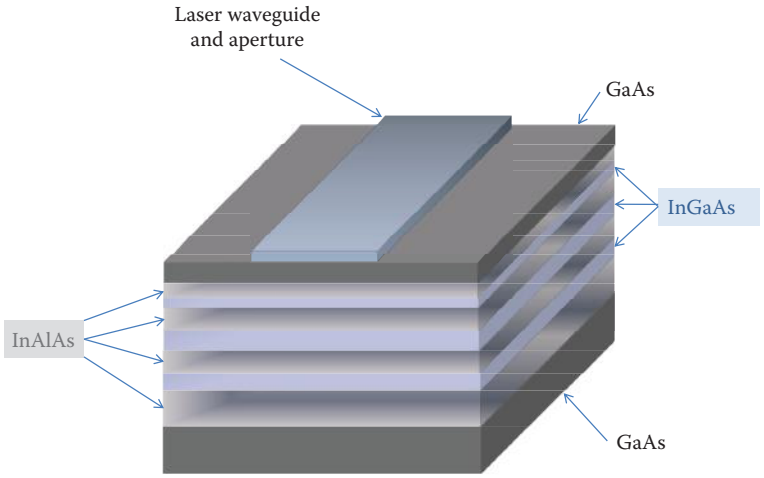


FIGURE 3.13 The layers of an active region of a quantum cascade laser.

has enabled a crystal of alternating mercury telluride and cadmium telluride layers to be built, because diffusion of one layer into the other is slowed by the lower temperature.

The production of single crystals with a carefully controlled varying composition is vital to making semiconductor devices. The production of crystals for quantum-well lasers illustrates how carefully such syntheses can be controlled.

A *molecular beam* is a narrow stream of molecules formed by heating a compound in an oven with a hole, which is small compared to the mean free path of the gaseous molecules produced. Very thin layers can be built up by directing a beam of precursor molecules onto the substrate. The system is kept under ultra-high vacuum. Because of the very low pressure, the reactants need not be as volatile as in other vapour deposition methods. An application of this method is the growth of single crystals for **quantum cascade lasers**, semiconductor lasers that emit in the far infrared. The first quantum cascade laser crystal contained nanometre thickness layers of $In_{0.52}Al_{0.48}As$ alternating with $In_{0.53}Ga_{0.47}As$, as shown in Figure 3.13. To make this crystal, beams of aluminium, gallium, arsenic and indium are directed onto a substrate InP crystal. The substrate needs to be heated to allow the atoms deposited from the beams to migrate to their correct lattice position. The relative pressures of the component beams are adjusted for each layer to give the desired compositions.

3.7.2 Chemical Vapour Transport

In CVD, solids are formed from gaseous compounds. In **chemical vapour transport**, a solid or solids interact with a volatile compound and a solid product is deposited in a different part of the apparatus. It is used both for preparing compounds and for growing crystals from powders and for the purification of less pure crystalline

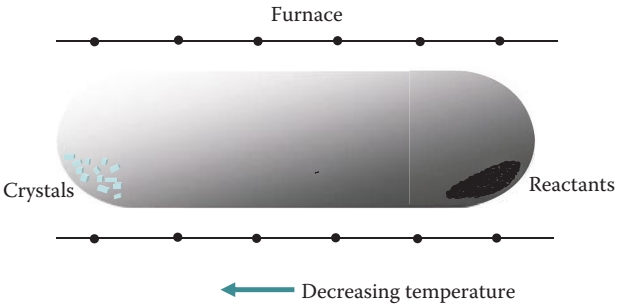


FIGURE 3.14 The growth of magnetite crystals using chemical vapour transport.

material. The reactant solids are placed in a sealed quartz tube together with a transport/reactant gas, which is heated by a tube furnace; the tube is kept hotter at one end than the other (Figure 3.14); the pure crystals deposit at the cooler end.

Magnetite crystals can be grown using the reaction of magnetite with hydrogen chloride gas:

$$Fe_3O_4(s) + 8HCl(g) = FeCl_2(g) + 2FeCl_3(g) + 4H_2O(g).$$

Powdered magnetite is placed at one end of the reaction vessel and the tube is evacuated and filled with HCl gas. The tube is sealed and then heated. The reaction is endothermic ($\Delta H_m^{\ominus}$ positive), so the equilibrium moves to the right as the temperature is raised. Thus, at the hotter end of the tube, magnetite reacts with the hydrogen chloride gas and is transported down the tube as gaseous iron chlorides. At the cooler end of the tube, the equilibrium shifts to the left and magnetite is deposited.

Crystals of *transition metal sulfides and selenides* can be prepared from the pure metal and chalcogenide and a transport agent such as iodine. The initial reaction at the hotter end produces the metal iodide, which subsequently reacts with the sulfur or selenium at the cooler end to give, for instance, TiS_2.

Bismuth chalcogenides have thermoelectric properties, that is, they can turn heat into electricity and vice versa. Two-dimensional films have a possible application in microheating and cooling, and it has proved possible to make the selenides and tellurides, Bi_2Se_3, Bi_2Te_3 and $Bi_2Te_{2.7}Se_{0.3}$, by a chemical vapour transport method. A film of nanocrystalline bismuth is placed at the centre of the reaction vessel and heated to about 320°C. The pure chalcogenide, placed a few centimetres away, is only heated to 250°C and a stream of argon and hydrogen carries its vapour over the heated metal, forming the crystalline chalcogenide film. Using this method, the films are found to preferentially develop the *001* faces.

3.7.3 Melt Methods

Melt methods depend on both the compound being stable in the liquid phase and the availability of a high-temperature, inert, containing vessel.

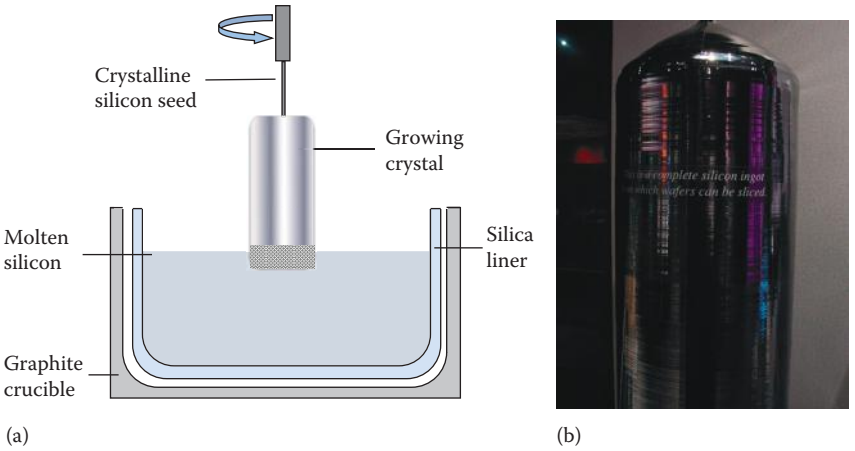


FIGURE 3.15 (a) The Czochralski process for producing a very pure single crystal of silicon. (b) Silicon boule. (Part (b) from Tonie van Ringelestijn.)

The silicon required by the electronics industry for semiconductor devices has to have levels of key impurities, such as phosphorus and boron, of less than 1 atom in 10^{10} Si. Silicon is first converted to the highly volatile trichlorosilane ($SiHCl_3$), which is then distilled and decomposed on rods of high-purity silicon at 1000°C to give high-purity polycrystalline silicon. This is made into large single crystals by the **Czochralski process**. The silicon is melted in an atmosphere of argon, and then a single-crystal rod is used as a seed that is dipped into the melt and then slowly withdrawn, pulling with it an ever-lengthening single crystal in the same orientation as the original seed (Figure 3.15). In addition to silicon, this method is also used for preparing other semiconductor materials such as Ge and GaAs and for ceramics such as perovskites and garnets. Dopant impurity atoms can be added to the molten silicon to make *n*-type or *p*-type semiconductors. The ingot is sliced into thin wafers for electronics and photovoltaics.

The **float-zone process** is also used for producing large single crystals of silicon, which can then be sliced. A polycrystalline rod of silicon is held vertically and a single-crystal seed crystal is attached at one end. A moving heater then heats the area around the seed crystal, and a molten zone is caused to traverse the length of the rod, producing a single crystal as it moves (Figure 3.16). This process also has the advantage of refining the product—the so-called **zone refining**—as impurities tend to stay dissolved in the melt and are thus swept in front of the forming crystal to the end, where they can be cut off and rejected.

Related methods include the **Bridgman** and **Stockbarger methods**, where a temperature gradient is maintained across a melt so that crystallisation starts at the cooler end on a seed crystal; this can be achieved either by using a furnace with a temperature gradient or by pulling the sample through a furnace. The crystallisation is in the same orientation as the seed. These methods are used when the Czochralski process proves difficult, for instance, for GaAs.

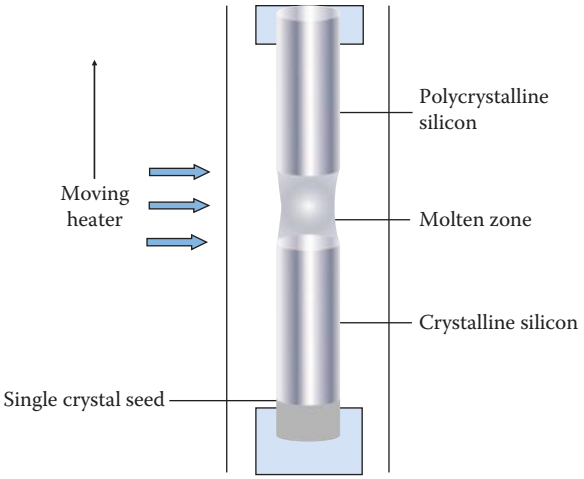


FIGURE 3.16 The float-zone method for producing very pure crystals of silicon.

The old method used originally for producing artificial ruby and sapphire, the **Verneuil method**, is still widely used for making artificial gemstones. The powdered sample is melted in a high-temperature oxyhydrogen flame and droplets are allowed to fall onto a seed crystal where they crystallise. Using the same method but using a plasma torch to melt the powder can achieve even higher temperatures.

Skull melting is a method used for growing large oxide crystals, which melt at such high temperatures that normal containment vessels are not viable. The crucible is made of copper water-cooled fingers and base, which makes an outer skin on the molten material, sufficiently solid to hold it in place. The container can be evacuated and filled with an appropriate atmosphere. A power supply of up to 100 kW, 4 MHz produces a radio frequency that melts the sample. Temperatures of up to 3300°C can be reached, allowing the growth of large crystals (several centimetres) of refractory oxides, such as ZrO_2, ThO_2, CoO and Fe_3O_4.

3.7.4 Solution Methods

Crystallisation from solution takes place either by cooling a saturated hot solution or by allowing evaporation to take place. Many crystalline substances have traditionally been grown from saturated solutions in a solvent; hot solutions are prepared and then cooled and crystals precipitate from the supersaturated solution.

Various other techniques can be used to induce crystallisation and these include: evaporating the solvent by heating; leaving in the atmosphere; placing in a desiccator or using reduced pressure; freezing the solvent; and addition of other components (solvent or salt) to reduce solubility.

For solids such as oxides that are very insoluble in water, it may be possible to dissolve them in melts of borates, fluorides or even metals, in which case the solvent

is generally known as a **flux** because it brings down the melting temperature of the solute. The melt is cooled slowly until the crystals form, and then the flux is poured off or dissolved away. This method has been successfully used for preparing crystalline silicates, quartz, and alumina, among many others.

3.8 INTERCALATION

Many layered solids form intercalation compounds, where a neutral molecule is inserted between weakly bonded layers. These solids, produced by the reversible insertion of guest molecules into lattices, are known as **intercalation compounds**, and although originally applied to layered solids, it is now taken to include other solids with similar host–guest interactions. Intercalation compounds have importance as catalysts, as conducting solids and therefore electrode materials, as a means of encapsulating molecules potentially for drug delivery systems, as a method of synthesising composite solids, and as reducing agents.

3.8.1 Graphite Intercalation Compounds

When every layer in graphite is intercalated, the larger metals (K, Rb, Cs) form MC_8 compounds and the smaller metals (Li, Sr, Ba, Ca) form MC_6. For example, when potassium vapour reacts with graphite under argon at 150°C–200°C, it forms a bronze-coloured compound (KC_8) in which the potassium atoms sit between the graphite layers, and the interlayer spacing is increased by 200 pm (Figure 3.17). The potassium donates an electron to the graphite (forming K^+) and the in-layer conductivity of the graphite increases; it is also a superconductor. KC_8 is a powerful reducing agent and is increasingly employed in organic syntheses. CaC_6 is made by immersing graphite in a liquid Li–Ca alloy for several days; it is also a superconductor. Graphite electron-acceptor intercalation compounds have also been made with NO_3^-, CrO_3, Br_2, $FeCl_3$ and AsF_5, and some of these compounds have electrical conductivity approaching that of aluminium.

Interest in the lithium intercalation of graphite and of oxides has increased in recent years due to the use of the lightweight, rechargeable, lithium-ion batteries in laptops and mobile devices (see Chapter 5): The driving reaction of these cells is that of Li with a transition metal oxide such as CoO_2 to form an intercalation compound, Li_xCoO_2. The anode is usually made of LiC_6 graphite intercalated with lithium that easily releases Li^+; this travels through the electrolyte to the cathode synthesising Li_xCoO_2. The Li^+ ions thus move between the two intercalation compounds. Various cathode materials, such as spinels, CoO_2, NiO_2, TiS_2 and MnO_2, can be used, but the hunt is always on for better materials with higher conductivity, intercalation, reversibility and kinetics. Recently, TiO_2 nanowires have been made, which are found to intercalate high proportions of Li with potential for new anode materials. The method is to produce the TiO_2-B polymorph of titania, which is less dense than anatase or rutile, having a more open structure which is conducive to intercalation. A suspension of anatase in conc.NaOH is heated in an autoclave for 3 days, washed with acid, dried and the resulting titanate is calcined in air at 400°C to give the

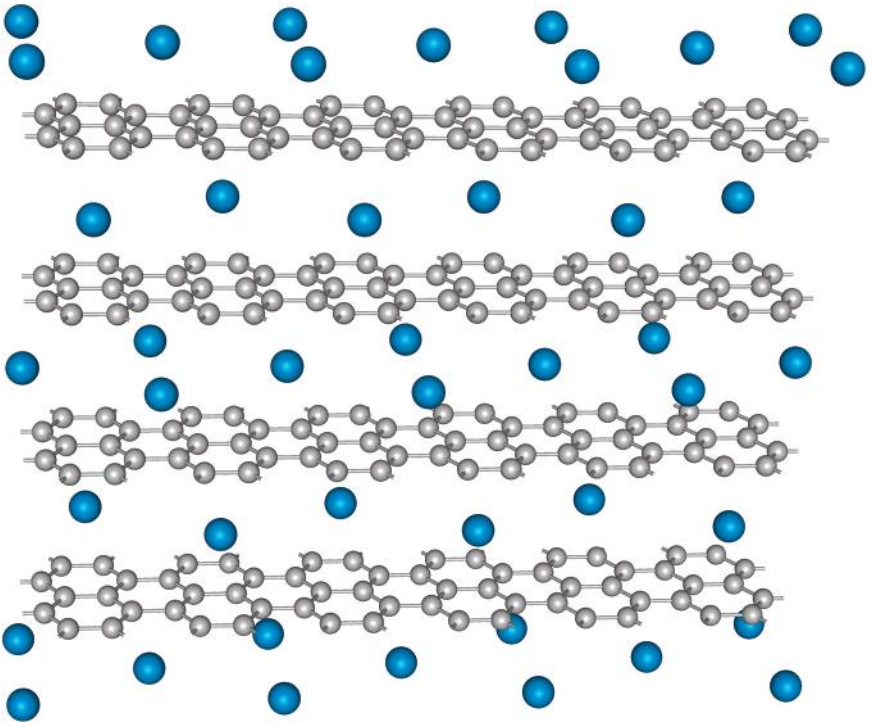


FIGURE 3.17 The structure of KC_8 (K, blue; C, grey).

nanowires. The quantity of lithium intercalation has been found to be very high and extremely reversible.

Layered structures are also found for many oxides and sulfides of transition metals. They can be intercalated with alkali metals (Li, Na, K) to give superconducting solids and conducting solids that are useful for solid-state battery materials. **Titanium disulfide** has a CdI_2 structure (Chapter 1). The solid is golden-yellow and has a high electrical conductivity along the titanium layers. The conductivity of titanium disulfide can be increased by forming intercalation compounds with electron donors, the best-known example being with lithium, Li_xTiS_2. This compound is synthesised in the cathode material of the rechargeable battery described in Chapter 5 and can also be synthesised directly by the lithiation of TiS_2 with a solution of butyl lithium:

$$x C_4H_9Li + TiS_2 = Li_xTiS_2 + \tfrac{x}{2} C_8H_{18} .$$

Pillaring is a technique for building strong molecular pillars that prop the layers of a clay sufficiently far apart to create a two-dimensional space between the layers where molecules can interact. In a typical reaction, a clay such as montmorillonite

undergoes ion exchange with the large inorganic cation $[Al_{13}O_4(OH)_{24}(H_2O)_{12}]^{7+}$ by slurrying with an alkaline solution of aluminium ions. After washing with water, the material is calcined, the hydroxyls and water are lost and alumina pillars are created between the silicate layers of the clay, creating a permanently microporous substance.

Many other inorganic pillars have been incorporated into clays, for instance, zirconia pillars can be created by calcining after ion exchange with zirconium oxychloride.

3.9 SYNTHESIS OF NANOMATERIALS

The methods of making nanostructures is becoming increasingly ingenious and it is impossible to do more here than give a flavour of the many methods that are being used. The methods of synthesis of nanoparticles are often divided into **top-down** and **bottom-up syntheses**. The meaning of these terms is becoming increasingly blurred, but the original meaning of 'top-down' referred to making products in bulk and then manipulating them into the form required, for example, casting, recrystallising and etching. 'Bottom-up' was coined to refer to the building up of a structure, atom by atom, in a controlled fashion, so that each atom is placed in its specific place and the structure is built with no defects.

3.9.1 Top-Down Methods

Top-down methods for nanoparticles apply to taking the already synthesised bulk structure and rendering it into nano-sized particles; for instance, the electronics industry routinely uses lithography and etching methods to form nanoscale patterns on circuits. A few other examples are considered in the following sections.

Data storage using atomic force microscope lithography involves using a large array of atomic force microscope (AFM) tips (64×64) and a polymer-coated storage disk. To record data, a current is passed through the tip, raising its temperature to 400°C. The polymer in the immediate vicinity of the tip is softened, producing a small indentation. To read the data back, the tip is heated to 300°C, too low to cause softening during the brief time that the tip is in contact, but enough to cause a change in resistance when the tip drops into an indentation.

The original technique has now been extended to a room temperature process where a special copolymer coating is used that needs lower pressure and temperature to form the indentations. A high density of data storage at >1 terabytes (TB) per square inch is achievable.

Mechanical grinding, such as **high-energy milling**, can be used to form nanoparticles from bulk solids. There are various kinds of mills in use, but the principle behind the technique is the same, in that mechanical attrition is used to produce uniformly sized particles. In ball milling, the solid is rotated in a cylindrical device with a grinding medium such as ceramic or metal balls. Care must be taken to avoid contamination of the product from the grinding medium.

The top-down method **vaporisation** is to take the bulk solid, heat to vaporise it and condense the ultrafine particles on a surface for collection; often an inert gas is also added. Alternatively, evaporation under reduced pressure and condensation can be used.

Mechanochemical synthesis of nanoparticles uses a milling process to induce a chemical reaction between the components, such that the reaction produces nanoparticles embedded in a matrix that can be subsequently washed away. In this way, the particles are kept well apart from one another, so one of the big problems of synthesising nanoparticles is avoided—that of agglomeration. The reactions occur at the interfaces of the particles generated in the milling and are constantly renewed. Reactions take place at much lower temperatures than would otherwise be required, but must not be allowed to go too fast or to combust. With highly exothermic reactions, diluents such as NaCl are added to reduce the contact between the components and slow the reactions down.

This technique has been successfully used to produce:

- Fine metal powders, such as Fe, Co, Ni, Cu and Ag, for example, $NiCl_2 + 2Na = Ni + 2NaCl$, produces 5 nm Ni particles embedded in NaCl. A top-down milling approach gives particles that are several thousand times larger.
- Oxides, such as Al_2O_3, ZrO_2, Cr_2O_3, SnO_2 and ZnO, for example, $ZrCl_4 + 2CaO = ZrO_2 + 2CaCl_2$.
- Sulfides, such as ZnS and CdS, for example, $CdCl_2 + Na_2S = CdS + NaCl$, produces 4 nm particles of the CdS II–VI semiconductor.

Microemulsions can be used to synthesise nanoparticles. When preparing nanoparticles, the simultaneous control of several factors is important: particles need to be both in the nanometre range and homogeneous, and they have to be prevented from growing larger and from agglomeration. Reactions in microemulsions have proved to be very efficient for controlling these conditions in nanoparticle preparation. A microemulsion is defined as a dispersion of droplets of either oil in water or water in oil, of a diameter <100 nm; the droplets are stabilised by an interfacial film of surfactant molecules. A cosurfactant such as an alcohol or amine may also be used. When the droplets consist of a dispersion of aqueous droplets in an organic solvent, inverted micelles (also known as inverse and reverse micelles) are formed so that the hydrophilic heads of the surfactant lie in the aqueous core of the droplets and the hydrophobic tails are in the organic solvent (Figure 3.18).

In a typical preparation, two separate microemulsions of aqueous solutions containing the reactant salts are prepared in oil (A and B in Figure 3.18) and these are then mixed together. As the droplets collide, they fuse to make a larger micelle with which the salts interact. This larger micelle is less stable and breaks back down into two droplets containing the product that has precipitated out.

This micelle method has been used to prepare a range of useful nanoparticles including: metals; oxides such as CeO_2 used in fuel cells and gas sensors (Chapter 5); titanates and zirconates, such as $SrTiO_3$ and $SrZrO_3$, which are important dielectric oxides used in electronic applications (Chapter 8); SnO_2, an important wideband gap

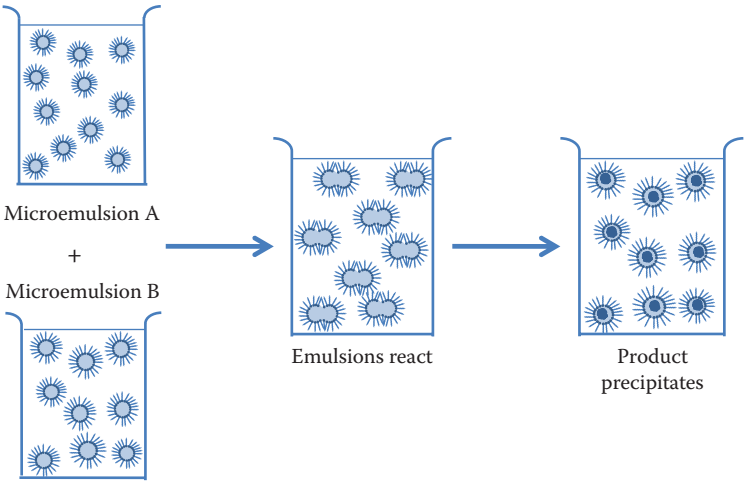


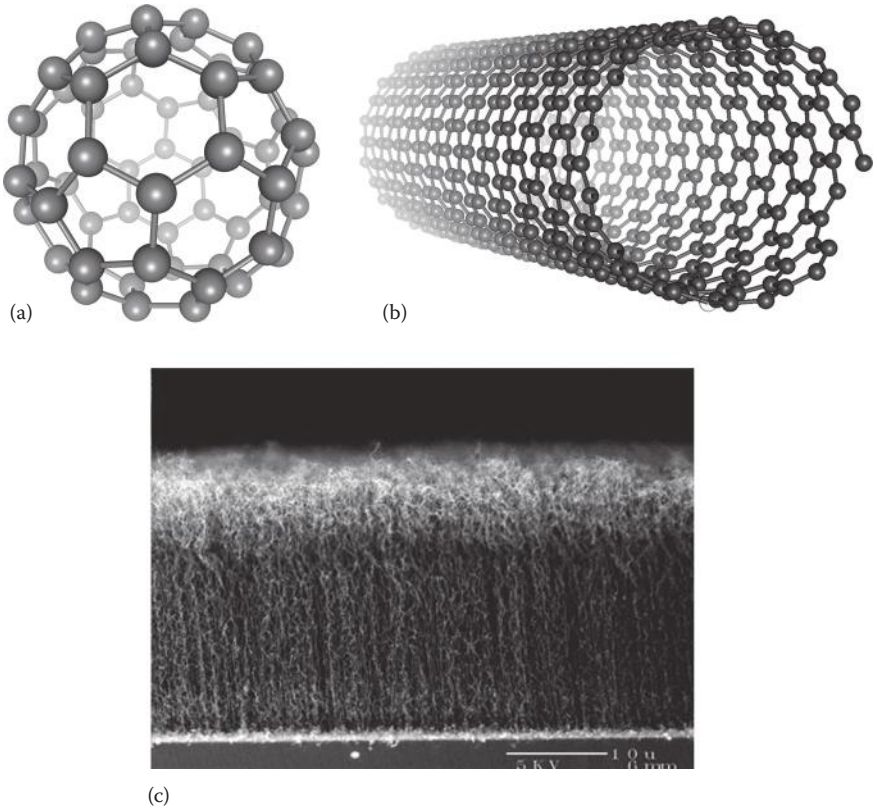
FIGURE 3.18 Microemulsion synthesis, showing the mixing and interaction of two microemulsions, A and B, to give a solid precipitate.

(3.6 eV) *n*-type semiconductor used in gas sensors (Chapter 4); magnetic particles; and superconductors such as $YBa_2Cu_3O_{7-x}$ (Chapter 8).

Fullerenes are allotropes of carbon, which are graphene sheets rolled into spheres and ellipsoids, often known as **buckyballs**, and tubes known as buckytubes or **carbon nanotubes (CNT)** (Figure 3.19a). The fullerenes will be discussed fully in Chapter 10. They form a class of nanoparticles of their own, and a huge amount of research is devoted to their manufacture due to their properties, which range from strength, conductivity, superconductivity and heat resistance, to hydrogen storage.

Since their original discovery in 1985, they have been found to form in the soot of candle flames, in lightning strikes and most recently (2010), they have been identified in a young planetary nebula. The original method of making them, which is still used, is to pass a large electrical current between two carbon electrodes in an inert atmosphere, forming an arc discharge and thus heating the electrodes to >1700°C; many different fullerenes can be extracted from the resulting soot, C_{60} and C_{70} being the major products. This method also produces both single-walled and multi-walled CNTs. The combustion of hydrocarbons in fuel-rich flames is also used industrially to produce fullerenes.

There are several common methods for synthesising CNTs, which are nowadays manufactured in very large quantities (hundreds of tonnes) and are used to strengthen polymers and carbon fibres. They can either be grown in situ or synthesised separately and then incorporated into the matrix. When the arc discharge method is used for producing CNTs, a metal catalyst, such as Fe, Ni or Co, is incorporated into one of the electrodes; the addition of diamond powder to the catalyst increases the yield. The laser ablation method is very similar to the arc discharge, only now a high-power laser is used to vaporise a graphite pellet containing the metal catalyst.



(a) (b) (c)

FIGURE 3.19 (a) The structure of the C_{60} buckyball, (b) the structure of a single-walled carbon nanotube and (c) a carbon nanotube forest. (Part (c) from Professor Ray Baughman, University of Texas at Dallas. With permission.)

CVD techniques are usually used for producing nanotubes when growing them in position. In this method, carbon-containing gases are deposited onto a substrate that contains a metal catalyst. If the catalyst has been deposited in nanoparticle arrays, then forests of vertically aligned nanotubes can be formed (Figure 3.19c). This method has the advantage of using much lower temperatures than the laser and arc discharge techniques, usually between 700°C and 900°C. Templating methods are also used to make nanotubes and nanowires (see Section 3.9.3).

The search for a substance harder than diamond has resulted in the finding of **nanodiamond**, which is harder and less compressible. Nanostructured diamond materials have been prepared in several ways: Nanodiamond with a nanorod structure has been made from both graphite and fullerenes, either by using extreme pressure conditions (up to 40 GPa) or at lower pressures combined with heating; nanodiamond with a grain size of about 5 nm has also been made by a detonation method when two carbon-containing explosive materials, TNT (2,4,6-trinitrotoluene) and RDX (cyclotrimethylenetrinitramine), are set off together. The diamond

material is extracted from the debris of the explosion and unwanted by-products, by prolonged boiling in acid. High-energy lasers and ultrasound methods have also been used to make nanodiamond. Nanodiamonds are finding use in polishing, fillers, additives, and drug delivery.

3.9.2 Bottom-Up Methods: Manipulating Atoms and Molecules

The term bottom-up is now often used to mean structures that are built from constituent atoms, ions or molecules or that self-assemble. Many of the techniques covered earlier in this chapter have been used to prepare nano-sized particles in this way, cf. sol–gel, hydrothermal, combustion, detonation, CVD, ultrasound cavitation and MBE methods and will not be revisited here. Another example of bottom-up synthesis is the use of scanning probe tips to move atoms into specific sites. One of the drivers for this research has been the production of ever yet smaller electronic circuits. Traditionally, integrated circuits have been assembled using techniques in which a pattern is laid down on the surface of a crystal and the uncovered parts are treated to form a differently-doped semiconductor, an insulating oxide or a metallic connector. The patterns are produced by photolithography, that is, shining light through a mask onto a light-sensitive chemical on a surface. The spacing of the components using this technique is limited by diffraction effects to about 100 nm. To produce integrated circuits in which the components are nanometre scale requires new techniques. It is possible to use variations on the existing approach by, for example, using UV radiation, X-rays or electron beams in place of visible light. Another approach is to use scanning probe microscopy (SPM; see below) to print nanoscale patterns onto a substrate. For example, if a current is passed through the probe tip to raise its temperature, it can soften a thermoplastic polymer in the immediate vicinity of the tip, producing a small indentation. A completely different approach is to build up the required nanoscale structures, atom by atom, or molecule by molecule.

The operation of the **scanning tunnelling microscope** was explained in Chapter 2. In addition to giving us a picture of the surface and any molecules on it, we noted then, that STM can also be used to move atoms and molecules across a surface and to make molecules react. If the STM tip is brought closer to the surface than usual, the force between it and the atoms/molecules on the surface increases. Repulsive forces between the tip and the surface (produced by reversing the sign of the voltage) can push molecules away from the tip across the surface. If this force is attractive, moving the tip across the surface pulls the molecule or atom along after it. The famous picture of xenon atoms forming the letters IBM on a nickel surface was produced by increasing the voltage so that the Xe became attached to the tip, sliding the atom across the surface and then decreasing the voltage again at the position required. Figure 3.20 shows another image created at IBM by moving CO molecules around on a Pt surface.

Molecules can also be changed by the STM in single-molecule chemical reactions when the STM tip is used to channel electrons into a molecule. The energy can both dissociate bonds and provide the activation energy for bond formation.

Atomic force microscopy (AFM) (see Chapter 2), like STM, relies on a very sharp tip, but in this case the tip is brought close enough to the surface such

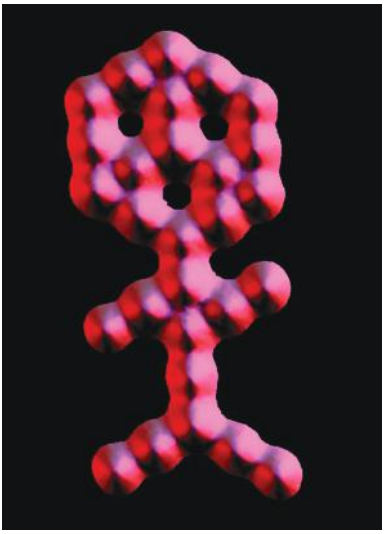


FIGURE 3.20 'Carbon dioxide man'; an STM manipulated image of carbon monoxide molecules on a Pt *111* surface. (Originally created by IBM Corporation.)

that the intermolecular forces between the tip and the surface can be measured. Nonconducting and semiconducting solids form essential components in circuits and in order to make nanocircuits on an atomic scale, the ability to manoeuvre single atoms into place is an essential tool. An AFM tip has been used to move a single atom on a silicon surface simply by pushing on it (Figure 3.21). The freed atom attached itself to the tip and could be removed from the surface. When the tip was lowered, the atom slipped back into place.

By attaching a particular molecule to the tip, the AFM can be made responsive to certain molecules or groups of molecules and not others. AFM is mostly used for

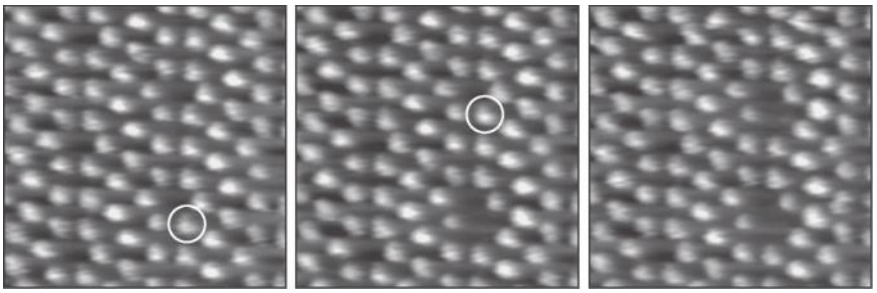


FIGURE 3.21 Two mechanical AFM single-atom vertical manipulation processes performed successively over the selected atomic positions marked with circles in (a) and (b) on an Si *111* surface, creating the vacancies in (c) (image size 66×66 Å^2). (Reprinted with permission from Custance, O., et al., *Physical Review Letters*, **90**, 176102, 2003. Copyright 2003 by the American Physical Society.)

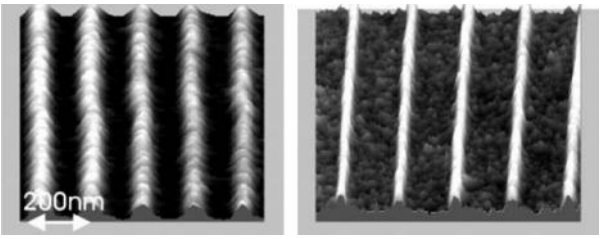


FIGURE 3.22 Two grids of oxide lines written using a tip bias of –10 V and an ambient humidity of 61% (*left*) and 14% (*right*). (With kind permission from Springer Science+Business Media: *Applied Physics A*, **66**, 1998, S659, Avouris, P.)

the study of biological molecules, and modified tips have been used for extracting particular molecules, for example, proteins from cell membranes.

By scanning AFM tips along a surface or tapping them, it is possible to *induce* chemical reactions along a line or to form a pattern of dots. One example is the application of a voltage to an AFM tip to oxidise a silicon semiconductor with lines of oxide of nanometre dimensions. The silicon surface is cleaned of any oxide on the surface with dilute hydrofluoric acid (HF) and then a negative voltage is applied to the AFM tip. The width of the SiO_2 structures produced is of the order of tens of nanometres (Figure 3.22). Some water vapour must be present in the surrounding atmosphere.

In **dip-pen nanolithography** using a molecular 'ink', the molecules are placed on an AFM tip and are delivered to a substrate surface via a water meniscus; examples include nanoscale patterns of thiols on gold and of silanes on silicon. The tip is loaded either by dipping in a solution and allowing to dry, or by vapour deposition. Under ambient conditions, a water meniscus naturally forms between the tip and the sample, due to the humidity of the air, so that when the tip is close to the substrate, the molecules are delivered to the surface via the meniscus and form a self-assembled monolayer. Surface tension holds the tip at a fixed distance from the surface, as it moves across it. The size of the deposit is dependent on the tip radius of curvature, the relative humidity, which controls the size of the water droplet, and any diffusion of the molecules across the surface. Nanoscale structures of 50 nm can be generated. The technique has the advantage that the same instrument can be used to both build the nanostructure and image it. Huge arrays of AFM tips linked together can be used (more than 200 × 200); one tip is positioned slightly further (0.4 nm) above the surface than the others and is used for imaging and hence guidance on moving the array across the surface. The other tips all lay down identical structures.

Liquid 'inks' that interact directly with the surface can also be used. The scale of the features created in this way is much larger—of the order of 1–10 μm. This technique allows patterning with biological molecules, such as proteins, peptides and DNA—useful in biological research investigating the interaction of particular molecules with cells.

'Optical tweezers' can be formed from the interaction between the electric dipole of light and the polarisability of molecules or dielectric particles producing a force of the order of piconewtons. By using one or more tightly focussed laser beams, this

tiny force can be exploited to move molecules or nanoparticles and arrange them into particular structures. It is commonly used to study the interaction of enzymes and proteins with DNA.

3.9.3 Synthesis Using Templates

Nanoparticles and nanowires can be made using template methods. There are broadly three main ways of doing this and these are summarised in a scheme for the formation of one-dimensional structures in Figure 3.23:

- Porous solids and membranes
- Organic and biological templates
- Using a preexisting nanostructure

In the first method, porous solids that possess nano-sized cavities or channels are impregnated with a solution of the reactants so that the cavities are filled. Sufficient

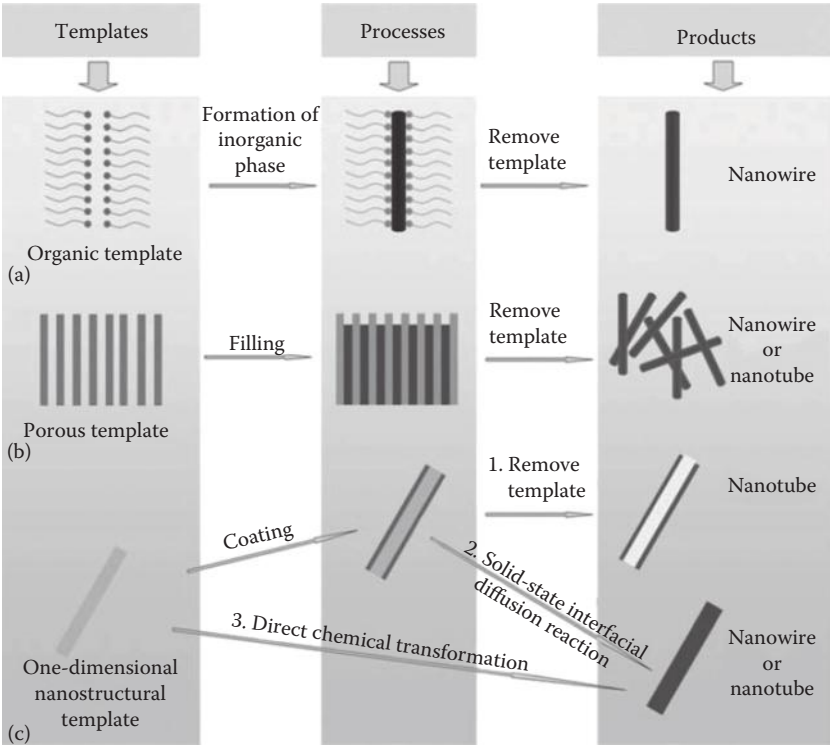


FIGURE 3.23 Schematic illustrations showing three different template processes for the synthesis of one-dimensional inorganic nanostructures: (a) cylindrical micelle, (b) porous membrane and (c) preexisting nanostructure templates. (Liang, H-W., et al.: *Advanced Materials*. 2010, **22**, 3925. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reprinted with permission.)

heat is then applied for the reaction to take place within the cavities, thus restricting the size of the product. The final step removes the original template scaffold by dissolving (silica-based scaffolds) or reaction (oxidation of carbon-based scaffolds). If the pores of the template scaffold interconnect, then an inverse replica structure of the template will be formed. If the cavities are discrete, as in some zeolites, for instance, then nanoparticles result (Figure 3.24).

Porous solids such as silica gels, mesoporous silica and zeolites (see Chapter 6) can be used for the templates. Mesoporous silica has a long-range order with specific pore sizes and geometry and has been used to form nanostructures or nanoparticles of such diverse solids as mesoporous carbon, metal oxides, sulfides, nitrides, spinels and phosphates. Mesoporous carbon, made by this method, can then itself be used as a template to form metal oxides.

A nanoporous aluminium oxide membrane can be formed on a substrate by the anodisation of a film of aluminium (when aluminium is used as the anode in an electrochemical cell). Such membranes can be created with highly organised close-packed arrays of straight pores arranged vertically to the substrate, with controlled density and length and pore sizes of about 30–200 nm diameter. These anodic aluminium oxide (AAO) membranes can be used as templates to form *ordered arrays* of nanotubes and nanowires—the negative of the original membrane. For instance, arrays of CNTs can be formed in this way using a predeposited metal catalyst on

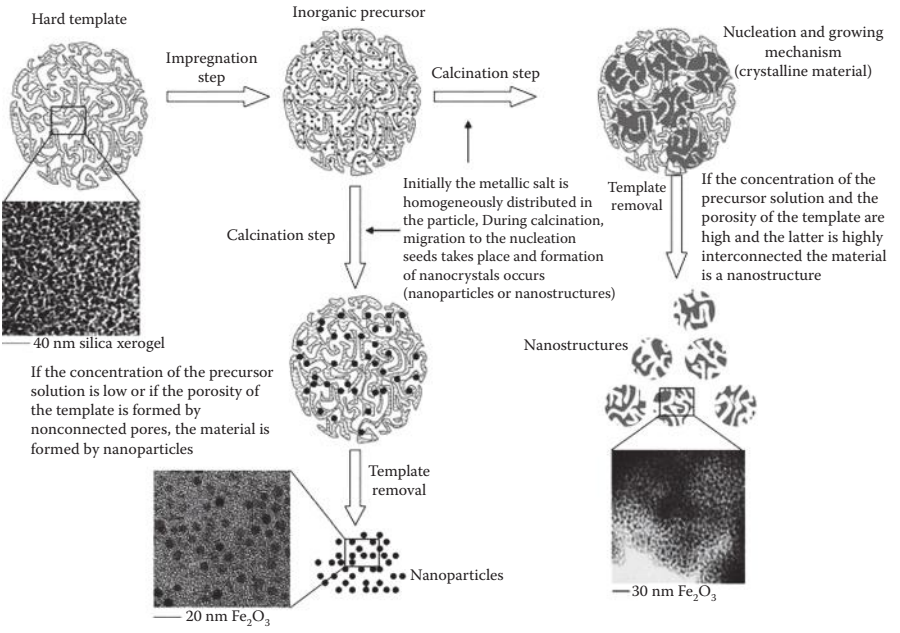


FIGURE 3.24 A schematic showing the use of a silica xerogel to form either nanoparticles or a nanostructure of Fe_2O_3. (Reprinted from *Materials Research Bulletin*, **41**, Valdés-Solís, T., and Fuertes, A.B., High-surface area inorganic compounds prepared by nanocasting techniques, 2187–2197, Copyright 2006, with permission from Elsevier.)

the AAO, followed by the thermal breakdown of carbon-containing precursors. The aluminium oxide template is then etched away using acids. Metal nanowires, such as Pd, and metal oxide nanotubes, such as CuO, MnO_2, ZnO and TiO_2, have also been made using an AAO template. The process can be continued to remake a positive. For instance, an AAO template has been used to make an array of PMMA polymer nanopillars. Nickel is then electroplated onto these pillars. When the pillars are dissolved away using an organic solvent, the original AAO template is replicated in metallic Ni (Figure 3.25).

Organic surfactants are used to make self-assembling micelles or inverted micelles, which encapsulate the inorganic material. Cylindrical micelles will form nanowires or nanorods. The reaction takes place within the micelle, which is subsequently removed to leave the nanoparticle or wire. Gold and metal oxide nanowires have been made this way.

Many different biological molecules and structures have been used as templates to make nanoparticles. These include DNA, peptides, the tobacco mosaic virus (TMV), pollen grains, diatoms and even human hair. The TMV is a long tube with an external diameter of 18 nm, and it has been successfully coated with various metal oxides

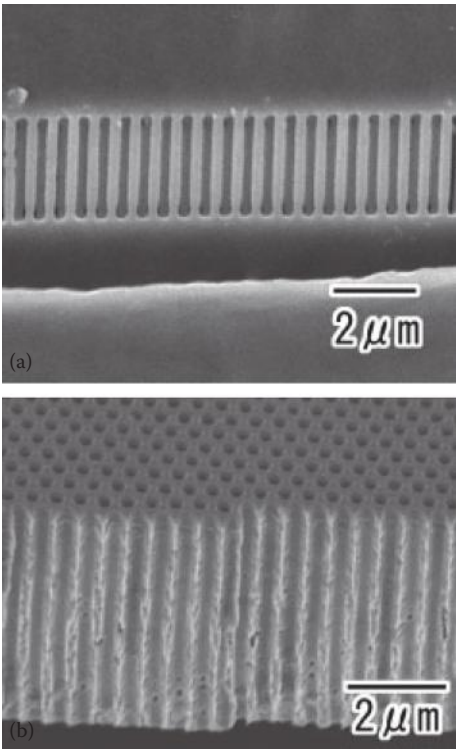


FIGURE 3.25 (a) An SEM of a PMMA nanopillar array after removing the alumina template. (b) An SEM of an Ni hole array membrane. (Masuda, H., et al.: *Advanced Materials.* 2005, **17**, 2241–2243. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reprinted with permission.)

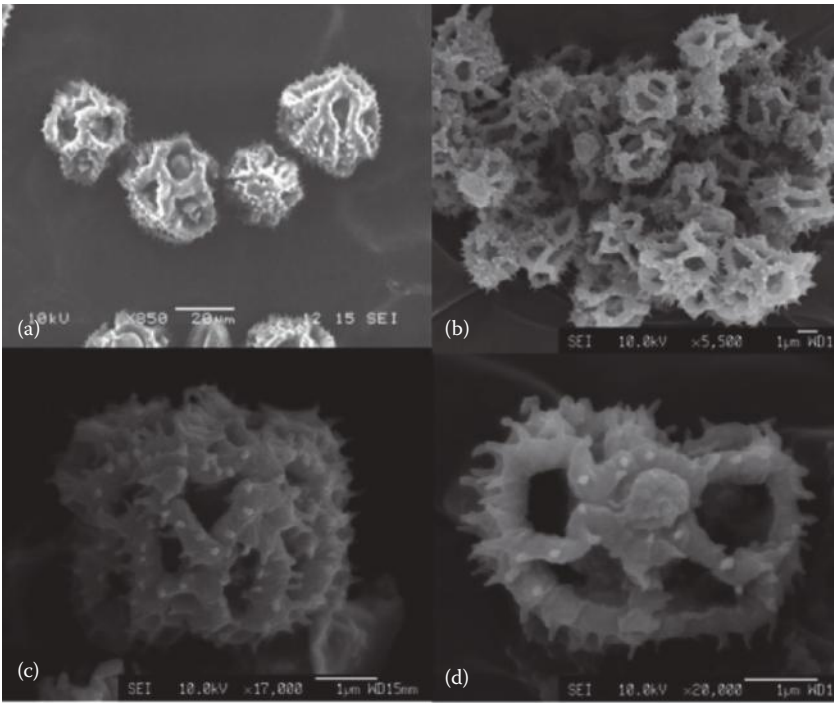


FIGURE 3.26 SEM images of (a) native dandelion (*Taraxacum*) pollen grains (scale bar: 20 μm). (b–d) SEM images of titania replicas produced by pollen grain templating and calcinations (scale bars, 1 μm). (b) Low magnification image. (c, d) Single titania particles. (Reprinted with permission from Hall, S.R., et al., *Chemistry of Materials*, **18**, 598, 2006. Copyright 2006 American Chemical Society.)

and sulfides to form composite nanotubes. Porous TiO_2 has been prepared by replicating pollen grains (Figure 3.26), and TiO_2 nanotubes by replicating human hair.

An unusual approach to constructing electronic circuits is to use the coding properties of DNA. By modifying the DNA so that a compound only attaches to some sequences in a DNA strand, we can ensure that a following reaction takes place only where that compound is present or only where it is absent. In one example, DNA was reacted with glutaraldehyde, which produced aldehyde groups on the DNA. A protein was then polymerised on a probe DNA molecule. The resulting nucleoprotein filament attached to a specific sequence on the original DNA. Thus, specific lengths of DNA were covered by the protein and others were left bare. The DNA was then reacted with silver nitrate. Where aldehyde groups were free (i.e., not covered by the filament), the silver nitrate was reduced to silver. The silver clusters formed catalysed the formation of gold from a solution of $KAuCl_4$, KSCN and hydroquinone. This produced a gold wire of a few micrometres long and 50–100 nm wide.

A preformed nanostructure can be used as a template in several ways. One method is to deposit a new compound on the outside of a nanotube, for instance. CVD or sol–gel methods can be used to do this and an outer shell is formed, the template is

then dissolved or oxidised away to leave a hollow core. For example, a range of metal oxide nanotubes, SiO_2, Al_2O_3 and TiO_2, among others, has been synthesised using CNTs as the template. The oxide precursor is deposited using a sol–gel method, followed by calcination. The CNT is then removed by oxidation at high temperature.

A direct chemical reaction between a nanostructure and a deposited outer shell can create a new nanostructure. Metal nanowires can be oxidised to metal oxide, sulfides and selenide nanowires, for instance. Metal carbide nanorods have been made by reacting CNTs with metal halides. Complex oxide nanotubes have been made by the solid-state reaction between two oxides. Al_2O_3 was deposited onto ZnO nanowires and heated; because the diffusion rate of the ZnO is faster, $ZnAl_2O_4$ nanotubes are formed.

3.10 CHOOSING A METHOD

Our choice of methods is not by any means exhaustive. Our aim here, therefore, is not to provide a way of choosing the method for a particular product. (Indeed, several of the examples given in this chapter show that several methods can be suitable for one substance.) Instead, we hope to give a few pointers to deciding whether a particular method is suitable for a particular material.

It is important that you consider the stability of the compounds under the reaction conditions and not at normal temperature and pressure. As we have illustrated in this chapter, a particular method may be chosen because the desired product is stable only at raised pressures or because it decomposes if the reaction temperature is too high.

What form do you want the product to be in? You might, for example, choose vapour phase epitaxy because an application requires a single crystal, or a method such as the precursor method or hydrothermal synthesis because you need a homogeneous product.

How pure must your product be? If you require high purity, you could choose a method that involves a volatile compound as a starting material because such compounds are generally easier to purify than solids.

You should consider the availability of the reactants required for a particular method. If you are considering a precursor method, is there a suitable precursor with the right stoichiometry? The CVD method needs reactants of similar volatility; do your proposed reactants meet this requirement? In microwave synthesis, does at least one of your starting materials absorb microwaves strongly?

QUESTIONS

1. The Chevrel phase $CuMo_6S_8$ was prepared by a ceramic method. What would be suitable starting materials and what precaution would you have to take?
2. Which synthetic methods would be suitable for producing the following characteristics:

   a. A thin film of material?
   b. A single crystal?

c. A single crystal containing layers of different material with the same crystal structure?
d. A powder of homogeneous composition?

3. What are the advantages and disadvantages of using the sol–gel method to prepare barium titanate for use in a capacitor?
4. A compound $(NH_4)_2Cu(CrO_4)_2{\cdot}2NH_3$ is known. How could this be used to prepare $CuCr_2O_4$? What would be the advantage of this method over a ceramic method? Suggest which solvent was used to prepare the ammonium compound.
5. β-TeI is a metastable phase formed at 465–470 K. Suggest an appropriate method of preparation.
6. In hydrothermal processes involving alumina (Al_2O_3), such as the synthesis of zeolites, alkali is added to the reaction mixture. Suggest a reason for this addition.
7. The zeolite ZSM-5 is prepared by heating a mixture of silicic acid, $SiO_2{\cdot}nH_2O$, NaOH, $Al_2(SO_4)_3$, water, *n*-propylamine and tetrapropylammonium bromide in an autoclave for several days at 160°C. The product from this reaction is then heated in air. Why is tetrapropylammonium bromide used in the reaction, and what is the effect of the subsequent oxidation reaction?
8. Which of the following oxides would be good candidates for microwave synthesis?—$CaTiO_3$, $BaPbO_3$, $ZnFe_2O_4$, $Zr_{1-x}Ca_xO_{2-x}$, KVO_3.
9. Crystals of silica can be grown using the chemical vapour transport method with hydrogen fluoride as a carrier gas. The reaction involved is $SiO_2(s) + 4HF(g) = SiF_4(g) + 2H_2O(g)$. This reaction is exothermic. In a temperature gradient method, will the silica crystals grow at the hotter or cooler end of the reaction vessel?
10. In the preparation of lithium niobate by CVD, argon-containing oxygen is used as a carrier gas. In the preparation of mercury telluride, the carrier gas is hydrogen. Suggest reasons for these choices of carrier gas.

4 Solids
Their Bonding and Electronic Properties

4.1 INTRODUCTION

In Chapter 1, you were introduced to the physical structure of solids—how their atoms and ions are arranged in space. We now turn to a description of the bonding in solids—the electronic structure. Some solids consist of molecules bound together by very weak forces. We shall not be concerned with these because their properties are essentially those of the molecules. In the final section we shall be concerned with 'purely ionic' solids bound by electrostatic forces between ions, as discussed in Chapter 1. The solids considered here are mostly those in which all the atoms can be regarded as bound together. We shall be looking at the basic bonding theories of these solids and how the theories account for the very different electrical conductivities of different groups of solids. We shall cover both theories based on the free-electron model: a view of a solid as an array of ions held together by a sea of electrons; and on the molecular orbital theory: a crystal seen as a giant molecule. Of these solids, some of the most important are the semiconductors. Many solid-state devices—transistors, photocells, light-emitting diodes (LEDs), solid-state lasers and solar cells—are based on semiconductors. We shall introduce examples of some devices and explain how the properties of semiconductors make them suitable for these applications. We start with the free-electron theory of solids and its application to metals and their properties.

4.2 BONDING IN SOLIDS: FREE-ELECTRON THEORY

Traditionally, bonding in metals has been approached through the idea of free electrons, a sort of electron gas.

The free-electron model regards a metal as a box in which the electrons are free to roam unaffected by the atomic nuclei or by each other. The nearest approximation to this model is provided by the metals on the far left of the periodic table—Group 1 (Na, K, etc.), Group 2 (Mg, Ca, etc.)—and aluminium. These metals are often referred to as simple metals.

The theory assumes that the nuclei stay fixed on their lattice sites surrounded by the inner or core electrons, while the outer or valence electrons travel freely through the solid. If we ignore the cores, then the quantum mechanical description of the outer electrons becomes very simple. Taking just one of these electrons, the problem becomes the well-known 'particle in a box'. We start by considering an electron in a one-dimensional solid.

The electron is confined to a line of length a (the length of the solid), which we shall call the x axis. Since we are ignoring the cores, there is nothing for the electron to interact with, so it experiences zero potential within the solid. The Schrödinger equation for the electron is

$$\frac{-\hbar^2}{2m_e}\frac{d^2\psi}{dx^2} = (E - V)\psi, \tag{4.1}$$

where $\hbar$ is Planck's constant divided by 2π, m_e is the mass of the electron, V is the electrical potential, ψ is the wave function of the electron and E is the energy of an electron with wave function. When $V = 0$, the solutions to this equation are simple sine or cosine functions, and we can verify this by substituting $\psi = \sin\sqrt{2m_e E/\hbar^2}\,x$ into Equation 4.1, as follows:

If $\psi = \sin\sqrt{2m_e E/\hbar^2}\,x$, then differentiating once gives

$$\frac{d\psi}{dx} = \sqrt{\frac{2m_e E}{\hbar^2}}\cos\sqrt{\frac{2m_e E}{\hbar^2}}\,x.$$

Differentiating twice, we get

$$\frac{d^2\psi}{dx^2} = -\frac{2m_e E}{\hbar^2}\sin\sqrt{\frac{2m_e E}{\hbar^2}}\,x,$$

which is $-(2m_e(E/\hbar^2))\psi$, thus

$$\frac{d^2\psi}{dx^2} = -\frac{2m_e E}{\hbar^2}\psi.$$

Multiplying this by $-(\hbar^2/2m_e)$ gives Equation 4.1 with $V = 0$.

The electron is not allowed outside the box and to ensure this, we put the potential to infinity outside the box. Since the electron cannot have infinite energy, the wave function must be zero outside the box and since it cannot be discontinuous, it must be zero at the boundaries of the box. If we take the sine wave solution, $\psi = \sin\sqrt{2m_e E/\hbar^2}\,x$, then this is zero at $x = 0$. To be zero at $x = a$ as well, there must be a whole number of half waves in the box. Sine functions have a value of zero at angles of $n\pi$ radians, where n is an integer, therefore $a\sqrt{2m_e E/\hbar^2} = n\pi$. The energy is thus quantised, $E = n^2h^2/8m_e a^2$, with quantum number n. Therefore, as n can take all integral values, this means that there are an infinite number of energy levels with larger gaps between each level. Most solids, of course, are three-dimensional (although we shall meet some later where conductivity is confined to one or two dimensions), so we now need to extend the free-electron theory to three dimensions.

For three dimensions, the metal can be taken as a rectangular box, $a \times b \times c$. The appropriate wave function is now the product of three sine or cosine functions and the energy is given by

$$E = \frac{h^2}{8m_e}\left(\frac{n_a^2}{a^2} + \frac{n_b^2}{b^2} + \frac{n_c^2}{c^2}\right). \tag{4.2}$$

Each set of quantum numbers, n_a, n_b and n_c, will give rise to an energy level. However, in three dimensions, there are many combinations of n_a, n_b and n_c that will give the same energy, whereas the one-dimensional model has only two levels of each energy (n and −n). For example, the sets of numbers in Table 4.1 all give $(n_a^2/a^2 + n_b^2/b^2 + n_c^2/c^2) = 108$, and hence the same energy. The number of states with the same energy is known as the degeneracy. For small values of the quantum numbers, it is possible to write out all the combinations that will give rise to the same energy, but when we are dealing with a crystal of say 10^{20} atoms, it becomes difficult to work out all the combinations. However, we can estimate the degeneracy of any level in a band of this size by introducing a quantity called the **wave vector** and assuming that this is continuous. If we substitute k_x, k_y and k_z for $n_a\pi/a$, $n_b\pi/b$ and $n_c\pi/c$, then the energy becomes

$$E = \left(k_x^2 + k_y^2 + k_z^2\right)\hbar^2/2m_e, \tag{4.3}$$

where k_x, k_y and k_z can be considered as the components of a vector **k**; the energy is proportional to the square of the length of this vector. The vector **k** is called the **wave vector** and is related to the momentum of the electron wave, as can be seen by comparing the classical expression $E = p^2/2m$, where p is the momentum and m is the mass, with the expression above. This gives the electron momentum as $\pm\mathbf{k}\hbar$.

All the combinations of quantum numbers giving rise to one particular energy correspond to a wave vector of the same length $|\mathbf{k}|$. Thus, all possible combinations leading to a given energy, produce vectors whose ends lie on the surface of a sphere of radius $|\mathbf{k}|$. The total number of wave vectors with energies up to and including that of the given energy is given by the volume of the sphere, which is $4k^3\pi/3$, where $|\mathbf{k}|$ is written as k. To convert this to the number of states with energies up to the given

TABLE 4.1

Sets of Values of n_a/a, n_b/b and n_c/c Such That $(n_a^2/a^2 + n_b^2/b^2 + n_c^2/c^2) = 108$

n_a/a	n_b/b	n_c/c
6	6	6
2	2	10
2	10	2
10	2	2

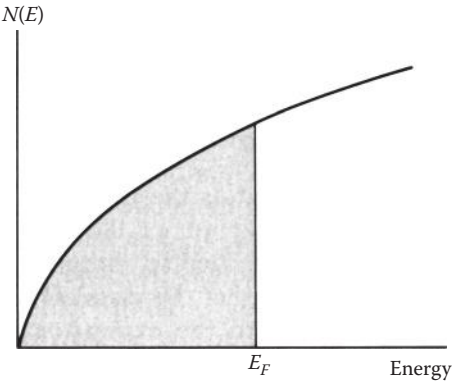


FIGURE 4.1 A density of states curve based on the free-electron model. The levels occupied at 0 K are shaded. Note that later in this book, energy is plotted on the vertical axis. In this figure, energy is plotted along the horizontal axis for comparison with the experiment.

energy, we have to use the relationships between the components of **k** and the quantum numbers n_a, n_b and n_c given earlier. This comparison shows that we have to multiply the volume by abc/π^3. Now we have the number of states with energies *up to* a particular energy, but it is more useful to know the number of states *with* a particular energy. To find this, we define the number of states in a narrow range of k values, dk. The number of states up to and including those of wave vector length $k + \mathrm{d}k$ is $(4/3\pi^2)V(k + \mathrm{d}k)^3$, where V (= abc) is the volume of the crystal. Therefore, the number of states with values between k and $k + \mathrm{d}k$ is $(4/3\pi^2)V((k + \mathrm{d}k)^3 - k^3)$, which when $(k + \mathrm{d}k)^3$ is expanded, gives a leading term $(4/\pi^2)Vk^2\mathrm{d}k$. This quantity is called the **density of states, *N*(*k*)d*k***. In terms of the more familiar energy, the density of states $N(E)\mathrm{d}E$ is given by $\sqrt{(2m_e)^3 E} \times V/2\pi^2\hbar^3\,\mathrm{d}E$. A plot of $N(E)\mathrm{d}E$ against E is given in Figure 4.1. Note that the density of states increases with increasing energy; therefore, the higher the energy, the greater is the number of states. In metals, the valence electrons fill up the states with paired spins, starting from the lowest energy up. For sodium, for example, each atom contributes one 3s electron and the electrons from all the atoms in the crystal occupy the levels in Figure 4.1 until all the electrons are used. The highest-occupied level is called the **Fermi level**.

Now let us see how this theoretical density of states compares with reality. Experimentally, the density of states can be determined by X-ray emission spectroscopy. A beam of electrons or high-energy X-rays hitting a metal is able to remove core electrons. In sodium, for example, the valence electron is 3s, so the 2s or 2p electrons might be removed. The core energy levels are essentially atomic levels, so electrons have been removed from a discrete, well-defined energy level. The electrons from the conduction band can now jump down to the vacated energy level, emitting X-rays as they do so. The X-ray energy will depend on the level of the conduction band from which the electron has come. A scan across the emitted X-rays will correspond to a scan across the filled levels. The intensity of the radiation emitted will depend on the number of electrons with that particular energy, that is, the intensity depends on the density of states of the conduction band. Figure 4.2 shows some X-ray emission

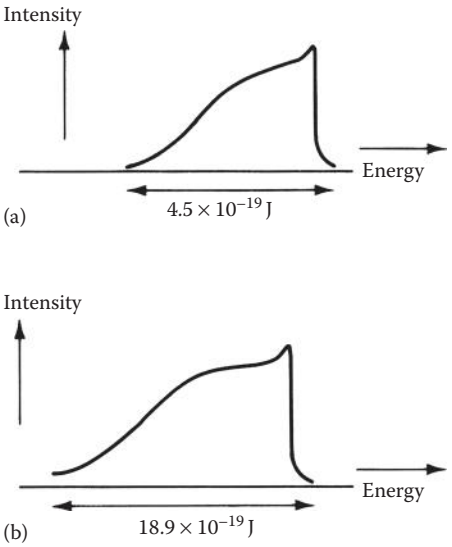


FIGURE 4.2 X-ray emission spectra obtained from (a) sodium metal and (b) aluminium metal when conduction electrons fall into the 2p level. The slight tail at the high-energy end is due to thermal excitation of the electrons close to the Fermi level.

spectra for sodium and aluminium, and we can see that the shape of these curves approximately resembles the occupied part of Figure 4.1, so that the free-electron model appears to describe these bands quite well.

If we look at the density of states found experimentally for metals with more electrons per atom than the simple metals, however, the fit is not as good. We find that instead of continuing to increase with energy, as in Figure 4.1, the density reaches a maximum and then starts to decrease with energy. This is shown in the 3d band of Ni in Figure 4.3.

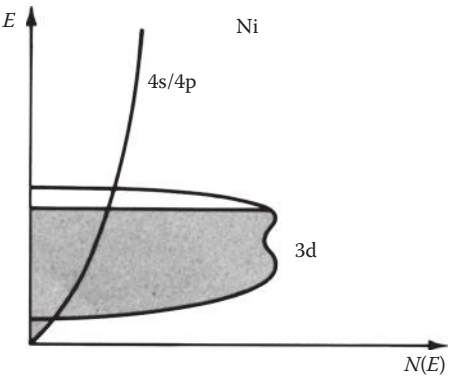


FIGURE 4.3 The band structure of nickel. Note the shape of the 3d band.

Extensions of this model, in which the atomic nuclei and core electrons are included by representing them by a potential function (V) in Equation 4.1 (plane wave methods), can account for the density of states in Figure 4.3 and can be used for semiconductors and insulators as well. However, we shall use a different model to describe these solids, one based on the molecular orbital theory of molecules, which is described in the next section. We conclude this section by using our simple model to explain the electrical conductivity of metals.

4.2.1 Electronic Conductivity

The wave vector (**k**) is the key to understanding electrical conductivity in metals. For this purpose, it is important to note that a vector has direction as well as magnitude. It follows that there may be many different energy levels with the same value of $|\mathbf{k}|$ and hence the same value of the energy, but with different components, k_x, k_y and k_z, giving a different direction to the momentum. In the absence of an electric field, all directions of the wave vector **k** are equally likely, so there are equal numbers of electrons moving in all directions (Figure 4.4).

If we now connect our metal to the terminals of a battery producing an electric field, then an electron travelling in the direction of the field will be accelerated and the energies of those levels with a net momentum in this direction will be lowered. Electrons moving in the opposite direction will have their energies raised, so some of these electrons will drop down into levels of lower energy corresponding to the momentum in the opposite direction. Thus, more electrons will move in the direction of the field than in other directions. This net movement of the electrons in one direction is an electric current. The net velocity in an electric field is shown in Figure 4.5.

An important point to note about this explanation is that it assumes that there are empty energy levels available that are close in energy to the Fermi level. You will see later that the existence of such levels is crucial in explaining the conductivity of semiconductors.

The model as we have described it so far, explains why metals conduct electricity, but it does not account for the finite resistance of metals. The current (i) flowing through a metal for a given applied electric field (V) is given by Ohm's law, $V = iR$, where R is the resistance. It is characteristic of a metal that R increases with

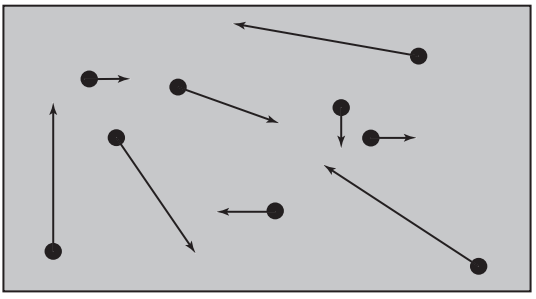


FIGURE 4.4 Electrons in a metal in the absence of an electric field. They move in all directions, but overall there is no net motion in any direction.

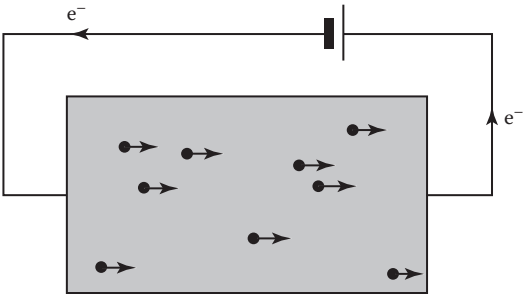


FIGURE 4.5 The sample of Figure 4.4 in a constant electric field, established by placing the rod between the terminals of a battery. The electrons can move in all directions, but now their velocities are modified so that each also has a net movement or drift velocity in the left to right direction.

increasing temperature (or putting it another way, the conductance (σ) decreases with increasing temperature); that is, for a given field, the current decreases as the temperature is raised. So far, there is nothing in our theory that will impede the flow of electrons. To account for electrical resistance, it is necessary to introduce the ionic cores. If these were arranged periodically on the lattice sites of a perfect crystal and did not move, then they would not interrupt the flow of electrons. However, most crystals contain some imperfections and these can scatter electrons. If the component of the electron's momentum in the field direction is reduced, then the current will drop. In addition, even in perfect crystals, the ionic cores will vibrate. There are a set of crystal vibrations in which the ionic cores vibrate together, the normal modes of the lattice. These vibrations are called **phonons**. An example of a crystal vibrational mode would be one in which each ionic core would be moving out of phase with its neighbours along one axis. (If they moved in phase, the whole crystal would move.) Similar to vibrations in small molecules, each crystal vibration has its set of quantised vibrational levels. The conduction electrons moving through the crystal are scattered by the vibrating ionic cores and lose some energy to the phonons. As well as reducing the electron flow, this mechanism increases the crystal's vibrational energy. The effect then is to convert electrical energy to heat. This ohmic heating effect is put to good use in, for example, the heating elements of kettles.

4.3 BONDING IN SOLIDS: MOLECULAR ORBITAL THEORY

We know that not all solids conduct electricity; however, no explanation for this has been offered by the simple free-electron model previously discussed. To understand semiconductors and insulators, we turn to another description of solids, molecular orbital theory. In the molecular orbital approach to bonding in solids, we regard solids as a very large collection of atoms bonded together and try to solve the Schrödinger equation for a periodically repeating system. For chemists, this has the advantage that solids are not treated as totally different species from small molecules.

However, solving such an equation for a solid is something of a tall order since exact solutions have not yet been found for small molecules and even a small crystal

could well contain of the order of 10^{16} atoms. An approximation often used for smaller molecules is that the molecular wave functions can be formed by combining atomic wave functions. This **linear combination of atomic orbitals (LCAO)** approach can also be applied to solids.

We shall start by reminding the reader how to combine atomic orbitals for a very simple molecule, H_2. For H_2, we assume that the molecular orbitals are formed by combining 1s orbitals on each of the hydrogen atoms. These can combine in phase to give a bonding orbital or out of phase to give an antibonding orbital. The bonding orbital is lower in energy than the 1s orbitals and the antibonding orbital is higher in energy. The amount by which the energy is lowered for a bonding orbital depends on the amount of overlap of the 1s orbitals on the two hydrogens. If the hydrogen nuclei are pulled further apart, for example, the overlap decreases and the decrease in energy is less. (If the nuclei are pushed together, the overlap will increase but the electrostatic repulsion of the two nuclei becomes important and counteracts the effect of the increased overlap.)

Suppose we form a chain of hydrogen atoms. For N hydrogen atoms, there will be N molecular orbitals. The lowest energy orbital will be that in which all the 1s orbitals combine in phase, and the highest energy orbital will be that in which the orbitals combine out of phase. In between are $(N - 2)$ molecular orbitals, in which there is some in-phase and some out-of-phase combinations. Figure 4.6 shows a plot of the energy levels as the length of the chain increases. Note that as the number of atoms increases, the number of levels increases; however, the spread of energies seems to increase only slowly and levels off for long chains. Extrapolating to crystal-length chains, we can see that there would be a very large number of levels within a comparatively small range of energies. A chain of hydrogen atoms is a very simple and artificial model; as an estimate of the energy separation of the levels, let us take a typical band in an average-size metal crystal. A metal crystal might contain 10^{16} atoms and the range of energies might be only 10^{-19} J. The average separation between the levels would thus be only 10^{-35} J. The lowest energy levels in the hydrogen atom are separated by energies of the order of 10^{-18} J, so we can see that the energy separation in a crystal is minute. The separation is in fact so small that, as in the free-electron model, we can think of the set of levels

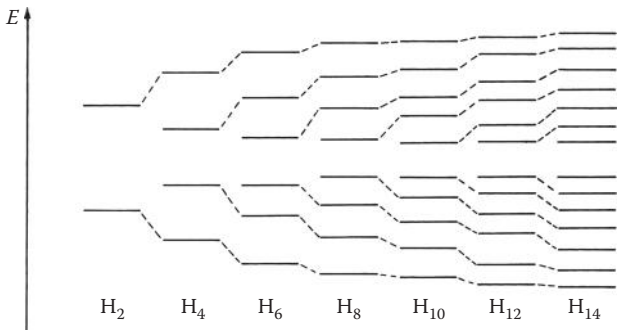


FIGURE 4.6 The orbital energies for a chain of N hydrogen atoms as N increases.

as a continuous range of energies. Such a continuous range of allowed energies is known as an **energy band**.

In actual calculations on crystals, it is impractical to include all 10^{16} atoms, so we use the periodicity of the crystal. We know that the electron density and the wave function for each unit cell are identical, so we form combinations of orbitals for the unit cell that reflect the periodicity of the crystal. Such combinations have patterns like the sine waves that we obtained from the particle in the box calculation. For small molecules, the LCAO expression for molecular orbitals is

$$\Psi(i) = \sum_n c_{ni}\phi_n,$$

where the sum is over all n atoms, $\Psi(i)$ is the ith orbital and c_{ni} is the coefficient of the orbital on the nth atom for the ith molecular orbital.

For solids, we replace this with

$$\Psi(\mathbf{k}) = \sum_n e^{i\mathbf{k}.\mathbf{r}_n} a_{n\mathbf{k}}\phi_n,$$

where $\mathbf{k}$ is the wave vector, $\mathbf{r}$ is the position vector of a nucleus in the lattice and $a_{n\mathbf{k}}$ is the coefficient of the orbital on the nth atom for a wave vector of $\mathbf{k}$.

The hydrogen chain orbitals were made up from only one sort of atomic orbital—1s—and one energy band was formed. For most of the other atoms in the periodic table, it is necessary to consider other atomic orbitals in addition to the 1s orbitals, and we find that the allowed energy levels form a series of energy bands separated by ranges of forbidden energies. The ranges of forbidden energy between the energy bands are known as band gaps. Aluminium, for example, has the atomic configuration $1s^2 2s^2 2p^6 3s^2 3p^1$ and would be expected to form a 1s band, a 2s band, a 2p band, a 3s band and a 3p band, all separated by band gaps. In fact, the lower energy bands, those formed from the core orbitals 1s, 2s and 2p, are very narrow and for most purposes they can be regarded as a set of localised atomic orbitals. This arises because these orbitals are concentrated very close to the nuclei, so there is little overlap between the orbitals on the neighbouring nuclei. In small molecules, the greater the overlap, the greater is the energy difference between the bonding and the antibonding orbitals. Likewise for continuous solids, the greater the overlap, the greater is the spread of energies or **bandwidth** of the resulting band. For aluminium, then, 1s, 2s and 2p electrons can be taken to be core orbitals and only 3s and 3p bands are considered.

Just as in small molecules, the available electrons are assigned to the levels in the energy bands starting with the lowest level. Each orbital can take two electrons of opposed spin. Thus, if N atomic orbitals were combined to make the band orbitals, then $2N$ electrons are needed to fill the band. For example, the 3s band in a crystal of aluminium containing N atoms can take up to $2N$ electrons, whereas the 3p band can accommodate up to $6N$ electrons. As aluminium has only one 3p electron per atom, however, there would only be N electrons in the 3p band and only $N/2$ levels would be occupied. As in the free-electron model, the highest-occupied level at 0 K

is the Fermi level. Now let us see how this model applies to some real solids and how we can apply the concept of energy bands to understanding some of their properties. First, we revisit the simple metals.

4.3.1 Simple Metals

The crystal structures of the simple metals are such that the atoms have high coordination numbers. For example, the Group 1 elements have body-centred cubic structures with each atom surrounded by eight others (see Chapter 1). This high coordination number increases the number of ways in which the atomic orbitals can overlap. The ns and np bands of the simple metals are very wide because of the large amount of overlap, and, since the ns and np atomic orbitals are relatively close in energy, the two bands merge. This can be shown even for a small chain of lithium atoms, as can be seen in Figure 4.7.

For the simple metals, then, we do not have an ns band or an np band, but rather one continuous band, which we shall label ns/np. For a crystal of N atoms, this ns/np band contains $4N$ energy levels and can hold up to $8N$ electrons.

However, the simple metals have far fewer than $8N$ electrons available; they have only N (Group 1), $2N$ (Group 2) or $3N$ (Al). Thus, the band is only partly full. There are empty energy levels near the Fermi level and the metals are good electronic conductors.

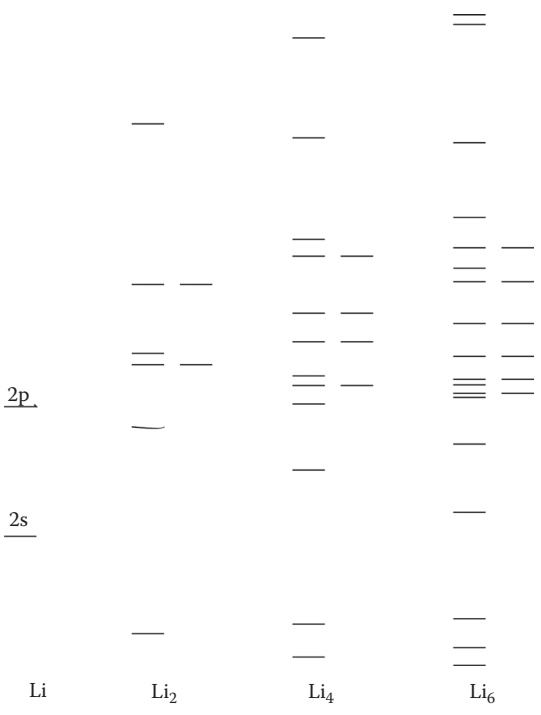


FIGURE 4.7 2s and 2p levels for an Li atom and for chains of 2, 4 and 6 Li atoms.

If we now move to the right of the periodic table, we find elements that are semiconductors or insulators in the solid state; we now go on to consider these solids.

4.4 SEMICONDUCTORS: Si AND Ge

Carbon (as diamond polymorph), silicon and germanium, instead of forming structures of high coordination like the simple metals, form structures in which the atoms are tetrahedrally coordinated (see Chapter 1). With these structures, the ns/np bands still overlap, but the ns/np band splits into two. Each of the two bands contains $2N$ orbitals and can accommodate up to $4N$ electrons. We can think of the two bands as an analogy of bonding and antibonding; the tetrahedral symmetry not giving rise to any nonbonding orbitals. Carbon, silicon and germanium have the electronic configurations ns^2np^2, thus they have available $4N$ electrons—exactly the correct number to fill the lower band. This lower band is known as the **valence band**, the electrons in this band essentially bonding the atoms in the solid together.

One question that may have occurred to you is 'Why do these elements adopt the tetrahedral structure rather than one of the higher coordination structures?'. If these elements adopted a structure like that of the simple metals, then the ns/np band would be half full. The electrons in the highest-occupied levels would be virtually nonbonding. In the tetrahedral structure, however, all $4N$ electrons would be in the bonding levels. This is illustrated in Figure 4.8. Elements with few valence electrons will thus be expected to adopt high coordination structures and be metallic. Those with larger numbers (4 or more) will be expected to adopt lower coordination structures in which the ns/np band is split and only the lower bonding band is occupied.

Tin (in its most stable room temperature form) and lead, although in the same group as silicon and germanium, are metals. For these elements, the atomic s–p energy separation is greater, and the overlap of s and p orbitals is much less than in silicon and germanium. For tin, the tetrahedral structure would have two s/p bands, but

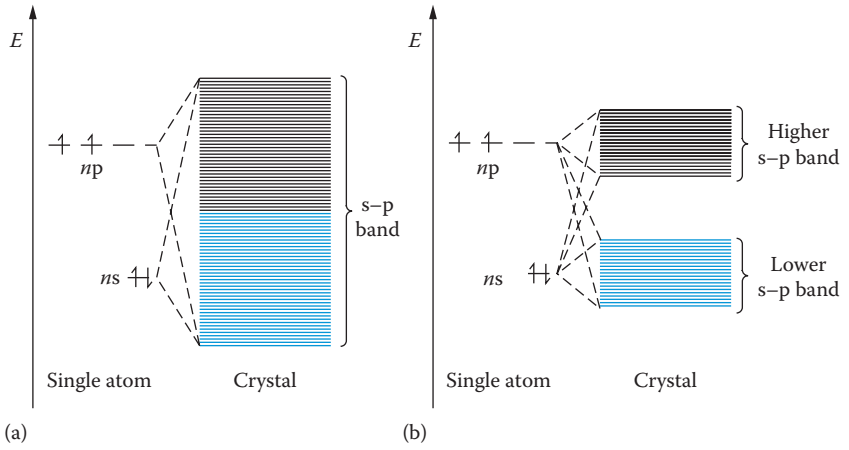


FIGURE 4.8 Energy bands formed from ns and np atomic orbitals for (a) a body-centred cubic crystal and (b) a crystal of diamond structure, showing filled levels for $4N$ electrons.

the band gap is almost zero. Below 291 K, tin undergoes a transition to the diamond structure, but above this temperature, it is more stable for tin to adopt a higher coordination structure. The advantage of having no nonbonding levels in the diamond structure is reduced by the small band gap. In lead, the diamond structure would give rise to an s band and a p band rather than ns/np bands, since the overlap of s and p orbitals is even smaller. It is more favourable for lead to adopt a cubic close-packed structure and, because lead has only $2N$ electrons to go in the p band, it is metallic. This is an example of the **inert pair effect** on the chemistry of lead, in which the s electrons act as core electrons. Another example of the inert pair is the formation of divalent ionic compounds containing Pb^{2+}, rather than tetravalent covalent compounds like those of silicon and germanium.

Silicon and germanium thus have a completely full valence band and are expected to be insulators. However, they belong to a class of materials known as **semiconductors**.

The electronic conductivity (σ), is given by the expression:

$$\sigma = nZe\mu,$$

where n is the number of charge carriers per unit volume, Ze is their charge (in the case of an electron, this is simply e, the electronic charge) and μ, the mobility, is a measure of the velocity in the electric field.

The conductivity of metallic conductors *decreases* with temperature. As the temperature is raised, the phonons gain energy and the lattice vibrations have larger amplitudes. The displacement of the ionic cores from their lattice sites is thus greater and the electrons are scattered more, and the net current is reduced because the mobility (μ) of the electrons is reduced.

The conductivity of **intrinsic semiconductors** such as silicon or germanium *increases* with temperature. In these solids, conduction can only occur if electrons are promoted to the higher s/p band, the **conduction band**, because only then will there be a partially full band. The current in the semiconductors depends on n, which in this case is the number of electrons free to transport charge. The number of electrons that are able to transport charge is given by the number promoted to the conduction band plus the number in the valence band that have been freed to move into empty levels by this promotion. As the temperature increases, the number of electrons promoted increases, so the current increases. The increase depends on the size of the band gap. At any one temperature, more electrons will be promoted in a solid with a small band gap, than in one with a large band gap, so that the solid with the smaller band gap will be a better conductor. The number of electrons promoted varies with temperature in an exponential manner, so that small differences in band gap lead to large differences in the number of electrons promoted and hence the number of current carriers. For example, tellurium has a band gap of about half that of germanium, but because of the exponential variation at a given temperature, the ratio of electrons promoted in tellurium to that in germanium is of the order of 10^6. Germanium has an electrical resistivity of 0.46 Ω m at room temperature compared with that of tellurium of 0.0044 Ω m or with that of a typical insulator of around 10^{12} Ω m.

The change in resistivity with temperature is used in thermistors, which are used as thermometers and also in fire alarm circuits.

4.4.1 Photoconductivity

Electrons can also be promoted across a semiconductor band gap by forms of energy other than heat energy, for example, by light. If the photon energy ($h\nu$) of light shining on a semiconductor is greater than the energy of the band gap, then valence band electrons will be promoted to the conduction band and conductivity will increase. The promotion of electrons by light is illustrated in Figure 4.9.

Semiconductors with band gap energies corresponding to photons of visible light are **photoconductors** because they are, essentially, nonconducting in the dark, but conduct electricity in the light. One use of such photoconductors is in **electrophotography**. In the xerographic process, there is a positively charged plate covered with a film of semiconductor. The semiconductor used is not silicon, but a solid with a more suitable energy band gap, such as selenium, the compound As_2Se_3 or a conducting polymer. During copying, the light reflected from the white parts of the page to be copied hits the semiconductor film; the parts of the film receiving the light become conducting, and an electron is promoted to the conduction band. This electron cancels the positive charge on the film, the positive hole in the valence band being removed by an electron from the metal backing plate entering the valence band. Now the parts of the film that received light from the original are no longer charged, but the parts underneath the black lines are still positively charged. In the next stage of the process, tiny, negatively-charged, plastic capsules of ink (toner) are spread onto the semiconductor film, but are only attracted to the charged parts of the film, which they cling to. A piece of positively charged white paper removes the toner from the semiconductor film and hence

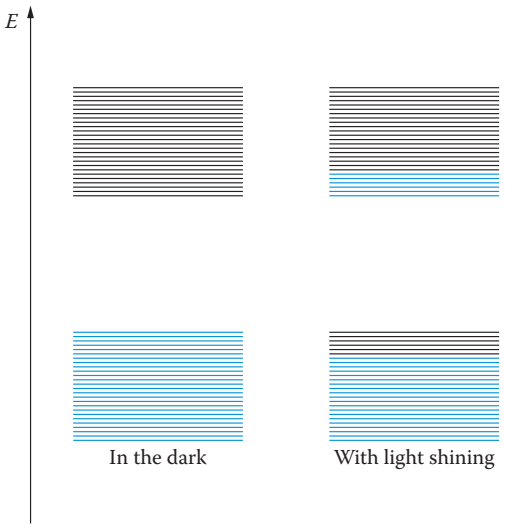


FIGURE 4.9 The promotion of electrons from the valence band to the conduction band by light.

acquires an image of the black parts of the original. Finally, the paper is heated to melt the plastic coating and fix the ink.

4.4.2 Doped Semiconductors

The properties of semiconductors are extremely sensitive to the presence of impurities at concentrations as low as 1 part in 10^{10}. For this reason, silicon manufactured for transistors and other devices must be very pure. The deliberate introduction of a very low concentration of certain impurities into the very pure semiconductor, however, alters the properties in a way that has proved invaluable in constructing semiconductor devices. Such semiconductors are known as **doped** or **extrinsic semiconductors** as opposed to natural semiconductors such as Si and Ge, which are called **intrinsic semiconductors**. Consider a crystal of silicon containing boron as an impurity. Boron has fewer valence electrons, one less than silicon; therefore, for every silicon replaced by boron, there is an electron missing from the valence band (see Figure 4.10), that is, there are **positive holes** in the valence band and these enable the electrons near the top of the band to conduct electricity, thus the doped solid will be a better conductor than pure silicon. A semiconductor like this, doped with an element with fewer valence electrons than the bulk of the material, is called a **p-type semiconductor** because its conductivity is related to the number of positive holes (or empty electronic energy levels) produced by the impurity.

Now suppose that instead of boron, the silicon is doped with an element with more valence electrons than silicon—phosphorus, for example. The doping atoms form a set of energy levels that lie in the band gap between the valence and conduction bands and close to the conduction band. Since the atoms have more valence electrons than silicon, these energy levels are filled. Thus, there are electrons present close to the bottom of the conduction band and they are easily promoted into the band. This time, the conductivity increases because of extra electrons entering

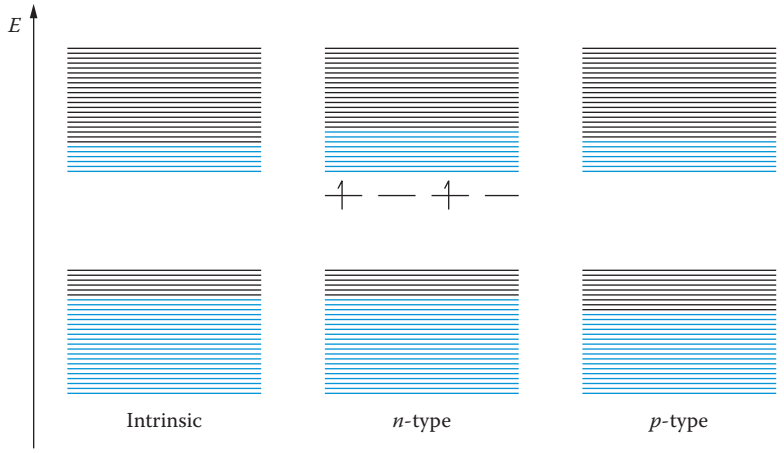


FIGURE 4.10 Intrinsic, *n*-type and *p*-type semiconductors showing negative charge carriers (electrons in the conduction band) and positive holes.

the conduction band. Such semiconductors are called ***n*-type**; ***n*** for negative charge carriers or electrons. Figure 4.10 schematically shows the energy bands in intrinsic, *p*-type and *n*-type semiconductors.

The *n*-type and *p*-type semiconductors in various combinations make up many electronic devices, such as rectifiers, transistors, photovoltaic cells and LEDs.

4.4.3 *p–n* Junction and Field Effect Transistors

p–n junctions are prepared either by doping different regions of a single crystal with different atoms or by depositing one type of material on top of another, using techniques such as chemical vapour deposition (see Chapter 3). The use of these junctions stems from what happens where the two differently doped regions meet. In the region of the crystal where *n*-type and *p*-type semiconductors meet, there is a discontinuity in the electron concentration. Although both *n*-type and *p*-type semiconductors are electrically neutral, the *n*-type has a greater concentration of electrons than the *p*-type. In order to try and equalise the electron concentrations, electrons drift from *n*-type to *p*-type. However, this drift produces a positive electric charge on the *n*-type and a negative electric charge on the *p*-type. The electric field thus set up encourages electrons to drift back to the *n*-type. Eventually, a state is reached in which the two forces are balanced and the electron concentration varies smoothly across the junction, as in Figure 4.11.

The regions immediately on either side of the junction are known as **depletion regions** because there are fewer current carriers (electrons or empty levels) here. Applying an external electric field across such a junction disturbs the equilibrium and the consequences of this are exploited in LEDs and transistors. In LEDs, which we discuss in Chapter 7, a voltage is applied so that the *n*-type semiconductor becomes negative relative to the *p*-type. An important feature of most transistors is a voltage applied in the reverse direction; that is, the *n*-type is positive with respect to the *p*-type.

The 1956 Nobel Prize in Physics was awarded to Bardeen, Brattain and Shockley for their work on developing transistors. Here, we briefly consider an important class of

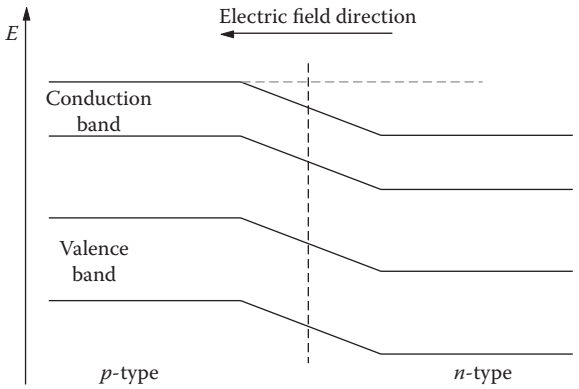


FIGURE 4.11 The bending of energy levels across a *p–n* junction.

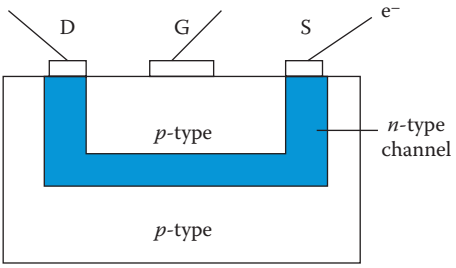


FIGURE 4.12 A schematic diagram of a field effect transistor showing source (S), drain (D) and gate (G) electrodes.

transistors, the field effect transistors (FET). These have many uses, including as amplifiers in radios and in other electronic equipment and as gates in computer circuits.

A simple *n*-channel FET consists of a block of heavily doped *p*-type semiconductor with a channel of *n*-type semiconductor (Figure 4.12). Electrodes are attached to the *p*-type block and to the *n*-type channel. The electrodes attached to the *n*-type channel are known as the source (negative electrode that provides electrons) and the drain (positive electrode). The electrode attached to the *p*-type block is known as the gate electrode. A low voltage (typically 6 V) is applied across the source and the drain electrodes. To fulfil the transistor's role as an amplifier or a gate, a current is applied to the gate electrode. Electrons flow into the *p*-type semiconductor, but cannot cross the *p*–*n* junction because the valence band in the *n*-type is full. Therefore, the electrons fill the valence band in the *p*-type. Because of the charge produced in the *p*-type semiconductor, the electrons in the *n*-type channel move towards the centre of this channel. The net result is an increased depletion zone and because the *p*-type is more heavily doped, this effect is greater for the *n*-type channel. The depletion zones contain fewer charge carriers than the bulk semiconductor, so the current in the *n*-type channel is greatly reduced. If the voltage on the gate electrode is reversed, then electrons flow out of the *p*-type, the depletion zones shrink and the current through the *n*-type channel increases. By varying the sign of the voltage across the gate electrode, for example, by using an alternating current, the current in the *n*-type channel is turned on or off.

Transistor amplifiers consist of an electronic circuit and other components such as resistors. The signal to be amplified is applied to the gate electrode. The output signal is taken from the drain. Under the conditions imposed by the rest of the circuit, the current through the *n*-type channel increases linearly with the magnitude of the incoming voltage. This current produces a drain voltage that is proportional to the incoming voltage but with a magnitude that is greater by a constant factor.

Metal oxide semiconductor field effect transistors (MOSFETs) are FETs with a thin film of silicon dioxide between the gate electrode and the semiconductor. The charge on the silicon dioxide controls the size of the depletion zone in the *p*-type semiconductor. MOSFETs are easier to mass-produce and are used in integrated

circuits and microprocessors for computers. Traditionally, transistors have been silicon-based, but a recent development is FETs based on organic materials.

4.5 BANDS IN COMPOUNDS: GALLIUM ARSENIDE

Gallium arsenide (GaAs) is a rival to silicon in some semiconductor applications, including solar cells, and it is also used for LEDs and in solid-state lasers, as you will see in Chapter 7. It has a diamond-type structure and is similar to silicon, except that it is composed of two kinds of atoms. The valence orbitals in Ga and As are the 4s and 4p and these form two bands, each containing 4*N* electrons as in silicon. Because of the different 4s and 4p atomic orbital energies in Ga and As, however, the lower band has a greater contribution from As and the conduction band has a higher contribution from Ga. Thus, GaAs can be thought of as having partial ionic character because there is a partial transfer of electrons from Ga to As. The valence band has more arsenic than gallium character, so all the valence electrons end up in orbitals in which the possibility of being near an As nucleus is greater than that of being near a Ga nucleus. The band energy diagram for GaAs is shown in Figure 4.13. GaAs is an example of a class of semiconductors known as III/V semiconductors, in which an element with one more valence electron than the silicon group is combined with an element with one less valence electron. Many of these compounds are semiconductors, for example, GaSb, InP, InAs and InSb. Moving farther along the periodic table, there are II/VI semiconductors such as CdTe and ZnS. Towards the top of the periodic table and further out towards the edges, for example, AlN and AgCl, the solids tend to adopt different structures and become more ionic. For the semiconducting solids, the band gap decreases down a group, for example, GaP > GaAs > GaSb and AlAs > GaAs > InAs.

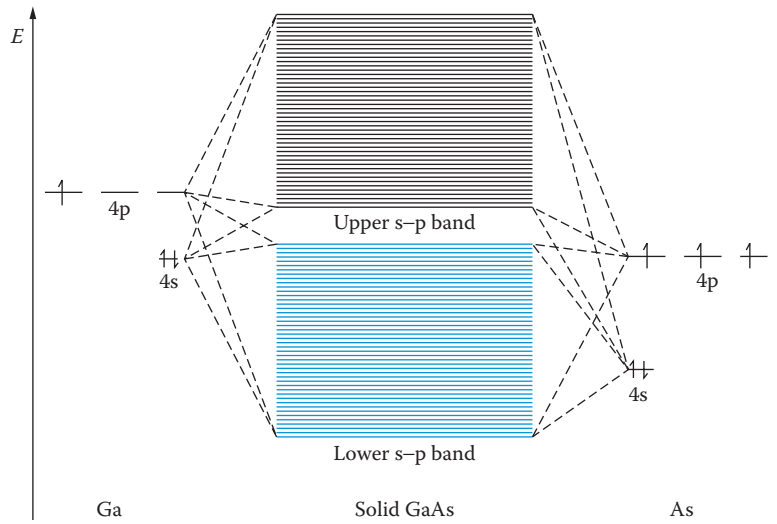


FIGURE 4.13 An orbital energy level diagram for gallium arsenide.

4.6 BANDS IN d-BLOCK COMPOUNDS: TRANSITION METAL MONOXIDES

Monoxides (MO) with structures based on sodium chloride* are formed by the first-row transition elements Ti, V, Mn, Fe, Co and Ni. TiO and VO are metallic conductors, and the others are semiconductors. The O 2p orbitals form a filled valence band. The 4s orbitals on the metal form another band. What of the 3d orbitals?

In the sodium chloride structure, the symmetry enables three of the five d orbitals on different atoms to overlap. Because the atoms are not nearest neighbours, the overlap is not as large as in pure metals and the bands are thus narrow. The other two d orbitals overlap with the orbitals on the adjacent oxygens. Thus, there are two narrow 3d bands, the lower one, labelled t_{2g}, can take up to $6N$ electrons and the upper one, labelled e_g, can take up to $4N$ electrons. Divalent titanium has two d electrons, so there are $2N$ electrons to fill the $3N$ levels of the lower band. Similarly, divalent vanadium has three d electrons, so the lower band is half full. As in the case of pure metals, a partly filled band leads to metallic conductivity. For FeO (Fe^{2+}, d^6), the t_{2g} band would be full, so it is not surprising to find that it is a semiconductor, but MnO with only five electrons per manganese (Mn^{2+}, d^5) is also a semiconductor. CoO and NiO, which should have partially full e_g levels, are also semiconductors. It is easier to understand these oxides using a localised d electron model.

Across the first transition series, there is a contraction in the size of the 3d orbitals; the 3d orbital overlap therefore decreases and the 3d band narrows. In a wide band such as the s/p bands of the alkali metals, the electrons are essentially free to move through the crystal, keeping away from the nuclei and from each other. In a narrow band, by contrast, the electrons are more tightly bound to the nuclei. Interelectron repulsion in these circumstances becomes important; in particular, the repulsion between electrons on the same atom. Consider an electron in a partly filled band, moving from one nucleus to another. In an alkali metal, the electron would already be in the sphere of influence of the surrounding nuclei and would not be greatly repelled by the electrons on these nuclei. The 3d electron moving from one nucleus to another adds an extra electron near to the second nucleus, which already has 3d electron density near it. Thus, electron repulsion on the nucleus is increased. For narrow bands, therefore, we have to balance the gains in energy made by band formation, against the electron repulsion felt. For MnO, FeO, CoO and NiO, electron repulsion wins and it becomes more favourable for the 3d electrons to remain in localised orbitals than to be delocalised.

The band gap between the oxygen 2p band and the metal 4s band is sufficiently wide for the pure oxides to be considered insulators. However, they are almost invariably found to be nonstoichiometric, that is, their formulae are not exactly MO, and this leads to semiconducting properties, which will be discussed in Chapter 5.

The MOs are not unique in displaying a variation of properties across the transition series. The dioxides form another series and CrO_2 will be discussed later because of

* As explained in Chapter 3, TiO and VO have structures based on sodium chloride but with one-sixth of each element missing in an ordered manner. The other oxides, when stoichiometric, adopt the sodium chloride structure.

its magnetic properties. Several classes of mixed oxides also exhibit a range of electronic properties; for instance, the perovskites $LaTiO_3$, $SrVO_3$ and $LaNiO_3$ are metallic conductors, $LaRhO_3$ is a semiconductor and $LaMnO_3$ is an insulator. Sulfides also show progression from metal to insulator. In general, compounds with broader d bands are metallic. Broad bands tend to occur for elements at the beginning of the transition series and for second and third row metals, for example, NbO and WO_2. Metallic behaviour is also more common among lower oxidation state compounds and with less electronegative anions.

4.7 CLASSICAL MODELLING

Although advances in computer power mean that many solids, including those with complex structures, can be studied using quantum theory, there are occasions when a much simpler model proves sufficient. This approach is based on lattice energy calculations and has proved useful for studying the structures of zeolites (zeolites are discussed in Chapter 6), for studying the effect of isolated point defects, and for tracking ions as they move through a solid. It allows random occupancy of sites in such calculations which is not generally possible with quantum mechanics. In addition this approach is often used with methods that follow the random movement of atoms over time as such methods are very demanding of computer resource.

Classical modelling methods start by calculating the electrostatic interaction between the ions, then terms are added to represent intermolecular forces such as van der Waals forces and to allow for the polarisability of the ions. If hydrogen is present or there are molecules or complex ions in the solid, then it may be necessary to add terms that describe covalent bonds.

Let us take some examples.

4.7.1 Intrinsic Defects in Alkali Halide Crystals

In Chapter 5, you will see that many solids are not perfect; they contain defects, that is, they have missing ions, ions that have been oxidised or reduced, or ions occupying sites other than those of the perfect lattice. An early application of classical methods was the estimation of the energy of the defect energies of alkali halides. Such calculations can tell us which type of defect is favoured and can give an indication of the concentration of the defects expected. To do this, we compare the values obtained for the lattice energy for the perfect solid and for the solid containing a defect.

Alkali metal halides are a good approximation to ionic solids, so the calculation of lattice energy can start by assuming that the ions are point charges and by applying Coulomb's law. The energy of interaction then has to be summed over an infinite array of ions. In Chapter 1, we saw that this gives a series that is slow to converge. In most calculations, this is overcome using the **Ewald method**. The Ewald method divides the summation into near and far contributions. For the near contribution, a charge distribution taking the form of a Gaussian function, $\exp[-r^2/a^2]$, where a determines the width or spread of the Gaussian, is added around each point charge. These distributions have a net charge equal and opposite to that of

the point charges. This ensures that beyond the width of the Gaussian, other point charges see a net charge of zero and hence no interaction. This effectively reduces the Coulomb interaction to a short-range interaction, which can be evaluated. The interaction between the Gaussian charge distributions forms the far contribution, which can be calculated via a Fourier transform and subtracted from the near contribution.

Even these simple solids are not entirely ionic, so other terms are added to the calculation. First, we have to allow for repulsion between the electron clouds of the ions at short distances. In Chapter 1, an expression for this, by Born, was introduced. In these calculations, it is common instead to use an expression by Buckingham, $E = A\exp(-r/\rho) - (C/r^6)$, where r is the distance from the centre of the ion and A, ρ and C are the adjustable parameters. Secondly, the polarisability of the ion (its tendency to deform as other ions approach) must be allowed for, especially for larger ions such as Cs^+ and I^-. A simple but effective model is used in which the ion is divided into a core and a shell. The shell can expand and contract around the core under the influence of a simple harmonic force, $F = -k(r - r_0)^2$, where r is the radius of the shell, k is a force constant (known as a spring constant) and r_0 is the equilibrium bond length. The smaller the force constant, the easier it is for the shell to alter its size. It is also possible to allow the shell to move off-centre to mimic the deformation of the ion.

Using Coulomb's law, Buckingham potentials to represent polarisability, and a core-shell model, the lattice energies of the solids can be reproduced to a high degree of accuracy.

For defect solids, the lattice energy calculation is then repeated with either one ion missing or one ion moved from its lattice site to an interstitial position. A sphere of ions around the defect is allowed to alter position in response to the defect, but beyond a certain distance, the lattice is assumed to be unaffected. The difference between the lattice energies with and without the defect is known as the **defect energy**.

Values of defect energies for some ionic solids are given in Chapter 5, along with the method for obtaining the concentration of defects from these values.

4.7.2 Zeolites

Zeolites are silica-based framework materials with a wide variety of uses, such as molecular sieves and catalysts. The unit cells of these solids are large and, until recently, they were difficult to model using quantum mechanical methods. Classical methods have proved amazingly successful in the study of zeolites, predicting possible new structures, determining the position of ions and molecules within the cages formed by these solids, and throwing light on the mechanism of catalysis by zeolites.

Surprisingly, the ionic model turns out to be a good approximation. The lattice energy calculations can be performed using the same method as for the alkali halides, assuming that the framework is composed of Si^{4+} and O^{2-} ions. One additional term is needed, and it is a term that constrains the OSiO angles. This term is another simple harmonic term that tends to pull the angle back towards an equilibrium value if it deviates.

4.7.3 Ionic Conductors for Fuel Cells

Solid oxide fuel cells are currently (2011) a very active area of research (see Chapter 5). One aspect of the performance of these cells that is difficult to measure experimentally is the pathway that the oxide ions take through the solid. Although experiments can provide a value for the activation energy for the path, it is difficult to capture the ions in mid-flight in order to ascertain the path that they take. Following the path with quantum mechanical calculations is also difficult as these are in a periodic arrangement. Thus, an oxide ion moved from its site in one unit cell would also be moved in all unit cells. Large supercells are then required to ensure that the displaced oxide does not interact with those in the neighbouring cells.

Classical calculations are widely used to determine such pathways. The same types of terms used for the alkali halides are used to describe the solid. When an oxide ion is moved from its normal site to a nearby interstitial site or onto an oxide ion vacancy in a neighbouring cell, the migrating oxide ion is moved in a series of steps along a proposed path and the surrounding ions are allowed to relax at each step along the path. As with defect calculations, the bulk of the solid remains unaffected. In this manner, different pathways can be explored and the activation energy for any path can be calculated. This technique can be applied to diffusion of any ion through an ionic solid. Examples of such pathways are given in Chapter 5.

When performing calculations on solids, there is thus a choice of methods. When choosing the most appropriate method, you need to consider the type of property you are interested in, the accuracy needed, the computer resources available and the size of the unit cell.

QUESTIONS

1. In the free-electron model, the electron energy is entirely kinetic. Using the formula $E = 1/2\ mv^2$, calculate the velocity of electrons at the Fermi level in sodium metal.
2. The density of magnesium metal is 1740 kg m^{-3}. A typical crystal has a volume of 10^{-12} m^3 (corresponding to a cube of side 0.1 mm). How many atoms would such a crystal contain?
3. An estimate of the total number of occupied states can be obtained by integrating the density of states from 0 to the Fermi level.

$$N = \int_0^{E_F} N(E)\,dE = \int_0^{E_F} E^{1/2}\left(2m_e\right)^{3/2} V/2\pi^2\hbar^3\,dE = \left(2m_eE_F\right)^{3/2} V/3\pi^2\hbar^3.$$

   Calculate the total number of occupied states for a magnesium crystal of volume (a) 10^{-12} m^3, (b) 10^{-6} m^3 and (c) 10^{-29} m^3 (approximately atomic size). Compare your results with the number of electrons available and comment on the different answers to (a), (b) and (c).
4. The energy associated with one photon of visible light ranges from 2.4 to 5.0×10^{-19} J. The band gap in selenium is 2.9×10^{-19} J. Explain why

selenium is a good material to use as a photoconductor in applications such as photocopiers.

5. The band gaps of several semiconductors and insulators are given below. Which substances would be photoconductors over the entire range of visible wavelengths?

Substance	Si	Ge	CdS
Band gap (10^{-19} J)	1.9	1.3	3.8

6. Which of the following doped semiconductors will be *p*-type and which will be *n*-type? (a) arsenic in germanium, (b) germanium in silicon, (c) indium in germanium, (d) silicon on antimony sites in indium antimonide (InSb) and (e) Mg on gallium sites in gallium nitride (GaN).
7. Would you expect carborundum (SiC) to adopt a diamond structure or one of a higher coordination? Explain why.
8. Comment on the electrical properties of the perovskites $LaTiO_3$ (metallic), $LaVO_3$ (semimetal) and $LaMnO_3$ (insulator) in terms of trends in the width of the d bands across the transition series.
9. Give an example of where an advantage would be obtained by using classical calculations rather than quantum mechanical calculations.

5 Defects and Nonstoichiometry

5.1 POINT DEFECTS: AN INTRODUCTION

In a perfect crystal, all atoms would be in their correct lattice positions in the structure. This situation only exists at the absolute zero of temperature, 0 K. Above 0 K, **defects** occur in the structure. These defects may be **extended defects**, such as **dislocations**. The strength of a material depends very much on the presence (or absence) of extended defects, such as dislocations and **grain boundaries**; however, discussion of these types of phenomena lie more in the realm of materials science and will not be discussed further. Defects can also occur at isolated atomic positions and are known as **point defects**, which can be due to the presence of a foreign atom at a particular site, or to a vacancy where normally one would expect an atom to be present. Point defects can have significant effects on the chemical and physical properties of a solid, for instance, the beautiful colours of many gemstones are due to defect impurity atoms in the crystal structure. Another example is ionic solids, which are able to conduct electricity by a mechanism that is due to the movement of *ions* through vacant ion sites within the lattice (this is in contrast to the electronic conductivity explored in the previous chapter, which depends on the movement of *electrons*).

5.2 DEFECTS AND THEIR CONCENTRATION

Defects fall into two main categories: **intrinsic defects**, which are integral to the crystal in question—they do not change the overall composition and because of this, they are also known as **stoichiometric defects**; and **extrinsic defects**, which are created when a foreign atom is inserted into the lattice.

5.2.1 Intrinsic Defects

Intrinsic defects fall into two categories: **Schottky defects**, which consist of vacancies in the lattice; and **Frenkel defects**, where a vacancy is created by an atom or an ion moving into an interstitial position.

For a 1 : 1 solid, MX, a Schottky defect, consists of a *pair* of vacant sites, a cation vacancy and an anion vacancy. This is shown in Figure 5.1 for an alkali halide–type structure: the number of cation vacancies and that of anion vacancies have to be equal in order to preserve electrical neutrality. A Schottky defect for an MX_2-type structure will consist of the vacancy caused by the M^{2+} ion together with *two* X^- anion vacancies, thereby balancing the electrical charges. Schottky defects are more

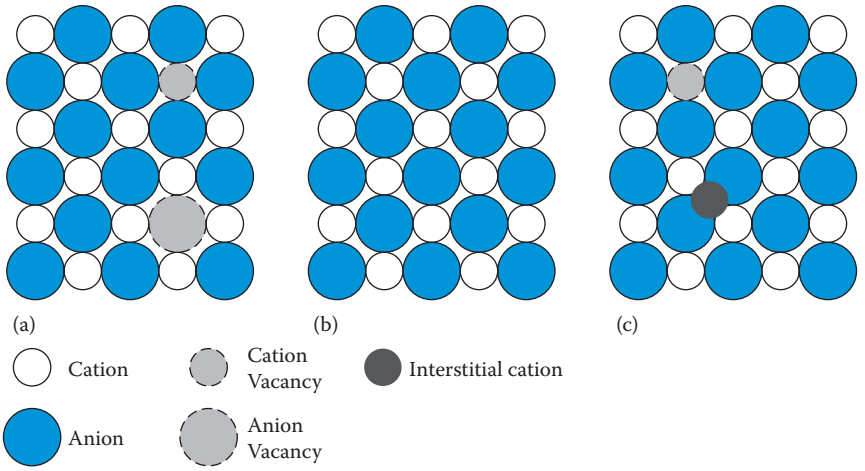


FIGURE 5.1 Schematic illustration of intrinsic point defects in a crystal of composition MX: (a) Schottky pair; (b) perfect crystal; (c) Frenkel pair.

common in 1 : 1 stoichiometry, and examples of crystals that contain them include rock salt (NaCl), wurtzite (ZnS) and caesium chloride (CsCl).

A Frenkel defect usually occurs only on one sublattice of a crystal and consists of an atom or an ion moving into an interstitial position, thereby creating a vacancy. This is illustrated in Figure 5.1c for an alkali halide–type structure such as NaCl, where one cation is shown as having moved out of the lattice and into an interstitial site. This type of behaviour is seen, for instance, in AgCl, where we observe such a *cation Frenkel defect* when Ag^+ ions move from their octahedral coordination sites into tetrahedral coordination, and this is illustrated in Figure 5.2. The formation of this type of defect is important in the photographic film process, when they are formed in the light-sensitive silver bromide (AgBr) used in photographic emulsions.

It is less common to observe an *anion Frenkel defect* when an anion moves into an interstitial site. This is because the anions are commonly *larger* than the cations in the structure, therefore it is more difficult for them to enter a crowded low-coordination interstitial site.

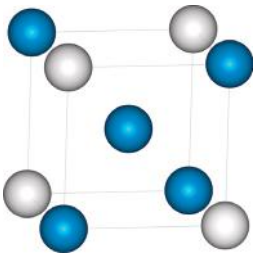


FIGURE 5.2 The tetrahedral coordination of an interstitial Ag^+ ion in AgCl.

An important exception to this generalisation lies in the formation of anion Frenkel defects in compounds with the fluorite structure, such as calcium fluoride (CaF_2; see Chapter 1). (Other compounds adopting this structure are strontium and lead fluorides [SrF_2, PbF_2] and also thorium uranium and zirconium oxides [ThO_2, UO_2, ZrO_2], which we will discuss again later in this chapter.) One reason for this is that the anions have a lower electrical charge than the cations and therefore can move nearer to each other. The other reason lies in the nature of the fluorite structure, which is shown again in Figure 5.3. You may recall that we can think of this structure as based on a cubic close-packed (*ccp*) array of Ca^{2+} ions with all the tetrahedral holes occupied by the F^- ions. This leaves all of the larger octahedral holes unoccupied, giving a very open structure. This can be seen clearly if we redraw the structure as in Figure 5.3c based on a simple cubic array of F^- ions. The unit cell now consists of eight small **octants** with the Ca^{2+} ions now occupying every other octant. The two different views are completely equivalent, but the cell shown in Figure 5.3c shows the possible interstitial sites more clearly.

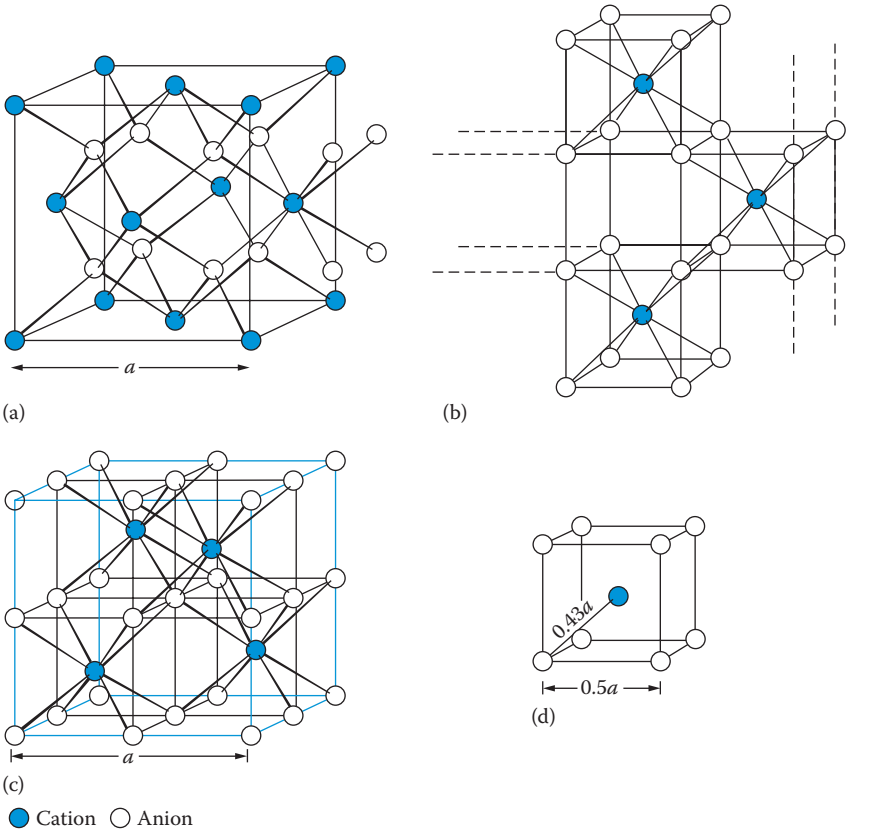


FIGURE 5.3 The crystal structure of fluorite MX_2. (a) A unit cell as a *ccp* array of cations. (b, c) The same structure redrawn as a simple cubic array of anions: the unit cell is marked by a coloured outline in (c). (d) Cell dimensions.

5.2.2 Concentration of Defects

Energy is required to form a defect: this means that the formation of defects is always an **endothermic process**. It may seem surprising that defects exist in crystals at all, and yet they do, even at low temperatures, albeit in very small concentrations. The reason for this is that although it costs energy to form defects, there is a commensurate *gain* in entropy. The enthalpy of formation of the defects is thus balanced by the gain in entropy such that, at equilibrium, the overall change in the free energy of the crystal due to the defect formation is zero according to the equation:

$$\Delta G = \Delta H - T\Delta S$$

The interesting point is that, thermodynamically, we do not expect a crystalline solid to be perfect, contrary, perhaps, to our 'commonsense' expectation of symmetry and order. At any particular temperature, there will be an equilibrium population of defects in the crystal.

The number of Schottky defects in a 1 : 1 crystal of composition MX is given by

$$n_S \approx N \exp\left(\frac{-\Delta H_S}{2kT}\right) \tag{5.1}$$

where n_S is the number of Schottky defects per unit volume, at T/K, in a crystal with N cation and N anion sites per unit volume; ΔH_S is the enthalpy required to form one defect. It is relatively straightforward to derive these equations for the equilibrium concentration of Schottky defects by considering the change in entropy of a perfect crystal due to the introduction of defects. The change in entropy will be due to the change in the frequency of the normal mode vibrations of atoms around the defects (the phonons) and also to the arrangement of the defects, the **configurational entropy**. It is possible to estimate this latter quantity, using the methods of statistical mechanics.

If the number of Schottky defects is n_S per unit volume at T/K, then there will be n_S cation vacancies and n_S anion vacancies in a crystal containing N possible cation sites and N possible anion sites per unit volume. The **Boltzmann formula** tells us that the entropy of such a system is given by

$$S = k \ln W \tag{5.2}$$

where W is the number of ways of distributing n_S defects over N possible sites at random and k is the Boltzmann constant (1.380662×10^{-23} J K^{-1}). The number of cation vacancies (W_c), equals the number of anion vacancies (W_a), and both are given by probability theory, which shows that

$$W_c = W_a = \frac{N!}{(N - n_S)!\, n_S!} \tag{5.3}$$

where $N!$ is called 'factorial N' and is mathematical shorthand for $N \times (N-1) \times (N-2) \times (N-3) \cdots \times 1$.

The total number of ways of distributing these defects (W) is given by the product of W_c and W_a:

$$W = W_c W_a$$

It follows that when a perfect crystal has no defects, that is, $n_S = 0$, then $W = 1$ and $S = 0$; a perfect crystal has zero configurational entropy. The *change* in configurational entropy due to introducing defects into a perfect crystal is thus

$$\Delta S = k \ln W = k \ln \left\{ \frac{N!}{(N-n_S)!\,n_S!} \right\}^2$$

$$= 2k \ln \left\{ \frac{N!}{(N-n_S))!\,n_S!} \right\}$$

We can simplify this expression using **Stirling's approximation** that, for values of $N \gg 1$,

$$\ln N! \approx N \ln N - N$$

and the expression becomes (after manipulation)

$$\Delta S = 2k \left\{ N \ln N - (N-n_S) \ln (N-n_S) - n_S \ln n_S \right\}$$

If the enthalpy change for the formation of a single defect is ΔH_S and we assume that the enthalpy change for the formation of n_S defects is $n_S \Delta H_S$, then the Gibbs free energy change is given by

$$\Delta G = n_S \Delta H_S - 2kT \left\{ N \ln N - (N-n_S) \ln (N-n_S) - n_S \ln n_S \right\}$$

At equilibrium, at constant T, the Gibbs free energy of the system must be a minimum with respect to changes in the number of defects n_S; thus,

$$\left(\frac{\mathrm{d}\Delta G}{\mathrm{d}n_S} \right) = 0$$

Therefore,

$$\Delta H_S - 2kT \frac{\mathrm{d}}{\mathrm{d}n_S} \left[N \ln N - (N-n_S) \ln (N-n_S) - n_S \ln n_S \right] = 0$$

$N \ln N$ is a constant and hence its differential is zero; the differential of $\ln x$ is $1/x$ and of $(x \ln x)$ is $(1 + \ln x)$. On differentiating, we get

$$\Delta H_S - 2kT\left[\ln\left(N - n_S\right) + 1 - \ln n_S - 1\right] = 0$$

hence, $\Delta H_S = 2kT \ln[(N - n_S)/n_S]$ and $n_S = (N - n_S) \exp(-\Delta H_S/2kT)$. As $N \gg n_S$, we can approximate $(N - n_S)$ by N, finally giving

$$n_S \approx N \exp\left(\frac{-\Delta H_S}{2kT}\right) \tag{5.1}$$

If we express this equation in molar quantities, it becomes

$$n_S \approx N \exp\left(\frac{-\Delta H_S}{2RT}\right) \tag{5.4}$$

where now ΔH_S is the enthalpy required to form one mole of Schottky defects and R is the universal gas constant, 8.314 J K mol^{-1}; the units of ΔH_S are in joules per mole.

By a similar analysis, we find that the number of Frenkel defects present in a crystal MX is given by the expression:

$$n_F \approx \left(NN_i\right)^{1/2} \exp\left(\frac{-\Delta H_F}{2kT}\right) \tag{5.5}$$

where n_F is the number of Frenkel defects per unit volume, N is the number of lattice sites and N_i is the number of interstitial sites available. ΔH_F is the enthalpy of formation of one Frenkel defect. If ΔH_F is the enthalpy of formation of 1 mol of Frenkel defects, the expression becomes

$$n_F \approx \left(NN_i\right)^{1/2} \exp\left(\frac{-\Delta H_F}{2RT}\right) \tag{5.6}$$

Note that in these derivations, we have ignored the *phonon correction* due to the lattice vibrations. This correction tends to increase the value of n_S and n_F a little, because the presence of vacancies tends to lower the frequency and thus the energy required by the lattice vibrations.

The exponential form of these equations means that n_S and n_F are very temperature-dependent.

Table 5.1 lists some enthalpy of formation values for Schottky and Frenkel defects in various crystals. Using the information in Table 5.1 and Equation 5.1, we can now get an idea of how many defects are present in a crystal. If we assume that ΔH_S has

TABLE 5.1
Formation Enthalpy of Schottky and Frenkel Defects in Some Selected Compounds

Compound	ΔH (10^{-19} J)	ΔH (eV)[a]
Schottky Defects		
MgO	10.57	6.60
CaO	9.77	6.10
LiF	3.75	2.34
LiCl	3.40	2.12
LiBr	2.88	1.80
LiI	2.08	1.30
NaCl	3.69	2.30
KCl	3.62	2.26
Frenkel Defects		
UO_2	5.45	3.40
ZrO_2	6.57	4.10
CaF_2	4.49	2.80
SrF_2	1.12	0.70
AgCl	2.56	1.60
AgBr	1.92	1.20
β-AgI	1.12	0.70

[a] The literature often quotes values in electron volts (eV), so these are included for comparison; 1 eV = 1.60219×10^{-19} J.

a middle-of-the-range value of 5×10^{-19} J and substitute in Equation 5.1, we find that the proportion of vacant sites n_s/N at 300 K is 6.12×10^{-27}. This illustrates how few Schottky defects are present at room temperature. At $T = 1000$ K, this rises to 1.37×10^{-8}, a huge increase of a factor of ~10^{-19}, but still only one or two vacancies per hundred million sites.

Whether Schottky or Frenkel defects are found in a crystal depends mainly on the value of ΔH, the defect with the lower ΔH value predominating. In some crystals, it is possible for *both* types of defects to be present.

In a later section, we shall see that in order to change the properties of crystals, particularly their ionic conductivity, we may wish to introduce more defects into the crystal. It is important, therefore, at this stage, to consider how this might be done.

First, we have seen from the calculation above that raising the temperature introduces more defects. We would have expected this to happen because defect formation is an endothermic process and **Le Chatelier's principle** tells us that increasing the temperature of an endothermic reaction will favour the products—in this case, defects. Second, if it were possible to decrease the enthalpy of formation of a defect, ΔH_S or ΔH_F, this would also increase the proportion of defects present. A simple calculation as we did before, again using Equation 5.1, but now with a lower value for ΔH_S, say 1×10^{-19} J, allows us to see this; Table 5.2 compares the results. The effect on the

TABLE 5.2
Values of n_S/N

T/K	$\Delta H_S = 5 \times 10^{-19}$ J	$\Delta H_S = 1 \times 10^{-19}$ J
300	6.12×10^{-27}	5.72×10^{-6}
1000	1.37×10^{-8}	2.67×10^{-2}

numbers of defects is dramatic—at 1000 K, there are now *approximately three defects for every hundred sites*. It is difficult to see how the value of ΔH could be manipulated within a crystal, but we do find crystals where the value of ΔH is lower than usual due to the nature of the structure, and this can be exploited. This is true for one of the systems that we shall look at in detail later, that of α-AgI. Third, if we introduce impurities selectively into a crystal, we can increase the defect population.

5.2.3 Extrinsic Defects

We can introduce vacancies into a crystal by **doping** it with a selected impurity: for instance, if we add $CaCl_2$ to an NaCl crystal, in order to preserve electrical neutrality, each Ca^{2+} ion replaces *two* Na^+ ions, so *one* cation vacancy is created. Such created vacancies are known as **extrinsic**. An important example that we will meet later in the chapter is that of **zirconia (ZrO_2)**: this structure can be stabilised by doping with CaO, where the Ca^{2+} ions replace the Zr(IV) atoms in the lattice. The charge compensation here is achieved by the production of anion vacancies on the oxide sublattice.

5.3 IONIC CONDUCTIVITY IN SOLIDS

One of the most important aspects of point defects is that they make it possible for atoms or ions to move through the structure. If a crystal structure were perfect, it would be difficult to envisage how the movement of either atoms by **diffusion** through the lattice or ions through **ionic conductivity**, that is, ion transport under the influence of an external electric field, could take place. Setting up equations to describe either diffusion or conductivity in solids is a very similar process, therefore we have chosen to concentrate here on conductivity, because many of the examples later in this chapter are of solid electrolytes.

Two possible mechanisms for the movement of ions through a lattice are sketched in Figure 5.4. In Figure 5.4a, an ion hops or jumps from its normal position on the lattice to a neighbouring equivalent but vacant site. This is called the **vacancy mechanism** (it can equally well be described as the movement of a vacancy rather than the movement of the ion). Figure 5.4b shows the **interstitial mechanism**, where an interstitial ion jumps or hops to an adjacent equivalent site. These simple pictures of movement in an ionic lattice are known as the **hopping model**, and they ignore the more complicated cooperative motions.

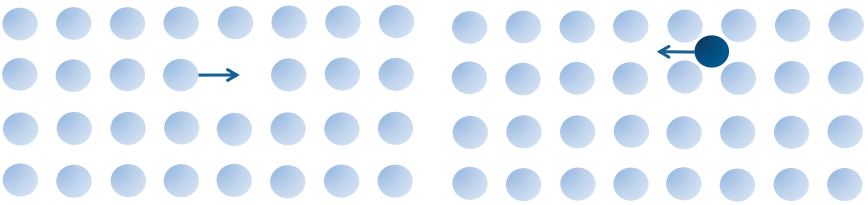


FIGURE 5.4 A schematic representation of ionic motion by (a) a vacancy mechanism and (b) an interstitial mechanism.

Ionic conductivity (σ) is defined in the same way as electronic conductivity:

$$\sigma = nZe\mu \tag{5.7}$$

where n is the number of charge carriers per unit volume, Ze is their charge (expressed as a multiple of the charge on an electron, $e = 1.602189 \times 10^{-19}$ C), and μ is their **mobility**, a measure of the drift velocity in a constant electric field. Table 5.3 shows the sort of conductivity values one might expect to find for different materials. As we might expect, ionic crystals, although they do conduct, are poor conductors compared with metals. This is a direct reflection of the difficulty that the charge carrier (in this case an ion, although sometimes an electron) has in moving through the crystal lattice.

Equation 5.7 is a general equation defining conductivity in all conducting materials. In order to understand why some ionic solids conduct better than others, it is useful to look at the definition more closely in terms of the hopping model that we have set up. First, we have said that an electric current is carried in an ionic solid by the defects. In the cases of crystals, where the ionic conductivity is carried by the vacancy or interstitial mechanism, n, the concentration of charge carriers, will be closely related to the concentration of defects in the crystal, n_S or n_F; thus, μ will refer to the mobility of these defects in such cases.

TABLE 5.3
Typical Values of Electrical Conductivity

Material	Conductivity (S m^{-1})
Ionic Conductors	
Ionic crystals	$<10^{-16}$–10^{-2}
Solid electrolytes	10^{-1}–10^{3}
Strong (liquid) electrolytes	10^{-1}–10^{3}
Electronic Conductors	
Metals	10^{3}–10^{7}
Semiconductors	10^{-3}–10^{4}
Insulators	$<10^{-10}$

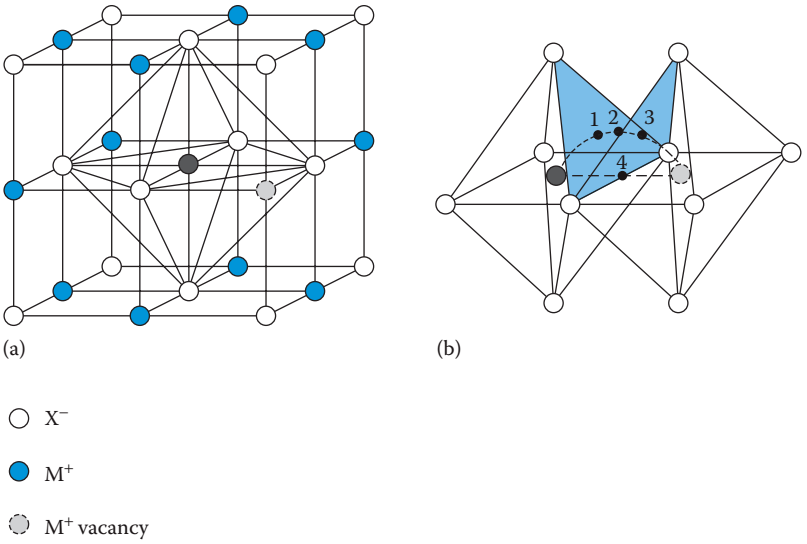


FIGURE 5.5 A sodium chloride-type structure, showing (a) a coordination octahedron of the central cation and (b) a coordination octahedra of the central cation and an adjacent vacancy.

Let us look more closely at the mobility of the defects. Take the case of NaCl, which contains Schottky defects. The Na^+ ions are the smallest and therefore move most easily; however, they also meet quite a lot of resistance, as shown in Figure 5.5. We have used dotted lines to illustrate two possible routes that the Na^+ could take from the centre of the unit cell to an adjacent vacant site. The direct route (labelled **4**) is clearly going to be very unlikely as it leads directly between two Cl^- ions, which, in a close-packed structure such as this, are going to be very close together. The other pathway first passes through one of the triangular faces of the octahedron (point **1**), then through one of the tetrahedral holes (point **2**) and finally through another triangular face (point **3**), before arriving at the vacant octahedral site. The coordination of the Na^+ ion taking this path changes from $6 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 6$ as it jumps from one site to the other. While there is clearly going to be an energy barrier to this happening, it will not be as large as for the direct path where the Na^+ ion becomes two-coordinate. In general, we would expect the ion to follow the lowest energy path available: a schematic diagram of the energy changes involved in such a pathway is shown in Figure 5.6; notice that the energy of the ion is the same at the beginning and the end of the jump. The energy required to make the jump (E_a) is known as the **activation energy**. The temperature dependence of the mobility of the ions can be expressed by an **Arrhenius equation**:

$$\mu \propto \exp\left(\frac{-E_a}{kT}\right) \tag{5.8}$$

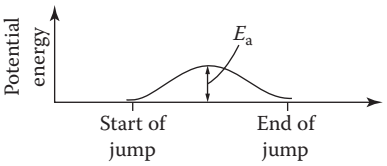


FIGURE 5.6 A schematic representation of the change in energy during the motion of an ion along the lowest energy path.

or

$$\mu = \mu_0 \exp\left(\frac{-E_a}{kT}\right) \tag{5.9}$$

where μ_0 is a proportionality constant, known as a pre-exponential factor, which depends on several factors: the number of times per second that the ion attempts the move (v), called the **attempt frequency** (this is the frequency of the lattice vibration, of the order of 10^{12}–10^{13} Hz); the distance moved by the ion; and the size of the external field. If the external field is small (up to about 300 V cm^{-1}), a temperature dependence of $1/T$ is introduced into the pre-exponential factor.

If we combine all this information in Equation 5.7, we arrive at an expression for the variation of the ionic conductivity with temperature that has the form:

$$\sigma = \frac{\sigma_0}{T} \exp\left(\frac{-E_a}{T}\right) \tag{5.10}$$

The term σ_0 now contains n and Ze, as well as the information on the attempt frequency and the jump distance. This expression accounts for the fact that ionic conductivity *increases* with temperature. If we now take logs of Equation 5.10, we get

$$\ln \sigma T = \ln \sigma_0 \div \left(\frac{E_a}{T}\right)$$

Plotting $\ln \sigma T$ against $1/T$ should produce a straight line with a slope of $-E_a$. The expression in Equation 5.10 is sometimes plotted empirically as

$$\sigma = \sigma_0 \exp\left(\frac{-E_a}{T}\right)$$

because plotting either $\ln \sigma T$ or $\ln \sigma$ makes little difference to the slope; both types of plot are found in the literature. The results of doing this for several compounds are shown in Figure 5.7.

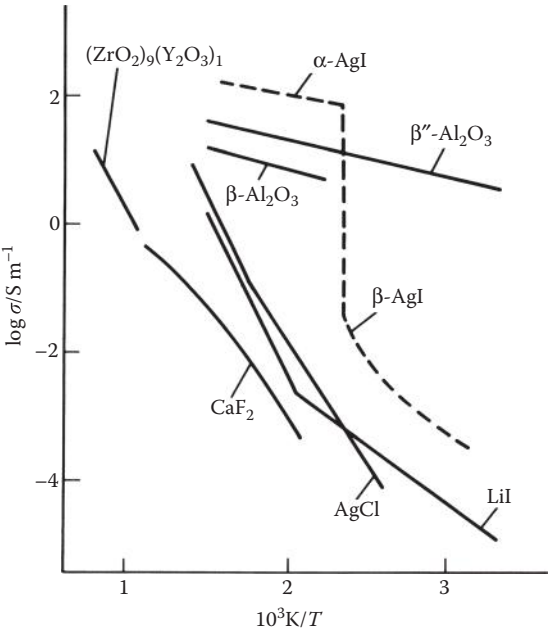


FIGURE 5.7 The conductivities of selected solid electrolytes over a range of temperatures.

Ignore the AgI line for the moment, which we will discuss in the next section. The other lines on the plot are straight lines apart from the one for LiI where we can clearly see two lines of differing slope. So, it looks as though the model we have set up does describe the behaviour of many systems. But how about LiI? In fact, other crystals also show this kink in the plot (some experimental data is shown for NaCl in Figure 5.8, where it can also be seen). Is it possible to explain this using the equations that we have just set up?

The explanation for the two slopes in the plot lies in the fact that even a very pure crystal of LiI contains some impurities and the line corresponding to low temperatures (on the right of the plot) is due to the extrinsic vacancies: at low temperatures, the concentration of intrinsic vacancies is so small that it can be ignored because it is dominated by the defects created by the impurity. For a particular amount of impurity, the number of vacancies present will be essentially constant. In this **extrinsic region**, μ thus depends only on the cation mobility due to these extrinsic defects, whose temperature dependence is given by

$$\mu = \mu_0 \exp\left(\frac{-E_a}{kT}\right) \tag{5.9}$$

However, at the higher temperatures on the left-hand side of the graph, the concentration of the intrinsic defect increases to such an extent that it is now similar to or greater than the concentration of the extrinsic defects. The concentration of the

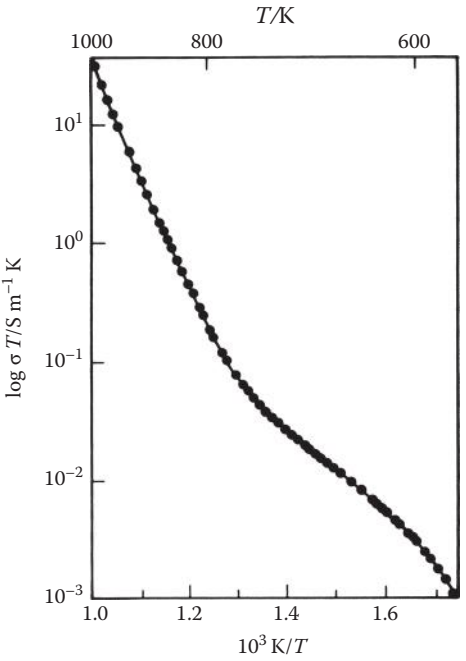


FIGURE 5.8 The ionic conductivity of NaCl plotted against reciprocal temperature.

intrinsic defects, unlike that of the extrinsic defects, will *not* be constant; indeed, it will vary according to Equation 5.1, as we showed earlier:

$$n_S \approx N \exp\left(\frac{-\Delta H_S}{2kT}\right) \tag{5.1}$$

so the conductivity in this **intrinsic region** on the left-hand side of the plot is given by

$$\sigma = \frac{\sigma'}{T} \exp\left(\frac{-E_a}{kT}\right) \exp\left(\frac{-\Delta H_S}{2kT}\right) \tag{5.11}$$

A plot of $\ln \sigma T$ versus $1/T$ in this case will give a greater value for the activation energy, which we label E_S because it will actually depend on two terms, the activation energy for the cation jump (E_a) and the enthalpy of the formation of a Schottky defect:

$$E_S = E_a + \tfrac{1}{2}\Delta H_S \tag{5.12}$$

Similarly, for a system with Frenkel defects:

$$E_F = E_a + \tfrac{1}{2}\Delta H_F \tag{5.13}$$

From plots such as these, we find that the activation energies lie in the range 0.05–1.1 eV, rather lower than the enthalpies of the defect formation. As we have seen, raising the temperature increases the number of defects, thereby increasing the conductivity of a solid. But better than increasing the temperature to increase the conductivity is to find materials that have low activation energies, less than about 0.2 eV. We find such materials in the top right-hand corner of Figure 5.7.

5.4 SOLID ELECTROLYTES

Much of the research in solid-state chemistry is related to the ionic conductivity properties of solids, and new electrochemical cells and devices are constantly being developed that contain solid, rather than liquid, electrolytes. Solid-state batteries are potentially useful because they can perform over a wide temperature range, they have a long shelf life, it is possible to make them very small and they are spill-proof. We use batteries all the time—to start cars, in toys, in watches, in cardiac pacemakers and for electric vehicles. Increasingly, we need lightweight, small, but powerful batteries for a variety of uses, such as computer memory chips, laptop computers and mobile phones and devices. Once a **primary battery** has discharged, the reaction cannot be reversed and it has to be thrown away, so there is also interest in solid electrolytes in the production of **secondary** or **storage batteries**, which are reversible because once the chemical reaction has taken place, the reactant concentrations can be restored by reversing the cell reaction using an external source of electricity. If storage batteries can be made to give sufficient power and they are not too heavy, they become useful as alternative fuel sources, to power cars for instance. Whenever possible, batteries need to be made out of nontoxic, recyclable materials. All these goals give plenty of scope for research. Chemical sensors, electrochromic devices and fuel cells that depend on the conducting properties of solids are also being developed and we look at examples of all of these later in this chapter.

5.4.1 Fast-Ion Conductors: Silver Ion Conductors

Some ionic solids have been discovered that have a much higher conductivity than is typical for such compounds and these are variously known as **fast-ion conductors**, **superionic conductors** and **solid electrolytes**. One of the earliest to be noticed, in 1913, by Tubandt and Lorenz, was a high-temperature phase of **silver iodide**.

Below 147°C, there are two phases of AgI: γ-AgI, which has the zinc blende structure, and β-AgI with the wurtzite structure; both are based on a close-packed array of iodide ions with half of the tetrahedral holes filled. However, above 147°C, a new phase, α-AgI, is observed where the iodide ions now have a body-centred cubic (*bcc*) lattice. If we look at Figure 5.7, we can see that a dramatic increase in conductivity is observed for this phase: the conductivity of α-AgI is very high, 131 S m^{-1}, a factor of 10^4 higher than that of β- or γ-AgI and comparable with the conductivity of the best conducting liquid electrolytes. How can we explain this startling phenomenon?

The explanation lies in the crystal structure of α-AgI. The structure is based on a *bcc* array of I^- ions, as shown in Figure 5.9a. Each I^- ion in the array is surrounded by eight equidistant I^- ions. In order to see where the Ag^+ ions fit into the structure,

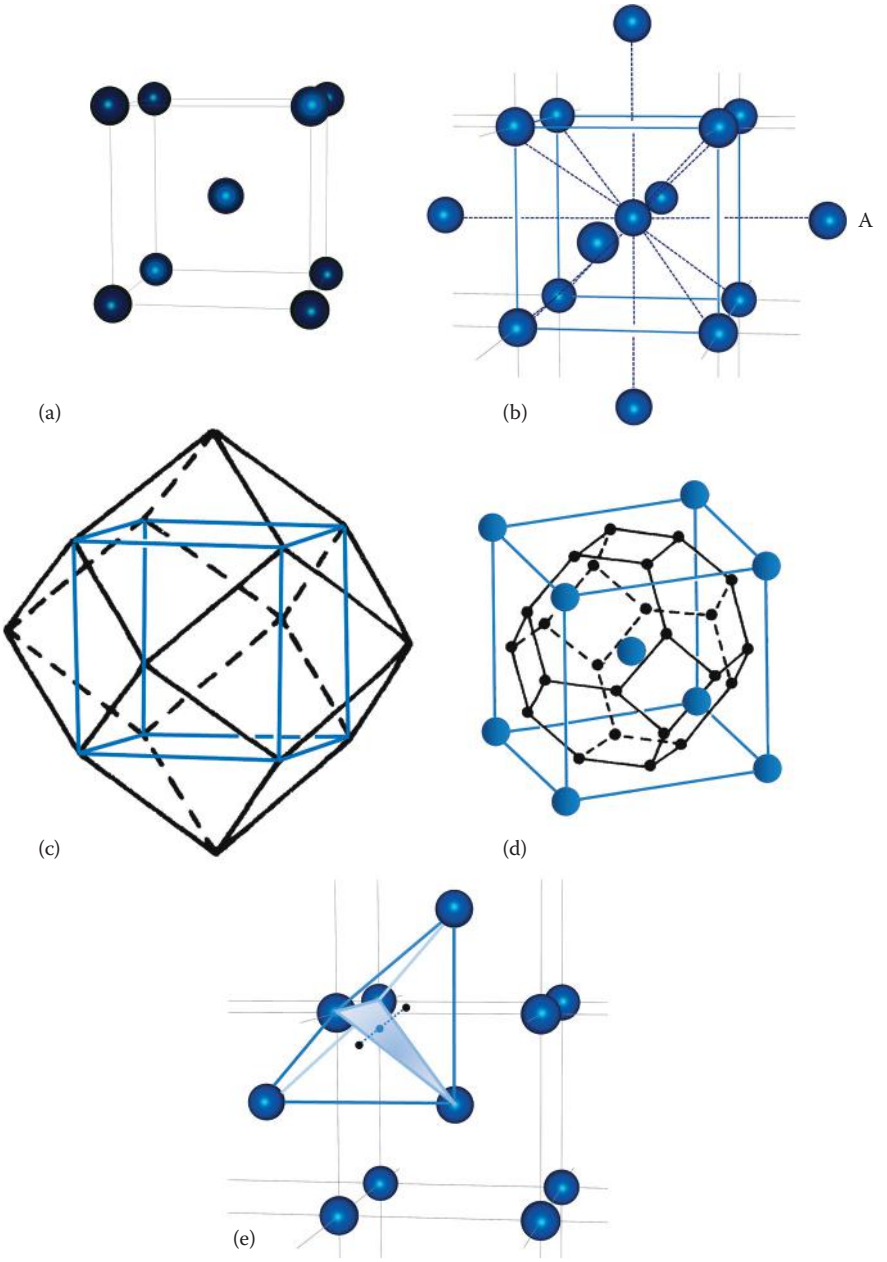


FIGURE 5.9 Building up the structure of AgI. (a) The body-centred cubic array of I^- ions. (b) The *bcc* array extended to the next-nearest neighbours. (c) A rhombic dodecahedron. (d) The *bcc* array with enclosed truncated octahedron—the 24 vertices (black dots) lie at the centres of the distorted tetrahedra. (Because these 24 sites lie on the unit cell faces, they are each shared between two unit cells, making 12 such sites per unit cell on average.) (e) The positions of two tetrahedral sites and a trigonal site between them.

we need to look at the *bcc* structure in a little more detail. Figure 5.9b shows the same unit cell but with the next-nearest neighbours added in, these are six at the body-centres of the surrounding unit cells and are only 15% further away than the eight immediate neighbours. This means that the atom marked **A** is effectively surrounded by *14* other identical atoms, although not in a completely regular way: these 14 atoms lie at the vertices of a **rhombic dodecahedron** (Figure 5.9c). It can also be convenient to describe the structure in terms of the space-filling **truncated octahedron**, shown in Figure 5.9d, which has six square faces and eight hexagonal faces corresponding to the two sets of neighbours. This is called the **domain** of an atom. Each vertex of the domain lies at the centre of an interstice (like the tetrahedral and octahedral holes found in the close-packed structures), which in this structure is a distorted tetrahedron. Figure 5.9e shows two such adjacent distorted tetrahedral holes, and we can see that each of these has a 'tetrahedral' site lying on the face of the unit cell. Where the two tetrahedra join, they share a common triangular face—and there is a trigonal site at the centre of this face, which is also marked in Figure 5.9e. A third type of site at the centre of each face and at each edge of the unit cell can also be defined: these sites have distorted octahedral coordination. The structure thus possesses the unusual feature that there are a great variety of positions that can be adopted by the Ag^+ ions, and Figure 5.10 sketches many of them. Each AgI unit cell possesses two I^- ions ($[8 \times \frac{1}{8}]$ at the corners and one at the body centre), so only two positions for the counterbalancing Ag^+ ions need to be available. From the possible sites that we have described, there are 6 distorted octahedral sites, 12 'tetrahedral' sites and 24 trigonal sites—42 possible sites—the Ag^+ ions have a huge choice of positions open to them. The structure determinations indicate that the Ag^+ are statistically distributed among the 12 tetrahedral sites, all of which have the same energy; therefore, counting these sites only, we find that there are *five* spare sites per Ag^+ ion. We can visualise the Ag^+ ions moving from tetrahedral site to tetrahedral site by jumping through the vacant trigonal site, following the paths marked with solid lines in Figure 5.10, continually creating and destroying the Frenkel defects and able to move easily through the lattice. The paths marked with thin dotted lines in Figure 5.10 require higher energy, as they pass through the vacant 'octahedral' sites that are more crowded. The jump that we have described from tetrahedral → trigonal → tetrahedral only changes the coordination number from 4 → 3 → 4 and, experimentally, the activation energy is found to be very low, 0.05 eV. This easy movement of the Ag^+ ions through the lattice has often been described as a **molten sublattice** of Ag^+ ions.

The very high conductivity of α-AgI seems to arise because of a conjunction of favourable factors, and we can list the features that have contributed to this:

- The charge on the ions is low, the mobile Ag^+ ions are monovalent.
- There are a large number of vacant sites for the cations to move into.
- The structure has an open framework with pathways that the ions can move through.
- The coordination around the ions is also low, so that when they jump from one site to another, the coordination only changes a little, affording a route through the lattice with a low activation energy.

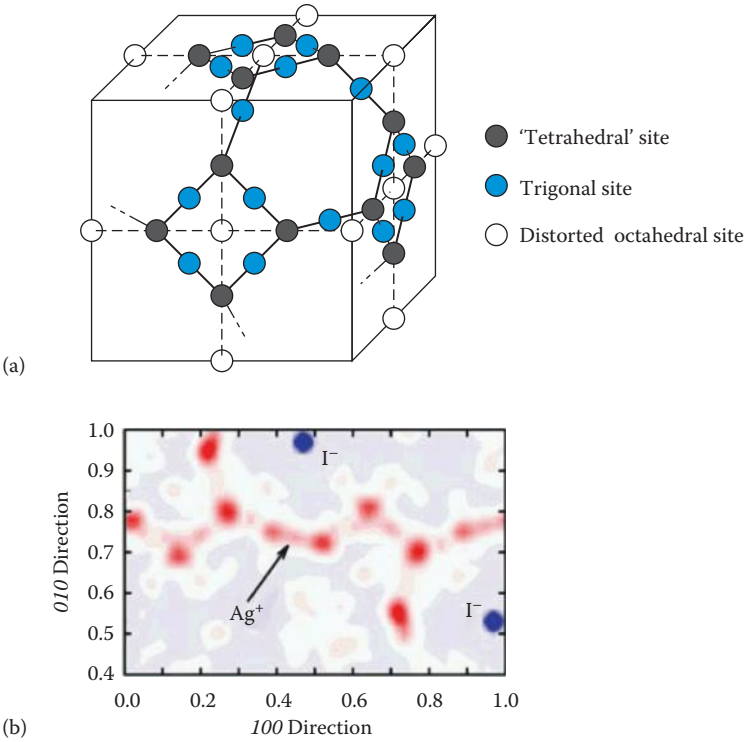


FIGURE 5.10 (a) Possible cation sites in the *bcc* structure of α-AgI. The thick solid and dashed lines mark possible diffusion paths. (b) The crystal structure of $RbAg_4I_5$ highlighting the time-averaged distribution of Ag^+ density in the ionic conductivity channels. (Reprinted from *Journal of Solid State Chemistry*, **165**, Hull, S., et al., Crystal structures and ionic conductivities of ternary derivatives of the silver and copper monohalides, 363–371, Copyright 2002, with permission from Elsevier.)

- The anions are rather **polarisable**; this means that the electron cloud surrounding an anion is easily distorted, making the passage of a cation past an anion rather easier.

These properties are important when looking for other fast-ion conductors.

The special electrical properties of α-AgI inevitably led to a search for other solids exhibiting high ionic conductivity, preferably at temperatures lower than 147°C. The partial replacement of Ag by Rb forms the compound $\mathbf{RbAg_4I_5}$. This compound has an ionic conductivity at room temperature of 25 S m^{-1}, with an activation energy of only 0.07 eV. The crystal structure is different from that of α-AgI, but similarly the Rb^+ and I^- ions form a rigid array while the Ag^+ ions are randomly distributed over a network of tetrahedral sites through which they can move (Figure 5.10b).

If a conducting ionic solid is to be useful as a solid electrolyte in a battery, not only must it possess a high conductivity, but also must have *negligible electronic conductivity*. This is to stop the battery short-circuiting: the electrons must only pass

TABLE 5.4
α-AgI-Related Ionic Conductors

Anion Structure			
bcc	***ccp***	***hcp***	**Other**
α-AgI	α-CuI	β-CuBr	$RbAg_4I_5$
α-CuBr	$\alpha\text{-}Ag_2Te$		
$\alpha\text{-}Ag_2S$	$\alpha\text{-}Cu_2Se$		
$\alpha\text{-}Ag_2Se$	$\alpha\text{-}Ag_2HgI_4$		

through the external circuit, where they can be harnessed for work. $RbAg_4I_5$ has been used as the solid electrolyte in batteries with electrodes made of Ag and rubidium triiodide ($Rb[I_3]$). Such cells operate over a wide temperature range (from −55°C to +200°C), have a long shelf-life and can withstand mechanical shock.

A table of ionic conductors that behave in a similar way to α-AgI is given in Table 5.4. Some of these structures are based on a close-packed array of anions and this is noted in the table; the conducting mechanism in these compounds is similar to that in α-AgI. The chalcogenide structures, such as silver sulfide and selenide, tend to show electronic conductivity as well as ionic conductivity, although this can be quite useful in an electrode material as opposed to an electrolyte.

5.4.2 Fast-Ion Conductors: Oxygen Ion Conductors

Technologically, oxide ion conductors have become very important. They are used in

- Oxygen sensors such as in the exhausts of cars to monitor O_2 partial pressures and ensure clean burning.
- Membranes for oxygen separation.
- Solid oxide fuel cells (SOFCs), that produce electricity cleanly and efficiently from the reaction of fuels such as hydrogen and air, which use solid electrolytes to allow the transport of O^{2-}.

The fluorite structure, as we know from Figure 5.3, has plenty of empty space, which allows an F^- ion to move into an interstitial site. If the activation energy for this process is sufficiently low, we might expect the compounds with this structure to show ionic conductivity. Indeed PbF_2, which has this structure, has a low ionic conductivity at room temperature, which increases smoothly with temperature to a limiting value of ~500 S m^{-1} at 500°C. Uranium, thorium and cerium readily form oxides with the fluorite structure, UO_2, ThO_2 and CeO_2, respectively, but zirconium is different: the cubic (fluorite) form of ZrO_2 only forms at high temperatures or when doped with another element. ZrO_2 is found as the mineral baddeleyite, which has a monoclinic structure. Above 1000°C, this changes to a tetragonal form and at high temperatures ZrO_2 adopts the cubic fluorite structure.

At the turn of the last century, Nernst found that mixed oxides of yttria (Y_2O_3) and zirconia (ZrO_2) glowed white hot if an electric current was passed, and he attributed

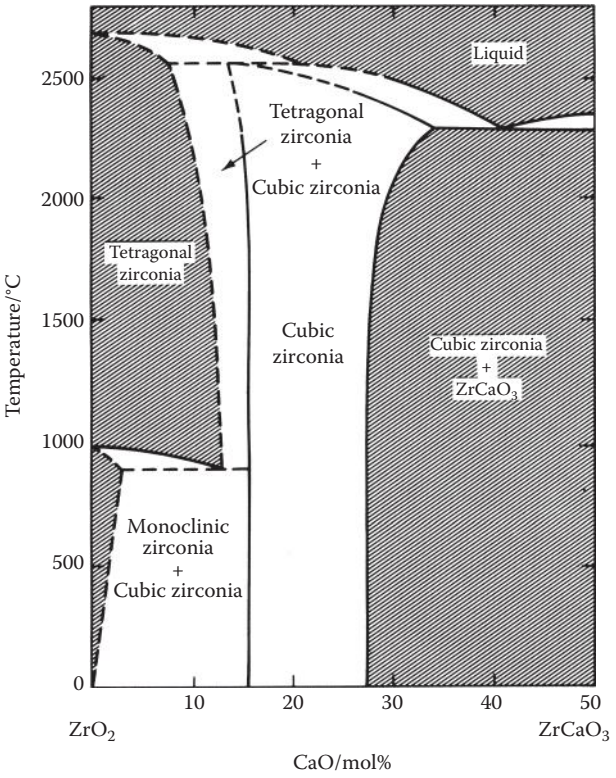


FIGURE 5.11 A phase diagram of the pseudobinary CaO–ZrO_2 system. The cubic calcia-stabilised zirconia phase occupies the central band in the diagram and is stable till about 2400°C.

this to the conduction of oxide ions. Nernst used this doped zirconia for the filaments in his 'glower' electric lights, which started to replace candles, oil and gas lamps, making his fortune before the arrival of tungsten filament bulbs. It was found that the cubic fluorite form of ZrO_2 can be stabilised at room temperature by the addition of other oxides such as yttria (Y_2O_3), forming a compound known as **yttria-stabilised zirconia (YSZ)**. CaO also forms a new phase with ZrO_2 after heating at about 1600°C; the phase diagram in Figure 5.11 shows that at about 15 mol% CaO, the new phase, known as **calcia-stabilised zirconia**, appears alone and persists until about 28 mol% CaO. Above 28 mol%, another new phase appears, finally appearing pure as $CaZrO_3$. If the Ca^{2+} ions occupy the Zr^{4+} sites, compensating vacancies are created in the O^{2-} sublattice: thus, for every Ca^{2+} ion taken into the structure, *one* anion vacancy is created. Consequently, these stabilised zirconias are exceptionally good fast-ion conductors of O^{2-} anions (Figure 5.7).

Ionic conductivity in solids maximises at relatively low concentrations of the dopants; not at the 50% one might expect; this is because there is an elastic interaction between the substituted ion and the vacancy created. The best conductivity seems to be achieved when the crystal lattice is distorted as little as possible, that is, when

the dopant ion is similar in size to the cation it is replacing. Consequently, two of the best oxygen-ion conductors are scandia-stabilised zirconia doped with Sc_2O_3 (ScSZ) and ceria doped with gadolinia (Gd_2O_3; GDC); these solid electrolytes are, however, rather expensive. Many other materials are made with this type of structure based on oxides such as CeO_2, ThO_2 and HfO_2 as well as ZrO_2, and doped with rare earth or alkaline earth oxides: these are collectively known as stabilised zirconias and are widely used in electrochemical systems.

YSZ together with ScSZ and GDC are the most commonly used materials for the solid electrolyte in **solid oxide fuel cells**. Another interesting application of stabilised zirconia is in the detection of oxygen, where it is used in both **oxygen meters** and **oxygen sensors**, which are based on a specialised electrochemical cell (see Section 5.5.3).

The **perovskite structure, ABX_3** (Chapter 1, Figure 1.44), has two different metal sites that could be substituted with lower valence metal cations, leading to oxygen vacancies. Materials based on lanthanum gallate ($LaGaO_3$) have been successfully doped with strontium and magnesium to produce $La_{1-x}Sr_xGa_{1-y}Mg_yO_{3-\delta}$ (LSGM) with similar conductivities to the stabilised zirconias, but at lower temperatures. Using them in oxide-ion conducting devices, such as fuel cells, has the advantage of bringing down the operating temperature.

For some applications, such as the cathode materials in SOFCs (Section 5.4.4), a material is needed that can conduct both ions and electrons. The strontium-doped perovskites, such as $LaMnO_3$ (LSM) and $LaCrO_3$ (LSC), have both these properties.

The development of other oxide conductors continues to take place.

- The LAMOX family of oxide conductors, based on $La_2Mo_2O_9$, have high conductivity above 600°C, but tend to be susceptible to reduction by hydrogen.
- The BIMEVOX family of oxide conductors, based on doped bismuth vanadate $Bi_4V_2O_{11}$, have high conductivity above 600°C.
- The apatite structures, $La_{10-x}M_6O_{26+y}$ (M = Si or Ge), conduct well at very high temperatures.

5.4.3 Fast-Ion Conductors: Sodium Ion Conductors

β-Alumina is the name given to a series of compounds that show fast-ion conducting properties. The parent compound is sodium β-alumina ($Na_2O{\cdot}11Al_2O_3$ [$NaAl_{11}O_{17}$]), and it is found as a by-product of the glass industry. (The compound was originally thought to be a polymorph of Al_2O_3 and was named as such; it was only later found to contain sodium ions, but the original name has stuck.) The general formula for the series is $M_2O{\cdot}nX_2O_3$, where n can range from 5 to 11; M is a monovalent cation, such as (alkali metal)$^+$, Cu^+, Ag^+ or NH_4^+; and X is a trivalent cation $A1^{3+}$, Ga^{3+} or Fe^{3+}. The real composition of β-alumina actually varies quite considerably from the ideal formula and the materials are always found to be rich in Na^+ and O^{2-} ions to a greater or lesser extent.

The interest in these compounds started in 1966 when research at the Ford Motor Co. showed that the Na^+ ions were very mobile both at room temperature and above.

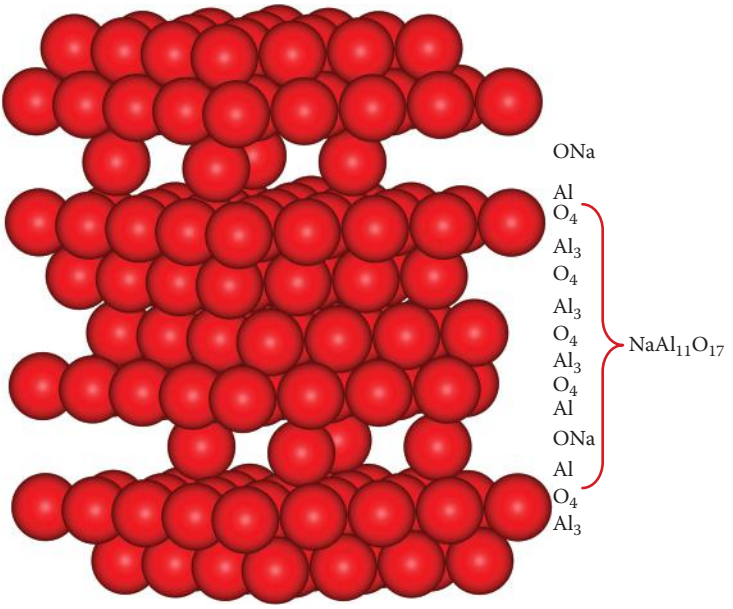


FIGURE 5.12 The oxide layers in β-alumina with the ratio of atoms in each layer of the structure.

The high conductivity of the Na^+ ions in β-alumina is due to the crystal structure. This can be thought of as close-packed layers of oxide ions, but in every fifth layer, three-quarters of the oxygens are missing (Figure 5.12). The four close-packed layers contain the Al^{3+} ions in both octahedral and tetrahedral holes (they are known as the 'spinel blocks' because of their similarity to the crystal structure of the mineral spinel $MgAl_2O_4$—discussed in Chapter 1 and illustrated in Figure 1.43). The groups of four close-packed oxide layers are held apart by a rigid Al–O–Al linkage; this O atom constituting the fifth oxide layer contains only a quarter of the number of oxygens of each of the other layers. The Na^+ ions are found in these fifth oxide layers, which are mirror planes in the structure. The overall stoichiometry of the structure is also shown layer-by-layer in Figure 5.12: the sequence of layers is

$$B(ABCA)C(ACBA)B\ldots$$

where the brackets enclose the four close-packed layers and the intermediate symbols refer to the fifth oxide layer.

The crystal structure of β-alumina is shown in Figure 5.13. The Na^+ ions can move around easily as there are plenty of vacancies, so there is a choice of sites. Conduction in the β-aluminas only occurs within the planes containing the oxygen vacancies; these are known as the **conduction planes**. The alkali-metal cations cannot penetrate the dense 'spinel blocks', but can move easily from one site to another within the plane. β-Alumina is not found in the stoichiometric form—it is usually Na_2O rich, and the sodium-rich compounds have a much higher conductivity than stoichiometric

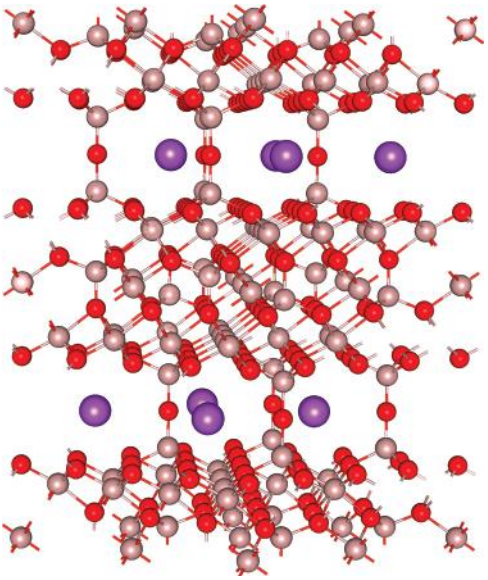


FIGURE 5.13 The structure of stoichiometric β-alumina (Na ● Al ● O ●).

β-alumina. The extra sodium ions have to be compensated by a counter defect in order to keep the overall charge on the compound at zero. There is more than one possibility for this, but in practice it is found that extra oxide anions provide the compensation and the overall formula can be written as $(Na_2O)_{1+x}{\cdot}11Al_2O_3$. The extra sodium and oxide ions both occupy the fifth oxide layer; the O^{2-} ions are locked into position by an Al^{3+} moving out from the spinel block, and the Na^+ ions become part of the mobile pool of ions. The Na^+ ions are so fluid that the ionic conductivity in β-alumina at 300°C is close to that of typical liquid electrolytes at ambient temperature.

β-Aluminas are used as electrolytes in the manufacture of molten salt electrochemical cells, particularly for power supplies.

In 1968, Hagman and Kierkegaard found an exciting new material, sodium zirconium phosphate ($NaZr_2(PO_4)_3$), known as **NZP** (Figure 5.14). The structure consists of corner-linked ZrO_6 octahedra joined by PO_4 tetrahedra, each of which corner-shares to four of the octahedra. This creates a three-dimensional system of channels through the structure containing two types of vacant site: Type I, a single distorted octahedral site (occupied by Na^+ ions in NZP), and three larger Type II sites (vacant in NZP). This has proved to be an incredibly versatile and stable structure with the general formula $A_xM_2((Si,P)O_4)_3$, which is adopted by hundreds of compounds. The balancing cations (A) are usually alkali or alkaline earth metals, and the structural metal(s) (M) is a transition metal such as Ti, Zr, Nb, Cr or Fe. The phosphorus can be substituted by silicon. The most famous member of this family is known as **NASICON** (from Na SuperIonic CONductor). This has proved to be a very good Na^+ fast-ion conductor with a conductivity of 20 S m^{-1} at 300°C. It has the formula $Na_3Zr_2(PO_4)(SiO_4)_2$ and three out of the four vacant sites are occupied by Na^+, allowing a correlated motion as the ions diffuse through the channels.

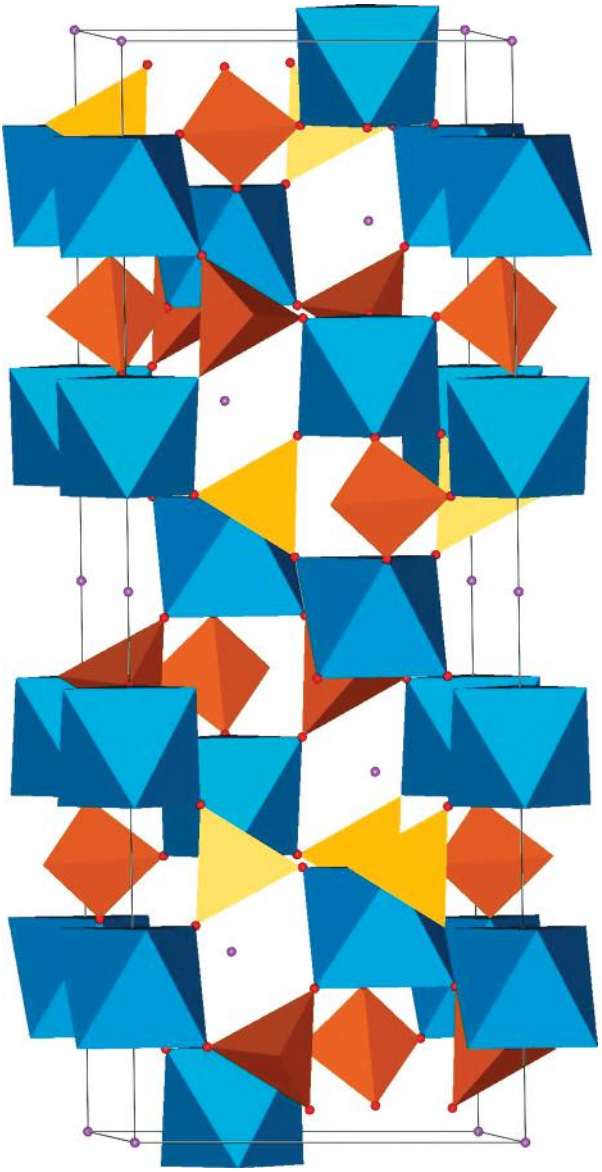


FIGURE 5.14 The structure of NZP (Na Si P O).

5.5 APPLICATIONS OF SOLID ELECTROLYTES

The development of solid-state conducting solids which are on a par with liquid electrolytes has revolutionised the design of batteries and other solid-state devices, and in this section we will explain the operating principles behind some of these devices. Figure 5.15 is a simple schematic diagram that we can use to explain the operation of several

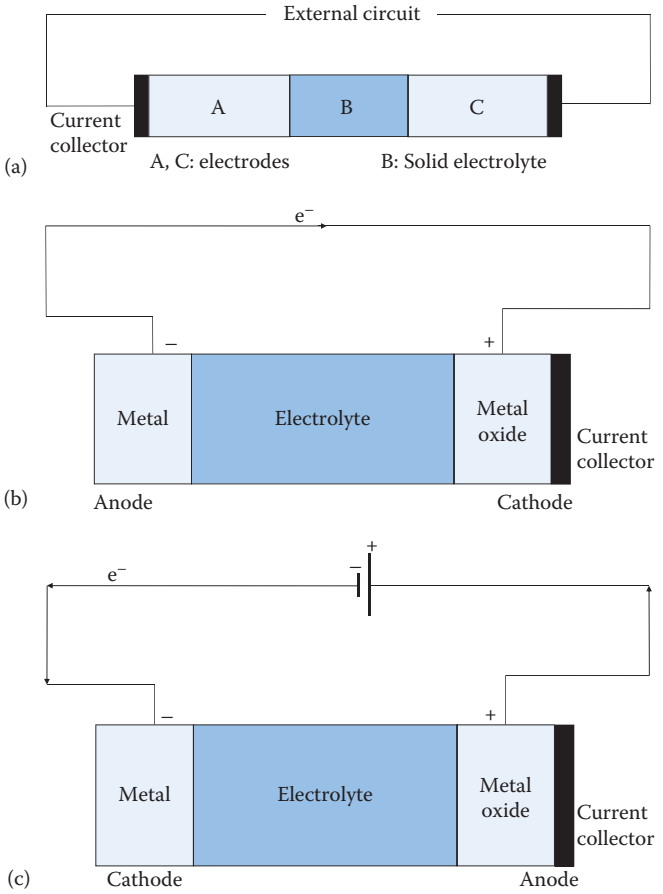


FIGURE 5.15 (a) A diagram of a general electrochemical device using a solid electrolyte. (b) A general electrochemical cell based on a metal and its oxide in the discharge mode. (c) The same cell under recharging.

different types of electrochemical device. The solid electrolyte (B), (which is able to conduct ions but *not* electrons), is sandwiched between two electrodes (A and C), which are connected by an external circuit around which electrons can flow.

5.5.1 Batteries

A **battery** or **electrochemical cell** produces an electric current at a constant voltage as a result of a chemical reaction and consists of two **electrodes** separated by an **electrolyte**. The ions taking part in the reaction pass through the electrolyte and are then either oxidised or reduced at one of the electrodes. The electrode at which oxidation takes place, usually a metal, is called the **anode**, and reduction takes place at the **cathode**; this usually consists of a metal current collector together with an

active component which is in a high oxidation state and can be reduced, such as a metal oxide (Figure 5.15b).

A battery only has a discharge mode and once discharged, the reaction cannot be reversed. However, secondary or storage batteries are reversible, and the reactant concentrations can be restored by reversing the cell reaction using an external source of electricity (Figure 5.15c). In the recharging mode, note that the electrons in the external circuit are now being driven *towards* the negatively charged electrode where reduction is now taking place; the negatively charged electrode becomes the cathode.

To be useful as a solid electrolyte in a battery, the conducting solid must have a high ionic conductivity, but it must be an electronic insulator, so that it can separate the two reactants of the device allowing only ions, and not electrons, to travel through the solid—or the device would short-circuit. When the cell is in the discharge mode, electrons are released at the negatively charged electrode and they travel around the external circuit where they can be harnessed for useful work. The electromotive force (emf) or voltage produced by the cell under standard open-circuit conditions is related to the standard Gibbs free energy change for the reaction by the following equation:

$$\Delta G^{\ominus} = -nE^{\ominus}F \tag{5.14}$$

where n is the number of electrons transferred in the reaction, $E^{\ominus}$ is the standard emf of the cell (the voltage delivered under standard, zero-current conditions) and F is the Faraday constant (96,485 C mol^{-1} or 96,485 J V^{-1}). The energy stored in a battery is a factor of the energy generated by the cell reaction and the amount of materials used; it is usually expressed in watt-hours (Wh; current × voltage × discharge time).

The reason that solid-state batteries are potentially useful is that they can perform over a wide temperature range, they have a long shelf life and it is possible to manufacture them so that they are extremely small. Lightweight rechargeable batteries can now be made to give sufficient power to maintain mobile phones for several days and laptop computers for several hours; they are used for backup power supplies and as fuel sources in electric cars.

Lithium-containing batteries are used in pacemakers, mobile devices, and electric cars. Because of the need for small, lightweight batteries, the energy stored in a battery is not always the most useful merit indicator. More important can be the **energy density** (watt-hours divided by battery volume in litres) or **specific energy** (watt-hours divided by battery weight in kilograms). Light elements such as lithium are obvious candidates for such battery materials.

Although LiI has a fairly low ionic conductivity (see Figure 5.7), the **LiI battery** used in heart pacemaker batteries for implantation was first proposed by Wilson Greatbatch, after the earlier use of mercury cells, and they went into production in the early 1970s (Figure 5.16). These batteries only need to provide a low current, but can be made very small, last a long time, produce no gas so they can be hermetically sealed and, above all, they are very reliable. LiI is the solid electrolyte that separates a lithium electrode from an electrode of iodine embedded in a conducting polymer (poly-2-vinyl-pyridine), which, based on Figure 5.15b, we can depict as

FIGURE 5.16 An early LiI Greatbatch heart pacemaker battery (1972), approximate size 5 × 8 cm, next to the much smaller 1996 model. (From Wilson Greatbatch Technologies, Inc. With permission.)

A		**B**		**C**
Li	//	LiI	//	I_2 and polymer

The electrode reactions are:

***Anode A*:** $2Li(s) = 2Li^+(s) + 2e^-$
***Cathode C*:** $I_2(s) + 2e^- = 2I^-(s)$

Because LiI contains intrinsic Schottky defects, the small Li^+ cations are able to pass through the solid electrolyte, while the released electrons go around an external circuit.

Making reliable rechargeable lithium batteries has proved to be a very difficult problem because the lithium re-deposits in a finely divided state that is very reactive and can even catch fire. Lithium reacts vigorously with water to form lithium hydroxide and liberating hydrogen, so batteries must be sealed absolutely from leakage.

The technology of lithium-containing batteries is constantly being developed; in the 1990s, the Sony Corporation in Japan developed rechargeable **lithium-ion batteries**, which can now be made to recycle reliably many times. Their lightness makes them popular for use in all electronic mobile devices (Figure 5.17a). The driving reaction of these cells is that of Li with a transition metal oxide, such as MnO_2 or CoO_2, to form an intercalation compound, for example Li_xCoO_2. One electrode is made of lithium embedded in graphite, forming an intercalation compound, typically C_6Li; in discharge mode, this easily releases Li^+. The lithium ions travel through a Li^+-containing electrolyte to the CoO_2 cathode, where it intercalates. The cathode can also be made of other materials that intercalate lithium, notably $FePO_4$, NiO_2 and TiS_2. The electrolyte is a nonaqueous solvent, such as ethylene carbonate,

FIGURE 5.17 (a) A laptop computer powered by lightweight lithium-ion rechargeable battery. (From Apple.) (b) The Tesla Roadster has a range of 211 miles and an estimated battery life of up to 7 years. (From Tesla Motors.) (c) The NaS battery installed in Presidio, Texas in 2010 can provide up to 4 MW of power for up to 8 h. (Courtesy of Electric Transmission Texas.)

mixed with a lithium complex salt; lithium hexafluorophosphate ($LiPF_6$), lithium tetrafluoroborate ($LiBF_4$) and lithium triflate ($LiCF_3SO_3$), being commonly used. The Li^+ ions 'rock' between the two intercalation compounds and no lithium metal is ever present, eliminating many of the hazards associated with lithium batteries. Advances have increased battery performance by increasing the surface area using electrodes formed from nanostructured materials, such as nanotubes.

In the discharge mode, we can write the cell reaction as

A		**B**		**C**
Li/C	//	Li^+ electrolyte	//	CoO_2

***Anode A*:** $Li_xC_6(s) = xLi^+ + 6C + xe^-$
***Cathode C*:** $xLi^+ + CoO_2(s) + xe^- = Li_xCoO_2(s)$

Once the cell has been discharged, it can be regenerated by passing an electric current that breaks down the complex, and the lithium ions migrate back to the negative electrode to reform the carbon intercalation compound (Figure 5.15c).

The safety of these batteries is ensured by the use of a nonaqueous electrolyte and a carefully sealed container that completely excludes water.

Lithium-ion polymer batteries have evolved from the lithium-ion battery. The cell reaction is the same, but the electrolyte is no longer suspended in a nonaqueous solvent; instead, the Li salt is suspended in a conducting polymer gel such as polyethylene oxide (PEO) (see Chapter 10) or polyacrylonitrile. These batteries can be enclosed in a flexible foil casing that makes them very light, easily shaped and very thin. Great care has to be taken with not overcharging them as they can overheat and expand.

Electric cars are using lithium-ion batteries to power them. They have several advantages over internal combustion engine cars in that they are more efficient (of the order of 90% compared with 30%), have zero emissions and, for the economy as a whole, they are less dependent on imported oil. They are currently still expensive compared with internal combustion engine cars due to the cost of the batteries, but many countries are now offering subsidies and tax incentives to buy them. As battery technology improves, the batteries will last longer and become cheaper and their range will become greater (Figure 5.17b).

Sodium-containing batteries are used in electrical storage, and electric cars. Na^+ conduction has been put to good use in a secondary battery that operates at high temperatures—the **sodium sulfur battery**. This system uses either NASICON or β-alumina as the electrolyte. From Equation 5.14, we know that in order to obtain a large voltage from a cell, we must have a cell reaction with a large negative Gibbs free energy change, such as a reaction between an alkali metal and a halogen might give. In terms of specific energy, such a reaction can yield up to 150 Wh kg^{-1} of material because it incorporates a highly energetic reaction between light substances. Such reactive materials have to be separated by an electrolyte that is impermeable to electrons, but can be crossed by ions. The electrolyte separates molten sodium from molten sulfur, and at the sulfur/electrolyte interface a complex reaction forming polysulfides of sodium takes place. The heat required to maintain the cell at the operating temperature of 300°C is supplied by the cell reaction itself. The cell reaction is given here for one such reaction (Figure 5.15):

A		**B**		**C**
Na(l)	//	β-Alumina	//	S(l) and graphite

***Anode A*:** $2Na(l) = 2Na^+ + 2e^-$
***Cathode C*:** $2Na^+ + 5S(l) + 2e^- = Na_2S_5(l)$
Overall reaction: $2Na(l) + 5S(l) = Na_2S_5(l)$

Later, low polysulfides are formed, and the discharge is terminated at a composition of about Na_2S_3. Despite the complexity of the reactions, the electrode process can be reversed by applying a current from an external source.

These cells have the advantages of high power and low weight, and in many countries, such as Japan and the United States, the components of these cells are abundantly available and relatively cheap. However, they require stringent safety features as they contain highly reactive and corrosive chemicals at 300°C, and this has meant that they are not well suited for their original development purpose, that of use in the electric car, but are mainly used for grid, wind farm and solar energy storage using banks of batteries sealed in a vacuum-insulated box.

Interest has also been directed towards a similar high-temperature system, the **ZEBRA battery** or the Na–$NiCl_2$ battery, which also uses β-alumina as an Na^+ ion conductor. The sulfur electrode is replaced by nickel chloride. The contact between the $NiCl_2$ electrode and the solid electrolyte is poor as they are both solids, and current flow is improved by adding a second liquid electrolyte (molten $NaAlCl_4$) between this electrode and the β-alumina. The overall cell reaction is now:

$$2Na + NiCl_2 = Ni + 2NaCl$$

The high specific energy for this cell of >100 Wh kg^{-1} gives electric vehicles powered by these batteries a range of up to 250 km, sufficient for daily use in a city. These batteries are fully rechargeable, robust and safe, and they have been found to need no maintenance over 100,000 km. They are a keen rival technology to the Li-ion batteries for powering electric vehicles, and they have the advantage of being made from much cheaper and more available raw materials, but the disadvantage of the high operating temperature.

The **sodium-ion battery** is analogous to the lithium-ion battery that uses the interaction with a graphite anode, in this case, to form Na_xC_6. This battery is currently (2011) at a research stage as it does not maintain its capacity after many cycles of charging and recharging. If it can be developed to perform consistently, it would have many advantages over ZEBRA and Na–S batteries as it operates at low temperatures.

5.5.2 Fuel Cells

Fuel cells work on the same principles as a battery (Figure 5.18)—the difference is that instead of the 'fuel' for the reaction being contained in the electrode materials

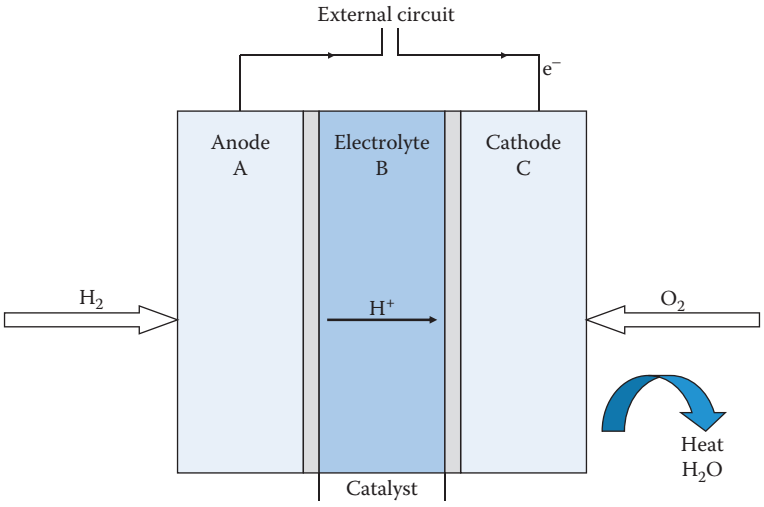


FIGURE 5.18 A schematic diagram of a fuel cell.

as in a battery, now it is fed in externally to the electrodes. This has the huge advantage that the cell can operate continuously as long as the fuel is available, unlike a battery, which either has to be thrown away once the battery material is exhausted (primary) or has to be recharged by plugging into an electricity supply for several hours (secondary/rechargeable) to reverse the cell reaction. In a fuel cell, the fuels used are usually hydrogen and oxygen (air), which react together electrochemically to produce water, electricity and heat. H_2 is fed to the anode where it is oxidised to H^+ ions and electrons. The electrons travel around the external circuit, and the H^+ ions travel through the electrolyte to the cathode, where they react with a supply of O_2. Another great advantage of fuel cells is that the by-products of water and heat are both potentially useful and nonpolluting. The fact that hydrogen and oxygen combine at low temperatures compared with normal combustion means that side-reactions producing NO_x pollution are avoided. Fuel cells operate with an efficiency of upwards of 50%, and this compares with typically 20%–30% for internal combustion engines and 35%–45% for diesel engines.

Currently, fuel cells are being intensively developed as they have the potential to provide power in a relatively nonpolluting fashion. As legislation in many countries, including Europe and the United States, is demanding fewer emissions from vehicles, including the greenhouse gases, there is a great deal of interest in producing the so-called *zero emission vehicles* (although this term is difficult to define and varies from one country to another). The internal combustion engine cannot meet such stringent demands, so alternatives have to be found. The main contenders are electric cars that run on either batteries or fuel cells or a combination of the two, both of which push the pollution problems to the sites of fuel or battery production and away from the roads. Current fuel cell developments include fuel cell-driven buses and cars, hybrid vehicles, and also power sources for homes and factories, and micro fuel cells for mobile phones and laptops (Figure 5.19). 2004 saw the first fuel cell-powered buses in London, and since then many other cities have followed suit. Most car manufacturers now have prototype zero-emission vehicles, and the first

FIGURE 5.19 A zero-emission hybrid fuel cell/Li^+-ion battery-powered bus from UTC Power, operating in Hartford, CT, USA. (From UTC Power. With permission.)

hydrogen fuel cell consumer vehicle—a Nissan 4 × 4 pickup truck—went on sale in the United States in 2005, but the commercial prospects are not good until the price can be made more competitive.

Notwithstanding all the advantages of fuel cells, there are also complications and drawbacks to their use. The reduction of oxygen at the cathode is rather slow at low temperatures, and in order to increase the rate of reaction, an expensive Pt catalyst is incorporated into the carbon electrodes. If the electrolyte allows the passage of H^+ ions, the cell reaction can be written

	A	**B**	**C**	
$H_2(g)$	//Pt/C electrode//	Hydrogen electrolyte	//Pt/C electrode//	$O_2(g)$

***Anode A*:** $H_2 = 2H^+ + 2e^-$
***Cathode C*:** $\frac{1}{2}O_2 + 2H^+ + 2e^- = H_2O$

The theoretical emf for this cell, calculated from the Gibbs function for the decomposition of water, is $E = 1.229$ V at 298 K, but this decreases with temperature to about 1 V at 500 K. Therefore, a compromise in cell design is always needed between the voltage generated and an operating temperature high enough to maintain a fast reaction.

The oxygen gas supply comes simply from the air, but a major drawback is the supply, transportation and storage of the hydrogen gas. The current infrastructure of most fuel stations is for the supply of liquid fuel—petrol (gasoline) and diesel. Hydrogen, a gas at room temperature, is the least dense of all the elements and has to be compressed or liquefied in order to store it in a manageable volume. Compressed hydrogen needs a strong tank that can withstand the high pressures involved and such cylinders are very heavy. Liquefied hydrogen requires cooling and must be kept very cold (the boiling temperature is −253°C), so the tanks have to be extremely well insulated. Research is underway developing materials, such as carbon nanotubes, that will store large amounts of hydrogen and release it when needed. Although hydrogen is probably less dangerous than a liquid hydrocarbon fuel, public perception about its safety needs to be overcome: many people have seen old footage of the hydrogen-filled Hindenburg airship exploding on landing.

Hydrogen also has to be generated; very pure hydrogen, required by some fuel cells, is electrolytically produced, which is an expensive process unless cheap sources of electricity can be used, such as wind turbines, solar energy or hydroelectric power. The alternative is to use a **reforming reaction**, where hydrogen-rich sources such as methane or methanol are reacted with steam to produce hydrogen and carbon dioxide. To avoid the transport and storage problems, hydrogen can also be produced in situ using a reformer and fed directly into the fuel cell. The reforming reaction has its own problems, in that the catalyst is poisoned by sulfur in the fuel; also small quantities of carbon monoxide are produced along with the hydrogen in the reforming reaction, and this also poisons the Pt catalyst in the fuel cell, thereby reducing its efficiency.

Fuel cells have been around for quite a long time; in 1839, William Grove, a Welsh physicist, made the first working fuel cell, but it was not until 1959 that Tom Bacon at Cambridge University produced a stack of 40 alkaline fuel cells (AFCs) that produced 5 kW of power. Around the same time, Willard Grubb and Leonard Niedrach at General Electric developed a conducting membrane fuel cell with a Pt catalyst on a Ti gauze, which was used in the Gemini earth-orbit space programme. By 1965, Pratt and Whitney had improved (longer life) the AFC for use in the Apollo missions, where they provided both power and drinking water for the astronauts. In 1983, Ballard Power (Canada) was established, and in 1993, they unveiled the first fuel cell buses, which went into service in Vancouver and Chicago.

There are several types of fuel cell currently under development, using different electrolyte systems, namely: phosphoric acid (PAFC), alkaline, molten carbonate (MCFC), regenerative, zinc-air, protonic ceramic (PCFC), proton exchange membrane (PEM), direct methanol (DMFC) and solid oxide (SOFC); the last four contain solid electrolytes.

The **solid oxide fuel cells** (SOFC) employ a ceramic oxide (typically scandia-doped or yttria-doped zirconia [M_2O_3/ZrO_2] or gadolinium-doped ceria) electrolyte, which becomes O^{2-} conducting at very high temperatures (800°C–1000°C); this system has the disadvantage of needing energy to heat the cell, but the advantage that at this temperature, reforming and H^+ production can take place internally without the need for expensive Pt catalysts. Because of the long time needed to get the cell up to its operating temperature, it is not useful for powering vehicles, but can be used for power generation in buildings and industry (Figure 5.20a). The cell reaction is

	A	**B**	**C**	
$H_2(g)$	//Electrode//	Solid oxide electrolyte	//Electrode//	$O_2(g)$

***Anode A*:** $H_2 + O^{2-} = H_2O + 2e^-$
***Cathode C*:** $\frac{1}{2}O_2 + 2e^- = O^{2-}$

Unlike the fuel cell illustrated in Figure 5.18, it is now the oxide ions created at cathode C that travel through the electrolyte, forming water at A and releasing two electrons that travel around the external circuit and the heat from the reaction.

It is vital that the solid oxide electrolyte can withstand the extreme conditions of hydrogen at the anode at 800°C or above. Under these conditions, many oxides would be reduced, liberating electrons, and thereby leading to unwanted electronic conductivity.

The cathode materials used have to conduct both oxide ions and electrons satisfactorily, but also for compatibility, have similar thermal expansion coefficients as the electrolyte. The strontium-doped perovskite lanthanum strontium manganate (LSM) (Section 5.4.2) is one of the materials of choice.

(a)

(b)

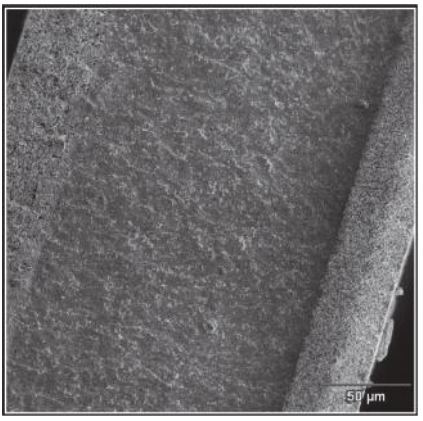


FIGURE 5.20 (a) 1 MW fuel cell. Direct FuelCell® (DFC®) power plant from FuelCell Energy, Inc., installed at California State University, Northridge, CA, USA. (From FuelCell Energy. With permission.) (b) An SEM of an SOFC cross section showing the porosity of the electrodes.

The anode is usually made from Ni doped into YSZ. Figure 5.20b shows the scanning electron micrograph of an SOFC fuel cell with the two porous electrodes enclosing the denser solid electrolyte; the whole structure is only millimetres across. Hundreds of these cells are then connected in a series to form a **stack** and systems are made that can provide outputs from 100 kW to 2 MW.

Research centres around improving performance and reliability, reducing the operating temperature and allowing for the use of different fuels.

The **PEM fuel cell** consists of two catalyst-coated carbon electrodes with a conducting polymer electrolyte in between. The fuel used is pure hydrogen. Usually, the catalysts are promoted Pt or other noble metal. The PEM fuel cells operate at a much lower temperature, ~80°C. The conducting polymer membrane (see Chapter 10) is usually Nafion, a sulfonated fluoropolymer made by Dupont, strengthened by Gore-Tex. The strongly acidic $-SO_2OH$ group allows the passage of H^+ ions, but not of atoms or electrons. Output is typically 1 V at 80°C; with a current flow

of 0.5 A cm^{-2}, this drops to 0.5 V because of ohmic losses. A membrane of 1 m^2 provides about 1 kW. To produce the correct power output, a number of cells are placed together to form a stack. Research centres on reducing the cost of producing the cells, particularly the precious metal catalysts, and on reducing the poisoning of the catalysts by impurities in the fuel gases. Reports of using much cheaper, doped carbon-nanotubes are encouraging. The PEM cells are being developed for vehicles, portable fuel cells and stationary power supplies.

PEM fuel cells tend to be smaller and lighter than SOFCs, providing up to 250 kW, which provides enough heat and electricity to power industry, and indeed for establishments needing high reliability and backup power supplies, such as banks and hospitals. Smaller systems (7 kW), about the size of a fridge, are produced that can provide all the power necessary for a house and the heat produced can also be harnessed to provide hot water.

Various types of fuel cells have been used to power the Gemini and Apollo space missions and the space shuttle. As a result of an EU initiative trial, DaimlerChrysler built fuel cell buses for use in 11 European cities, 3 for each city, which run on compressed hydrogen stored in tanks in the roof with a power of 200 kW. Each has a range of up to 300 km, a top speed of 80 km h^{-1} and carries 70 passengers. On a smaller scale, micro fuel cells can replace batteries in portable electrical equipment, such as mobile phones and palmtop computers. These cells can be replenished easily and quickly by adding more fuel in the form of a methanol capsule.

Although fuel cells offer exciting possibilities for the future, the current reality is that the electricity produced by them is still expensive, anything up to eight times that produced in a traditional generating plant with gas-fired turbines. Fuel cells also need to improve in reliability with better catalysts that are not poisoned by emissions from the reformer. There is no doubt, however, that they are strong contenders for the technology of the future, offering as they do the possibility to move towards a greener 'hydrogen economy', thereby eliminating greenhouse gas emissions.

5.5.3 Oxygen Sensor

Calcia-stabilised zirconia is used in the detection of oxygen, in both **oxygen meters** and **oxygen sensors**. In Figure 5.21, a slab of calcia-stabilised zirconia acts as the solid electrolyte (B) which separates two regions containing oxygen at different pressures. Gas pressures tend to equalise if they can, so, if $p' > p''$, oxygen ions, which are able to pass through the stabilised zirconia, tend to pass through the solid from the right-hand side to the left. This tendency produces a potential difference (because the ions are charged), indicating that oxygen is present (in the sensor) and the measurement of this potential gives a measure of the oxygen pressure difference (in the oxygen meter).

Oxygen gas is reduced to O^{2-} at the right-hand electrode (C), the oxide ions are able to pass through the doped zirconia and are oxidised to oxygen gas at the left-hand electrode (A). The equation for the cell reaction is

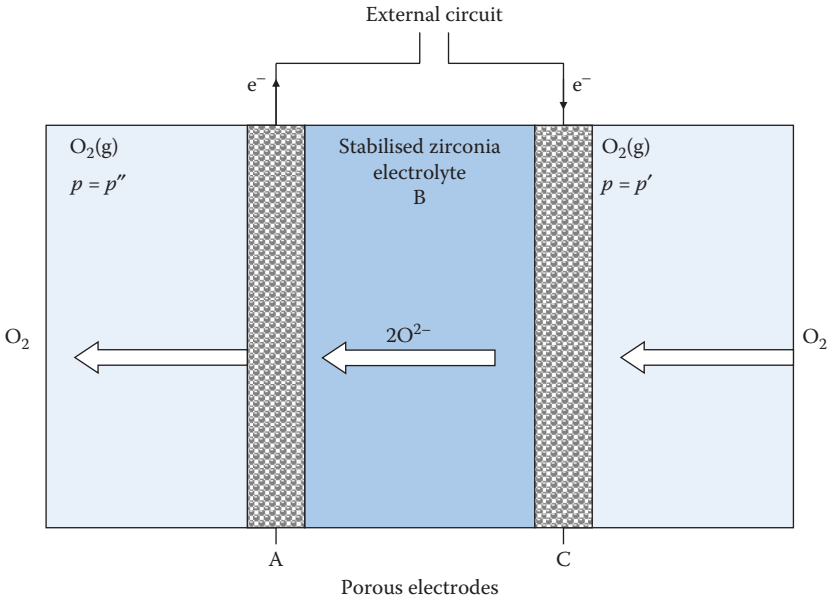


FIGURE 5.21 A schematic representation of an oxygen meter.

$$\mathbf{A:}\ 2O^{2-} \rightarrow O_2\left(p''\right) + 4e^-$$

$$\mathbf{C:}\ O_2\left(p'\right) + 4e^- \rightarrow 2O^{2-}$$

$$\text{Overall: } O_2\left(p'\right) \rightarrow O_2\left(p''\right)$$

Under standard conditions, we can relate the change in Gibbs free energy for the above reaction to the standard emf of the cell:

$$\Delta G^{\ominus} = -nE^{\ominus}F \tag{5.14}$$

The **Nernst equation** allows the calculation of the cell *emf* under nonstandard conditions, E. If the cell reaction is given by a general equation:

$$aA + bB + \cdots + ne = xX + yY + \cdots$$

then

$$E = E^{\ominus} - \frac{2.303RT}{nF}\log\left\{\frac{a_X^x a_Y^y \ldots}{a_A^a a_B^b \ldots}\right\} \tag{5.15}$$

where the quantities a_X, etc., are the activities of the reactants and the products. Applying the Nernst equation to the cell reaction in the oxygen meter, we get

$$E = E^{\ominus} - \frac{2.303RT}{4F} \log \frac{p''}{p'}$$

In this particular case, $E^{\ominus}$ is actually zero because under standard conditions, the pressure of the oxygen on each side would be atmospheric and there would be no net potential difference. The pressure of the oxygen on one side of the cell (say p'') is set to be a known reference pressure, usually either pure oxygen at 101.325 kPa or atmospheric oxygen pressure (~0.21×10^5 Pa). Making these two changes, we get

$$E = \frac{2.303RT}{4F} \log \frac{p'}{p_{\text{ref}}}$$

All the quantities in this equation are now known or can be measured, affording a direct measure of the unknown oxygen pressure p'. For this cell to operate, there must be no *electronic* conduction through the electrolyte.

Oxygen meters find industrial uses in the detection of oxygen in waste gases from chimneys or exhaust pipes, in investigating the operation of furnaces and in measuring the oxygen content in molten metals during the production process. Divers use oxygen meters to measure the partial pressure of oxygen in their gas supplies.

The principal use of oxygen meters is in the control of fuel economy in internal combustion engines. The sensor measures the amount of oxygen in the exhaust gas. If sufficient oxygen has been supplied to completely oxidise the fuel (a 'lean' mixture), not much oxygen is in the exhaust gas and a low voltage is detected. If too little O_2 has been supplied (a 'rich' mixture), then unburnt fuel is emitted, which is both wasteful and polluting; the pressure difference between reference and exhaust oxygen pressure is now greater and a higher voltage is detected. A feedback loop can then be used to adjust the mixture for optimum performance.

This same principle has been employed in developing sensors for other gases by using different electrolytes; examples include H_2, F_2, Cl_2, CO_2, SO_x and NO_x.

5.5.4 Oxygen Separation Membranes

The same principles used in the oxygen sensor are also being used for developing a system for separating oxygen from air, producing pure oxygen. This system would have the advantage that oxygen can be made at the point of use, such as for hospitals, and saves the huge refrigeration costs of cryogenic manufacture and also the transport in heavy pressurised steel cylinders.

The air on the high pressure side (p') is pressurised to about 30 atm, which then passes through the porous cathode and travels through the heated YSZ membrane as O^{2-} anions, reforming as O_2 at the anode (Figure 5.21). The transport of oxide

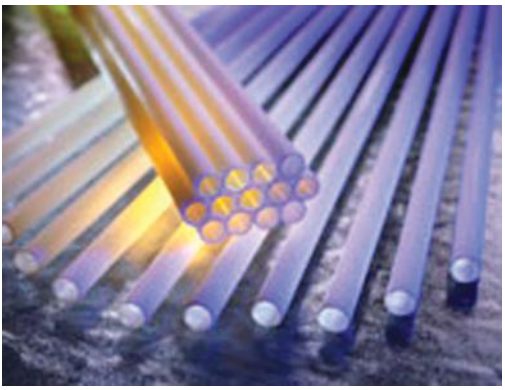


FIGURE 5.22 Tubular oxygen separation membranes. (From Commonwealth Scientific and Industrial Research Organisation, Australia.)

ions across the membrane can also be assisted by applying a voltage difference across it.

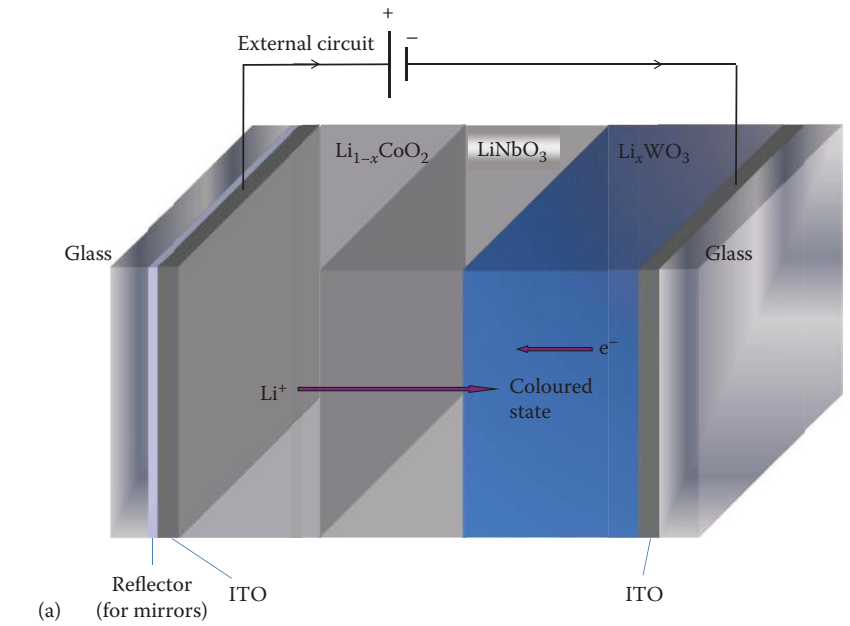
If the solid oxide electrolyte is formed into a tube with the anode on the outside and the cathode inside the tube, then the generated oxygen can be separated and collected under pressure (Figure 5.22).

5.5.5 Electrochromic Devices

On a similar principle to the oxygen separation membrane are **electrochromic devices** that work in the opposite sense to an electrochemical cell: instead of harnessing the chemical reaction between the electrodes to give an electric current, an electric current is applied *to* the cell, causing the movement of ions through the electrolyte. In an electrochromic device (Figure 5.23), this movement must change a colourless compound into a coloured compound at one of the electrodes. The device illustrated consists of a layered sandwich of different materials: on the outside are glass plates coated with a layer of indium tin oxide, which is both transparent and conducting, that forms the electrodes. The anode, lithium cobalt oxide ($LiCoO_2$), is able to release lithium easily and quickly, and the cathode, tungsten oxide (WO_3), is able to intercalate lithium ions, changing colour as it does so. Lithium ions travel through the solid lithium niobium oxide electrolyte.

A		**B**		**C**
$LiCoO_2$	//	$LiNbO_3$	//	WO_3

Oxidation occurs at the anode where electrons are released and Li^+ ions flow through the colourless electrolyte to form Li_xWO_3 at the tungsten oxide (WO_3) cathode, changing it from almost colourless to deep blue. The reaction is readily reversible when the current is reversed and has been found to cycle reliably for tens of



(b)

FIGURE 5.23 (a) A diagram of an electrochromic device. (b) Electrochromic office windows. (From Sage Electrochromics Inc.)

millions of cycles. Notice that the device does not create any light itself, but merely absorbs the incident light, thereby reducing its intensity.

The electrochromic effect gives the potential for interesting 'smart' devices such as windows that can be switched from clear to coloured in order to control the temperature and light in modern buildings (Figure 5.23b). The solid-state electrochromic material is particularly suitable for large devices such as windows and can also be used to change transparent windows to translucent for interior partitioning, for

instance in meeting room windows and public toilet doors: the fast IntercityExpress (ICE) trains running between Germany and its neighbouring countries use them for the glass panel separating the driver and the passengers. Electrochromic glass is used to dim rearview mirrors in cars from glare at night when the car behind does not dip its headlights; in these devices, a reflecting metal film (Ag, Pt, Al) is coated onto the glass.

5.6 COLOUR CENTRES

During early research in Germany, it was noticed that if the crystals of the alkali halides were exposed to X-rays, they became brightly coloured. The colour was thought to be associated with a defect, known then as a **Farbenzentre** (colour centre), now abbreviated to **F-centre**. Since then, it has been found that many forms of high-energy radiation (UV, X-rays, neutrons) will cause F-centres to form. The colour produced by the F-centres is always characteristic of the host crystal, for instance, NaCl becomes deep yellowy-orange, KCl becomes violet and KBr becomes blue-green.

Subsequently, it was found that F-centres can also be produced by heating a crystal in the vapour of an alkali metal: this gives a clue to the nature of these defects. The excess alkali metal atoms diffuse into the crystal and settle on cation sites; at the same time, an equivalent number of anion-site vacancies are created and ionisation gives an alkali-metal cation with an electron trapped at the anion vacancy (Figure 5.24). In fact, it does not even matter which alkali metal is used; if NaCl is heated with potassium, the colour of the F-centre does not change because it is characteristic of the electron trapped at the anion vacancy in the host halide. Work with ESR has confirmed that F-centres are indeed unpaired electrons trapped at vacant lattice (anion) sites.

The trapped electron provides a classic example of an 'electron in a box'. A series of energy levels are available for the electron, and the energy required to transfer from one level to another falls in the visible part of the electromagnetic spectrum—hence the colour of the F-centre. There is an interesting natural example of this phenomenon: the mineral fluorite (CaF_2) can be found in Derbyshire, UK, where it is known as 'Blue John', and its beautiful blue-purple colouration is due to the presence of F-centres.

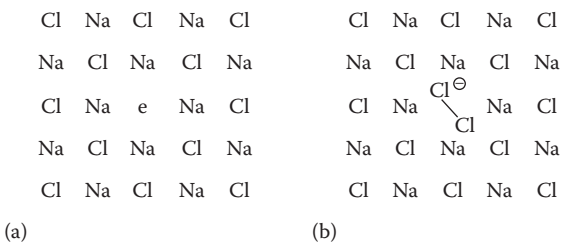


FIGURE 5.24 (a) The F-centre, an electron trapped on an anion vacancy. (b) H-centre.

Many other colour centres have now been characterised in alkali-halide crystals. The **H-centre** (halogen centre) is formed by heating, for instance, NaCl in Cl_2 gas. In this case, a $[Cl_2]^-$ ion is formed and it occupies a single anion site (Figure 5.24b). F-centres and H-centres are perfectly complementary—if they meet, they cancel one another out.

Another interesting natural example of colour centres lies in the colour of smoky quartz and amethyst. These semiprecious stones are basically crystals of silica (SiO_2) with some impurity present. In the case of smoky quartz, the silica contains a little aluminium impurity. The Al^{3+} substitutes for the Si^{4+} in the lattice and the electrical neutrality is maintained by H^+ present in the same amount as Al^{3+}. The colour centre arises when ionising radiation interacts with an $[AlO_4]^{5-}$group, liberating an electron that is then trapped by H^+:

$$\left[AlO_4\right]^{5-} + H^+ = \left[AlO_4\right]^{4-} + H.$$

The $[AlO_4]^{4-}$ group is now electron-deficient and can be considered to have a 'hole' trapped at its centre. This group is the colour centre, absorbing light and producing the smoky colour. In crystals of amethyst, the impurity present is Fe^{3+}, and on irradiation $[FeO_4]^{4-}$ colour centres are produced that absorb light to give the characteristic purple colouration.

5.7 NONSTOICHIOMETRIC COMPOUNDS

5.7.1 Introduction

We saw in the previous sections in this chapter that it is possible to *introduce* defects into a perfect crystal by adding an impurity. Such an addition causes point defects of one sort or another to form, but they no longer occur in complementary pairs. Impurity-induced defects are said to be **extrinsic**. We have also noted that when assessing which defects have been created in a crystal, it is important to remember that the overall charge on the crystal must always be zero.

Colour centres are formed if a crystal of NaCl is heated in sodium vapour; sodium is taken into the crystal, and the formula becomes $Na_{1+x}Cl$. The sodium atoms occupy cation sites, creating an equivalent number of anion vacancies; they subsequently ionise to form a sodium cation with an electron trapped at the anion vacancy (see Section 5.6). The solid so formed is a **nonstoichiometric compound** because the ratio of the atomic components is no longer the simple integer that we have come to expect from the valencies demonstrated in well-characterised compounds. A careful analysis of many substances, particularly inorganic solids, shows that it is common for the atomic ratios to be nonintegral: uranium dioxide, for instance, can range in composition from $UO_{1.65}$ to $UO_{2.25}$, certainly not the perfect UO_2 that we might expect. There are many other examples, some of which we discuss in some detail.

What kind of compounds are likely to be nonstoichiometric? 'Normal' covalent compounds are assumed to have a fixed composition where the atoms are

usually held together by strong covalent bonds formed by the pairing of two electrons. Breaking these bonds usually takes quite a lot of energy, and therefore under normal circumstances, a particular compound does not show a wide range of composition; this is true for most molecular organic compounds, for instance. Additionally, ionic compounds are usually stoichiometric, because to remove or add ions requires a considerable amount of energy. We have seen, however, that it is possible to make ionic crystals nonstoichiometric by doping them with an impurity, as with the example of Na added to NaCl. There is also another mechanism whereby ionic crystals can become nonstoichiometric: if the crystal contains an element with a variable valency, then a change in the number of ions of that element can be compensated for by changes in ion charge; this maintains the charge balance but alters the stoichiometry. Elements with a variable valency mostly occur in the transition elements, the lanthanoids and actinoids, and the heavier main group elements.

In summary, nonstoichiometric compounds can have formulae that are not simple integer ratios of atoms; they also usually exhibit a range of composition. They can be made by introducing impurities into a system, but are frequently a consequence of the ability of the metal to exhibit variable valency. Table 5.5 lists a few nonstoichiometric compounds together with their composition ranges.

TABLE 5.5

Approximate Composition Ranges for Some Nonstoichiometric Compounds

Compound		Composition Range[a]
TiO_x	[≈TiO]	$0.65 < x < 1.25$
	[≈TiO_2]	$1.998 < x < 2.000$
VO_x	[≈VO]	$0.79 < x < 1.29$
Mn_xO	[≈MnO]	$0.848 < x < 1.000$
Fe_xO	[FeO]	$0.833 < x < 0.957$
Co_xO	[≈CoO]	$0.988 < x < 1.000$
Ni_xO	[≈NiO]	$0.999 < x < 1.000$
CeO_x	[≈Ce_2O_3]	$1.50 < x < 1.52$
ZrO_x	[≈ZrO_2]	$1.700 < x < 2.004$
UO_x	[≈UO_2]	$1.65 < x < 2.25$
$Li_xV_2O_5$		$0.2 < x < 0.33$
Li_xWO_3		$0 < x < 0.50$
TiS_x	[≈TiS]	$0.971 < x < 1.064$
Nb_xS	[≈NbS]	$0.92 < x < 1.00$
Y_xSe	[≈YSe]	$1.00 < x < 1.33$
V_xTe_2	[≈VTe_2]	$1.03 < x < 1.14$

[a] Note that all composition ranges are temperature-dependent and that the figures here are intended only as a guide.

Until more sophisticated structure determination were possible, defects in both stoichiometric and nonstoichiometric crystals were treated entirely from the point of view that point defects are randomly distributed. However, isolated point defects are *not* scattered at random in nonstoichiometric compounds, but are often dispersed throughout the structure in some kind of regular pattern. In the following sections, we try to explore the relationship between stoichiometry and structure.

It is difficult to determine the structure of compounds containing defects and it is only very recently that much of our knowledge has been formed. X-ray diffraction is the usual method for the determination of the structure of a crystal; however, this method yields an *average* structure for a crystal. For pure, relatively defect-free structures, this is a good representation, but for nonstoichiometric and defect structures, it avoids precisely the information that we want to know. For this kind of structure determination, a technique that is sensitive to *local* structure is needed, and such techniques are scarce. Structures are often elucidated from a variety of sources of evidence: X-ray and neutron diffraction, density measurements, spectroscopy (when applicable) and high-resolution electron microscopy (HREM); magnetic measurements have also proved useful in the case of FeO. The electron microscopy techniques have done most to clarify our understanding of defect structures, as under favourable circumstances information on an atomic scale is now possible.

Nonstoichiometric compounds are of potential use to industry because their electronic, optical, magnetic and mechanical properties can be modified by changing the proportions of the atomic constituents. This is widely exploited and researched by the electronics and other industries. Currently, probably the best-known example of nonstoichiometry is that of oxygen vacancies in the high-temperature superconductors, such as YBCO and $YBa_2Cu_3O_{7-x}$. The structure of these is discussed in detail in Chapter 9.

5.7.2 Nonstoichiometry in Wustite (FeO) and MO-Type Oxides

Ferrous oxide is known as **wustite (FeO)**, and it has the NaCl (rock-salt) crystal structure. Accurate chemical analysis shows that it is nonstoichiometric: it is always deficient in iron. The FeO phase diagram (Figure 5.25) shows that the compositional range of wustite increases with temperature and that stoichiometric FeO is *not* included in the range of stability. Below 570°C, wustite disproportionates to α-iron and Fe_3O_4.

There are two ways in which an iron deficiency could be accommodated by a defect structure: there could either be *iron vacancies* giving a formula $Fe_{1-x}O$, or, alternatively, there could be an *excess of oxygen in interstitial positions*, with the formula FeO_{1+x}. A comparison of the theoretical and measured densities of the crystal distinguishes between the alternatives. The easiest method of measuring the density of a crystal is the flotation method. Liquids of differing densities (that dissolve in one another) are mixed together until a mixture is found that will just suspend the crystal so that it neither floats nor sinks. The density of that liquid mixture must then be the same as that of the crystal, and it can be found by weighing an accurately measured volume.

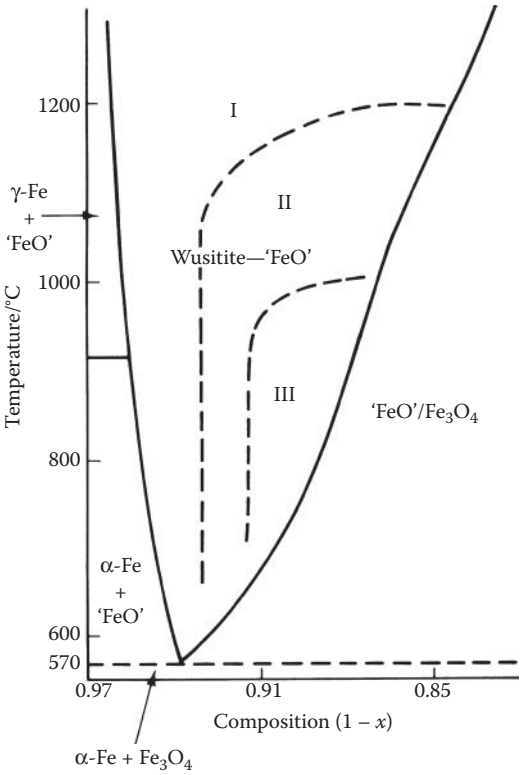


FIGURE 5.25 A phase diagram of the FeO system. I, II and III together comprise the wustite region.

The theoretical density of a crystal can be obtained from the volume of the unit cell and the mass of the unit cell contents. The results of an X-ray diffraction structure determination gives both of these data, as the unit cell dimensions are accurately measured and the type and number of formula units in the unit cell are also determined. An example of this type of calculation for FeO follows:

> A particular crystal of FeO was found to have a unit cell dimension of 430.1 pm, a measured density of 5.728 kg m^{-3} and an iron-to-oxygen ratio of 0.945. The unit cell volume (which is a cube) is thus $(430.1\ \text{pm})^3 = 7.956 \times 10^7\ (\text{pm})^3 = 7.956 \times 10^{-29}\ \text{m}^3$.

There are four formula units of FeO in a perfect unit cell with the rock-salt structure. The mass of these four units can be calculated from their relative atomic masses: Fe, 55.85 and O, 16.00. One mole of FeO weighs (55.85 + 16.00) g = 0.07185 kg, 4 moles weigh (4 × 0.07185) kg, and four formula units weigh $(4 \times 0.07185)/N_A$ kg $= 4.773 \times 10^{-25}$ kg; where Avogadro's number, $N_A = 6.022 \times 10^{23}\ \text{mol}^{-1}$.

The sample under consideration has an Fe/O ratio of 0.945. Assume, in the first instance, that it has *iron vacancies*: the unit cell contents in this

case will be $(4 \times 0.945) = 3.78$Fe and 4O. The mass of the contents will be $[(3.78 \times 55.85) + (4 \times 16.00)]/(N_A \times 10^3)$ kg. Dividing by the volume of the unit cell, we get a value of 5.742×10^3 kg m^{-3} for the density. If, instead, the sample possesses *interstitial oxygens*, the ratio of oxygens to iron in the unit cell will be given by $1/0.945 = 1.058$. The unit cell in this case will contain 4Fe and $(4 \times 1.058) = 4.232$O. The mass of this unit cell is given by: $[(4 \times 55.85) + (4.232 \times 16.00)]/(N_A \times 10^3)$ kg, giving a density of 6.076×10^3 kg m^{-3}. Comparing the two sets of calculations with the experimentally measured density of 5.728×10^3 kg m^{-3}, it is clear that this sample contains *iron vacancies* and that the formula should be written as $Fe_{0.945}O$. A table of densities is drawn up for FeO in Table 5.6.

It is found to be characteristic of most nonstoichiometric compounds that *their unit cell size varies smoothly with composition but the symmetry is unchanged*. This is known as **Vegard's law**.

In summary, nonstoichiometric compounds are found to exist over a range of composition, and throughout that range the unit cell length varies smoothly with no change in symmetry. It is possible to determine whether the nonstoichiometry is accommodated by vacancy or interstitial defects using density measurements.

So far, the discussion of the defects in FeO has been structural. Now we turn our attention to the balancing of the charges within the crystal. In principle, the compensation for the iron deficiency can be made either by the oxidation of some Fe(II) ions *or* by the reduction of some oxide anions. It is energetically more favourable to oxidise Fe(II). For each Fe^{2+} vacancy, *two* Fe^{2+} cations must be oxidised to Fe^{3+}. In the overwhelming majority of cases, defect creation involves changes in the cation oxidation state. In the case of metal excess in simple compounds, we would usually expect to find that neighbouring cation(s) would be reduced.

In a later section, we will look at some general cases of nonstoichiometry in simple oxides, but before we do that, we will complete the FeO story with a look at its detailed structure.

FeO has an NaCl structure with Fe^{2+} ions in the octahedral sites. The iron deficiency manifests itself as cation vacancies, and the electronic compensation made for this is that for every Fe^{2+} ion vacancy, there are two neighbouring Fe^{3+} ions, and this is confirmed by Mössbauer spectroscopy. If the rock-salt structure were preserved, the Fe^{2+}, Fe^{3+} and the cation vacancies would be distributed over the octahedral sites in the *ccp* O^{2-} array. However, Fe^{3+} is a high-spin d^5 ion and has no crystal field

TABLE 5.6

Experimental and Theoretical Densities (10^3 kg m^{-3}) for FeO

				Theoretical Density	
O:Fe Ratio	Fe:O Ratio	Lattice Parameter (pm)	Observed Density	Interstitial O	Fe Vacancies
1.058	0.945	430.1	5.728	6.076	5.742
1.075	0.930	429.2	5.658	6.136	5.706
1.087	0.920	428.5	5.624	6.181	5.687
1.099	0.910	428.2	5.613	6.210	5.652

stabilisation in either the octahedral or the tetrahedral sites and therefore no preference. Structural studies (X-ray, neutron and magnetic) have shown that some of the Fe^{3+} ions are in the *tetrahedral* sites. If a tetrahedral site is occupied by an Fe^{3+} ion, the immediately surrounding four Fe^{2+} octahedral sites must be vacant as they are too close to be occupied at the same time, and this type of defect is found for low values of x. At higher values of x, the structure contains various types of **defect cluster**, which are distributed throughout the crystal. A defect cluster is a region of the crystal where the defects form an *ordered structure*. One possibility is known as the **Koch–Cohen cluster** (Figure 5.26). This shows a standard NaCl-type unit cell at the centre, but with four interstitial, tetrahedrally coordinated Fe^{3+} ions in the tetrahedral holes; the thirteen immediately surrounding octahedral Fe^{2+} sites must be vacant. Surrounding this central unit cell, the other octahedral iron sites in the cluster are occupied, but they may contain either Fe^{2+} or Fe^{3+} ions. We therefore designate them simply as Fe_{oct}. The front and back planes have been cut away from the diagram in Figure 5.26 to make the central section more visible.

It is instructive to consider the composition of a cluster such as the one shown in Figure 5.26 in some detail. Clusters such as this are often referred to by the ratio of cation vacancies to interstitial Fe^{3+} in tetrahedral holes, in this case 13 : 4. The complete cluster, allowing for the back and front faces, which are not illustrated here, contains 8 NaCl-type unit cells, and thus 32 oxide ions. A stoichiometric array with no defects would also contain 32 Fe^{2+} cations on the octahedral sites: taking into account the 13 octahedral vacancies, there must be 19 Fe_{oct} ions and the 4 interstitial Fe^{3+}_{tet} ions, making 23 iron cations in all. The overall formula for the cluster is thus, $Fe_{23}O_{32}$, almost Fe_3O_4. It bears a strong resemblance to the structure of Fe_3O_4, the next highest oxide of iron.

Having determined the atomic contents of the cluster, we now turn our attention to the charges. There are 32 oxide ions, so to balance them, the Fe cations overall

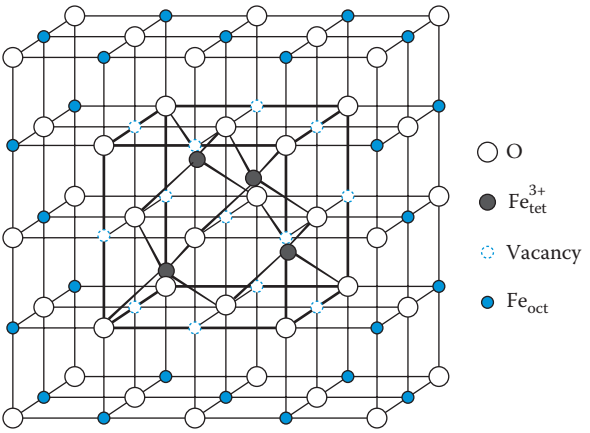


FIGURE 5.26 The Koch–Cohen cluster shown with the back and front planes cut away for clarity. The central section with four tetrahedrally coordinated Fe^{3+} ions is highlighted in bold.

must have 64 positive charges. Twelve are accounted for by the four Fe^{3+} ions in the tetrahedral positions, leaving 52 to find from the remaining 19 Fe_{oct} cations. This accounting is difficult to do by inspection. Suppose that there are $x Fe^{2+}$ ions and $y Fe^{3+}$ ions in the octahedral sites, we know that

$$x + y = 19$$

We also know that their total charges must equal 52, therefore,

$$2x + 3y = 52$$

giving two simultaneous equations. Solving gives $y = 14$ and $x = 5$. The octahedral sites surrounding the central (bold) unit cell are thus occupied by 5 Fe^{2+} ions and by 14 Fe^{3+} ions. By injecting such clusters throughout the FeO structure, the nonstoichiometric structure is built up. The exact formula of the compound (the value of x in $Fe_{1-x}O$) will depend on the average separation of the randomly injected clusters. Neutron scattering experiments indicate that as the concentration of the defects increases, the clusters order into a regular repeating pattern with its own unit cell of lower symmetry; the new structure is referred to as a **superstructure** or **superlattice** of the parent. In the oxygen-rich limit when the whole structure is composed of these clusters, there would be a new ordered structure of the formula $Fe_{23}O_{32}$ based on the structure of the parent defect cluster.

Manganese, cobalt and nickel oxides (MnO, CoO and NiO) have similar structures to FeO.

ZnO has the wurtzite crystal structure where half of the tetrahedral holes of the hexagonally packed oxygen anions are filled with Zn^{2+} ions. ZnO turns yellow when it is heated to 800°C; oxygen is lost, leaving nonstoichiometric $Zn_{1+x}O$ with oxide vacancies. To compensate, Zn^{2+} ions migrate to interstitial positions and are reduced to Zn^{+} ions or Zn atoms. The electron transfer between the Zn^{2+} and Zn^{+}/Zn results in the yellow colouration seen when ZnO is heated.

5.7.3 Uranium Dioxide

Above 1127°C, a single oxygen-rich nonstoichiometric phase of UO_2 is found with the formula UO_{2+x}, ranging from UO_2 to $UO_{2.25}$. Unlike FeO, where a metal-deficient oxide was achieved through cation vacancies, in this example the metal deficiency arises from interstitial anions.

$UO_{2.25}$ corresponds to U_4O_9, which is a well-characterised oxide of uranium known at low temperature. UO_2 has the fluorite structure. The unit cell is shown in Figure 5.27a and contains four formula units of UO_2. (There are four uranium ions contained within the cell boundaries; the eight oxide ions come from: $[8 \times \frac{1}{8}] = 1$ at the corners, $[6 \times \frac{1}{2}] = 3$ at the face centres, $[12 \times \frac{1}{4}] = 3$ at the cell edges and 1 at the cell body-centre.)

As more oxygen is taken into UO_2, the extra oxide ions go into interstitial positions. The most obvious site available is in the middle of one of the octants (the vacant octahedral holes), where there is no metal atom. This site, however, is not ideal for

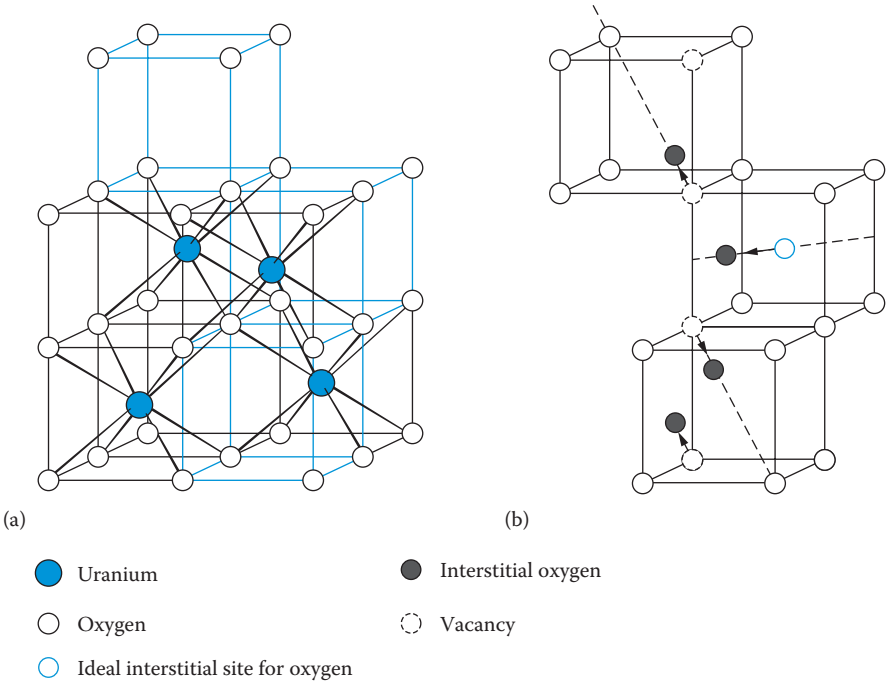


FIGURE 5.27 (a) The fluorite structure of UO_2. (b) Interstitial defect cluster in UO_{2+x}. The uranium positions (not shown) are in the centre of every other cube.

an extra oxide ion, as not only is it crowded, but it is also surrounded by eight ions of the same charge. Neutron diffraction shows that an interstitial oxide anion does not sit exactly at the centre of an octant, but it is displaced sideways; this has the effect of moving two other oxide ions from their lattice positions by a very small amount, leaving two vacant lattice positions. This is illustrated in Figure 5.27b, where three vacant octants are picked out and the positions of one additional interstitial oxide and the two displaced oxides with their vacancies are shown. This defect cluster can be thought of as two vacancies, one interstitial of one kind (O′) and two of another (O″), and this is called the 2 : 1 : 2 Willis cluster. The movement of the ions from the 'ideal' positions is shown by small arrows: the movement of the interstitial oxide O′ from the centre of an octant is along the direction of a diagonal of one of the cube faces (the *110* direction), whereas the movement of the oxide ions O″ on lattice positions is along cube diagonals (the *111* direction). The atomic composition of a fluorite unit cell (Figure 5.27a) when modified by this defect cluster is U_4O_9; the net oxygen gain is just the single new interstitial in the central octant of Figure 5.27b. A more commonly found defect cluster contains two vacancies, two interstitial oxides (O′) and two interstitial oxides (O″), leading to a 2 : 2 : 2 Willis cluster (not shown here). In the oxygen-rich limit for the UO_{2+x} nonstoichiometric structure, $UO_{2.25}$, these are ordered throughout the structure, and the structure has a very large unit cell based on $4 \times 4 \times 4$ fluorite unit cells (with a volume of 64 times that of the fluorite unit

cell). We can think of UO_{2+x} as containing **microdomains** of the U_4O_9 structure within that of UO_2. The electronic compensation for the extra interstitial oxide ions will most likely be the oxidation of the neighbouring U(IV) atoms to either U(V) or U(VI).

5.7.4 Titanium Monoxide Structure

Titanium and oxygen form nonstoichiometric phases that exist over a range of composition centred about the stoichiometric 1 : 1 value, from $TiO_{0.65}$ to $TiO_{1.25}$. We shall look at what happens in the upper range from $TiO_{1.00}$ to $TiO_{1.25}$.

At the stoichiometric composition of $TiO_{1.00}$, the crystal structure can be thought of as a NaCl-type structure with vacancies in both the metal and the oxygen

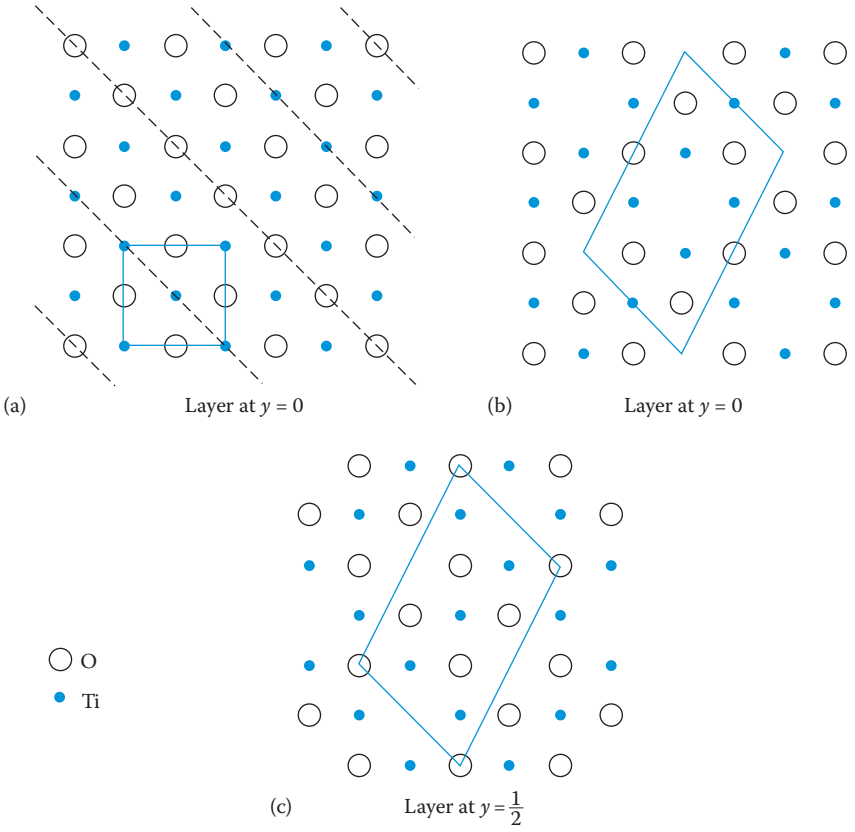


FIGURE 5.28 Layers parallel to the horizontal planes of Figure 1.31. (a) The hypothetical TiO structure of the NaCl type shown in Figure 1.31; the line of intersection of every third vertical diagonal plane is marked by a dashed line. (b) The same plane in the observed structure of TiO; every alternate atom is removed along the diagonal lines in (a). (c) The plane directly beneath the layer in (b); again, every alternate atom is removed along the cuts made by the planes whose intersection lines are shown in (a). In (b) and (c), the cross section of a monoclinic unit cell is indicated.

sublattices: one-sixth of the titaniums and one-sixth of the oxygens are missing. Above 900°C, these vacancies are randomly distributed, but below this temperature they are ordered, as shown in Figure 5.28.

In Figure 5.28a, we show a layer through an NaCl-type structure. Every *third* vertical diagonal plane has been picked out with a dashed line. In the $TiO_{1.00}$ structure, every other atom along those dashed lines is missing. This is illustrated in Figure 5.28b. If we consider that in these first two diagrams we are looking at the structure along the y axis, and that this layer is the top of the unit cells, $y = 0$ or 1, then the layer below this and parallel to it will be the central horizontal plane of the unit cell at $y = \frac{1}{2}$. This is drawn in Figure 5.28c, and again we notice that every other atom along every third diagonal plane is missing. This is true throughout the structure. In Figure 5.28, the unit cell of a perfect NaCl-type structure is marked on (a), whereas the boundaries of the new unit cell, taking the ordered defects into account, are marked on (b) and (c). The new unit cell of the superlattice is monoclinic (see Chapter 1) because the angle in the xz plane (β) is not equal to 90°. This structure is unusual in that it appears to be stoichiometric, but in fact it contains defect vacancies on both the anion and the cation sublattices.

As we noted in Chapter 4, unusually for a transition metal monoxide, $TiO_{1.00}$ shows metallic conductivity. The existence of the vacant sites within the TiO structure is thought to permit sufficient contraction of the lattice that the 3d orbitals on titanium overlap, thus broadening the conduction band and allowing electronic conduction.

When titanium monoxide has the limiting formula $TiO_{1.25}$, it has a different defect structure, still based on the NaCl structure, but with *all* the oxygens present, and one in every five titaniums missing (Figure 5.29). The pattern of the titanium vacancies

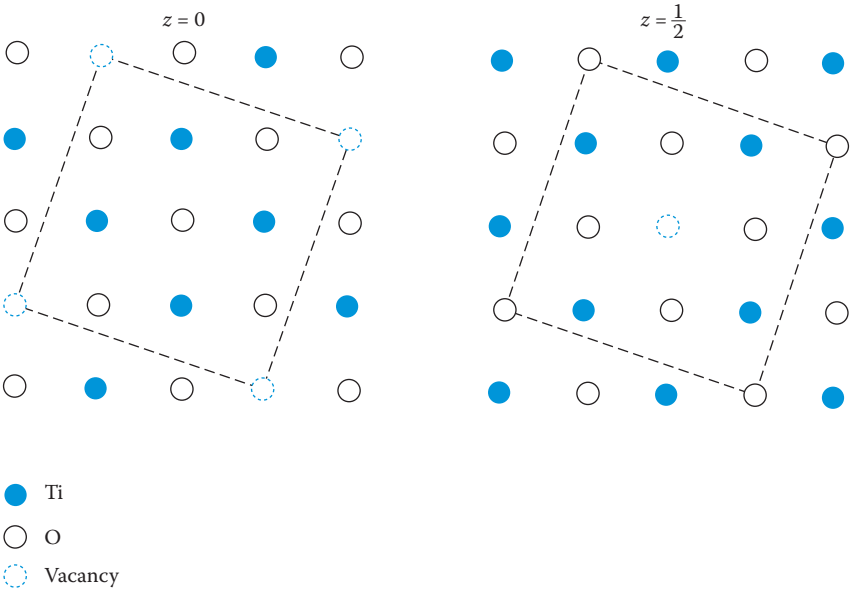


FIGURE 5.29 The structure of $TiO_{1.25}$ showing both O and Ti positions.

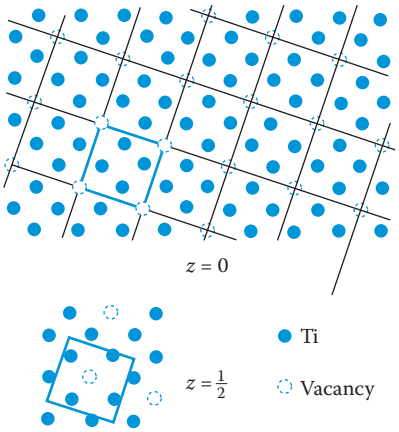


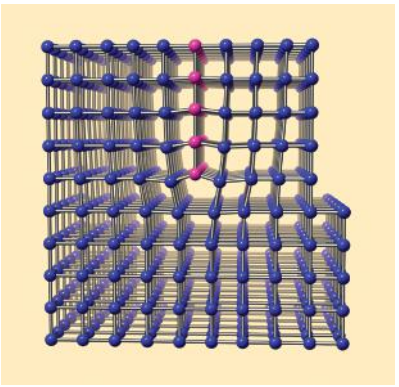
FIGURE 5.30 Successive Ti layers in the structure of $TiO_{1.25}$.

is shown in Figure 5.30, which shows a layer of the type in Figure 5.28 but with the oxygens omitted; only the titaniums are marked. If we draw lines through the titaniums, every fifth one is missing. The ordering of the defects has again produced a superlattice. Where samples of titanium oxide have formulae that lie between the two limits discussed here, $TiO_{1.00}$ and $TiO_{1.25}$, the structure seems to consist of portions of the $TiO_{1.00}$ and $TiO_{1.25}$ structures intergrown. Note that although most texts refer to the structure as $TiO_{1.25}$, as we have here, on the definitions that we have used previously when discussing 'FeO', it is more correctly written as $Ti_{0.8}O$ ($Ti_{1-x}O$), as this indicates that the structure contains titanium vacancies rather than interstitial oxygens.

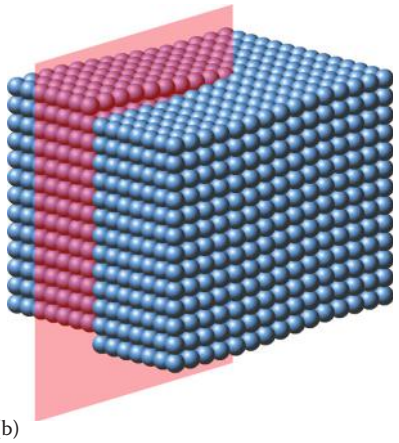
5.8 EXTENDED DEFECTS

We mentioned in the introduction to this chapter that crystals often contain **extended defects** as well as point defects. The simplest line or **linear defect** is an **edge dislocation** where there is a fault in the arrangement of the atoms in a line through the crystal lattice where an extra plane of atoms extends part way through the structure, thus distorting it (Figure 5.31a). Another linear defect is the **screw dislocation**, which occurs when a stress is applied to the crystal; the dislocation of the line of atoms is now perpendicular to the stress (Figure 5.31b). New material arriving at the surface of the crystal during crystallisation preferentially anchors itself to the edge formed and the planes of atoms now form a single helical surface (Figure 5.31b and 5.31c). Linear dislocations greatly reduce the strength of a crystal, as the planes can glide over one another more easily, the dislocation occurring along the densest planes of atoms, but they are the site of enhanced chemical activity and have been shown to be very important for the enhancement of the activity of catalysts.

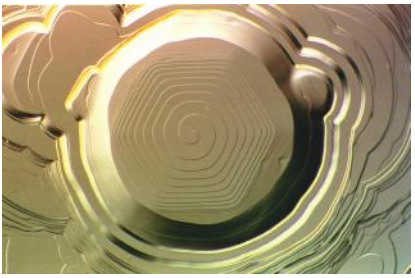
There are different types of **planar defects**, and examples include **grain boundaries**, **stacking faults**, and **chemical twinning** which can contain unit cells mirrored about the twin plane through the crystal.



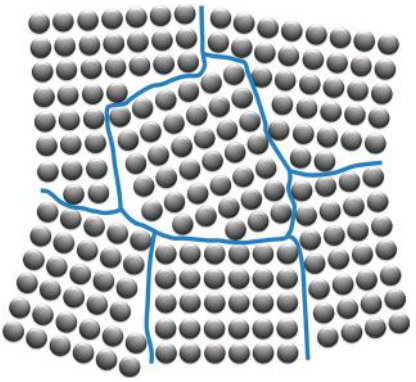
(a)



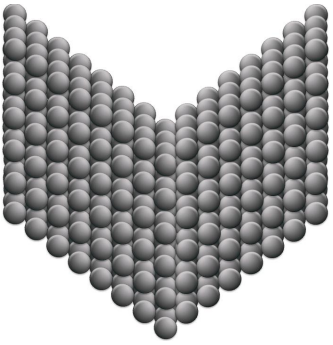
(b)



(c)



(d)



(e)

FIGURE 5.31 (a) An edge dislocation. (From Crystalmaker Software Ltd. With permission.) (b) A screw dislocation. (From IMPESS, European Space Agency.) (c) A screw dislocation emerging on the surface of a silicon carbide crystal. (From Gunter Wagner, Institute of Crystal Growth, Germany.) (d) Grain boundary. (e) Crystal twinning.

Polycrystalline substances are made of small crystallites orientated randomly to one another; the surface that separates each crystallite is known as the grain boundary (Figure 5.31d). Grain boundaries tend to hinder the movement of dislocations so that a polycrystalline metal is usually much stronger than a single crystal. To make metals stronger, interstitial atoms and grain boundaries may be deliberately introduced to 'pin' dislocations in place. Grain sizes vary hugely from the order of micometres to millimetres in diameter. Metal alloys are often stronger than the host metal—take steel for example—because the atoms 'foreign' to the structure have a tendency to congregate on the grain boundaries, thus hindering further the movement of dislocations. An **antiphase domain** can also occur with a crystal, where this grain has the reverse structure from the surrounding structure. At the antiphase boundary, each side of the boundary has the opposite phase: for instance, in a *ccp* (see Chapter 1), the sequence at the boundary would be ABCAB**CC**BACBA….

A stacking fault occurs in a crystal when there are either one or two extra planes or missing planes of atoms. This can occur quite easily in close-packed structures. Think of a *ccp* structure ABCABCABC…, if one of the planes is missing, an intrinsic fault, we get a faulty stacking sequence such as ABC**ACA**BCABC…. Similarly, in a *ccp* structure if an extra layer interposes itself, we could get the sequence ABCA**BAC**ABCABCA…, now called an extrinsic fault.

If stacking faults occur over many layers in a crystal, it can result in twinning. In a twinned crystal, two crystals meet at a common plane of atoms and are often mirror images of one another (Figure 5.31e).

Three-dimensional faults occur as small voids in a structure, where there are no atoms, or as small regions of a different phase, known as **precipitates**.

The above-mentioned defects are very important in the study of materials as they affect their physical properties, such as tensile strength and ductility. However, there are other planar defects in particular, that affect the chemical and crystallographic structure of a compound:

- Crystallographic shear (CS) planes, where the oxygen vacancies effectively collect together in a plane that runs through the crystal
- Intergrowth structures, where two different but related structures alternate throughout the crystal

5.8.1 Crystallographic Shear Planes

Nonstoichiometric compounds are found for the higher oxides of tungsten, molybdenum and titanium, WO_{3-x}, MoO_{3-x} and TiO_{2-x}, respectively. The reaction of these systems to the presence of point defects is entirely different from what has been discussed previously; in fact, the point defects are eliminated by a process known as **crystallographic shear (CS)**.

In these systems, a series of closely related compounds with very similar formulae and structure exist. The formulae of these compounds all obey a general formula, which for the molybdenum and tungsten oxides can be Mo_nO_{3n-1}, Mo_nO_{3n-2}, W_nO_{3n-1}

and W_nO_{3n-2} and for titanium dioxide it is Ti_nO_{2n-1}; n can vary taking values of 4 and above. The resulting series of oxides is known as a **homologous series** (as for the alkanes in organic chemistry). The first seven members of the molybdenum trioxide series are Mo_4O_{11}, Mo_5O_{14}, Mo_6O_{17}, Mo_7O_{20}, Mo_8O_{23}, Mo_9O_{26}, $Mo_{10}O_{29}$ and $Mo_{11}O_{32}$.

In these compounds, we find regions of corner-linked octahedra separated from each other by thin regions of a different structure known as the CS planes. The different members of a homologous series are determined by the fixed spacing between the CS planes. The structure of a shear plane is quite difficult to understand, and these structures are usually depicted by the linking of octahedra, as described in Chapter 1.

WO_3 has several polymorphs, but above 900°C the WO_3 structure is that of ReO_3, which is shown in Figure 5.32. (The structures of the other polymorphs are distortions of the ReO_3 structure.) ReO_3 is made up of $[ReO_6]$ octahedra that are linked

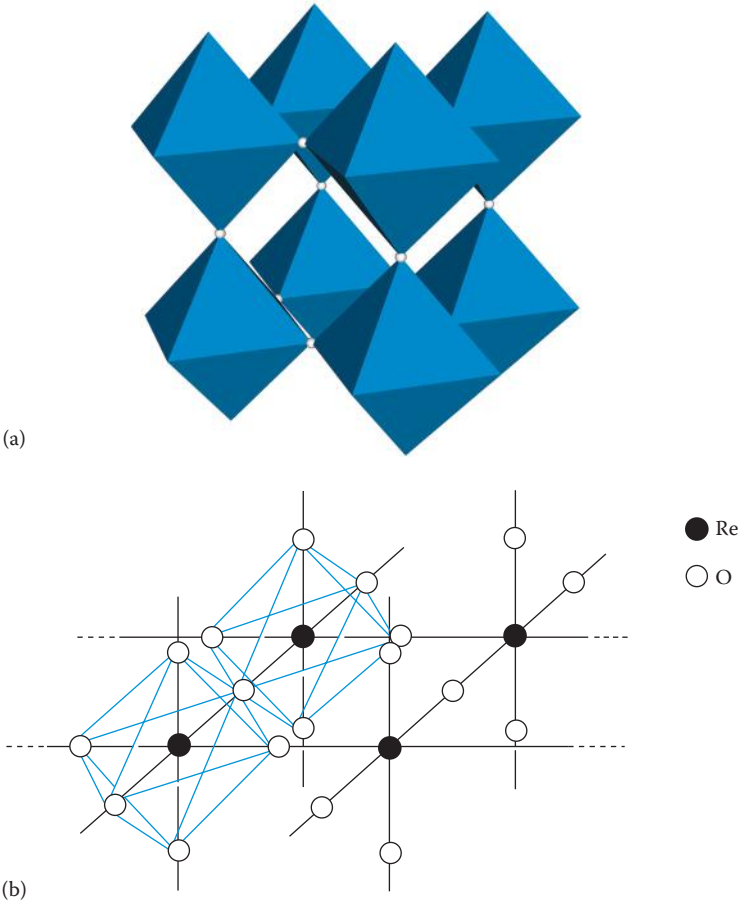


FIGURE 5.32 (a, b) Part of the ReO_3 structure showing the linking of octahedra through the corners.

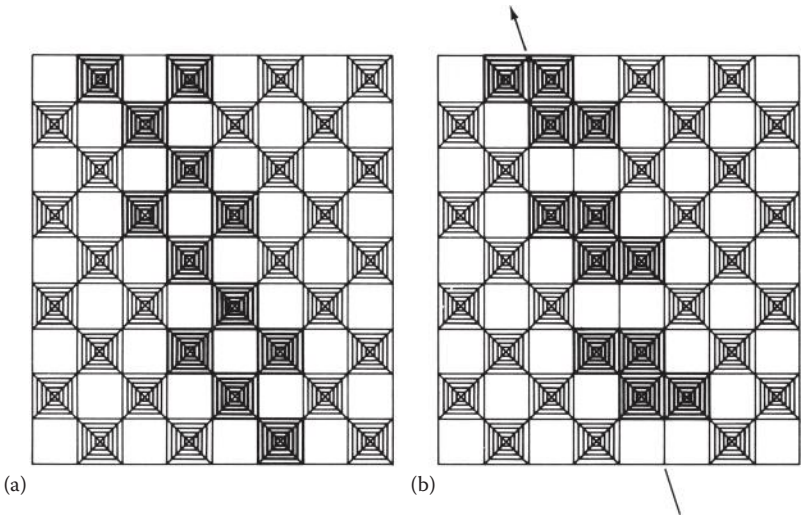


FIGURE 5.33 (a, b) The formation of a shear structure.

together through their corners; each corner of an octahedron is shared with another. Part of the ReO_3 structure is drawn in Figure 5.32b, and we can see that every oxygen atom is shared between two metal atoms: as six oxygens surround each Re, the overall formula is ReO_3. Figure 5.33a shows part of one layer of linked octahedra in the structure. Notice that within the layer, any octahedron is linked to four others; it is also linked, through its upper and lower corners, to octahedra in the layers above and below.

The nonstoichiometry in WO_{3-x} is achieved by some of the octahedra in this structure changing from corner-sharing to edge-sharing. Look now to the octahedra marked in bold in Figure 5.33a. The edge-sharing corresponds to shearing the structure so that the chains of bold octahedra are displaced to the positions in Figure 5.33b. This shearing occurs at regular intervals in the structure and is interspersed with slabs of the 'ReO_3' structure (corner-linked $[WO_6]$ octahedra). It creates groups of four octahedra that share edges. The direction of maximum density of the edge-sharing groups is called the CS plane and is indicated by an arrow in Figure 5.33b.

In order to see how CS alters the stoichiometry of WO_3, we need to find the stoichiometry of one of the groups of four octahedra that are linked together by sharing edges; part of the structure depicting one of these groups is shown in Figure 5.34a. The four octahedra consist of 4W atoms and 18O atoms (Figure 5.34b). Fourteen of the oxygen atoms are linked out to other octahedra (these bonds are indicated) so each are shared by two tungsten atoms, while the remaining four oxygens are only involved in the edge-sharing within the group. The overall stoichiometry is given by $[4W + (14 \times \frac{1}{2})O + 4O]$, giving W_4O_{11}.

Clearly, if groups of four octahedra with stoichiometry W_4O_{11} are interspersed throughout a perfect WO_3 structure, then the amount of oxygen in the structure

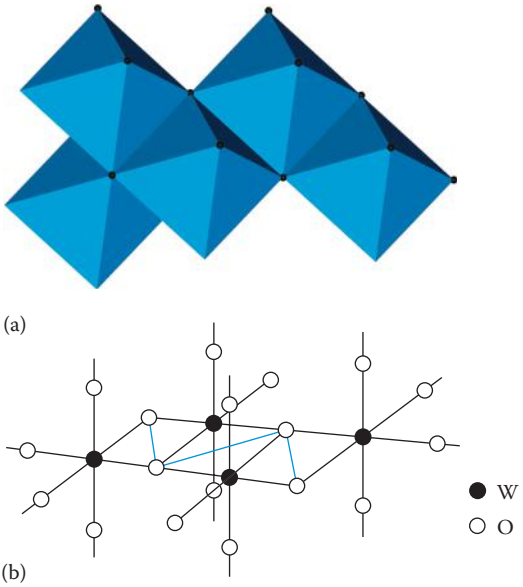


FIGURE 5.34 (a, b) A group of $[WO_6]$ octahedra-sharing edges formed by the creation of shear planes in W_nO_{3n-1}.

is reduced and we can write the formula as WO_{3-x}. The effect of introducing the groups of four in an ordered way can be quantified. If the structure sheared in such a way that the entire structure was composed of these groups, the formula would become W_4O_{11}. If there is one $[WO_6]$ octahedron for each group of four, then the overall formula becomes $[W_4O_{11} + WO_3] = W_5O_{14}$. Clearly, we can extend this process to any number of $[WO_6]$ octahedra regularly interspersed between the groups:

$$W_4O_{11} + 2WO_3 = W_6O_{17}; \qquad W_4O_{11} + 3WO_3 = W_7O_{20}$$
$$W_4O_{11} + 4WO_3 = W_8O_{23}; \qquad W_4O_{11} + 5WO_3 = W_9O_{26}$$
$$W_4O_{11} + 6WO_3 = W_{10}O_{29}; \qquad W_4O_{11} + 7WO_3 = W_{11}O_{32}$$

The basic formula of the group of four, W_4O_{11}, can be written as W_nO_{3n-1} where $n = 4$. This formula also holds for all the other formulae that are listed above, and we have produced the general formula for the homologous series simply by introducing set ratios of the edge-sharing groups in among the $[WO_6]$ octahedra.

The shear planes are found to repeat throughout a particular structure in a regular and ordered fashion, so any particular sample of WO_{3-x} will have a specific formula corresponding to one of those listed above. The different members of the homologous series are determined by the fixed spacing between the CS planes. An example of one of the structures is shown in Figure 5.35. A unit cell has been marked so that the ratio of $[WO_6]$ octahedra to the groups of four is clear. Within the marked

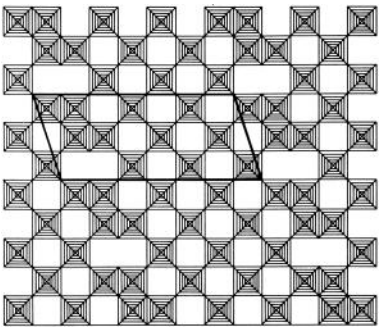


FIGURE 5.35 A member of the W_nO_{3n-1} homologous series with the projection of a unit cell marked.

unit cell there is one group of four and seven octahedra giving the overall formula $W_4O_{11} + 7WO_3 = W_{11}O_{32}$.

Members of the Mo_nO_{3n-1} series have the same structure as their W_nO_{3n-1} analog, even though unreduced MoO_3 does not have the ReO_3 structure, but a layer structure.

If the structures shear in such a way that groups of six octahedra share edges regularly throughout the structure, then homologous series with the general formula M_nO_{3n-2} are formed.

The homologous series for oxygen-deficient TiO_2 is given by the formula Ti_nO_{2n-1}. In this case, the octahedra along the CS planes are joined to each other by sharing *faces*, whereas in the unreduced parts of the TiO_2 structure, the octahedra share edges as in rutile.

5.8.2 Planar Intergrowths

Many systems show examples of intergrowth where a solid contains regions of more than one structure with clear solid–solid interfaces between the regions. Epitaxy (Chapter 3) is an example of such a phenomenon with great technological interest for the production of circuits and 'smart' devices. The zeolites ZSM-5 and ZSM-11 can form intergrowths (Chapter 6), as can the barium ferrites (Chapter 8). We will only look at one example here, that of intergrowths in the **tungsten bronzes**. The term *bronze* is applied to metallic oxides that have a deep colour, a metallic lustre, and are either metallic conductors or semiconductors. The sodium–tungsten bronzes (Na_xWO_3) have colours that range from yellow to red and deep purple, depending on the value of x.

We have already seen that WO_3 has the rhenium oxide (ReO_3) structure, with $[WO_6]$ octahedra joined through the corners. This is illustrated in Figures 5.32a and 5.33. The structure contains a three-dimensional network of channels throughout the structure and it has been found that alkali metals can be incorporated into the structure in these channels. The resultant crystal structure depends on the proportion of alkali metal in the particular compound. The structures are based on three main types: *cubic* phase where the alkali metal occupies the centre of the unit cell (similar to perovskite—see Chapters 1 and 9) and the *tetragonal* and *hexagonal* phases. The

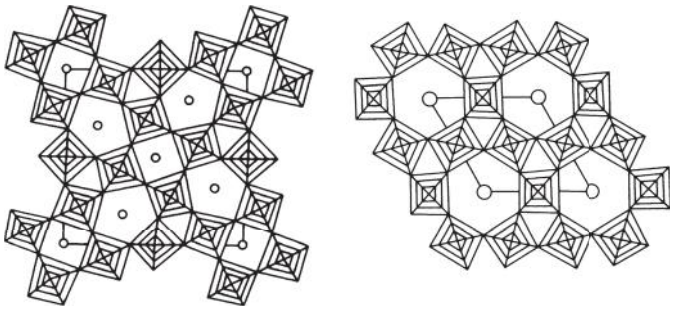


FIGURE 5.36 (a) The tetragonal tungsten bronze structure and (b) the hexagonal tungsten bronze structure. The shaded squares represent WO_6 octahedra, which are linked to form pentagonal, square and hexagonal tunnels. These are able to contain a variable population of metal atoms, shown as open circles.

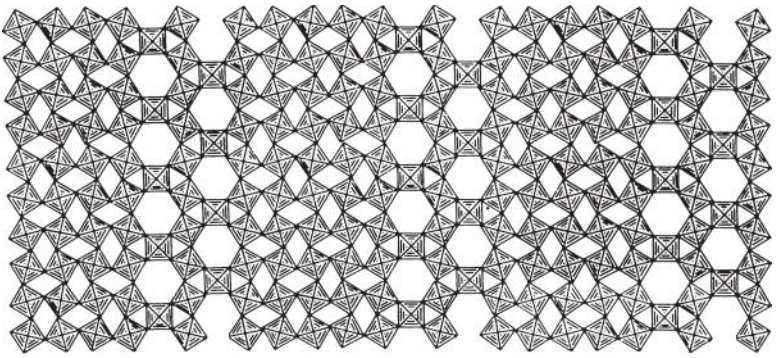
basic structures of two of these are illustrated in Figure 5.36. The electronic conductivity properties of the bronzes are because charge compensation has to be made for the presence of M^+ ions in the structure: this is achieved by the change in the oxidation state of some of the tungsten atoms from VI to V (such processes are discussed in more detail in the final section of this chapter).

The hexagonal bronze structure illustrated (Figure 5.36b) is formed when potassium reacts with WO_3 (K needs a bigger site) and the composition lies in the range $K_{0.19}WO_3$ to $K_{0.33}WO_3$. If the proportion of potassium in the compound is less than this, the structure is found to consist of WO_3 intergrown with the hexagonal structure in a regular fashion. The layers of the hexagonal structure can be either one or two tunnels wide, as shown in Figure 5.37. Similar structures are observed for tungsten bronzes containing metals other than potassium, such as Rb, Cs, Ba, Sn and Pb. A high-resolution electron micrograph of the barium tungsten bronze clearly shows its preferred single tunnel structure (Figure 5.38). In other samples, the separation of the tunnels increases as the concentration of barium decreases.

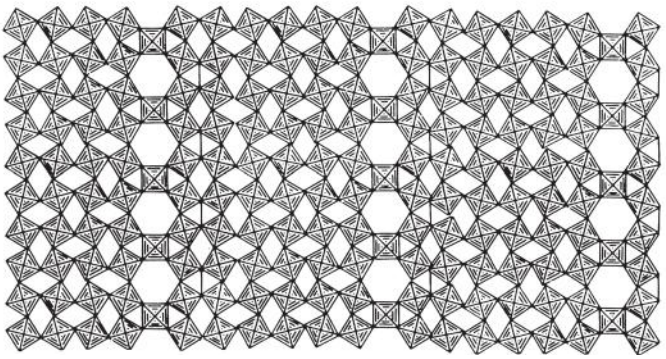
5.9 THREE-DIMENSIONAL DEFECTS

5.9.1 Block Structures

In oxygen-deficient Nb_2O_5 and in mixed oxides of Nb and Ti, and Nb and W, the CS planes occur in two sets at right angles to each other. The intervening regions of perfect structure now change from infinite sheets to infinite *columns* or *blocks*. These structures are known as **double shear** or **block structures** and are characterised by the cross-sectional size of the blocks: the block size is expressed as the number of octahedra sharing vertices. In addition to having phases built of blocks of one size, the complexity can be increased by having blocks of two or even three different sizes arranged in an ordered fashion. The block size(s) determines the overall stoichiometry of the solid. An example of a crystal showing two different block sizes is shown in Figure 5.39.



(a)



(b)

FIGURE 5.37 The idealised structures of two intergrowth tungsten bronze phases, (a) containing double rows of hexagonal tunnels and (b) containing single rows of tunnels. The tungsten trioxide matrix is shown as shaded squares and the hexagonal tunnels are shown empty, although in the known intergrowth tungsten bronzes the tunnels contain variable amounts of metal atoms.

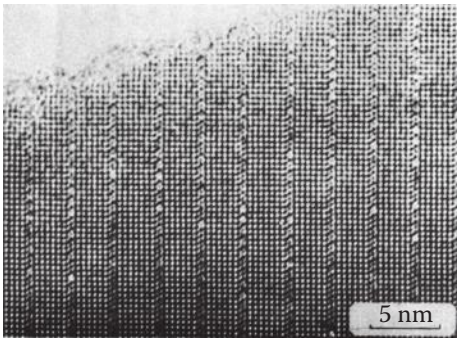


FIGURE 5.38 Electron micrograph of the intergrowth tungsten bronze phase Ba_xWO_3, clearly showing the single rows of tunnels. Each black spot on the image represents a tungsten atom, and many of the hexagonal tunnels seem to be empty or only partly filled with barium.

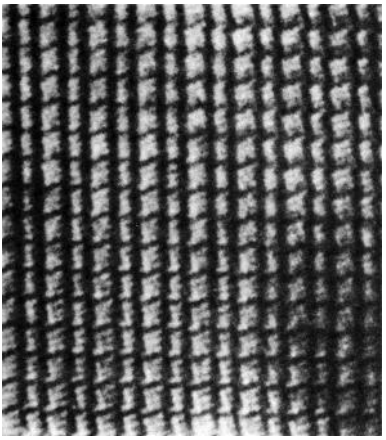


FIGURE 5.39 A high-resolution electron micrograph of the $W_4Nb_{26}O_{77}$ structure showing strings of (4×4) and (3×4) blocks. The CS planes between the blocks are shown as darker contrast. (From Dr. J. L. Hutchison.)

5.9.2 Pentagonal Columns

Three-dimensional faults also occur in the so-called **pentagonal column (PC) structures**. These structures contain the basic repeating unit shown in Figure 5.40a, which consists of a pentagonal ring of five $[MO_6]$ octahedra. When these stack on top of one another, a pentagonal column is formed that contains chains of alternating M and O atoms. These pentagonal columns can fit inside an ReO_3-type structure in an ordered way, and depending on the spacing, a homologous series is formed.

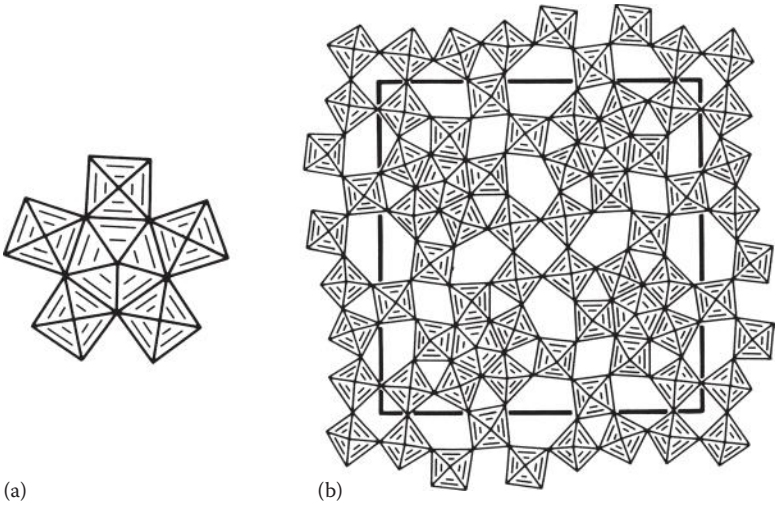


FIGURE 5.40 (a) A pentagonal column and (b) the structure of Mo_5O_{14}.

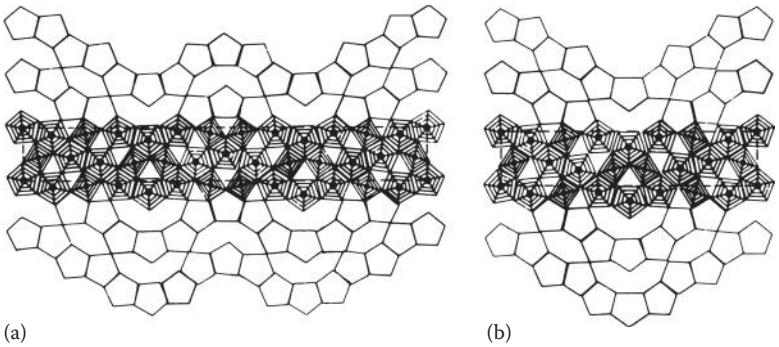


FIGURE 5.41 The idealised structures of (a) $Ta_{22}W_4O_{67}$ and (b) $Ta_{30}W_2O_{81}$. These phases are built from pentagonal columns (PCs), shown as shaded pentagons, and octahedra, shown as shaded squares. The wavelength of the chains of the PCs varies with composition in such a way that any given anion-to-cation ratio can be accommodated by an ordered structure.

An example is shown in Figure 5.40b for the compound Mo_5O_{14}. This type of structure is also found in the tetragonal tungsten bronzes.

5.9.3 Infinitely Adaptive Structures

In Section 5.9.1, we saw that the mixed oxides of niobium and tungsten could have a range of different compositions made by fitting together rectangular columns or blocks. The closely related Ta_2O_5–WO_3 system does something even more unusual. A large number of compounds form, but they are built up by fitting together pentagonal columns. The idealised structures of two of these compounds are shown in Figure 5.41: the structures have a *wavelike* skeleton of pentagonal columns. As the composition varies, so the wavelength of the backbone changes, giving rise to a huge number of possible ordered structures, known as **infinitely adaptive compounds**.

5.10 ELECTRONIC PROPERTIES OF NONSTOICHIOMETRIC OXIDES

Earlier, we considered the structure of nonstoichiometric FeO in some detail. If we apply the same principle to other binary oxides, we can define four types of compounds:

Metal excess (reduced metal)
Type A: anion vacancies present; formula MO_{1-x}
Type B: interstitial cations; formula $M_{1+x}O$

Metal deficiency (oxidised metal)
Type C: interstitial anions; formula MO_{1+x}
Type D: cation vacancies; formula $M_{1-x}O$

TABLE 5.7
Types of Nonstoichiometric Oxides (MO)

Metal Excess Reduced Metal M		Metal Deficiency Oxidised Metal M	
A Anion Vacancies MO_{1-x}	**B Interstitial Cations** $M_{1+x}O$	**C Interstitial Anions** MO_{1+x}	**D Cation Vacancies** $M_{1-x}O$
TiO, VO, (ZrS)	CdO, ZnO	—	TiO, VO, MnO, FeO, CoO, NiO

The four types are summarised in Table 5.7, and Figure 5.42 illustrates some of the structural possibilities for simple oxides with A-, B-, C- and D-type nonstoichiometry, assuming an NaCl-type structure.

Earlier, we looked in some detail at the structure of FeO; this falls into the **type D** category, with metal deficiency and cation vacancies resulting in oxidised metal. (Other compounds falling into this category are MnO, CoO and NiO.)

Type A oxides compensate for metal excess with anion vacancies. In order to maintain the overall neutrality of the crystal, two electrons have to be introduced for each anion vacancy. These can be trapped at a vacant anion site, as is shown in Figure 5.42a(i). However, it is an extremely energetic process to introduce electrons

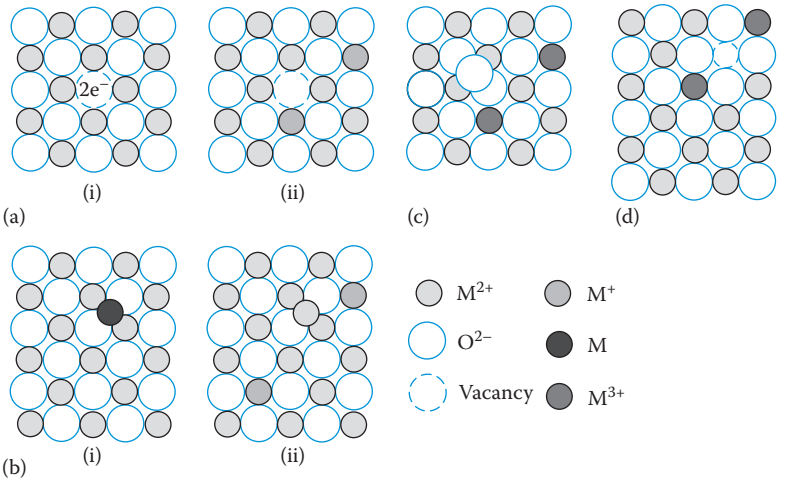


FIGURE 5.42 Structural possibilities for binary oxides. (a) Type A oxides: metal excess/anion vacancies. (i) This shows the two electrons that maintain charge neutrality, localised at the vacancy. (ii) The electrons are associated with the normal cations making them into M^+. (b) Type B oxides: metal excess/interstitials. (i) This shows an interstitial atom, whereas in (ii) the atom has ionised to M^{2+}, and the two liberated electrons are now associated with two normal cations, reducing them to M^+. (c) Type C oxides: metal deficiency/interstitial anions. The charge compensation for an interstitial anion is by way of two M^{3+} ions. (d) Type D oxides: metal deficiency/cation vacancies. The cation vacancy is compensated for by two M^{3+} cations.

into the crystal, so we are more likely to find them associated with the metal cations, as is shown in Figure 5.42a(ii), which we can describe as reducing those cations from M^{2+} to M^{+}. VO_{1-x} is an example of this type of system.

Type B oxides have a metal excess that is incorporated into the lattice in **interstitial** positions. This is shown in Figure 5.42b(i) as an interstitial atom, but it is more likely that the situation in Figure 5.42b(ii) will hold, where the interstitial atom has ionised and the two electrons so released are now associated with two neighbouring ions, reducing them from M^{2+} to M^{+}. Cadmium oxide (CdO) has this type of structure. Oxygen is lost when zinc(II) oxide is heated, forming $Zn_{1+x}O$, oxide vacancies form and to compensate, Zn^{2+} ions migrate to interstitial positions and are reduced to Zn^{+} ions or Zn atoms. Electron transfer can take place between the Zn^{2+} and Zn^{+}/Zn, resulting in the yellow colouration seen when ZnO is heated.

Type C oxides compensate for the lack of metal with **interstitial anions**. The charge balance is maintained by the creation of two M^{3+} ions for each interstitial anion (Figure 5.42c), each of which we can think of as M^{2+} associated with a positive hole.

Before considering the conductivity of these nonstoichiometric oxides, it is probably helpful to recap what we know about the structure and the properties of the stoichiometric binary oxides of the first-row transition elements. A summary of the properties of binary oxides is given in Table 5.8.

We discussed the conductivity of the stoichiometric oxides in Chapter 4, and we saw that their conductivity is dependent on two competing effects: on the one hand, the d orbitals overlap to give a band—the bigger the overlap the greater is the bandwidth—and electrons in the band are delocalised over the whole structure; on the other hand, interelectronic repulsion tends to keep the electrons localised on the individual atoms. TiO and VO behave as metallic conductors and must therefore have good overlap of the d orbitals producing a d electron band. This overlap arises partly because Ti and V are early in the transition series (the d orbitals have not suffered the contraction due to increased nuclear charge, seen later in the series) and partly because of the crystal structures (TiO has one-sixth of the titaniums and oxygens missing from an NaCl-type structure), which allow contraction of the structures and thus better d orbital overlap.

Further along the series, we saw that stoichiometric MnO, FeO, CoO and NiO are insulators. This situation is not easily described by band theory because the d orbitals are now too contracted to overlap much, typical bandwidths are 1 eV, and the overlap is not sufficient to overcome the localising influence of interelectronic repulsions. (It is this localisation of the d electrons on the atoms that gives rise to the magnetic properties of these compounds, which are discussed in Chapter 8.)

Going back now to the nonstoichiometric oxides, in the excess metal monoxides of type A and type B, we saw that extra electrons have to compensate for the excess metal in the structure. Figure 5.42 shows that these could be associated with an anion vacancy or, alternatively, they could be associated with the metal cations within the structure. Although we have described this association as reducing neighbouring cations, this association can be quite weak, and these electrons can be free to

TABLE 5.8
Properties of the First-Row Transition Element Monoxides

Element	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
Structure of stoichiometric oxide MO	NaCl structure	Does not exist	Defect NaCl 1/6 vacancies	Defect NaCl	Does not exist	NaCl structure	NaCl structure[a]	NaCl structure	NaCl structure	PtS structure	Wurtzite structure (NaCl at high pressure)
Defect structure			$Ti_{1-\delta}O$: Ti vacancies (intergrowths of $TiO_{1.00}$ and $TiO_{1.25}$ structures)	$V_{1-\delta}O$: V vacancies and tetrahedral V interstitials in defect clusters		$Mn_{1-\delta}O$: Mn vacancies	$Fe_{1-\delta}O$: Fe vacancies and tetrahedral Fe interstitials in defect clusters	$Co_{1-\delta}O$: Co vacancies	$Ni_{1-\delta}O$: Ni vacancies		$Zn_{1+\delta}O$: interstitial Zn
Conductivity of stoichiometric compound			Metallic	Metallic <120 K		Insulator	Insulator	Insulator	Insulator		Insulator
Conductivity of nonstoichiometric compound			Metallic	Metallic		*p*-Type hopping semiconductor	*p*-Type	*p*-Type	*p*-Type		*n*-Type
Magnetism (Chapter 8)			Diamagnetic	Diamagnetic		Paramagnetic $\mu = 5.5\mu_B$ (antiferromagnetic when cooled, $T_N = 122$ K)	Paramagnetic (antiferromagnetic when cooled, $T_N = 198$ K)	Antiferromagnetic ($T_N = 292$ K)	Antiferromagnetic (paramagnetic when cooled, $T_N = 530$ K)		

[a] Exactly stoichiometric FeO is never found.

move through the lattice; they are not necessarily strongly bound to particular atoms. Thermal energy is often sufficient to make these electrons move, so conductivity *increases* with temperature. We associate semiconductivity with such behaviour (metallic conductivity *decreases* with temperature).

In Chapter 4, we discussed semiconductors in terms of band theory. An intrinsic semiconductor has an empty conduction band lying close above the filled valence band. Electrons can be promoted into this conduction band by heating, leaving positive holes in the valence band; the current is carried by both the electrons in the conduction band and by the positive holes in the valence band. Semiconductors such as silicon can also be doped with impurities to enhance their conductivity. For instance, if a small amount of phosphorus is incorporated into the lattice, the extra electrons form impurity levels near the empty conduction band and are easily excited into it. The current is now carried by the electrons in the conduction band and the semiconductor is known as ***n*-type** (*n* for negative). Correspondingly, doping with Ga increases the conductivity by creating positive holes in the valence band and such semiconductors are called ***p*-type** (*p* for positive).

Compounds of type A and B would produce *n*-type semiconductors because the conduction is produced by electrons. Conduction in these nonstoichiometric oxides is not easily described by band theory, for the reasons given earlier for their stoichiometric counterparts—the interelectronic repulsions have localised the electrons on the atoms. Therefore, it is easiest to think of the conduction electrons (or holes) as localised or trapped at atoms or at defects in the crystal rather than delocalised in bands throughout the solid. Conduction then occurs by jumping or **hopping** from one site to another under the influence of an electric field. In a perfect ionic crystal where all the cations are in the same valence state, this would be an extremely energetic process. However, when two valence states, such as Zn^{2+} and Zn^{+}, are available, as in these transition metal nonstoichiometric compounds, the electron jump between them does not take much energy. Although we cannot develop this theory here, we can note that the conduction in these so-called **hopping semiconductors** can be described by the equations of diffusion theory in much the same way as we did earlier for ionic conduction: we find that the mobility of a charge carrier (either an electron or a positive hole), μ, is an activated process and we can write

$$\mu \propto \exp\left(\frac{-E_a}{kT}\right) \tag{5.16}$$

where E_a is the activation energy of the hop and is of the order of 0.1–0.5 eV. The hopping conductivity is given by the expression:

$$\sigma = ne\mu \tag{5.17}$$

where n is the number of mobile charge carriers per unit volume and e is the electronic charge. (Notice that these equations are analogous to Equations 5.8 and 5.7 in Section 5.3, describing ionic mobility (μ) and ionic conductivity (σ),

respectively.) The density of mobile carriers, n, depends only on the composition of the crystal and does not vary with temperature. From Equation 5.16, we can see that, as for ionic conductivity, the hopping electronic conductivity increases with temperature.

In the type C and D monoxides, we have shown the lack of metal as being compensated for by the oxidation of the neighbouring cations to M^{3+}. The M^{3+} ions can be regarded as M^{2+} ions associated with a positive hole. Accordingly, if sufficient energy is available, conduction can be thought to occur via the positive hole hopping to another M^{2+} ion and the electronic conductivity in these compounds will be *p*-type. MnO, CoO, NiO and FeO are materials that behave in this way. We saw this behaviour of hopping semiconduction described for NiO in Chapter 4, where it was described in terms of electron hopping. Regarding the charge carriers as positive holes is simply a matter of convenience and the description of a positive hole moving from Ni^{3+} to Ni^{2+} is the same as saying that an electron moves from Ni^{2+} to Ni^{3+}.

Nonstoichiometric materials that cover the whole range of electrical activity from metal to insulator can be listed. Here, we have considered some metallic examples that can be described by band theory (TiO, VO) and others (such as MnO) that are better described as hopping semiconductors. Other cases, such as WO_3 and TiO_2, fall in-between these extremes and a different description again is needed. We also met nonstoichiometric compounds such as calcia-stabilised zirconia and β-alumina, which are good ionic conductors. Indeed, stabilised zirconia exhibits *both* electronic and ionic conductivities though, fortunately for its industrial usefulness, electronic conduction only occurs at low oxygen pressures. It is thus difficult to make generalisations about this complex behaviour and each case is best treated individually.

The semiconductor properties are extremely important to the modern electronics industry, which is constantly searching for new and improved materials. Much of their research is directed at extending the composition range and therefore the properties of these materials. The composition range of a nonstoichiometric compound is often quite narrow (the so-called line phase), so to extend it (and thus extend the range of its properties also) the compound is doped with an impurity. To take one example: if we add Li_2O to NiO and then heat to high temperatures in the presence of oxygen, Li^+ ions become incorporated in the lattice and the resulting black material has the formula $Li_xNi_{1-x}O$, where x lies in the range 0–0.1. The equation for the reaction (using stoichiometric NiO for simplicity) is given by

$$\tfrac{1}{2}x\mathrm{Li_2O} + (1-x)\mathrm{NiO} + \tfrac{1}{4}x\mathrm{O_2} = \mathrm{Li}_x\mathrm{Ni}_{1-x}\mathrm{O}$$

To compensate for the presence of the Li^+ ions, the Ni^{2+} ions will be oxidised to Ni^{3+} or the equivalent of a high concentration of positive holes located at Ni cations.

This process of creating electronic defects is called **valence induction** and it increases the composition range of 'NiO' tremendously. Indeed, at high Li

concentrations, the conductivity approaches that of a metal (although it still exhibits semiconductor behaviour in that its conductivity increases with temperature).

5.11 CONCLUSIONS

In this chapter, we have tried to give some idea of the size and complexity of this subject and also its fascination, without it becoming overwhelming. The main point to emerge from our explorations is that the concept of random, isolated point defects does *not* explain the complex structures of nonstoichiometric compounds, but that there are many different ways in which defects cluster together and become ordered or even eliminated. It is these defects that can be exploited to give solids properties that are useful in many technologically useful devices.

QUESTIONS

1. If $\square H_m^{\ominus}$ for the formation of Schottky defects in a certain MX crystal is 200 kJ mol^{-1}, calculate n_S/N and n_S per mole for the temperatures 300, 500, 700 and 900 K.
2. Table 5.9 gives the variation of defect concentration with temperature for CsI. Determine the enthalpy of formation for one Schottky defect in this crystal.
3. If we increase the quantity of impurity (say $CaCl_2$) in the NaCl crystal, how will this affect the plot shown in Figure 5.8? How is the transition point in the graph affected by the purity of the crystal?
4. In NaCl, the cations are more mobile than the anions. What effect, if any, do you expect small amounts of the following impurities to have on the conductivity of NaCl crystals: (a) AgCl, (b) $MgCl_2$, (c) NaBr, (d) Na_2O?
5. Unlike fluorides, pure oxides with the fluorite structure (MO_2) show high anion conduction only at elevated temperatures, above about 2300 K. Suggest a reason for this.

TABLE 5.9
Defect Concentration Data for CsI

T/K	$\frac{n_S}{N}$
300	1.08×10^{-16}
400	1.06×10^{-12}
500	2.63×10^{-10}
600	1.04×10^{-8}
700	1.43×10^{-7}
900	4.76×10^{-6}

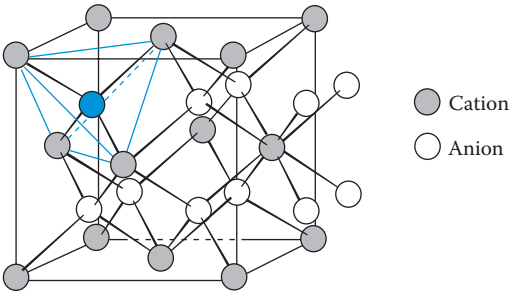


FIGURE 5.43 The fluorite structure, showing the coordination tetrahedron around one of the anions.

6. Figure 5.43 shows the fluorite structure with the tetrahedral environment of one of the anions shaded. (The anion behind this tetrahedron has been omitted for clarity.) Suppose this anion jumps to the octahedral hole at the body centre. Describe, and sketch, the pathway it takes in terms of the changing coordination by cations.
7. Make a simple estimate of the energy of the defect formation in the fluorite structure:
   a. Describe the coordination by nearest neighbours and next-nearest neighbours of an anion both for a normal lattice site and for an interstitial site at the centre of the unit cell shown in Figure 5.3a.
   b. Use the potential energy of two ions, given by

$$E = -\frac{e^2 Z}{4\pi\varepsilon_0 r}$$

$$= -\left(2.31\times10^{-28}\ \mathrm{J\,m}\right)\frac{Z}{r}$$

   where Z is the charge on the other ion and r is the distance between them, to estimate the energy of the defect formation in fluorite; $a = 537$ pm.
8. The zinc blende structure of γ-AgI, a low-temperature polymorph of AgI, is shown in Figure 5.44. Discuss the similarities and differences between this structure and that of α-AgI. Why do you think that the conductivity of the Ag^+ ions is lower in γ-AgI?
9. The compounds in Table 5.4 mostly contain either I^- ions or ions from the heavier end of Group 6. Explain.
10. Undoped β-alumina shows a maximum conductivity and minimum activation energy when the sodium excess is around 20–30 mol%. Thereafter, a further increase in the sodium content causes the conductivity to decrease. By contrast, β-alumina crystals doped with Mg^{2+} have a much higher conductivity than do undoped crystals. Explain these observations.

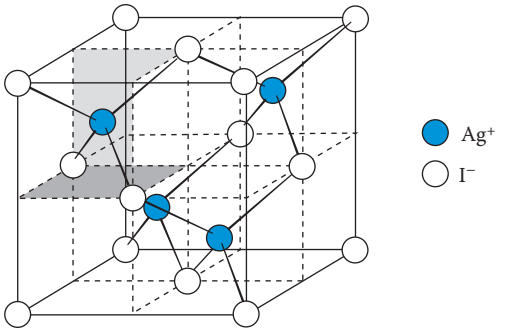


FIGURE 5.44 The zinc blende structure of γ-AgI.

11. Confirm the presence of iron vacancies for a sample of wustite that has a unit cell dimension of 428.2 pm, an Fe:O ratio of 0.910 and an experimental density of 5.613×10^3 kg m^{-3}.
12. How does the change in the lattice parameter of 'FeO' with iron content corroborate the iron vacancy model and refute an oxide interstitial model?
13. How would you expect the formation of colour centres to affect the density of the crystal?
14. Figure 5.45 shows the central section of a possible defect cluster for FeO.
    a. Determine the vacancy:interstitial ratio for this cluster.
    b. Assuming that this section is surrounded by Fe ions and oxide ions in octahedral sites as in the Koch–Cohen cluster, determine the formula of a sample made totally of such clusters.
    c. Determine the numbers of Fe^{2+} and Fe^{3+} ions in octahedral sites.
15. Use Figure 5.28 to confirm that TiO is a one-sixth defective NaCl structure, by counting up the atoms in the monoclinic cell.
16. Figure 5.29 shows layers in the $TiO_{1.25}$ structure with both the Ti and O sites marked. Use this and Figure 5.30 to demonstrate that the unit cell shown has the correct stoichiometry for the crystal.
17. How would you expect charge neutrality to be maintained in $TiO_{1.25}$?

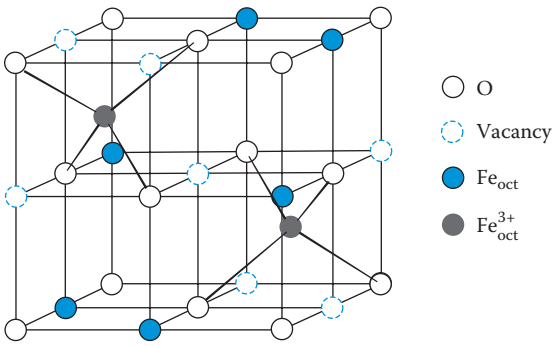


FIGURE 5.45 A possible cluster in $Fe_{1-x}O$.

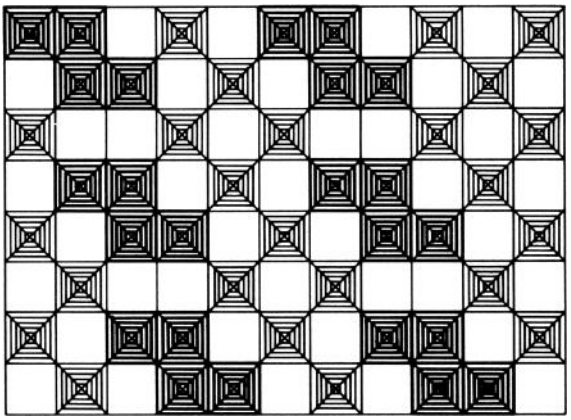


FIGURE 5.46 A member of the W_nO_{3n-1} homologous series.

18. Take a simple case where *two* metal oxide octahedra wish to eliminate oxygen by sharing. How does the formula change as they (i) share a corner, (ii) share an edge and (iii) share a face?
19. Figure 5.46 shows a member of the homologous series, W_nO_{3n-1}. To what formula does it correspond?
20. ZnO is a type B (excess metal) material. What do you expect to happen to its electronic properties if it is doped with Ga_2O_3 under reducing conditions?

6 Microporous and Mesoporous Solids

6.1 ZEOLITES

Over 60 naturally occurring zeolites have been characterised; however, in the quest for new catalysts, more than 150 entirely synthetic structures have been developed. Only seven (analcime, chabazite, clinoptilolite, erionite, ferrierite, mordenite and phillipsite) zeolites occur in large deposits, mostly in China and Cuba. Because of their many uses, zeolites are manufactured synthetically, and they represent a large proportion of the chemical industry.

Zeolites are used as absorbents for purification, as cation exchangers for water softening, as molecular sieves for separating molecules of different sizes and shapes (e.g., as drying agents and in catalysis) and as templating agents for the preparations of nanoparticles.

For many years, research has focused on the ability of the zeolites to act as catalysts in a wide variety of reactions, many of them highly specific, and they are used extensively in the industry for this purpose, particularly in petroleum refining, being worth many billions of pounds to the economies of many countries.

Natural zeolites are used mainly as lightweight building materials, but they also find use as adsorbers for cleaning wastewaters and for pet litter. About 80% of the synthetically produced zeolites are used in detergents for softening water by ion exchange (removing calcium ions mainly); specialised catalysts and adsorbers make up the rest of the market.

In 1756, the Swedish mineralogist Baron Axel Fredrik Cronstedt first described zeolites as a mineral group. They are a class of crystalline **aluminosilicates** based on rigid anionic frameworks with well-defined **pores** (**channels**) running through them, which intersect at **cavities** (**cages**). These cavities contain exchangeable metal cations (Na^+, K^+, etc.) and can also hold removable and replaceable guest molecules (water in naturally occurring zeolites). It is their ability to lose water on heating that has earned them their name: Cronstedt observed that on heating with a blowtorch, they hissed and bubbled as though they were boiling and thus named them zeolites from the Greek words *zeo*, to boil, and *lithos*, stone. This loss of water is completely reversible.

With cavity sizes of between 200 and 2000 pm, zeolites are classified as **microporous** solids.

6.1.1 Composition and Structure of Zeolites

The general formula for the composition of a zeolite is $M_{x/n}[(AlO_2)_x(SiO_2)_y]\cdot mH_2O$, where cations M of valence n neutralise the negative charges on the aluminosilicate framework.

6.1.2 Frameworks

The primary building units of zeolites are $[SiO_4]^{4-}$ and $[AlO_4]^{5-}$ tetrahedra (Chapter 1) linked together by **corner sharing**, forming oxygen bridges (Figure 6.1). The oxygen bridge is not usually linear, but the Si/Al-O-Si/Al linkage is very flexible and the angle can vary between 120° and 180°. The silicon–oxygen tetrahedra are electrically neutral when connected together in a three-dimensional network as in quartz (SiO_2) (Figure 6.2). The substitution of Si(IV) by Al(III) in such a structure, however, creates an electrical imbalance, and to preserve overall electrical neutrality, each $[AlO_4]$ tetrahedron needs a balancing positive charge. This is provided by exchangeable cations, such as Na^+, held electrostatically within the zeolite.

It is possible for the tetrahedra to link by sharing two, three or all four corners, thus forming a variety of different structures. The linked tetrahedra are usually illustrated by drawing a straight line to represent the oxygen bridge connecting two tetrahedral units. In this way, the six linked tetrahedra in Figure 6.3a and 6.3b are simply represented by a hexagon (Figure 6.3c). This is known as a **6-ring**, and a tetrahedrally coordinated Si or Al atom occurs at each intersection between two straight lines. As we will see later, many different ring sizes are found in the various zeolite structures.

Many zeolite structures are based on a secondary building unit that consists of 24 silica or alumina tetrahedra linked together; here we find 4-rings and 6-rings linked together to form a basket-like structure called the **sodalite unit** (also known as the **β-cage**), shown in Figure 6.4, which has the shape of a **truncated octahedron** (Figure 6.5). Several of the most important zeolite structures are based on the sodalite unit (Figure 6.6).

The mineral sodalite is composed of these units, with each 4-ring shared directly by two β-cages in a primitive array. Note that the cavity or cage enclosed by the eight sodalite units shown in Figure 6.6a is *itself* a sodalite unit, that is, sodalite units are space-filling. In this three-dimensional structure, a tetrahedral Si or Al atom is located at the intersection of four lines, as oxygen bridges are made by corner sharing from all four vertices of the tetrahedron. Sodalite is a highly symmetrical structure and the cavities link together to form the channels or pores, which run parallel to all three cubic crystal axes; the entrance to these pores is governed by the 4-ring window.

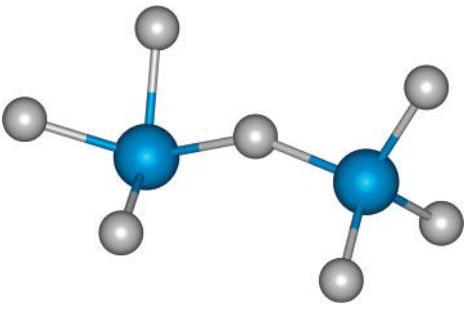


FIGURE 6.1 The zeolite building units. Two SiO_4/AlO_4 tetrahedra linked by corner sharing.

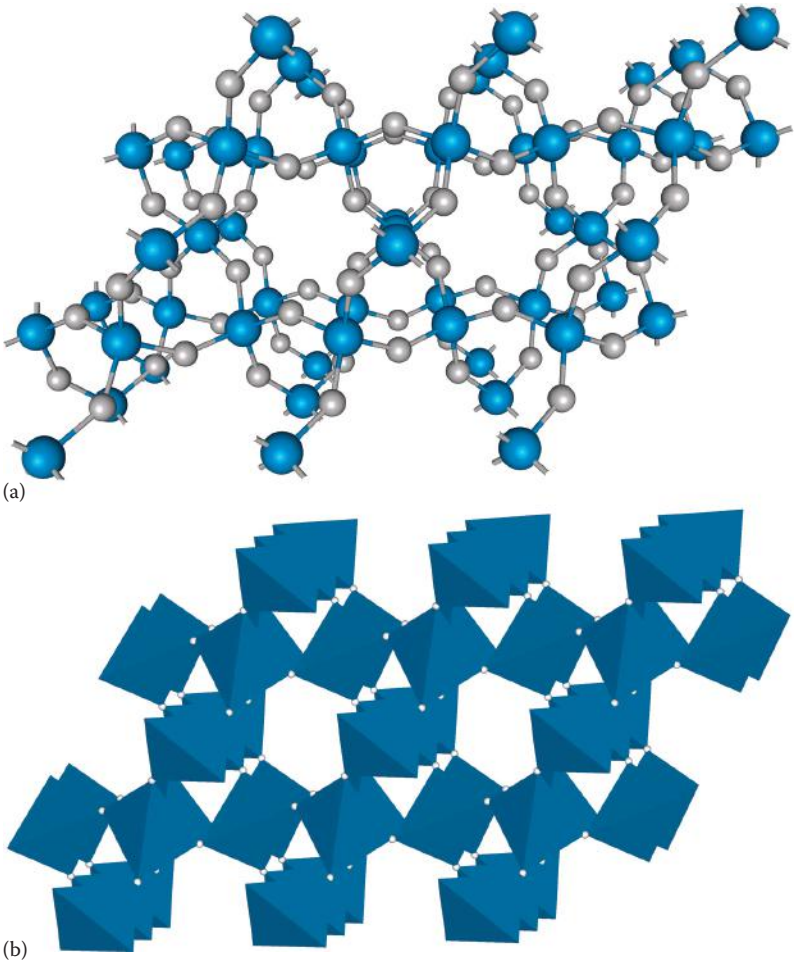


FIGURE 6.2 The structure of quartz (a) as a ball and stick representation and (b) as linked [SiO_4] tetrahedra.

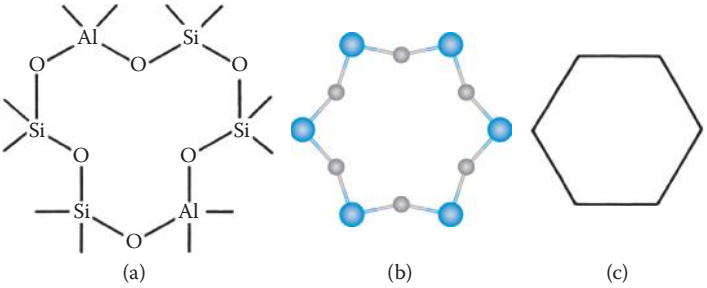


FIGURE 6.3 (a) A 6-ring containing two Al and four Si atoms, (b) a computer model of the 6-ring and (c) a shorthand version of the same 6-ring.

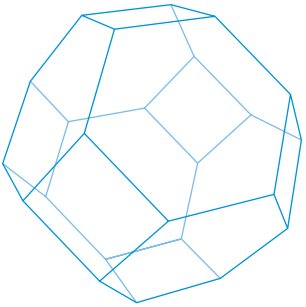


FIGURE 6.4 The sodalite unit.

A synthetic zeolite, **zeolite A** (also called Linde A), is shown in Figure 6.6b. Here, the sodalite units are again stacked in a primitive array, but now they are linked by oxygen bridges between the 4-rings. A three-dimensional network of linked cavities, each with a **truncated cuboctahedron** shape (Figure 6.5) through the structure, is thus formed; the truncated cuboctahedra are also space-filling (each shares its octagonal face with six others), forming channels that run parallel to the three cubic axial directions through these large cavities. The computer-drawn models in Figure 6.7 are of the zeolite A framework, showing the cavity and its 8-ring window more clearly and how it links to a sodalite cage. The formula of zeolite A is given by $Na_{12}[(SiO_2)_{12}(AlO_2)_{12}]\cdot 27H_2O$. In this fairly typical example, the Si/Al ratio is unity, and we find that in the crystal structure, the Si and Al atoms strictly alternate.

The structure of **faujasite**, a naturally occurring mineral, is shown in Figure 6.6c. The sodalite units are linked by oxygen bridges between four of the eight 6-rings in

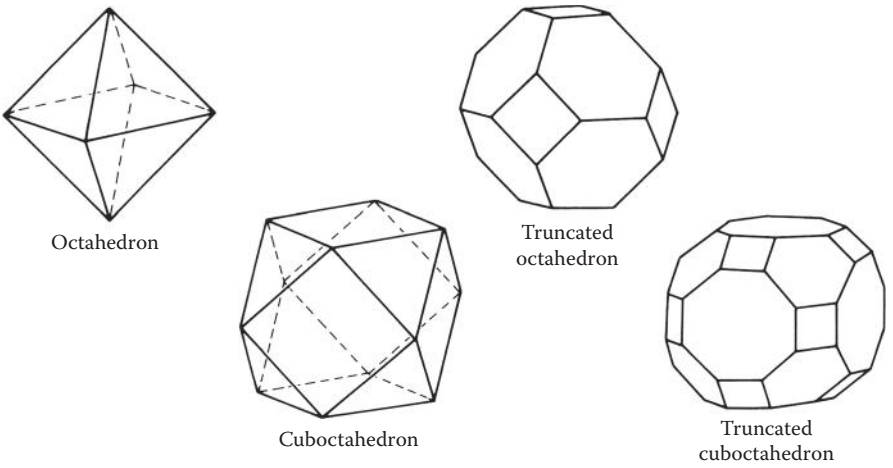


FIGURE 6.5 The relationship between an octahedron, a truncated octahedron, a cuboctahedron and a truncated cuboctahedron.

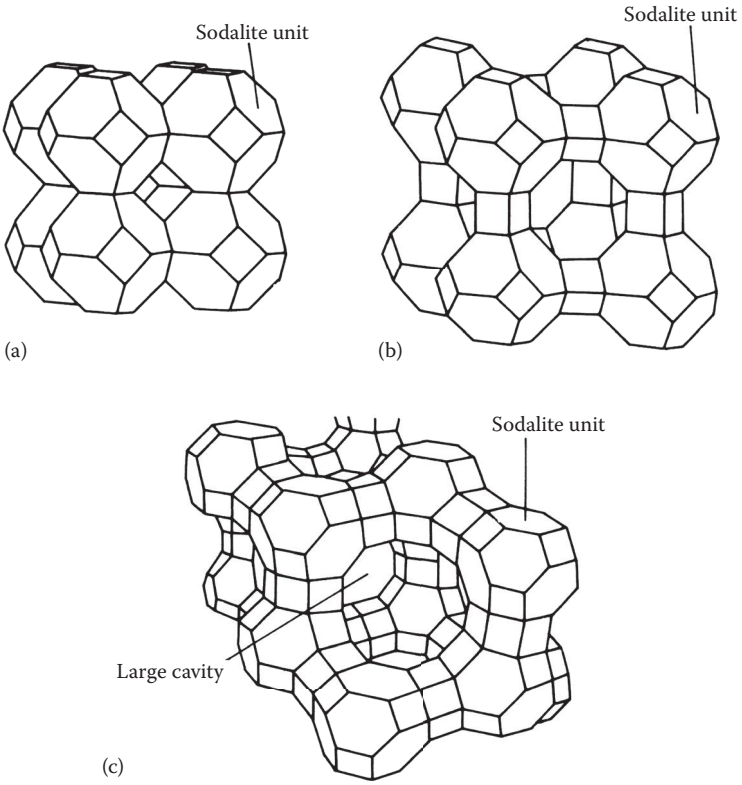


FIGURE 6.6 The zeolite frameworks built up from the sodalite units: (a) sodalite (SOD), (b) zeolite A (LTA) and (c) faujasite (zeolite X and zeolite Y) (FAU).

a tetrahedral array. The tetrahedral array encloses a large cavity (sometimes known as the α-cage) entered through a 12-ring window. The synthetic zeolites, **zeolite X** and **zeolite Y** (Linde X and Linde Y), also have this basic underlying structure. The zeolite X structures have an Si/Al ratio between 1 and 1.5, whereas the zeolite Y structures have Si/Al ratios between 1.5 and 3.

6.1.3 Synthesis of Zeolites

It was the pioneering work of Richard Barrer at the University of Aberdeen and Bob Milton at Union Carbide that led the way for the synthesis of zeolites. They were able to prepare zeolites using reactive silica and alumina reagents, such as sodium silicates and aluminates ($[Al(OH)_4]^-$), under hydrothermal conditions, at a high pH obtained by using an alkali metal hydroxide and/or an organic base. A gel forms by a process of copolymerisation of the silicate and aluminate ions. The gel is then heated gently (60°C–100°C) in an autoclave for several days, producing a condensed zeolite. The product obtained is determined by the synthesis conditions; temperature, time,

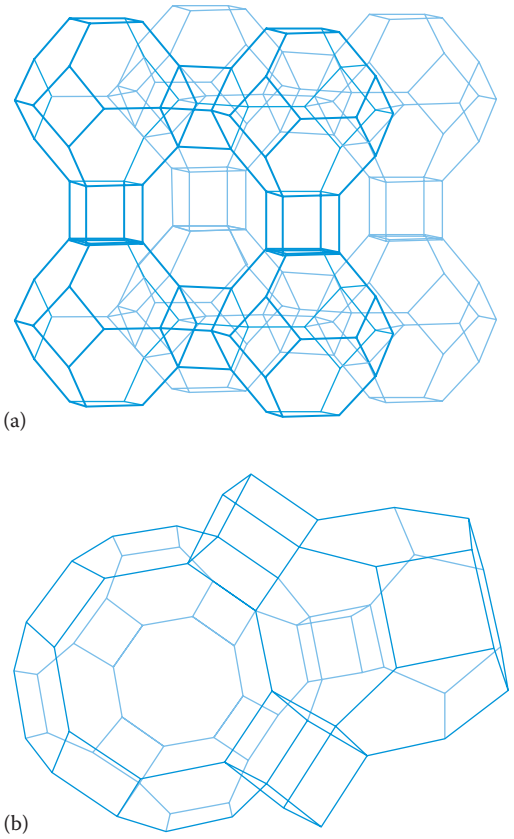


FIGURE 6.7 (a) The zeolite A framework and (b) a sodalite unit in zeolite A, showing the linkage to the truncated cuboctahedral cavity.

pH and mechanical movement are all possible variables. The presence of organic bases is useful for synthesising silicon-rich zeolites.

The formation of novel, silicon-rich synthetic zeolites has been facilitated by the use of **templates**, such as large quaternary ammonium cations instead of Na^+. For instance, the tetramethylammonium cation ($[(CH_3)_4N]^+$) is used in the synthesis of ZK-4. The aluminosilicate framework condenses around this large cation, which can subsequently be removed by chemical or thermal decomposition. ZSM-5 is produced in a similar way, using the tetra-*n*-propyl ammonium ion. Only a limited number of large cations can fit into the zeolite framework, and this severely reduces the number of $[AlO_4]$ tetrahedra that can be present, producing a silicon-rich structure.

The preparation of silicon-rich zeolites, such as zeolite Y, can be achieved by varying the composition of the starting materials, but it can also be done by subsequent removal of aluminium from a synthesised aluminosilicate framework using a chemical treatment. There are several different methods available; one being the extraction of the aluminium by mineral acid and another being the extraction using complexing agents.

6.1.4 Zeolite Nomenclature

The naming of zeolites and related structures has been somewhat unsystematic. Some structures are named after the parent minerals, for example, sodalite and faujasite, while others were named by researchers, or after the projects that synthesised them, for example, ZSM (Zeolite Socony Mobil). Unfortunately, this led to the same zeolites being synthesised by different routes, bearing different names—in some cases, up to 20 different trade names!

The International Zeolite Association (IZA), with the permission of the International Union of Pure and Applied Chemistry (IUPAC), introduced a three-letter structure code for unique framework topologies to try and simplify matters; zeolite A and the more silicon-rich zeolite ZK-4 have the same framework structure and are designated LTA. Similarly, ZSM-5 and its silicon-rich relation, silicalite, have the same framework and are both designated MFI.

The zeolites are also often written as M-[zeolite], where M refers to the particular cation in the structure, for example, Ca-zeolite A or even simply CaA or NaA.

6.1.5 Si/Al Ratios in Zeolites

The zeolites A and X are 'aluminium-rich' zeolites with Si/Al ratios of close to 1. These zeolites have very useful absorbent properties and are still used industrially, but are unstable and lose aluminium when attacked by acid, steam or water. The natural zeolite mordenite, which is the most siliceous of the naturally occurring zeolite minerals, was found to be more stable and because it has an Si/Al ratio of >5, research turned to synthesising zeolites with higher ratios in the hope of greater stability. Many of the new synthetic zeolites that have been developed for catalysis are thus highly siliceous: zeolite Y has an improved stability with a ratio of 2.25; zeolite ZK-4 (LTA), with the same framework structure as zeolite A, has a ratio of 2.5; and **ZSM-5** (MFI) can have an Si/Al ratio that lies between 20 and ∞ (the latter, called silicalite [see above], being virtually pure SiO_2), which far outstrips the ratio of 5.5 found in mordenite.

Clearly, changing the Si/Al ratio of a zeolite also changes its cation content; the fewer the aluminium atoms, the fewer cations will be present to balance the charges. The highly siliceous zeolites are inherently hydrophobic in character and their affinity is for hydrocarbons.

6.1.6 Exchangeable Cations

The zeolite Si/Al-O framework is rigid, but the cations are not an integral part of this framework and are often called **exchangeable cations**: they are fairly mobile and readily replaced by other cations (hence their use as cation-exchange materials).

The presence and position of the cations in zeolites are important for several reasons. The cross section of the rings and the channels in the structures can be altered by changing the size or the charge (and thus the number) of the cations, and this significantly affects the size of the molecules that can be adsorbed. A change in the cationic occupation also changes the charge distribution within the cavities and therefore the adsorptive behaviour and the catalytic activity.

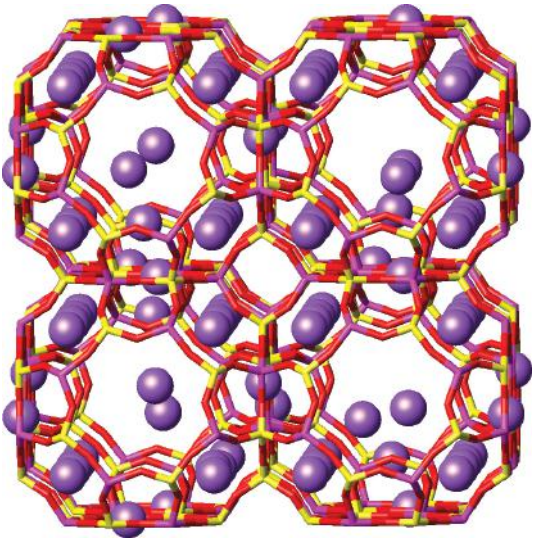


FIGURE 6.8 Framework and cation sites in the Na^+ form of zeolite A (LTA). (From Robert G. Bell, Royal Institution of Great Britain, London, UK. With permission.)

The balancing cations in a zeolite can have more than one possible location in the structure. Figure 6.8 shows the available sites in the Na^+ form of zeolite A. Some sites occupy most of the centres of the 6-rings, while other sites are in the 8-ring entrances to the β-cages. The presence of cations in these positions effectively reduces the size of the rings and the cages to any guest molecules that are trying to enter. In order to alter a zeolite to allow organic molecules, for instance, to diffuse into or through the zeolite, a divalent cation such as Ca^{2+}, can be exchanged for the univalent Na^+ or K^+ ions, not only halving the number of cations present but also replacing them with a smaller ion. As the divalent cations tend to occupy the sites in the 6-rings, this opens up the 8-ring windows, thus leaving the channels free for diffusion.

The normal crystalline zeolites contain water molecules that are coordinated to the exchangeable cations. These structures can be dehydrated by heating under vacuum, and in these circumstances, the cations move position at the same time, frequently settling on the sites with a lower coordination number. The dehydrated zeolites are extremely good drying agents, absorbing water to get back to their preferred hydrated condition.

The exchange of the metal cations for protons, usually by exchanging the metal ion with ammonium (NH_4^+), and then heating to drive off ammonia (NH_3), produces strongly acidic materials that have proved to be highly active acid catalysts.

6.1.7 Channels and Cavities

The important structural feature of zeolites, which can be exploited for various uses, is the network of linked cavities forming a system of channels throughout the

structure. These cavities are of molecular dimensions and can adsorb species small enough to gain access to them. A controlling factor in whether the molecules can be adsorbed in the cavities is the size of the **window** or the **port** into the channel; hence the importance of the number of tetrahedra forming the window, that is, the ring size. Figure 6.9 shows how the window sizes can vary.

The windows to the channels thus form a three-dimensional sieve with mesh widths of between about 300 and 1000 pm; hence the well-known name **molecular sieve** for these crystalline aluminosilicates. The zeolites thus have large internal surface areas and high sorption capacities for molecules small enough to pass through the window into the cavities. They can be used to separate mixtures such as straight-chain and branched-chain hydrocarbons.

The zeolites fall into three main categories. The channels may be parallel to (a) a single direction, so that the crystals are fibrous; (b) two directions arranged in planes, so that the crystals are lamellar; or (c) three directions, such as cubic axes, in which there is strong bonding in three directions. The most symmetrical structures have cubic symmetry. By no means do all zeolites fall neatly into this classification; some, ZSM-11,

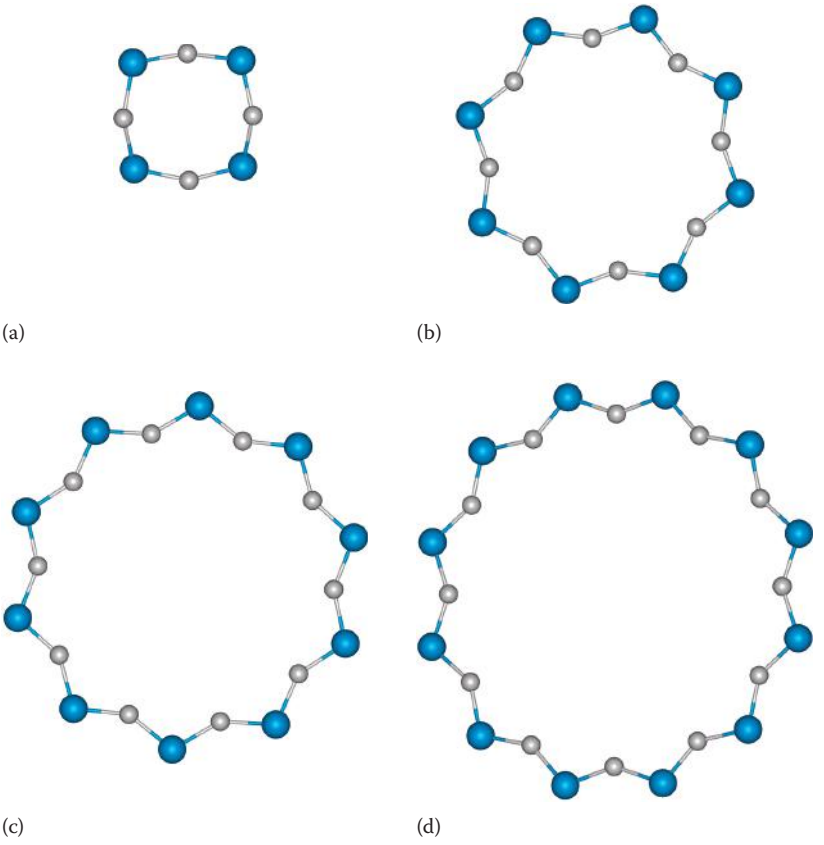


FIGURE 6.9 Computer models of various window sizes in zeolites, (a) 4-ring, (b) 8-ring, (c) 10-ring and (d) 12-ring.

for instance, have a dominant two-dimensional structure interlinked by smaller channels. A typical fibrous zeolite is edingtonite (EDI, $Ba[(AlO_2)_2(SiO_2)_3]4H_2O$), which has a characteristic chain formed by the regular repetition of five tetrahedra. The lamellar zeolites occur frequently in sedimentary rocks, for example, phillipsite (PHI, $(K/Na)_5[(SiO_2)_{11}(AlO_2)_5]\cdot 10H_2O$), which is a well-known example. In terms of their useful properties, zeolites are conveniently discussed based on their pore size.

Sodalite is a good example of a **small-pore zeolite**. A cavity in sodalite (the β-cage) is bounded by a 4-ring with a diameter of 260 pm; although this is a very small opening, it can admit water molecules (Figure 6.6) and can be used to dry gas streams.

The channels in zeolite A run parallel to the three cubic axial directions and are entered by a port of diameter 410 pm, determined by an 8-ring window; this is still considerably smaller than the diameter of the internal cavity, which measures 1140 pm across. The computer model of zeolite A in Figure 6.10 clearly shows the 8-ring windows, the channels running through the structure and the cavities created by their intersection.

Small-pore zeolites can accommodate linear-chain molecules, such as straight-chain hydrocarbons and primary alcohols and amines, but not branched-chain molecules. The port size can be enlarged to about 500 pm in diameter by replacing sodium ions with calcium ions. CaA was first used in 1959 by Union Carbide to absorb out the C_5 and C_6 alkanes in gasoline to leave it enriched with the higher octane hydrocarbons.

The values for channel and cavity sizes for various zeolites and zeotypes are shown in Table 6.1.

In the mid-1970s, some completely novel **medium-pore zeolite** structures were synthesised, which led to significant new developments. This family of framework structures, comprising the zeolites synthesised by the oil company Mobil, ZSM-5 and ZSM-11 (MEL), silicalite (MFI) and some closely related natural zeolites, have been given the generic name **pentasil**.

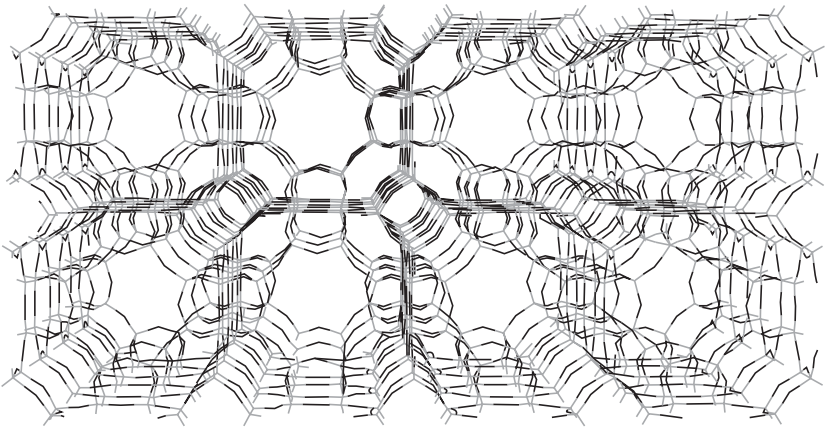


FIGURE 6.10 A computer model of the zeolite A (LTA) structure, showing the channels and the 8-ring windows.

TABLE 6.1
Window and Cavity Diameters in Zeolites

Zeolite/Zeotype	Structure Type Code	No. of Tetrahedra in Ring	Window Diameter/pm	Cavity Diameter/pm
Sodalite	SOD	4	260	600
Zeolite A	LTA	8	410	1140
Erionite-A	ERI	8	360 × 520	
ZSM-5	MFI	10	510 × 550	
			540 × 560	
Faujasite	FAU	12	740	1180
Mordenite	MOR	12	670 × 700	
			290 × 570	
Zeolite-L	LTL	12	710	
ALPO-5 (see Section 6.6)		12	800	
VPI-5		18	1200–1300	

ZSM-5 is a catalyst that is now widely used in the industrial world. Its structure is generated from the pentasil unit shown in Figure 6.11 (as are the others of this group). These units link into chains, which join to make layers. The appropriate stacking of these layers gives the various pentasil structures. Both ZSM-5 and ZSM-11 are characterised by channels controlled by 10-ring windows with diameters of about 550 pm. The pore systems in these zeolites do not link big cavities, but they do contain intersections where larger amounts of free space are available for molecular interactions to take place. Figure 6.12 shows computer models of the structures of ZSM-5 and ZSM-11. Figure 6.13a shows the pore system of ZSM-5 with nearly circular zigzag channels intersecting with straight channels of elliptical cross section; this contrasts with the ZSM-11 structure, which just has intersecting straight channels of almost circular cross section (Figure 6.13b); in both cases, the two-dimensional system of pores is linked by much smaller channels.

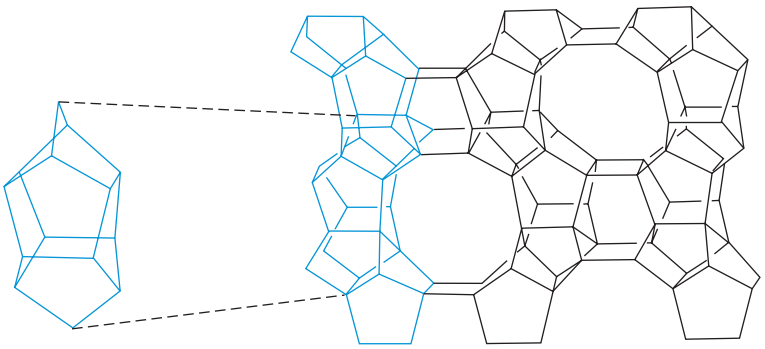


FIGURE 6.11 A pentasil unit together with a slice of the structure of ZSM-5, showing a linked chain of pentasil units highlighted.

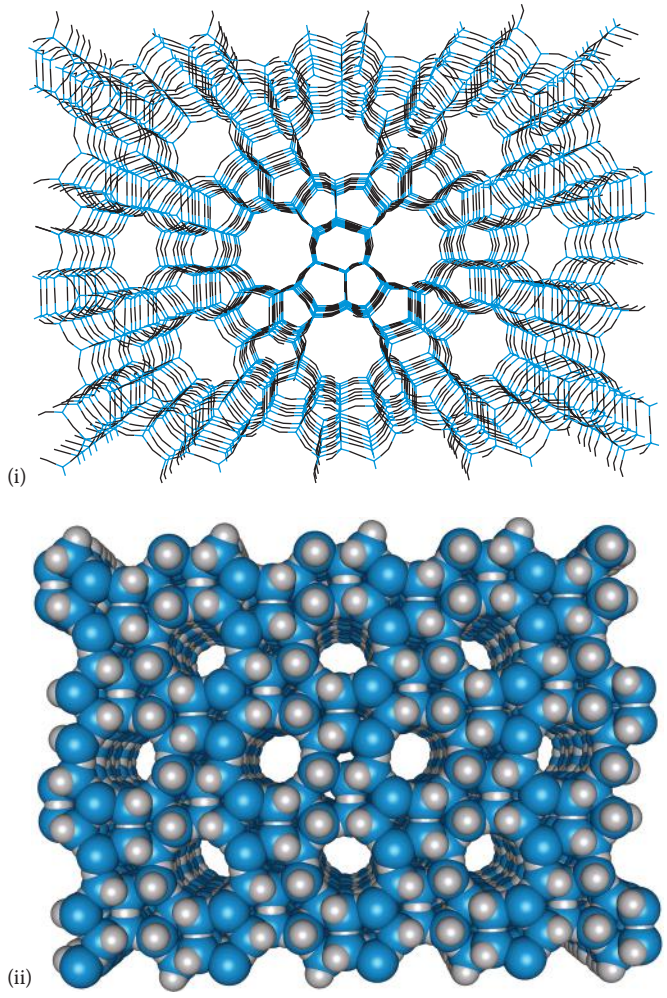


FIGURE 6.12 (a) (i, ii) Computer models of orthorhombic ZSM-5 showing the straight elliptical channels along the *y* direction and (b) (i, ii) computer models of ZSM-11 showing the near-circular straight channels that run along both the *x* and *y* directions in the tetragonal crystals. Both structures are shown as line drawings and with their van der Waals radii.

Faujasite is a **large-pore zeolite** based on four sodalite cages in a tetrahedral configuration, linked by the oxygen bridges through the 6-ring windows. This leads to a structure with large cavities of diameter 1180 pm entered by 12-ring windows of diameter 740 pm, resulting in a three-dimensional network of channels, as shown in Figure 6.14.

The channel system for **mordenite** (MOR) is shown in Figure 6.15. Mordenite has an orthorhombic structure and there are two types of channels, governed by 8-ring and 12-ring windows, respectively, all running parallel to each other and

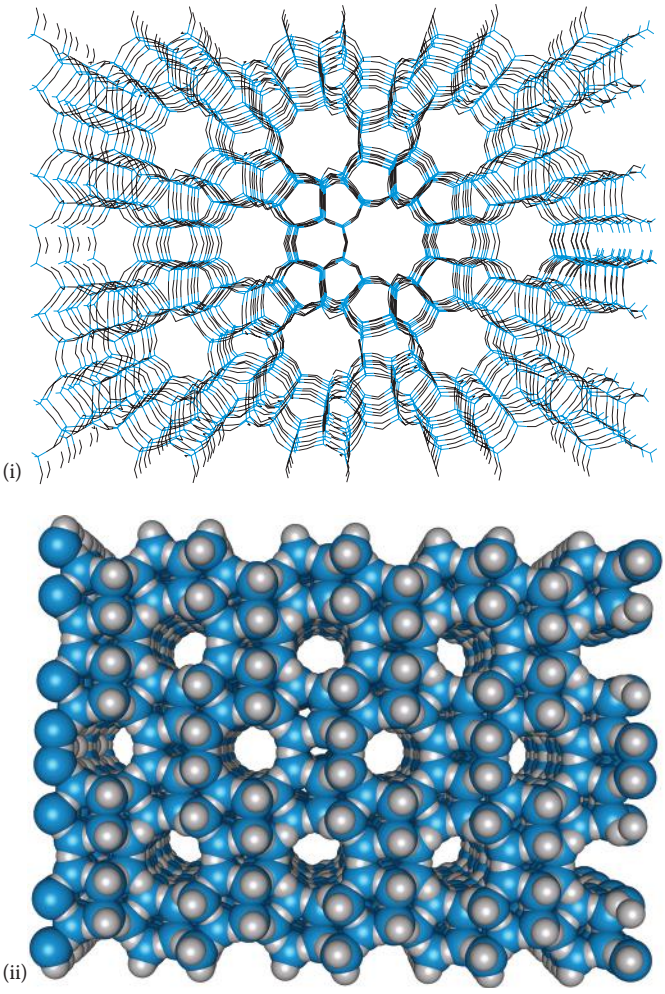


FIGURE 6.12 (Continued)

interconnected only by the smaller 5-ring and 6-ring systems. Interestingly, mordenite synthesised at low temperatures (<260°C) is found to have larger ports than natural mordenite or mordenite synthesised at higher temperatures, and this is due to crystal defects blocking some of the large pores.

MCM-22 was prepared from a precursor, MSM-22P, which has a layered structure. MCM-22 has a double-pore system accessed by 10-ring windows, one of sinusoidal channels and also a pore system of large supercages based on the 12-rings. However, its unusual catalytic activity is thought to rest in the 12-ring pockets on its surface. It is used in the production of ethylbenzene.

In 2003, the preparation of a thermally stable silicate zeolite with 18-ring pore openings, ECR-34, was reported.

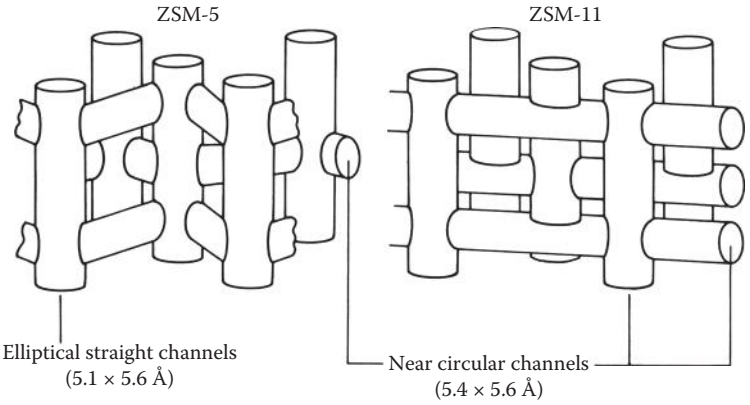


FIGURE 6.13 The interconnecting channel systems in ZSM-5 and ZSM-11.

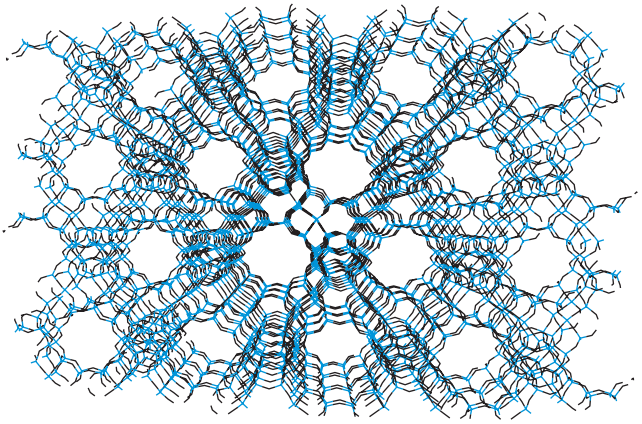


FIGURE 6.14 The cubic faujasite structure showing the channels that lie parallel to each of the face diagonals (oxygen bridges are included in this model).

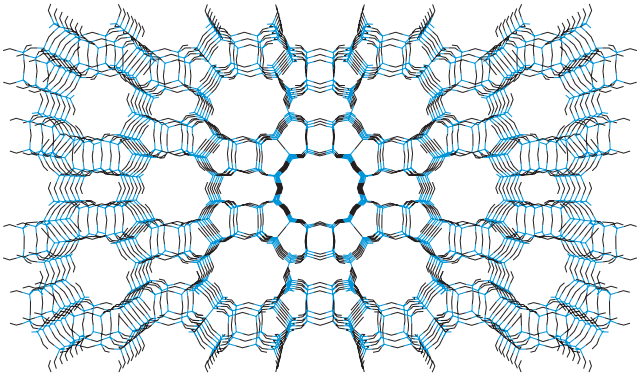


FIGURE 6.15 The channel structure of mordenite showing the larger channels running along the z direction.

6.1.8 Structure Determination

The structures of the zeolite frameworks have been determined by X-ray and neutron diffraction techniques. Some of the naturally occurring minerals were characterised in the 1930s, and the synthetic zeolites have been investigated from 1956 onwards. Unfortunately, it is extremely difficult for diffraction techniques to determine a structure unequivocally because Al and Si are next to each other in the periodic table and therefore have very similar atomic scattering factors (Chapter 2). It is possible to determine the overall shape of the framework with accurate atomic positions, but not to locate the Si and Al atoms precisely.

The positions of the Si and Al atoms have always been assigned by applying **Loewenstein's rule**, which forbids the presence of an Al-O-Al linkage in the structure. A corollary of this rule is that when the Si/Al ratio is 1, the amount of aluminium in a zeolite is maximum and the Si and Al atoms alternate throughout this structure.

When it comes to locating the cations, other problems arise. Not every cation site is completely occupied, so although the cation sites can be located, their occupancy is averaged. Furthermore, zeolites are usually microcrystalline and for successful diffraction studies, larger single crystals are needed (although Rietveld powder techniques have been successfully applied, especially with neutrons; see Chapter 2). One of the techniques used for successfully elucidating zeolite structures is magic angle spinning NMR spectroscopy (MAS-NMR) (see Chapter 2). Five peaks can be observed for the ^{29}Si spectra of various zeolites, which correspond to the five possible, different Si environments. Each Si is coordinated by four oxygen atoms, but each oxygen can then be attached either to an Si or to an Al atom, giving five possibilities:

- $Si(OA1)_4$
- $Si(OA1)_3(OSi)$
- $Si(OA1)_2(OSi)_2$
- $Si(OAl)(OSi)_3$
- $Si(OSi)_4$.

The characteristic ranges of these shifts are assigned to each coordination type—the more Al linkages, the more positive is the shift (Figure 2.34).

Interpreting the ^{27}Al solid-state NMR spectrum of minerals and zeolites is also now achievable; for instance, in determining the three different aluminium positions in kyanite (Al_2SiO_5). Diagnostically, it can also distinguish three different types of aluminium:

- Octahedrally coordinated $[Al(H_2O)_6]^{3+}$ trapped as a cation in the pores with a peak at about 0 ppm. ($[Al(H_2O)_6]^{3+}(aq)$ is used as the reference.)
- Tetrahedral Al ($Al(OSi)_4$), which gives a single resonance with characteristic Al chemical shift values for individual zeolites in the range 50–65 ppm.
- Tetrahedral $[AlCl_4]^-$, which gives a peak at about 100 ppm; such a peak can occur when a zeolite has been treated with $SiCl_4$ in order to increase the Si/Al ratio in the framework and should disappear with washing.

An important feature of ^{29}Si MAS-NMR spectra is that the measurement of the intensity of the observed peaks allows the Si/Al ratio of the framework of the sample to be calculated. This can be extremely useful when developing new zeolites for catalysis, as much research has concentrated on making highly siliceous varieties by replacing the Al in the framework. Conventional chemical analysis only gives an overall Si/Al ratio, which includes the trapped octahedral Al species and $[AlCl_4]^-$, which has not been washed away. At high Si/Al ratios, the ^{27}Al MAS-NMR results are more sensitive and accurate and are therefore preferred.

Even with this information, it is still an extremely complicated procedure to decide where each linkage occurs in the structure. The cation positions can give useful information as they tend to be as close as possible to the negatively charged Al sites.

High-resolution electron microscopy (HRTEM) is also used extensively for the structural examination of zeolites, particularly for intergrowths and faults. Extended X-ray absorption fine structure (EXAFS) has been used for determining the local coordination geometry of the exchangeable cations and how this changes on reaction or dehydration.

6.1.9 Zeolites as Dehydrating Agents

Normal crystalline zeolites contain water molecules that are coordinated to the exchangeable cations. As we noted previously, these structures can be dehydrated by heating under vacuum; in these circumstances, the cations move position, frequently settling on sites with a much lower coordination number. The dehydrated zeolites are very good drying agents, absorbing water to get back to the preferred high coordination condition. Zeolite A is a commonly used drying agent, and it can be regenerated by heating after use. Its commercial uses are in drying natural gas, refrigerant gases and hydrocarbons.

Highly siliceous zeolites have far fewer cations, so they tend to be much less hydrophilic, but can be used to absorb small organic molecules.

6.1.10 Zeolites as Ion Exchangers

The cations M^{n+} in a zeolite will exchange with others in a surrounding solution. In this way, the Na^+ form of zeolite A can be used as a water softener: the Na^+ ions exchange with the Ca^{2+} ions from the hard water. This was first exploited commercially in the early 1900s, and soon after, the Permutit Company was founded. Zeolite A was added to detergents as a water softener for many years, replacing the polyphosphates that have been of concern because of possible ecological damage. By the 1970s, both Henkel (Persil) and Procter & Gamble (Tide) had incorporated zeolite NaA into their formulas; the less dense zeolite MAP, the first new synthetic zeolite prepared by Barrer, is now widely used for this purpose, as is zeolite P (GIS), which has been developed for use in detergents.

When used as a water softener, the zeolite is reusable because it can be regenerated by running through a very pure saline solution; this is a familiar procedure to anyone who has used a dishwasher. It is possible to produce drinking water from

seawater by desalinating it through a mixture of Ag and Ba zeolites. This is such an expensive process, however, that it is only useful in an emergency.

Some zeolites have a strong affinity for particular cations. Clinoptilolite (HEU) is a naturally occurring zeolite that sequesters caesium and is used by British Nuclear Fuels (BNFL) to remove ^{137}Cs from radioactive waste, exchanging its own Na^+ ions for the radioactive Cs^+ cations. Similarly, zeolite A can be used for recovering radioactive strontium. Zeolites were heavily used in the clean-up operations after the Chernobyl and Three-Mile Island incidents.

6.1.11 Zeolites as Adsorbents

As dehydrated zeolites have very open, porous structures, they have large internal surface areas and are capable of adsorbing large amounts of substances other than water. The ring sizes of the windows leading into the cavities determine the sizes of the molecules that can be adsorbed. An individual zeolite has a highly specific sieving ability, which can be exploited for purification or separation. This was first noted for chabazite (CHA) as long ago as 1932, when it was observed that it would adsorb and retain small molecules such as formic acid and methanol, but would not adsorb benzene and larger molecules. CHA has been used commercially to adsorb polluting SO_2 emissions from chimneys. Similarly, the 410 pm pore opening in zeolite A (determined by an 8-ring and much smaller than the 1140 pm diameter of the cavity) can admit a methane molecule, but excludes a larger benzene molecule.

The zeolites that are useful as molecular sieves do not show any appreciable change in the basic framework structure on dehydration, although the cations move to positions of lower coordination. After dehydration, zeolite A and others are remarkably stable to heating and do not decompose below about 700°C. The cavities in dehydrated zeolite A amount to about 50% of the volume.

There are many uses of the selective sorbent properties of zeolites, of which we will describe just a few; Table 6.2 gives a brief summary of the industrial uses. The zeolites are regenerated after use, by heating, evacuation or flushing with pure gases.

It is possible to 'fine-tune' the pore opening of a zeolite to allow the adsorption of specific molecules. As discussed earlier, one method is to change the cation. This method can be used to modify zeolite A in order to separate the branched and cyclic hydrocarbons from the straight-chain alkanes (paraffins). When Na^+ ions are replaced by Ca^{2+} ions, the effective aperture increases. Once approximately one-third of the Na^+ ions have been replaced by the Ca^{2+} ions, many straight-chain alkanes can be adsorbed (Figure 6.16). However, all branched-chain, cyclic and aromatic hydrocarbons are excluded because their cross-sectional diameters are too large. This process can be useful industrially for separating the long straight-chain hydrocarbons required as the starting materials in the manufacture of biodegradable detergents. Petrol can also be upgraded by removing the shorter, straight-chain, low-octane-number alkane constituents that produce 'pinking' ('knocking'), the small explosions, which are harmful to engines.

At −196°C, oxygen is freely adsorbed by Ca–zeolite A, whereas nitrogen is essentially excluded. The two molecules are not very different in size: O_2 has a diameter of 346 pm, whereas that of N_2 is 364 pm. As the temperature rises, the adsorption of N_2

TABLE 6.2
Applications of Molecular Sieves in Industrial Adsorption Processes

Fields of Application	Uses		
	Drying	Purification	Separations
Refineries and the petrochemical industry	Paraffins, olefins, acetylenes, reformer gas, hydrocracking gas, solvents	Sweetening[a] of 'liquid petrol gas' and aromatics; removal of CO_2 from olefin-containing gases; purification of synthesis gas	Normal and branched-chain alkanes
Industrial gases	H_2, N_2, O_2, Ar, He, CO_2, natural gas	Sweetening and CO_2 removal from natural gas; removal of hydrocarbons from air; preparation of protective gases; removal of N_2 to enrich compressed air	Aromatic compounds
Industrial furnaces	Exogas, cracking gas, reformer gas	Removal of CO_2 and NH_3 from exogas and from ammonia fission gas	Nitrogen and oxygen

[a] 'Sweetening' is the removal of sulfur-containing compounds.

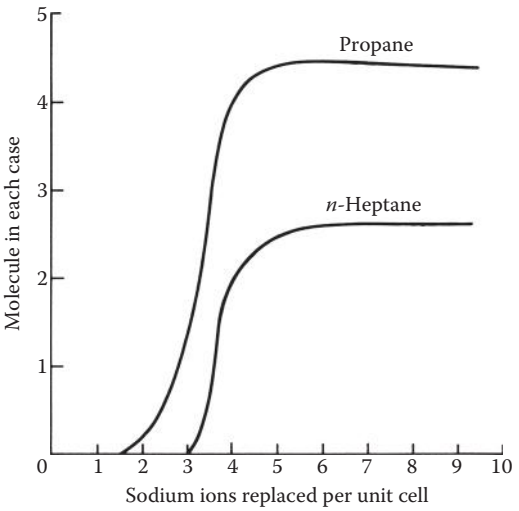


FIGURE 6.16 The effect of calcium exchange for sodium in zeolite A on hydrocarbon adsorption. The replacement of four sodium ions by two calcium ions permits easy diffusion of the *n*-alkanes into the zeolite channels.

increases to a maximum at around −100°C. The main reason is thought to be due to the thermal vibrations of the oxygen atoms in the window. Over a range of 80–300 K, a variation of vibrational amplitude of 10–20 pm could well be expected and therefore a variation of 30 pm in the window diameter would not be unreasonable. This would make the window just small enough at lower temperatures to exclude the N_2 molecules. Zeolite A can be used to separate or purify these gases. It has also been used commercially to purify argon by removing oxygen.

The other method used for 'fine-tuning' the pore openings is to change the Si/Al ratio. An increase in the proportion of Si will

- Slightly decrease the unit cell size and therefore the size of the cavities.
- Decrease the number of cations, thereby freeing the channels.
- Make the zeolite more hydrophobic (literally 'water hating') in character.

The hydrophobic zeolites can potentially be used to remove organic molecules from an aqueous solution; possible uses range from the removal of toxic materials from blood, to the production of nonalcoholic beverages by the selective removal of alcohol, and the decaffeination of coffee.

6.1.12 Zeolites as Catalysts

Zeolites are very useful catalysts, displaying several important properties that are not found in traditional amorphous catalysts. The amorphous catalysts have always been prepared in a highly divided state in order to give a high-surface area and thus a large number of catalytic sites. The presence of the cavities in zeolites provides a very large internal surface area that can accommodate as many as 100 times more molecules than the equivalent amount of amorphous catalysts. Zeolites are also crystalline and can be prepared with improved reproducibility, so they tend not to show the varying catalytic activity of the amorphous catalysts. Furthermore, their molecular sieve action can be exploited to control which molecules have access to (or which molecules can depart from) the active sites. This is generally known as **shape-selective catalysis**. The introduction of zeolite catalysts has also made a significant difference to the environment control as the liquid catalysts they replaced present industrial pollution problems.

The catalytic activity of decationised zeolites is attributed to the presence of acidic sites arising from the $[AlO_4]$ tetrahedral units in the framework. These acid sites may be Brønsted or Lewis in character. Zeolites that are normally synthesised usually have Na^+ ions balancing the framework charges, but these can be readily exchanged for protons by a direct reaction with an acid, giving surface Si–OH and Al–OH hydroxyl groups—the **Brønsted sites**. Alternatively, if the zeolite is not stable in an acid solution, it is common to form the ammonium (NH_4^+) salt, and then heat it so that the ammonia is driven off, leaving a proton. The oxygen of the hydroxyl was thought to be three-coordinate, forming a bridge between Si and Al. Further heating removes water from the Brønsted site, exposing a tricoordinated aluminium ion, which has electron-pair acceptor properties to hydroxyl groups on adjacent silicon atoms; this is identified as a **Lewis acid site**. A scheme for the

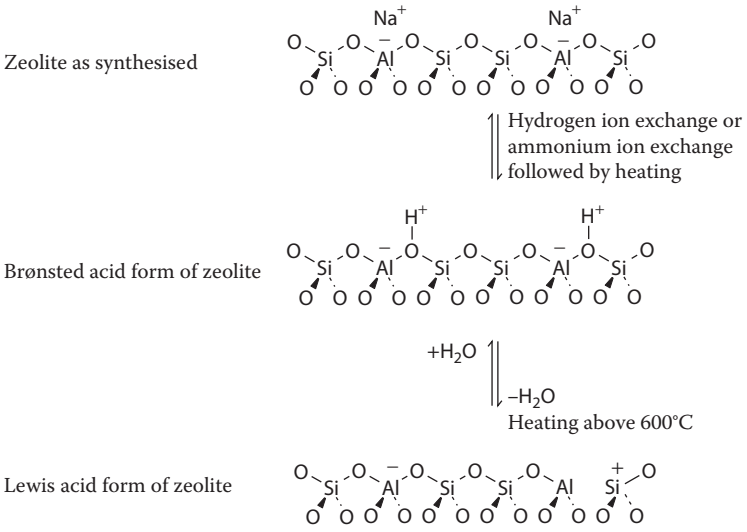


FIGURE 6.17 A scheme for the generation of Brønsted and Lewis acid sites in zeolites.

formation of these sites is shown in Figure 6.17. The surfaces of the zeolites can thus display either Brønsted or Lewis acid sites, or both, depending on how the zeolite is prepared. Brønsted sites are converted into Lewis sites as the temperature is increased above 600°C and water is driven off.

Not all zeolite catalysts are used in the decationised or acid form; it is also quite common to replace the Na^+ ions with lanthanoid ions such as La^{3+} or Ce^{3+}. These ions now place themselves so that they can best neutralise the three separated negative charges on the tetrahedral Al in the framework. The separation of the charges causes high electrostatic field gradients in the cavities which are sufficiently large to polarise the C–H bonds or even to ionise them, enabling a reaction to take place. This effect can be strengthened by a reduction in the aluminium content of the zeolite so that the $[AlO_4]$ tetrahedra are farther apart. If one thinks of a zeolite as a solid ionising solvent, the difference in the catalytic performances of various zeolites can be likened to the behaviour of different solvents in solution chemistry. A rare-earth substituted form of zeolite X became the first commercial zeolite catalyst for the cracking of petroleum in the 1960s. Crude petroleum is initially separated by distillation into fractions, and the heavier gas–oil fraction is cracked over a catalyst to give petrol (gasoline). A form of zeolite Y, which has proved to be more stable at high temperatures, is now used. These catalysts yield ~20% more petrol than earlier catalysts and do so at lower temperatures. The annual catalyst usage for catalytic cracking is worth billions of pounds sterling.

A third way of using zeolites as catalysts is to replace the Na^+ ions with other metal ions, such as Ni^{2+}, Pd^{2+} or Pt^{2+}, and then reduce them in situ so that the metal atoms are deposited within the framework. The resultant material displays the properties associated with a supported metal catalyst and extremely high dispersions of the metal can be achieved. A mixture of a CaA zeolite absorbent and a Pt–H–mordenite catalyst is used to remove the small straight-chain alkanes and to convert them to

branched isomers and so improve the octane number. This process (the total isomerisation process) operates worldwide.

Another technique for preparing a zeolite-supported catalyst involves the physical adsorption of a volatile inorganic compound followed by thermal decomposition: $Ni(CO)_4$ can be adsorbed on zeolite X and decomposed by gentle heating, leaving a dispersed phase of nearly atomic nickel in the cavities, and this has been found to be a good catalyst for the conversion of carbon monoxide to methane:

$$CO + 3H_2 = CH_4 + H_2O$$

There are several types of shape-selective catalysis:

- **Reactant shape-selective catalysis**: Only molecules with dimensions less than a critical size can enter the pores and reach the catalytic sites, so they react there. This is illustrated diagrammatically in Figure 6.18a, in which a straight-chain hydrocarbon is able to enter the pore and react, but the branched-chain hydrocarbon is not able to do so.

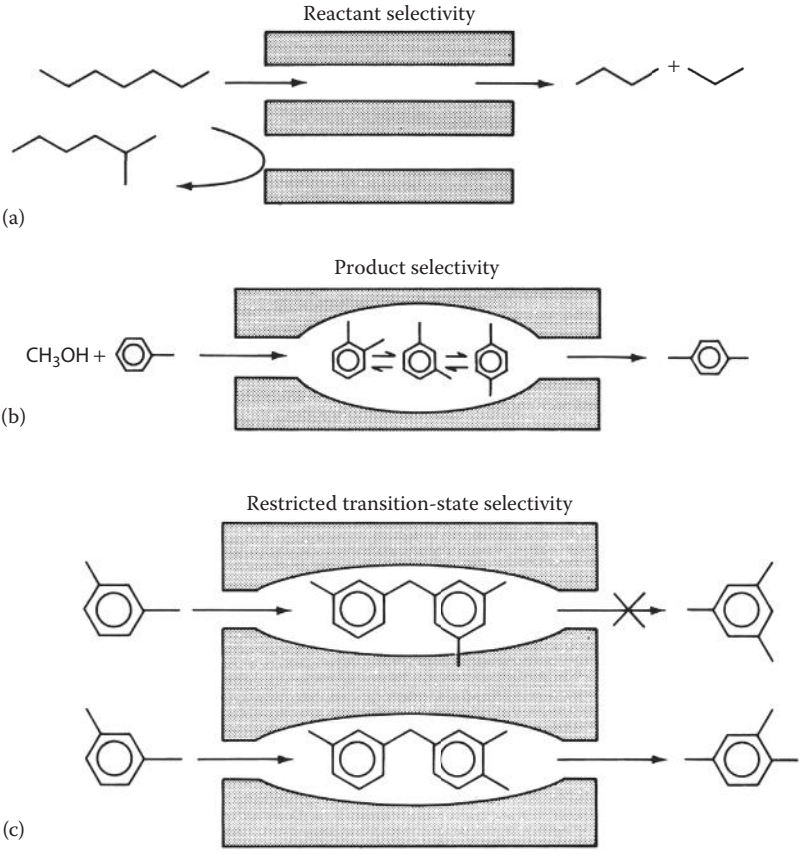


FIGURE 6.18 A shape-selective catalysis: (a) reactant, (b) product and (c) transition state.

- **Product shape-selective catalysis**: Only products less than a certain dimension can leave the active sites and diffuse out through the channels, as illustrated in Figure 6.18b for the preparation of xylene. A mixture of all three isomers is formed in the cavities, but only the para form is able to escape.
- **Transition-state shape-selective catalysis**: Certain reactions are prevented because the transition state requires more space than is available in the cavities, as shown in Figure 6.18c, for the transalkylation of dialkylbenzenes.

Reactant shape-selective catalysis is demonstrated in the dehydration of butanols. If butan-l-ol (*n*-butanol) and butan-2-ol (*iso*-butanol) are dehydrated either over Ca-zeolite A or over Ca-zeolite X, we see a difference in the products formed.

$$CH_3CH_2CH_2CH_2OH = CH_3CH_2CH{=}CH_2 + H_2O$$

$$CH_3CH_2\underset{\displaystyle OH}{\underset{|}{C}}HCH_3 = CH_3CH{=}CHCH_3 + H_2O$$

Zeolite X has windows large enough to admit both of the alcohols easily, and both undergo conversion to the corresponding alkene. Over zeolite A, however, the dehydration of the straight-chain alcohol is straightforward, but virtually none of the branched-chain alcohols are converted, as these are too large to pass through the smaller windows of zeolite A. These results are summarised in Figure 6.19. Notice that, at higher temperatures, curve d begins to rise. This is because the lattice vibrations increase with temperature, making the pore opening slightly larger and thus beginning to admit butan-2-ol. The very slight conversion at lower temperatures is thought to take place on external sites.

One of the industrial processes using ZSM-5 provides us with an example of **product shape-selective catalysis**: the production of 1,4-(para-)xylene. Para-xylene is

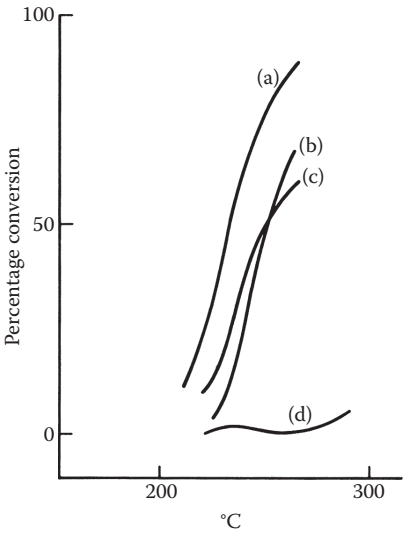


FIGURE 6.19 Dehydration of (a) *iso*-butanol on Ca-X, (b) *n*-butanol on Ca-X, (c) *n*-butanol on Ca-A and (d) *iso*-butanol on Ca-A.

used in the manufacture of terephthalic acid, the starting material for the production of polyester fibres such as 'Terylene'.

Xylenes are produced in the alkylation of toluene by methanol (Figure 6.20a). When ZSM-5 catalyst is impregnated with phosphoric acid (to produce the acid catalyst), the reaction is selective because of the difference in the rates of diffusion through the channels, of the different isomers produced. This is confirmed by the observation that selectivity increases with increasing temperature, indicating the increasing importance of diffusion limitation. The diffusion rate of *para*-xylene is approximately 1000 times faster than that of the other two isomers. The computer models in Figure 6.20b and c show why *meta*-xylene and *ortho*-xylene cannot enter a 10-ring window easily and hence cannot diffuse along the channels. The xylenes isomerise within the pores, so *para*-xylene diffuses out while the ortho-isomers and meta-isomers are trapped and take more time to convert to the para form before escaping. Up to 97% of selective conversion to *para*-xylene has been achieved by suitable treatment of this catalyst. X-ray studies have demonstrated that the *para*-xylene molecules are located both in the sinusoidal channels and at the junction of the straight and sinusoidal channels. Using a

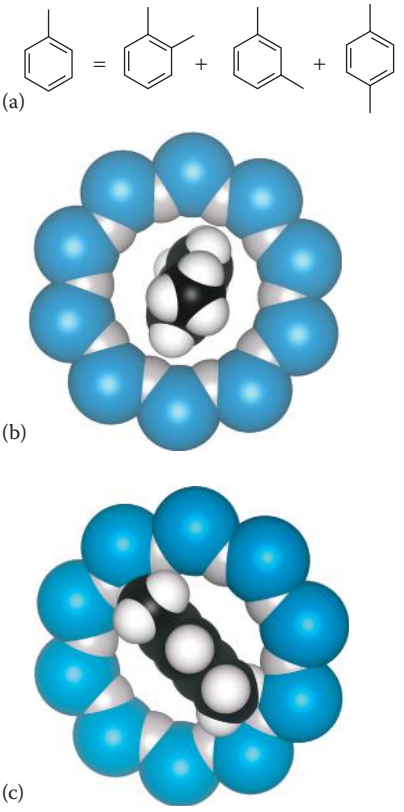


FIGURE 6.20 (a) The structures of xylene, ortho-xylene, meta-xylene and para-xylene. Computer models showing how (b) para-xylene fits neatly in the pores of ZSM-5, whereas (c) meta-xylene is too big to diffuse through.

zeolite catalyst makes this a much greener process than using the conventional alkylation reaction using CH_3I over an $AlCl_3$ or $FeCl_3$ catalyst, as a pure product is produced in good yield, and the zeolite catalyst can be regenerated.

ZSM-5 is used as the catalyst in the production of monoethyl benzene from benzene and ethene; it is highly selective for the monosubstituted product, which is the precursor of styrene and also does not use the harmful aluminium chloride, which the old Friedel–Crafts process employed.

ZSM-5 is also the catalyst that is used to convert methanol into hydrocarbons—the **methanol-to-gasoline** or **MTG** process. This research received a great impetus at the end of the 1970s when oil was in short supply and prices rose sharply. Subsequently, more oil was released, the prices dropped and much of the research was put into abeyance. However, the political instability of the Middle East and concern over long-term oil reserves make this an important process. New Zealand, with no oil reserves of its own, but with a supply of natural gas (CH_4), brought the first MTG plant onstream in 1986. Methane is first converted to methanol (CH_3OH), which is then partially dehydrated to dimethyl ether and water. This subsequently dehydrates further to give ethene and other alkenes, which react rapidly to give a range of hydrocarbons:

$$2CH_3OH = CH_3OCH_3 + H_2O \xrightarrow{-H_2O} C_2 \text{ to } C_5 \text{ alkenes} \longrightarrow \text{alkanes, cycloalkanes, aromatics}$$

The conversion of ZSM-5 to hydrocarbons mainly produces branched-chain and aromatic hydrocarbons in the C_9–C_{10} range, which is ideal for high-octane unleaded fuel. It is thought that the methanol molecules enter ZSM-5 through the more restricted zigzag channels, react at the cavities where the pores intersect and then the larger hydrocarbons exit through the straight channels—the so-called **molecular traffic control**.

Transition-state shape-selective catalysis occurs in the acid-catalysed transalkylation of dialkylbenzenes, where one of the alkyl groups is transferred from one molecule to another. This bimolecular reaction involves a diphenylbenzene transition state. When the transition state collapses, it can split to give either the 1,2,4-isomer or the 1,3,5-isomer, along with the monoalkylbenzene. When the catalyst used for this reaction is mordenite (Figure 6.18c), the transition state for the formation of the symmetrical 1,3,5-isomer is too large for the pores, and the 1,2,4-isomer is formed in almost 100% yield. (This compares with the equilibrium mixtures, in which the symmetrically substituted isomers tend to dominate.)

The synthesis of the open-framework zeolites dramatically improved the number of accessible active sites for catalysis. It is estimated that ZSM-5 has a turnover of more than 300 molecules per active site per minute during the cracking process and that other processes such as xylene isomerisation are even faster, with turnovers of up to 10^7 molecules per active site per minute.

Active research continues to find new zeolite structures and develop new catalysts. The later years of the twentieth century saw the commercial introduction of zeolite-rho (RHO), zeolite-L (LTL), zeolite-β (BEA), ferrierite (FER), MCM-22 (MWW) and frameworks similar to zeolites that also contain phosphorus (SAPOs see Section 6.3). Table 6.3 summarises the use of many zeolites in the industrial processes associated with the petrochemical and refining industries.

TABLE 6.3

Commercial Uses of Molecular Sieves (Refining and Petrochemicals)

Zeolite	Y	L	MOR	BETA	MCM-22	ZSM-5	SAPO-11	Ferrierite	Rho	SAPO-34	Other
IZA Code	FAU	LTL	MOR	BEA	MWW	MFI	AEL	FER	RHO	CHA	
Industrial Process											
Ethyl benzene					✓	✓					✓
Cumene	✓		✓	✓	✓	✓					✓
Other aromatics						✓					
Xylenes						✓					✓
C_4 isomer			✓								
C_4^{2-} isomer						✓		✓			
C_5^{2-} isomer						✓					
Iso-dewaxing							✓				
Amination						✓			✓	✓	
C_3, C_4/aromatic			✓			✓					
Naphtha aromatic		✓									
FCC (fluid catalytic cracking)	✓					✓					
Dewaxing						✓					
Hydrocracking	✓					✓					
MTG (methanol to gasoline)						✓					
MTO (methanol to olefins)						✓				✓	
Toluene											
Trans-alkylation						✓					✓

Source: Rabo, J.A. and Schoonover, M.W., *App. Catal. A General*, **222**, 261, 2001.

✓catalyst used commercially for this process.

6.2 OTHER MICROPOROUS FRAMEWORK STRUCTURES

Other framework structures based on zeolites have also been synthesised, which contain atoms other than aluminium and silicon, such as boron, gallium, germanium and phosphorus, which are tetrahedrally coordinated by oxygen. Such compounds are known as **zeotypes**. Pure aluminium phosphate, commonly called ALPO, and its derivatives, can take the same structural forms as some zeolites, such as sodalite (SOD), faujasite (FAU) and chabazite (CHA) (e.g., ALPO-20 is isostructural with sodalite and ALPO-17 with erionite), as well as novel structures, with many new compounds now synthesised. The metal–aluminium phosphates can be formed with metals such as Li, Be, Mg, Mn, Fe, Co, Ni and Zn, replacing some of the aluminium; these are usually designated MeALPOs, where Me stands for metal. If the compound contains silicon, or silicon and a metal partially replacing aluminium or phosphorus, then they are known as SAPOs. In the same way that the replacement of Si(IV) by Al(III) in the zeolite structures leads to the formation of Brønsted acid sites, so does the replacement of Al(III) by Me(II) in ALPOs; readily detachable protons are loosely bound to a framework oxygen atom near the divalent metal. Consequently, in recent years, there has been much interest in these compounds as possible heterogeneous catalysts that might be at least as good, if not better, than the zeolites. Indeed, SAPOs have been found to be very good catalysts for the conversion of methanol to olefins.

The aluminophosphate compounds are formed using templates in a similar fashion to many of the zeolites, but under more acidic conditions. Typical templates for forming large pores are the tetra-*n*-propyl ammonium ion and the tri-*n*-propylamine. One of the new channel structures formed is ALPO-5, shown in a projection in Figure 6.21a; ALPO-5 has 12-ring windows with a diameter of about 800 pm, similar to that found in zeolite Y. More recently, extra-large–pore materials have been prepared. The first of these was an ALPO, designated VPI-5 (Figure 6.21b), which has no zeolite analog, and is based on β-cages separated by double four-membered rings to give 18-ring windows and a pore diameter of between 1200 and 1300 pm. The hydrogenation of cyclooctene using rhodium-substituted VPI-5 has been reported.

Others structures such as cloverite, a gallophosphate with a mesporous 3000 pm supercage accessed through six cloverleaf-shaped channels (Figure 6.22), have followed. The interest in these large, channelled structures lies in their potential for

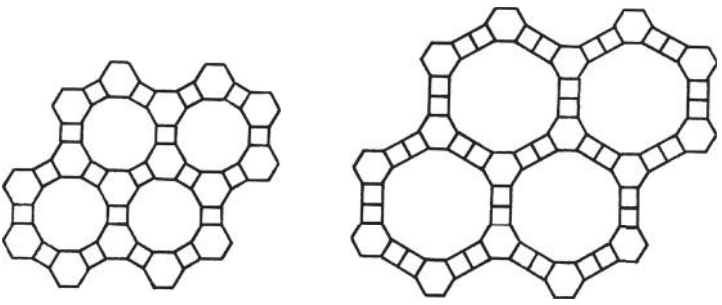


FIGURE 6.21 Schematic structural projection drawings of (a) ALPO-5 and (b) VPI-5.

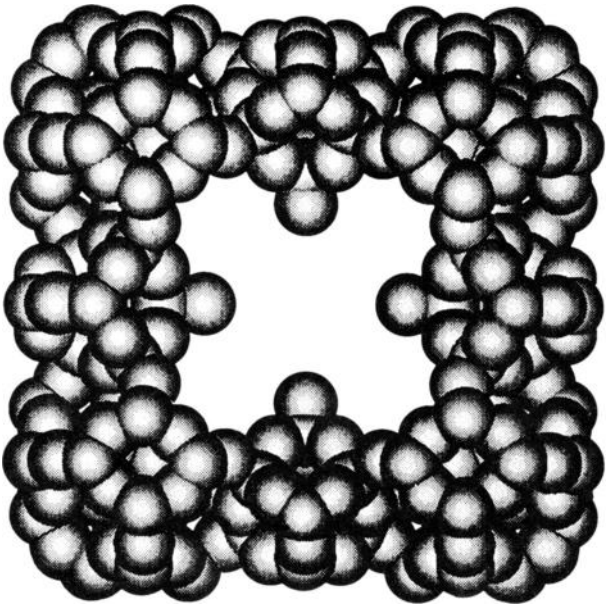


FIGURE 6.22 The framework of the synthetic gallophosphate cloverite, represented as shaded spheres of van der Waals radii. (From Catlow, C.R.A., ed., *Modelling of Structure and Reactivity in Zeolites*, Academic Press, London, 1992. Reproduced by permission of Academic Press Ltd.)

allowing catalytic conversions to take place using molecules that are too large to enter the conventional zeolitic-size channels. There is currently much interest in these mesoporous structures, which form the subject of the next section.

6.3 MESOPOROUS STRUCTURES

6.3.1 Mesoporous Aluminosilicate Structures

In 1992, scientists at Mobil Research and Development Corporation developed a family of silicate and aluminosilicate materials (M41S) that had much larger pores than conventional zeolites. The drive to synthesise such structures not only lay in their intrinsic interest but also in the need for catalysts to produce petroleum products from heavier crudes that have much larger hydrocarbon molecules. The pores in these materials lie in the range 2–50 nm and give rise to the name **mesoporous** solids. Porous materials are classified by IUPAC: microporous materials have pore diameters of less than 2 nm; macroporous materials have pore diameters of greater than 50 nm; mesoporous solids have pore diameters between these two.

There are many possible advantages in having larger pores: the surface areas of the materials are very large; diffusion through the material could be easier and faster; the pores do not block up so easily, and they can be used for liquid-phase reactions, unlike zeolites, which are usually used for gases and vapours. The pores are,

nevertheless, small enough to provide shape-selective effects, and the walls of the pores can be modified by adding functional groups. The possibilities for such solids are very exciting because, apart from the more obvious roles in shape-selective catalysis, sorption and molecular sieving, the incorporation of metals into the framework may lead to new and better catalysts. Applications as hosts for nanoclusters with useful magnetic, optical and electronic properties, and the formation of molecular wires in the channels are also being researched.

The mesoporous material that has received most attention so far is MCM-41 (Mobil Crystalline Materials). It has a highly ordered hexagonal array of uniformly sized mesopores and can have a huge surface area of 1200 $m^2\ g^{-1}$. It is made by a templating technique, where the silicate or aluminosilicate walls of the mesopores, instead of forming around a single molecule or ion, form around an assembly of molecules known as a **micelle**. In a solution of silicate or silicate and aluminate anions, cationic long-chain alkyl trimethylammonium surfactants ($[CH_3(CH_2)_n(CH_3)_3N^+]X^-$) form rod-like micelles (Figure 6.23), with the hydrophobic tails clustering together inside the rods and the cationic heads forming the outside; these silicate/aluminate ions form a cladding around the micelles. The silicate-coated micelles

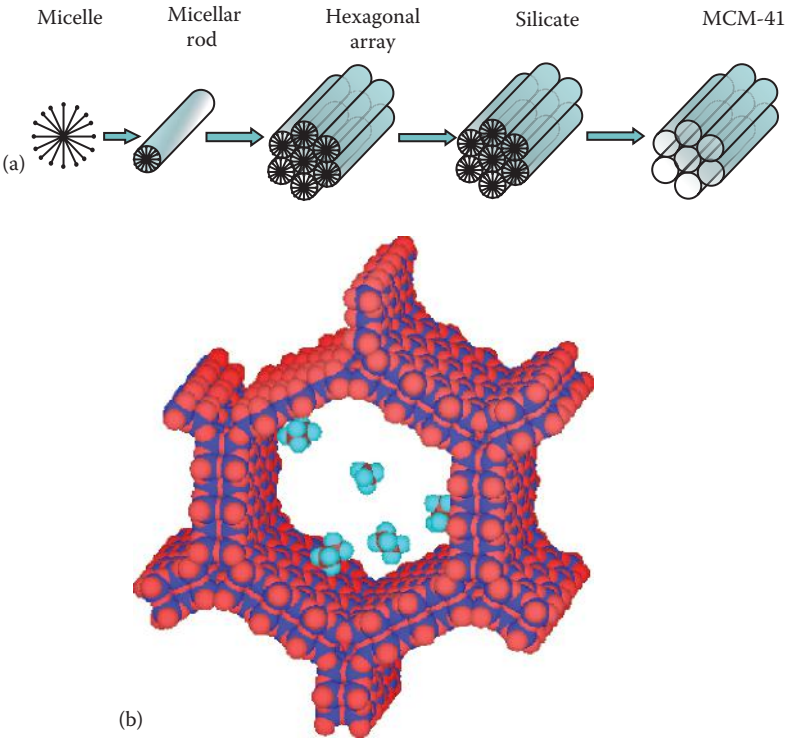


FIGURE 6.23 (a) A possible preparative route for the formation of MCM-41 by liquid crystal templating. (b) A computer graphic of ethane and methane molecules inside one of the hexagonal pores of MCM-41. (From Vladimir Gusev, Molecule of the Month: MCM-41, University of Bristol, http://www.chm.bris.ac.uk/motm/mcm41/mcm41.htm. With permission.)

pack together along the axes of the rods, earning the synthesis technique the name **liquid crystal templating**; under hydrothermal conditions, this mesoporous structure precipitates out of the solution. The calcination of the filtered solid in air at temperatures up to 700°C removes the template and produces the mesoporous solid. As we would expect, the alkyl chain length determines the size of the pores; where $n = 11$, 13 and 15, the pore diameters are 300, 340 and 380 nm, respectively, but when $n = 5$, micelles are not formed and zeolites such as ZSM-5 are formed around a single ion template instead. Adding trimethylbenzene to the synthesis mixture increases the pore diameter by 2 nm. The hexagonal arrangement of the pores in MCM-41 is found to be very ordered, although there is considerable disorder within the walls themselves.

A few years later, silica nanoparticles with much larger pores of 4.6–30 nm were produced at the University of California; SBA-15 (Santa Barbara Amorphous), a structure that also contains a hexagonal array of pores. Other mesoporous structures, such as the cubic MCM-48, and MCM-50, a lamellar structure, have also been produced.

The mesoporous solids have great potential for use as heterogeneous catalysts, which have the advantage of being simpler to use than homogeneous catalysts, as they can be separated by filtration at the end of a reaction.

6.3.2 Modified Mesoporous Aluminosilicates

A great deal of research has gone into various modifications of these mesoporous silicates to make them suitable for particular reactions. For instance, attached groups can make the surface more hydrophobic and more adaptable to absorbing hydrocarbons in pollution control, the groups can be suitable for binding to other guest molecules or they can endow unusual properties such as photoactivity. The walls of mesoporous materials can be modified using various methods:

- By direct synthesis, by co-condensing with a precursor.
- By grafting onto the surface using alcohols, silanols or organosilicon compounds.
- Further modification of a functionalised surface by chemical reaction or by heating.
- Selective grafting onto the outer surfaces can be achieved by coating before the template is removed or by blocking the pores. The internal surface can then be treated separately after the template is removed.

Organically modified MCM-41 can be prepared directly by using alkoxysilanes or organosiloxanes in the synthesis mixture, thereby coating the internal wall of the pores with functional groups. An example of a condensation reaction of an alcohol with the surface silanol groups to modify the pore wall is shown in Figure 6.24.

Transition metals and their complexes can be immobilised in the mesopores or incorporated into the structure to make silica-supported metal catalysts. For instance, titanium catalysts for selective oxidation can be formed by modifying the mesoporous structure with either Ti grafted onto the surface (Ti↑MCM-41) or Ti substituted into the framework (Ti→MCM-41). The grafted version makes a better catalyst for

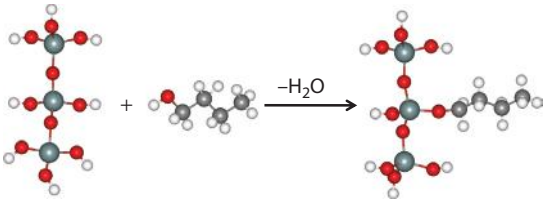


FIGURE 6.24 The formation of an Si–O–C bond at the silicate surface.

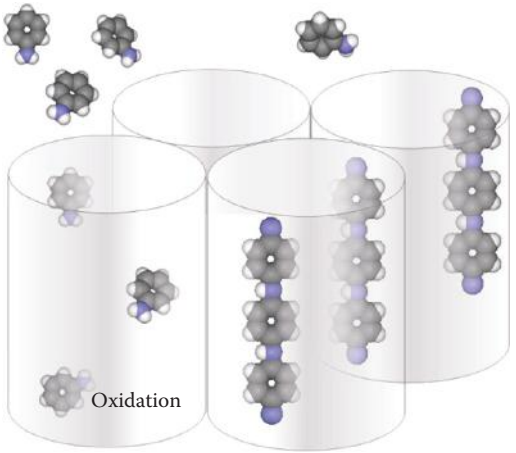


FIGURE 6.25 The encapsulation of aniline, and the formation of polyaniline in MCM-41.

the epoxidation of alkenes using peroxides and has good resistance to the leaching of the metal (Figure 6.25).

6.3.3 Periodic Mesoporous Organosilicas

Periodic mesoporous organosilicas (PMOs) were first reported in 1999 and are composed of silsesquioxanes that form Si–C bridges ($(RO)_3Si–R'–Si(OR)_3$), the organic spacer group that determines the pore size of the structure. Research has been partly driven by the novelty of the structures and by their potential usefulness as materials in the type of functions already discussed, but also for chromatography columns and silicone oils and resins. Their potential advantage lies both in their porosity and in the organic group(s) that are now actually an intrinsic part of the framework structure and not just grafted on as in the modified frameworks.

A large number of these poly(trialkyloxysilyl) structures have been made with different structures and pore sizes, although not many suitable precursors have been found because the frameworks tend to collapse when the template is removed. The synthetic method is similar to that used for the aluminosilicates, in that a silsesquioxane is mixed with a surfactant to form self-assembling micelles, which are then aged and heated to extract the template (Figure 6.26). The synthesis has many factors that

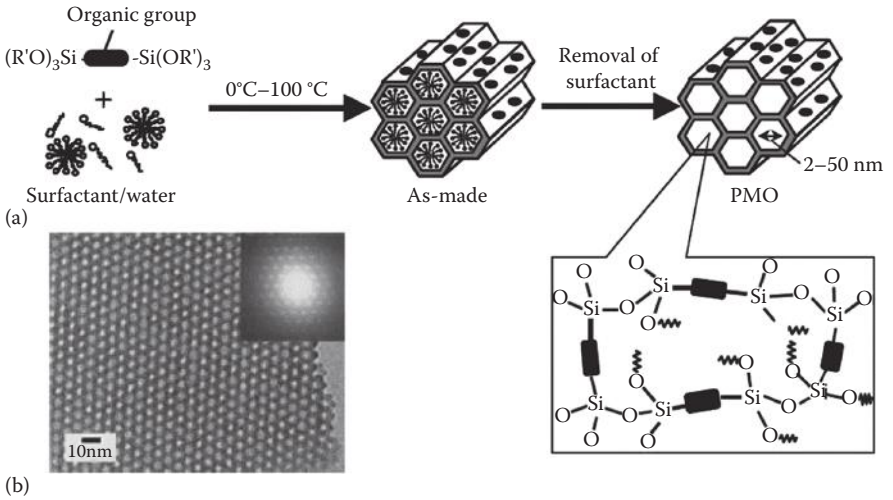


FIGURE 6.26 (a) The synthesis of periodic mesoporous organosilica from an organic-bridged silsesquioxane precursor. (b) A TEM image of the periodic pore structure of mesoporous ethane–silica. (From Kapoor, M.P. and Inagaki, S., *Bull. Chem. Soc. Japan*, **79**, 1463, 2006. With permission.)

can be varied, apart from the starting silicon precursor: temperature, pH, surfactant, ageing time and temperature. Of the precursor material ($(RO)_3Si–R'–Si(OR)_3$), the R group is usually $–OC_2H_5$, but the R′ group varies, based on, for instance, methane, ethane, benzene, thiophene and even ferrocene: some examples are $–(CH_2)_n–$, –(CH=CH)–, –(C≡C)–, *o*/*p*-$(C_6H_4)_n–$, $–(CH_2CH_2Si(OEt)_2CH_2CH_2)$, $–(S/O–(C_6H_4)_2)–$ and $–(Fe–(C_5H_4)_2)–$.

The future will undoubtedly bring new syntheses and new uses for these structures, and we would expect to hear of solids that can be processed into thin films, fibres and so on, with their properties tailored by different pore sizes and reactivity.

6.3.4 Metal Organic Frameworks

Metal organic frameworks (MOFs) have only come to prominence in the twenty-first century and form a very exciting class of porous materials with customisable adaptable frameworks with surface areas that are larger even than those of the zeolites. They basically consist of inorganic metal ions or metal atom clusters, joined together by (usually rigid) organic linker groups. Already, thousands of these compounds have been synthesised. Both the metal and its coordination and the type of organic linker dictate the shape, size and chemical attributes of the pores and therefore the properties and possible functions of the particular solid. MOFs have both the largest surface areas and the lowest densities known in crystalline solids, and the remarkable property that accounts for their exceptional behaviour is that all of the solids are accessible, and they contain no ‘dead space’. This leads to exceptionally free diffusion

through the channels. They can absorb large volumes of various gases, in particular hydrogen, and can selectively capture carbon dioxide. With the need for a clean fuel, such as hydrogen, to be used in fuel cells given the fast-depleting petroleum reserves (see Chapter 5) and the rapid increase of carbon dioxide in the atmosphere, this makes the storage and capture properties of MOFs of great interest in the fields of energy conservation and global warming. Most of the earlier research concentrated on their gas storage, gas separation and gas purification abilities, but they also have obvious potential in molecular sieving, heterogeneous catalysis, sensors and drug delivery. Their magnetic and luminescent properties are also of interest, and it is to all of these subjects that attention is increasingly turned.

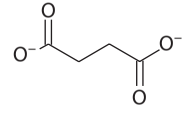
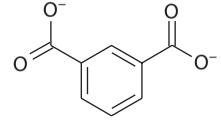
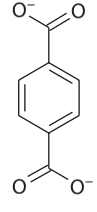
The synthesis of MOFs follows a very similar pattern to that of the zeolites; the crystals can be grown solvothermally or hydrothermally from a hot solution of the organic ligand and a metal precursor such as nitrate. Templating techniques can also be used to control the pore size and to give access to metal sites that are exposed when a templating solvent such as *N*,*N*-dimethylformamide or *N*,*N*-diethylformamide is removed; it is these metal sites that then provide sites for gases, particularly hydrogen, to adsorb. One of the problems with the synthesis of MOFs is their subsequent activation by removing the solvent used in their preparation from the pores by heating under vacuum, and it is at this stage that the channels may partially collapse, leading to a loss of surface area. Ball milling of a metal acetate precursor with the ligand avoids this problem, as it provides a solvent-free route to MOFs; unfortunately, this method only works in a limited range of compounds. Other methods for solvent removal without collapse, involve replacing the solvent with a lower boiling point solvent to give a less aggressive route to removal; replacing with benzene and then removing benzene by freeze drying; and removal using liquid CO_2 under supercritical conditions—the lack of surface tension in this instance prevents pore collapse.

A few examples of the types of organic ligands that are being used as linkers are given in Table 6.4; most commonly, they are derivatives of carboxylic acids and azoles.

One particular thread of huge interest in MOFs is **storage** and stems from the large quantity of hydrogen that MOFs are able to adsorb. Hydrogen has a large mass energy density, is abundantly available and produces no pollutants when burned. Currently, hydrogen is usually stored and transported as a liquid or compressed gas. It is not only the extra cost of hydrogen-fuelled vehicles that prevents more sales, but also the public perception of safety with general concern about using hydrogen as a fuel. Visions of the Hindenburg disaster, when a hydrogen-filled airship burst into flames, killing one-third of the passengers and crew, still loom large, even though it took place in 1937. A solid containing a very high volume of adsorbed hydrogen that can be easily and safely released becomes a very attractive option. Hydrogen physisorbs to the surface of an MOF with weak van der Waals forces, and thus the uptake increases with the surface area; it can also form stronger bonds by chemisorption. Usually, adsorption of hydrogen increases with decreasing temperature and increasing pressure up to about 20–40 bar (up to 100 bar is deemed safe for automotive uses). This means that MOFs are most efficient at hydrogen storage at low temperatures, mostly operating at liquid nitrogen temperatures

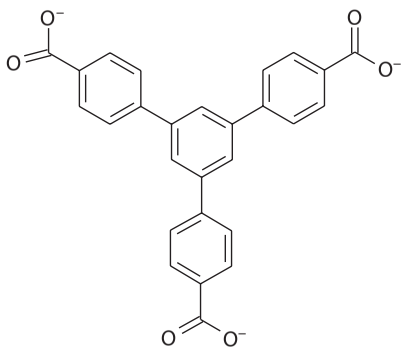
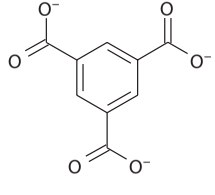
TABLE 6.4

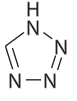
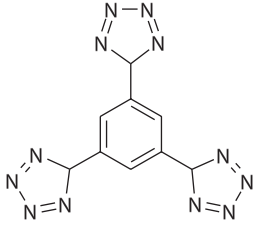
Some Ligands Used in MOF Synthesis

Common Name of Parent	IUPAC Name	Formula Anion	Ligand Abbreviation	MOF Example and Common Name	Structure Anion
Succinic acid	Butanedioate	$^{-}OOC–(CH_2)_2–COO^{-}$ $C_4H_4O_4$			
Isophthalic acid	Benzene-1,3-dicarboxylate	$C_6H_4(COO)_2^{2-}$ $C_8H_4O_4$			
Terephthalic acid	Benzene-1,4-dicarboxylate	$C_6H_4(COO)_2^{2-}$ $C_8H_4O_4$	BDC	$Zn_4O(BDC)_3$ MOF-5	

(*continued*)

TABLE 6.4 (Continued)
Some Ligands Used in MOF Synthesis

1,3,5-Tris(4-carboxyphenyl) benzene	Benzene-1,3,5-tribenzoate	$C_6H_3(C_6H_4COO)_3^{3-}$ $C_{27}H_{18}O_6$	BTB	$Zn_4O(BTB)_2$ MOF-177	
Trimesic acid	Benzene-1,3,5-tricarboxylate	$C_9H_3O_6$	BTC	$Cu_3(BTC)_2(H_2O)_3$ HKUST-1	

Imidazole 1,3-diazole	1*H*-1,3-Imidazole	$C_3H_4N_2$			
1,2,3-Triazole	1*H*-1,2,3-Triazole	$C_2H_3N_3$			
1,2,3,4-Tetrazole	1*H*-1,2,3,4-Tetrazole	CH_2N_4			
	Benzene-1,3,5-tris(1*H*-tetrazole)	$C_9H_6N_{12}$	BTT	$Mn_3[(Mn_4Cl)_3(BTT)_8]_2$ Mn-BTT	

(77 K), and finding MOFs that are efficient at room temperature is a very active area of research. Unfortunately, at room temperature, the thermal energy of the molecule becomes larger than the very weak van der Waals interaction of the physisorbed molecules. In order to create MOFs that adsorb hydrogen sufficiently well to be useful at room temperature, methods have to be found to strengthen the interaction between the hydrogen and the metal. One possible method is to decrease the pore size, thereby bringing them into closer contact; a pore size of diameter 0.7 nm allows two adjacent physisorbed layers to form, one on each surface. Widening the pore to a diameter of 1 nm allows a third layer to form in between. Another method is to create a high concentration of exposed metal sites on the surfaces and/or to increase the charge density of the metal ions so that a dipole is induced in the hydrogen molecule, resulting in stronger bonding. Two MOF structures that show very high hydrogen absorption at low temperatures are MOF-177 (surface area 5640 $m^2\ g^{-1}$) and MOF-5 (see Table 6.4). The structures of MOF-5 ($Zn_4O(BDC)_3$) and MOF-177 ($Zn_4O(BDC)_2$) are illustrated in Figure 6.27: MOF-5 has a square opening to the pores with widths of 1.38 and 0.92 nm. MOF-177 consists of six diamond-shaped channels, of diameter 1.08 nm, surrounding a pore that can accommodate a sphere of diameter 1.18 nm. MOF-5 uptakes 66 g l^{-1} of H_2 at 77 K and 100 bar, although this decreases to ~9 g l^{-1} at room temperature. It has been the subject of much computational modelling because of the partial charges on its surfaces. Hydrogen is not the only gas capable of storage in MOFs, although it has been the most studied, but carbon monoxide, oxygen and methane can also be stored in particular MOFs. The MOF PCN-14 synthesised from copper and an anthracene derivative is particularly good at methane storage. In 2008, the German adventurer Rainer Zietlow completed a round the world 'EcoFuel' trip of 28,000 miles in a specially adapted Volkswagen. The car ran only on methane fuel that was stored in 13 cylinders, some of them packed with MOFs manufactured by BASF. The van was able to drive for more than 1500 miles between fuelling stops.

The production of carbon dioxide by fossil fuel burning has produced an increased concentration of this greenhouse gas in the atmosphere and is giving rise to worldwide concern. The ability to clean up emission gases from hydrocarbon fuels, so-called carbon capture, has become increasingly urgent as has the safe storage and transport of hydrogen for hydrogen-fuelled vehicles. The chromium-containing MOF, with carboxylate linker, MIL-101, has a large surface area–to-weight ratio of 5900 $m^2\ g^{-1}$ and large pores of diameter between 3 and 3.5 nm. It shows record CO_2 storage ability, selectively absorbing ~400× its own volume of carbon dioxide. Other MOF structures with imidazolate linkers are known as ZIFs because they resemble zeolite structures; ZIF-100 and ZIF-69 selectively adsorb large volumes of CO_2.

There is also the possibility in the future that if a suitable MOF were found, CO_2 could be absorbed and then catalytically converted to a useful product—perhaps even to another useful fuel such as methanol (CH_3OH). One example of the imagination of the developers in the automobile industry was seen at the Beijing Auto Show in 2010, where Shanghai Automotive Industry Corporation (SAIC) unveiled their 'Ye Zi' (leaf) concept vehicle (Figure 6.28). The leaf-shaped roof contains photovoltaic cells which follow the Sun, on the wheels are tiny wind turbines, and the body is made from an MOF that absorbs carbon dioxide and water and produces electricity

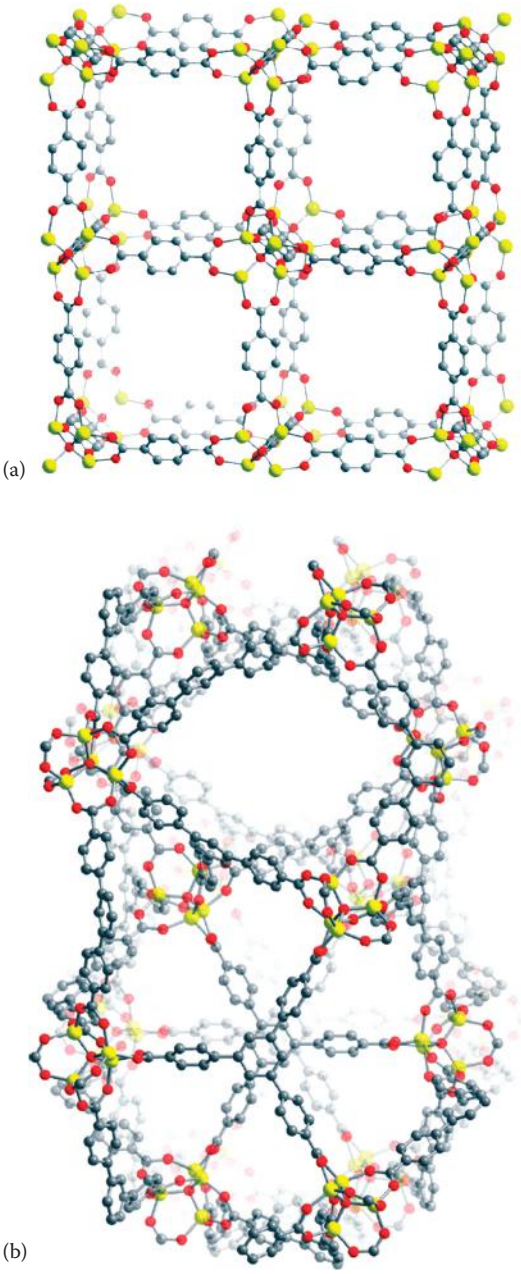


FIGURE 6.27 A portion of the crystal structure of (a) $Zn_4O(BDC)_3$ (MOF-5) and (b) $Zn_4O(BTB)_3$ (MOF-177). The yellow, grey and red spheres represent Zn, C and O atoms, respectively; H atoms are omitted for clarity. (From Murray, L.J., Dincă, M., and Long, J.R., *Chem. Soc. Rev.*, **38**, 1294, 2009. With permission.)

(a)

(b)

FIGURE 6.28 (a, b) The Shanghai Automotive Industry Corporation (SAIC) 'Ye Zi' (leaf) concept vehicle. (Courtesy of SAIC.)

and oxygen in emulation of photosynthesis. The electricity thus produced is stored in lithium-ion batteries (Chapter 5).

Drug delivery is another possible facet of the storage ability of MOFs, with the possibility of the slow release of painkillers, antitumour drugs, and so on.

The pore sizes in an MOF dictate which gaseous molecules are able to diffuse through them, making them useful candidates for **gas separation and purification**. Zn-based MOFs that retain CO_2 but allow through O_2 and N_2 have been made, whereas others will retain water but keep out larger molecules such as N_2, O_2, CO_2 and CH_4.

The metal centres in MOFs can act as Lewis acids to bases such as amines, sulfur-containing compounds, phosphines, etc. This makes them potentially useful for the clean-up of trace contaminants of these odoriferous compounds; also, of course, for the removal of poisonous gases, such as carbon dioxide, ammonia, chlorine and so on.

Heterogeneous catalysis is very important to the industry with about nine out of ten industrial processes using such catalysts, overall worth tens of millions of pounds to global economies. The combination of the high accessibility of the channels and the exposed metal sites on the surfaces makes MOFs very effective catalysts; the

surfaces can also be modified after synthesis. The pore structure enables size/shape-selective catalysis as in zeolites. However, an industrial catalyst has to be robust enough to withstand the temperature and pressure conditions of the reaction, and to be a viable economic concern, it must also cycle reliably hundreds of times. One of the major problems facing research into MOF catalysts has been their lack of stability at high temperatures compared with zeolites. Stability up to >500°C and for a pH range of 2–14 has been found for some benzene-1,3,5-tris(1*H*-imidazole) MOFs (Zn, Cu, Co, Ni). Others will certainly follow in the future in this very active area of research.

It is possible for an MOF to display **luminescence** for several different reasons, and these are depicted schematically in Figure 6.29:

- The organic linkers emit, or interactions between the linkers and a guest molecule produce, luminescence.
- Framework lanthanoid metal ions emit light that may be enhanced by the linker or by an adsorbed molecule in the pores.
- Luminescent molecules can be trapped in the pores or attached to a surface.

If the luminescence of a particular MOF can be activated by a guest molecule and/or the intensity of the luminescence is measured as a function of the concentration of the guest molecule, then this gives a perfect system for switching or for the sensor detection of the guest molecule.

In summary, MOFs are currently (2011) a very active area of research of great interest to industry, not only in the more developed areas of gas storage and separation but also in many other exciting areas. With a palette of metals across the

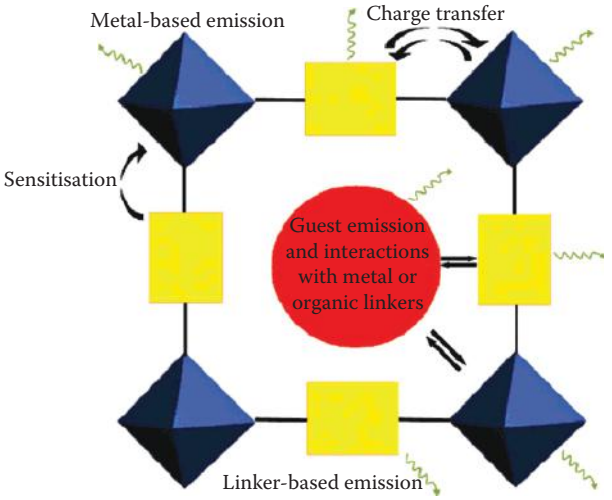


FIGURE 6.29 A representation of the emission possibilities in a porous MOF, wherein metal clusters (blue octahedra) are linked by organic linkers (yellow rectangles) with an incorporated guest (red circle). (From Allendorf, M.D., Bauer, C.A., Bhakta, R.K. and Houk, R.J.T., *Chem. Soc. Rev.*, **38**, 1330, 2009. With permission.)

periodic table to use and thousands of possible organic linker molecules, the possibilities for making new MOFs are boundless.

6.3.5 Metal Oxide Frameworks

The synthesis of crystalline, thermally stable, mesoporous transition metal oxides has proved difficult. Interest in this type of compound once again lies in the possibilities of utilising their catalytic, optical and magnetic properties. One successful approach has been to form the oxides around templates such as mesoporous silicates or block copolymers. When a hydrophobic block copolymer is used to direct the synthesis, heating converts the polymer to a rigid carbon support while the oxide crystallises around it; various metals, including niobium and titanium, have been synthesised by this technique. A complex synthesis of mesoporous γ-Fe_2O_3 was achieved by first forming mesoporous α-Fe_2O_3, reducing it to Fe_3O_4 and then reoxidising it, when the γ-form crystallised.

6.4 NEW MATERIALS

Research in zeolites has also branched out to try and prepare new and smart materials by incorporating various molecules and ions in the cages of these microporous and mesoporous structures, with possible use for bioencapsulation of proteins and enzymes, and for drug delivery. Another early example was the preparation of the pigment ultramarine, which is used in many paints and colourants. It is based on the zeolite SOD structure and contains S_3^- ions trapped in the cages; this is the same anion found in the mineral lapis lazuli, to which it imparts a beautiful deep blue colour. The treatment of zeolites such as Na-zeolite Y with sodium vapour, traps Na_4^{3+} ions in the cavities, which impart a deep red colour.

Small metal clusters can be incorporated into the pores of MCM-41 by encapsulating an organometallic compound by absorption and then decomposing it at low temperatures (2°C–300°C). Nanometre-size Sn–Mo clusters have been made in this way.

Polyaniline has been formed in the pores of Cu- or Fe-exchanged MCM-41 by the adsorption of aniline vapour and subsequent oxidative polymerisation (Figure 6.25), and these **molecular wires** show significant electronic conduction, although less than that of bulk polyaniline. Pyrolysis of polyacrylonitrile in the pores produces a graphite-like carbon chain, which exhibits microwave conductivity 10 times that of bulk carbonised polyacrylonitrile. Such materials have the potential for use in information processing as storage capacitors.

One area of this research has focussed on depositing semiconductor materials in zeolite cages. The resulting nanometre-size particles (sometimes called **quantum dots**) exhibit quantum size effects in their physical properties; that is, they have interesting electronic, magnetic and optical properties that are a consequence of their size rather than their chemical composition. One example of this is the narrow band gap semiconductor CdS, which in low concentrations forms discrete cubic $(CdS)_4$ clusters in the sodalite cages of zeolite A, zeolite X and zeolite Y. These clusters start to absorb in the UV at 350 nm, compared with 540 nm in bulk CdS,

and also they do not exhibit the intense emission of the bulk compound. As the pore-filling of the zeolite increases, the nanoparticles become interconnected forming superclusters, and materials are produced with properties intermediate between those of discrete particles and bulk semiconductors; the absorption edge shifts to 420 nm, and a strong emission also appears. Other semiconductors, such as Se, Te and GaP, have also been encapsulated in this way, with the hope that this research will lead to possibilities for optical transistors and optical data storage devices. Other guests that can be incorporated into zeolite cages and into mesoporous structures include the alkali metals, silver and silver salts, selenium and various conducting polymers.

Silicon nanoclusters have been formed in the cages by the decomposition of disilane in the pores to give photoluminescent composites, and a smart switching system has been made in MCM-41 by the photoreactivity of coumarin derivatives that were grafted onto the pore entrances. When irradiated, dimerisation takes place and the pores are closed, but irradiation with a shorter wavelength reverses the reaction and reopens the pores so that uptake and storage can take place.

Suffice it to say that in this rapidly developing field, new materials are being investigated with respect to their interesting physical properties, which then have the possibility of commercial exploitation—properties such as semiconductivity, photoconductivity and fast-ion conductivity, luminescence, colour and quantum size effects.

6.5 CLAY MINERALS

Like the zeolites, the smectite clays are a class of naturally occurring aluminosilicate minerals. They are often found as components of soils and sediments, and there are some large deposits, such as bentonite (consisting mainly of montmorillonite and beidellite) found in Cornwall, United Kingdom, and in Wyoming, United States. The smectite clays have a basic layer structure, which is shown in Figure 6.30; it consists of parallel layers of tetrahedral silicate [SiO_4] sheets and octahedral aluminate [$Al(O,OH)_6$] sheets. The silicate layers contain [SiO_4] tetrahedra, each linked through three corners to form an infinite layer. The aluminate layers contain a plane of octahedrally coordinated edge-linked aluminium ions, sandwiched between two inward pointing sheets of corner-linked [SiO_4] tetrahedra. If there is no replacement of Si or Al, then the layers are electrically neutral, producing the mineral pyrophyllite. In the smectite clays, different structures are formed because the substitution of silicon and aluminium by metal ions can take place in both the tetrahedral and the octahedral layers. The resulting negative charge is distributed on the oxygens of the layer surface, and any charge balance is restored by interlayer cations (usually Na^+ or Ca^{2+}); for instance, in montmorillonite, approximately one-sixth of the Al^{3+} ions have been replaced by Mg^{2+}, whereas in beidellite, about one-twelfth of the Si^{4+} have been replaced by Al^{3+}. The different members of the smectite group of clays are distinguished by the type and position of the cations in the framework.

The peculiar layer structure of these clays gives them cation exchange and intercalation properties that can be very useful. Molecules such as water, and polar organic

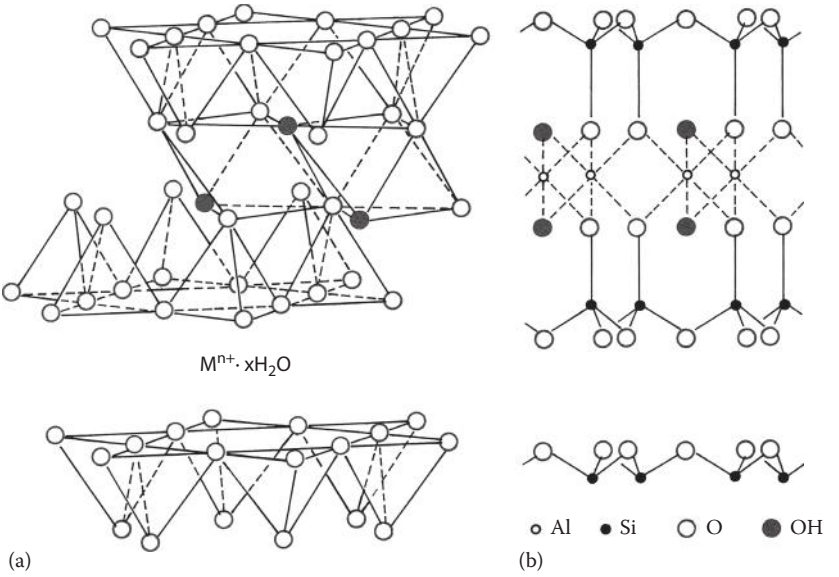


FIGURE 6.30 Idealised structure of the layers in a smectite clay, (a) showing only the oxygen/hydroxyl framework and (b) also showing the aluminium and silicon positions.

molecules such as glycol can easily intercalate between the layers and cause the clay to swell. Water enters the interlayer region as integral numbers of complete layers. Calcium montmorillonite usually has two layers of water molecules, but the sodium form can have one, two or three water layers; this causes the interlayer spacing to increase stepwise from about 960 pm in the dehydrated clay to 1250, 1550 and 1900 pm, as each successive layer of water forms.

The Na^+ and Ca^{2+} cations that make up the charge balance due to the substitution of Si and Al in the layers by other metals are usually hydrated and are located in the interlayer regions, loosely bound to the layer surfaces. They are known as **exchangeable cations** and can easily be replaced by other cations using ion exchange methods or by protons in the form of H_3O^+, to form acidic clays. Such acidic clays form very useful catalysts when the reactant molecules enter the interlayer regions. For many years, modified clays were used as the catalyst for petroleum cracking to produce petrol, although they have since been replaced by more thermally stable and selective zeolites. However, smectite clays can be used for the dimerisation of oleic acid, for the conversion of hexene to dihexyl ethers and for the formation of ethyl acetate from ethanoic acid and ethene. Recent research on clay catalysts has concentrated on the introduction of metal complexes into the interlayer region to act as the catalytic centre and also on the formation of robust molecular props to hold the interlayer regions apart—the so-called **pillaring**.

One of the problems with using clays as catalysts is that at high temperatures (>200°C) they start to dehydrate and the interlayer region collapses. A lot of effort has concentrated on forming temperature-stable pillars, which will hold the layers apart even at high temperature and in the absence of a swelling solvent, to form a

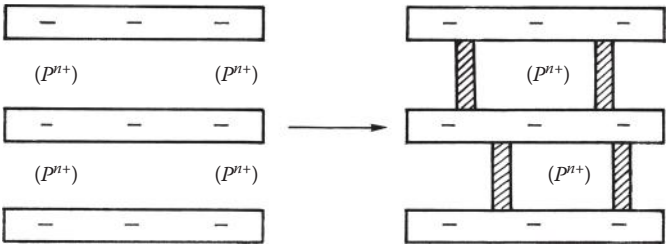


FIGURE 6.31 A representation of the formation of pillars in a smectite clay, where P^{n+} is the pillaring cation.

large interlayer region. The pillaring leads to a porous network between the layers analogous to the zeolite channels (Figure 6.31), and the hope is that by judicious adjustment of the size and spacings of the pillars, the pore sizes can be made bigger than those of the faujasitic zeolites. Because the layers are negatively charged, the molecular props that have been used for pillaring are large cations, such as tetraalkyl ammonium ions and the 'Al_{13}' Keggin ion $[Al_{13}O_4(OH)_{28}^{3+}]$; once the ion is incorporated, the clay is heated to dehydrate it, resulting in columns of small clusters of alumina (Al_2O_3), which are thermally stable up to 500°C. Other pillars can be formed from silica and the oxides of various metals, such as iron, zirconium and tin.

6.6 SUMMARY

The worldwide production of naturally occurring zeolites is about 4 million tonnes, predominantly clinoptilolite and chabazite from China and Cuba. The production of synthetic zeolites is less than half that of the natural zeolites at about 1.6 million tonnes; of this, 82% (zeolite A and zeolite X) are used for water softening in detergents, 5% (zeolite A, zeolite X, zeolite Y and mordenite) for absorbents and desiccants and 8% for catalysts—mostly zeolite Y but 12 other catalysts also have significant use. Table 6.3 summarises the main catalytic processes used in the petrochemical and refining industries.

6.7 POSTSCRIPT

We end this chapter with a quotation from a lecture given by Professor J. M. Thomas some years ago. It should please those of you who love the patterns and symmetry of crystalline solids.

> "… Figure 6.32 shows, on the right, a projected structure of the zeolite that we have been studying, a synthetic catalyst discovered in New Jersey in 1975. Its structure was elucidated some 6 years ago. On the left, I reproduce a pattern made on the wall of a mosque in Baku in the Soviet state of Azerbaijan in 1086 AD, the year that the *Domesday Book* appeared by order of William the Conqueror. These two structures have exactly the same pattern. 'There is nothing new under the sun'."

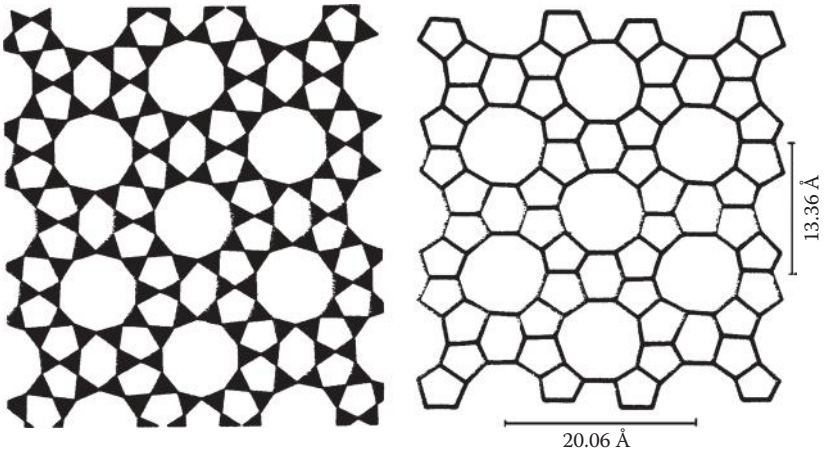


FIGURE 6.32 Around a central 10-membered ring in the synthetic zeolite catalyst, shown schematically on the right, there are eight five-membered and two six-membered rings. In exactly the same sequence, such five-membered and six-membered rings circumscribe a central 10-membered ring in the pattern on the left.

QUESTIONS

1. Zeolite A has a single peak in the ^{29}Si MAS NMR spectrum at 89 ppm and has an Si/Al ratio of 1. Comment on these observations.
2. A series of zeolites were synthesised with an increasing Si/Al ratio. It was found that the catalytic activity also increased until the ratio was about 15 : 1, after which it declined. Suggest a reason for this behaviour.
3. Zeolite A (Ca form), when loaded with platinum, has been found to be a good catalyst for the oxidation of hydrocarbon mixtures. However, if the mixture contains branched-chain hydrocarbons, these do not react. Suggest a possible reason.
4. Both ethene and propene can diffuse into the channels of a particular mordenite catalyst used for hydrogenation. Explain why only ethane is produced.
5. Explain why, when 3-methyl pentane and *n*-hexane are cracked over zeolite A (Ca form) to produce smaller hydrocarbons, the percentage conversion for 3-methyl pentane is less than 1%, whereas that for *n*-hexane is 9.2%.
6. When toluene is alkylated by methanol with a ZSM-5 catalyst, an increase in the crystallite size from 0.5 to 3 μm approximately doubles the amount of para-xylene produced. Suggest a possible explanation.
7. The infrared stretching frequency of the hydroxyl associated with the Brønsted sites in decationised zeolites falls in the range of 3600–3660 cm^{-1}. As the Si/Al ratio in the framework increases, this frequency tends to decrease. What does this suggest about the acidity of the highly siliceous zeolites?

7 Optical Properties of Solids

7.1 INTRODUCTION

Perhaps the most well-known example of solid-state optical devices is the laser and, with the widespread use of CDs, DVDs and laser printers, lasers have become commonplace. Of interest to the solid-state chemist are two types of laser, typified by the ruby laser and the gallium arsenide laser. Because laser light is more easily modulated than light from other sources, it is increasingly used for sending information; light travelling along optical fibres replacing electrons travelling along wires in, for example, telecommunications. To transmit light over long distances, the optical fibres must have particular absorption and refraction properties, and the development of suitable substances has become an important area of research.

Of course, the applications of solid-state optical properties were known and commercially important before lasers were developed and we also consider examples of these. Light-emitting diodes (LEDs) are used for displays, including those on digital watches and scientific instruments. The mechanism by which light is produced in LEDs is similar to that in the gallium arsenide laser. Another important group of solids are the light-emitting solids, known as phosphors, which are used on plasma television screens and in fluorescent light tubes.

Very broadly speaking, there are two situations that have to be considered in explaining the above-mentioned devices in which light is absorbed or emitted. In the first, which is relevant to the ruby laser and to phosphors for fluorescent lights, the light is emitted by an impurity ion in a host lattice. We are concerned here with what is essentially an atomic spectrum modified by the lattice. In the second case, which applies to LEDs and the gallium arsenide laser, the optical properties of the delocalised electrons in the bulk solid are important.

Some devices, such as optical fibres and 'cloaks of invisibility', depend on the refractive properties of materials; therefore, following on from discussions of devices that depend on the absorption/emission of light, the process of refraction is considered.

In the past three decades, a new type of material, photonic crystals, with the potential to form the basis of optical integrated circuits has been the subject of much interest. We shall look briefly at what these are and their potential uses. Finally, we consider the unusual optical properties of metamaterials, a class of solids that has attracted much attention due to their potential use in devices to render objects invisible. Photonic crystals and metamaterials differ from the other examples in that their characteristic properties are a result of their construction rather than their composition.

7.2 INTERACTION OF LIGHT WITH ATOMS

When an atom absorbs a photon of light of the correct wavelength, it undergoes a transition to a higher energy level. To a first approximation in many cases, we can think of one electron in the atom absorbing the photon and being excited. The electron will only absorb the photon if the energy of the photon matches that of the energy difference between the initial and final electronic energy levels, and if certain rules, known as **selection rules**, are obeyed. In light atoms, the electron cannot change its spin and its orbital angular momentum must change by one unit; in terms of quantum numbers: $\Delta s = 0$, and $\Delta l = \pm 1$; there is no restriction on the changes of the principal quantum number. One way of thinking about this is that the photon has zero spin and one unit of angular momentum. The conservation of spin and angular momentum then produces these rules. For a sodium atom, for example, the 3s electron can absorb one photon and go to the 3p level. The 3s electron will not, however, go to the 3d ($\Delta l = +2$) or 4s level ($\Delta l = 0$). Figure 7.1 illustrates allowed and forbidden transitions.

However, the spin and the orbital angular momenta are not entirely independent and coupling between them allows forbidden transitions to occur, although the probability of an electron absorbing a photon and being excited to a forbidden level is much less than the probability of it being excited to an allowed level. Consequently, spectral lines corresponding to forbidden transitions are less intense than those corresponding to allowed transitions.

An electron that has been excited to a higher energy level will sooner or later return to the ground state. It can do this in several ways. The electron may simply

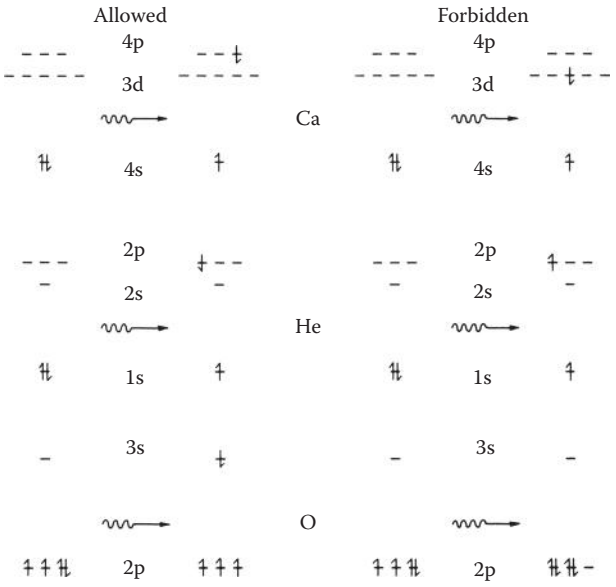


FIGURE 7.1 Allowed and forbidden atomic transitions.

emit a photon of the correct wavelength at random, some time after it has been excited. This is known as **spontaneous emission**. Alternatively, a second photon may come along and instead of being absorbed, may induce the electron to emit. This is known as **induced** or **stimulated emission** and plays an important role in the action of lasers. The emitted photon in this case is in phase with, and travelling in the same direction as, the photon inducing the emission; the resulting beam of light is said to be **coherent**. Finally, the atom may collide with another atom, losing energy in the process or giving energy to its surroundings in the form of vibrational energy. These are examples of **nonradiative transitions**. Spontaneous and stimulated emissions obey the same selection rules as absorption. Nonradiative transitions have different rules. In a crystal, as in a molecule, the atomic energy levels and the selection rules are modified. As an example, let us take an ion with one d electron outside a closed shell (Ti^{3+}, for example). This will help us understand the ruby laser.

In the free ion, the five 3d orbitals all have the same energy. In a crystal, these levels are split; for example, if the ion occupied an octahedral hole, the 3d levels would be split into a lower, triply degenerate (t_{2g}) level and a higher, doubly degenerate (e_g) level. This is shown in Figure 7.2. An electronic transition between these levels is now possible. In the free ion, a transition from one d level to another involves zero energy change, so it would not be observed even if it were allowed. In the crystal, the transition involves a change in energy, but it is still forbidden by the selection rules. Lines corresponding to such transitions can, however, be observed, albeit with low intensity, because the crystal vibrations mix different electronic energy levels. Thus, the 3d levels may be mixed with the 4p levels, giving a small fraction of 'allowedness' to the transition. Figure 7.3 shows an absorption band due to a transition from t_{2g} to e_g for the ion Ti^{3+}, which has one d electron. (This band is, in fact, two closely spaced bands because the excited state is distorted from a true octahedron and the e_g level is further split into two.)

We shall see now the role played by a similar forbidden transition in the operation of the ruby laser.

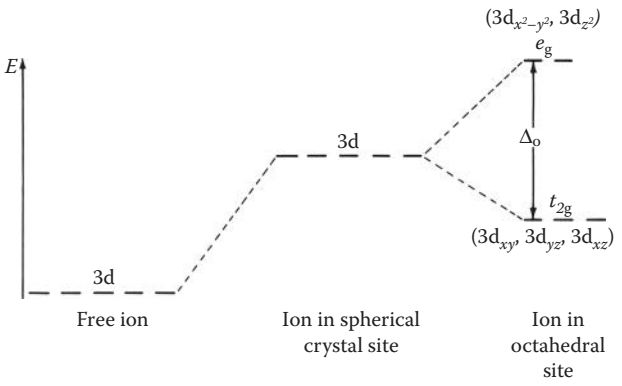


FIGURE 7.2 Splitting of d levels in an octahedral site in a crystal.

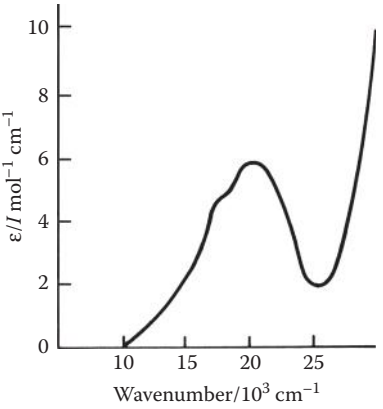


FIGURE 7.3 The t_{2g} to e_g transition of Ti^{3+}. The band is, in fact, two overlapping bands. This is due to a further splitting of the e_g levels.

7.2.1 Ruby Laser

Ruby is corundum (one form of Al_2O_3, see Chapter 1) with 0.04%–0.5% Cr^{3+} ions as an impurity replacing the aluminium ions. The aluminium ions, and hence the chromium ions, occupy distorted octahedral sites. Thus, as discussed above, the 3d levels of the chromium ion will be split. Cr^{3+} has three 3d electrons, and, in the ground state, these occupy separate orbitals with parallel spins. When light is absorbed, one of these electrons can undergo a transition to a higher energy 3d level; this is similar to the transition discussed in the previous section, but with three electrons it is necessary to consider the changes in the electron repulsion as well as the changes in orbital energies. When the changes in the electron repulsion are included, we find that there are two transitions at different energies, which correspond to the jump from the lower to the higher 3d level.

Having absorbed light and undergone one of these transitions, the chromium ion could now simply emit radiation of the same wavelength and return to the ground state. However, in ruby, there is a fast, radiationless transition in which the excited electron loses some of its energy and the crystal gains vibrational energy. The chromium ion is left in a state in which it can only return to the ground state by a forbidden transition in which an electron changes its spin. Such a transition is *doubly* forbidden because it also breaks the rule that forbids $3d \leftrightarrow 3d$, so it is even less likely to occur than the original absorption process. The states involved are shown schematically in Figure 7.4. The Cr^{3+} ions absorb light and go to states 3 and 4. They then undergo a radiationless transition to state 2. Because the probability of spontaneous emission for state 2 is low, and there is no convenient nonradiative route to the ground state, a considerable population of state 2 can build up. When eventually (about 5 msec later) some of the ions in state 2 return to the ground state, the first few spontaneously emitted photons interact with other ions in state 2 and induce these to emit. The resulting photons will be in phase and travelling in the same direction as the spontaneously emitted photons and will induce further emission as they travel through the ruby. In the laser, the ruby is enclosed by a reflecting cavity so that the

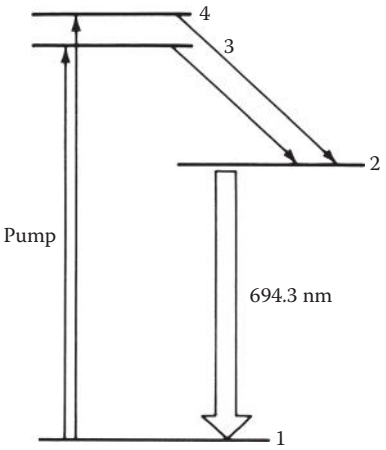


FIGURE 7.4 The states of the Cr^{3+} ion involved in the ruby laser transition.

photons are reflected back into the crystal when they reach the edge. The reflected photons induce further emission and by this means, an appreciable beam of coherent light is built up. The mirror on one end can then be removed and a pulse of light emitted. The name laser is a reflection of this build-up of intensity. It is an acronym standing for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation. (Similar devices producing coherent beams of microwave radiation are known as masers.) A typical arrangement for a pulsed ruby laser is shown in Figure 7.5. A high-intensity flash lamp excites the Cr^{3+} ions from level 1 to levels 3 and 4. The lamp can lie alongside the crystalline rod of ruby or can be wrapped around it. At one end of the reflective cavity surrounding the ruby crystal is a Q switch, which switches from being reflective to transmitting the laser light and can be as simple as a rotating segmented mirror, but it is usually a more complex device.

Ruby was the first material used for solid-state lasers, but now several other crystals are employed. The crystals used need to contain an impurity with an energy level such that a return to the ground state is only possible by a forbidden transition in the infrared, visible or near ultraviolet. It must also be possible to populate this

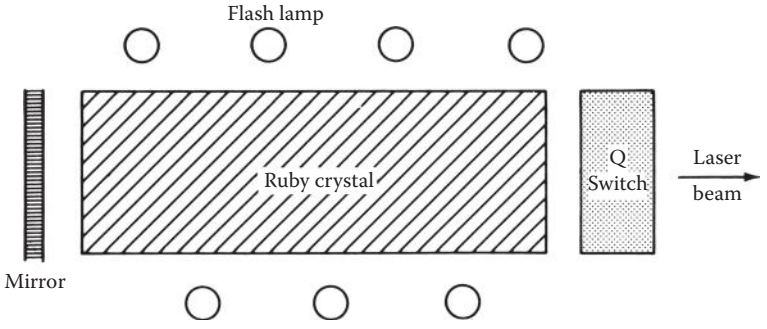


FIGURE 7.5 A sketch of a ruby laser.

TABLE 7.1
Examples of Crystals Used as Lasers

Ion	Host	Wavelength Emitted/nm
Ti^{3+}	Sapphire	650–1100 (tunable)
Nd^{3+}	Fluorite (CaF_2)	1046
Sm^{3+}	Fluorite	708.5
Ho^{3+}	Fluorite	2090
Nd^{3+}	Calcium tungstate ($CaWO_4$)	1060
Nd^{3+}	Yttrium vanadate (YVO_4)	1064
Nd^{3+}	Yttrium aluminium garnet (YAG, $Y_3Al_5O_{12}$) (Nd/YAG laser)	1064

level through an allowed (or at least less forbidden) transition. Research has tended to concentrate on transition metal ions and lanthanoid ions in various hosts, because these ions have suitable transitions of the right wavelength. Some examples are given in Table 7.1 with the wavelength of the laser emission.

7.2.2 Phosphors in Fluorescent Lights

Phosphors are solids that absorb energy and re-emit it as light. As in the lasers that we have just described, the emitter is usually an impurity ion in a host lattice. However, for the general uses of phosphors, it is not necessary to produce intense, coherent beams of light, and the emitting process is spontaneous rather than induced. There are many applications of phosphors, for example, the colours of plasma television screens are produced by phosphors that are bombarded with electrons from a beam. In terms of tonnage produced, one of the most important applications is in fluorescent light tubes, including low-energy bulbs.

Fluorescent lights produce radiation in the ultraviolet (254 nm) by passing an electric discharge through a low pressure of mercury vapour. The tube is coated inside with a white powder that absorbs the ultraviolet light and emits visible radiation. For a good fluorescent light, the efficiency of the conversion should be high and the emitted light should be such that the appearance of everyday objects viewed by it should resemble as closely as possible their appearance in daylight. Most phosphors for fluorescent lights have been based on alkaline earth halophosphates, such as $3Ca_3(PO_4)_2{\cdot}CaF_2$. As in lasers, the usual dopants are the transition metal or the lanthanoid ions, but more than one impurity ion is needed to approximate the whole visible spectrum. Not all the impurity ions need to be capable of absorbing the exciting radiation as the host lattice can act to transmit the energy from one site to another. For example, in a phosphor doped with Mn^{2+} and Sb^{3+} ions, the ultraviolet radiation emitted by the mercury atoms is only absorbed by the antimony (Sb^{3+}) ions. The excited antimony ion drops down to a lower excited state via a nonradiative transition. Emission from this lower state produces a broad band in the blue region of the visible spectrum. Some of the energy absorbed by the antimony travels through the host crystal and is absorbed by the manganese

ions. The excited Mn^{2+} ions emit yellow light and return to the ground state. The two emission bands together produce something close to daylight. Other phosphors have been introduced that are more efficient and give a closer approximation to daylight. For example, a good approximation is given by a combination of blue light from barium magnesium aluminate doped with divalent europium (Eu^{2+}); green light from an aluminate doped with cerium (Ce^{3+}) and terbium (Tb^{3+}) ions; and red light from yttrium oxide doped with trivalent europium (Eu^{3+}).

Fluorescent lights also emit broadband radiation in the near infrared, and there has been much interest in finding luminescent materials that will convert this to visible light. In this case, the incident light is of lower energy than the emitted light and the process is known as **upconversion**. Such processes are also exploited in upconversion lasers, where a phosphor is used to produce shorter wavelength light from a red laser. Obviously, a ground-state atom or ion cannot absorb a photon of one frequency radiation and then emit a photon of higher frequency radiation from the excited state reached by the absorption process. So how does upconversion work?

In upconversion systems, absorption takes place in two stages. An ion absorbs a photon of the incident radiation and goes to an excited state. It then transfers most of the energy either to another state of that ion or to the excited state of another ion. If this second excited state is metastable, it has time to absorb another photon before it spontaneously emits radiation and returns to a lower state. Either the absorbing ion or the one to which the energy has been transferred is now in a higher level, whose energy above the ground state is greater than the energy of the absorbed photon. Emission from this higher level produces shorter wavelength radiation than that absorbed. The ions doped into upconvertors are frequently lanthanoid ions or a transition metal ion and a lanthanoid ion. For example, low concentrations of Ho^{3+} in Y_2O_3 absorb red laser light and emit yellow-green light. Ho^{3+} in the ground state ($^5I_8$) absorbs a photon and is excited to an excited state ($^5F_5$) from where it drops to a succession of lower energy states, ending in the $^5I_7$ state (Figure 7.6). While in the $^5I_7$

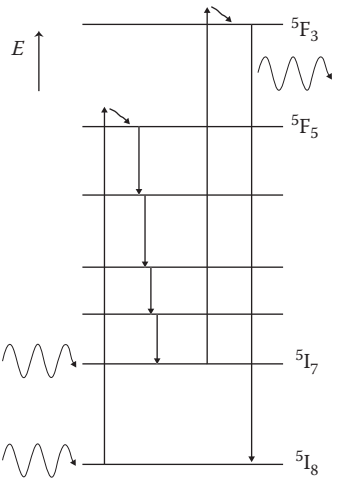


FIGURE 7.6 The atomic states of Ho^{3+}.

state, the ion absorbs another photon of red light and goes to the $^5F_3$ state from where it emits yellow-green light and returns to the ground state.

In phosphors and in the ruby laser, light is absorbed and emitted by electrons localised on an impurity site, but in other optical devices, the radiation is emitted by delocalised electrons. In the next section, therefore, we shall consider the absorption and emission of radiation in solids with delocalised electrons, particularly in semiconductors.

7.3 ABSORPTION AND EMISSION OF RADIATION IN CONTINUOUS SOLIDS

For continuous solids, the absorption and the emission of radiation involve transitions between energy bands rather than between discrete atomic energy levels. Radiation falling on an insulator or a semiconductor is absorbed by electrons in delocalised bands, in particular those near the top of the valence band, causing these electrons to be promoted to the conduction band. In metals, electrons can be promoted from the partially occupied conduction band to higher energy bands. Because there are many closely packed levels in an energy band, the absorption spectrum is not a series of lines as in atomic spectra, but a broad peak with a sharp threshold close to the band gap energy. The absorption spectrum of GaAs, for example, is shown in Figure 7.7.

Transitions to some levels in the conduction band are more likely than transitions to other levels. This is because transitions between levels in bands, like those between atomic energy levels, are governed by selection rules. The spin selection rule still holds; when promoted, the electron does not change its spin. However, the orbital angular momentum rules are not appropriate for energy bands, and the rule governing change in the quantum number l is replaced by a restriction on the wave vector (**k**). As seen in Chapter 4, the energy levels in a band are characterised by the wave vector, the momentum of the electron wave being given by $\mathbf{k}\hbar$. The momentum of a photon with a wavelength in the infrared, visible or ultraviolet region is very small compared with that of the electron in the band, so conservation of momentum produces the selection rule for transitions between bands: an electron cannot change its wave vector when it absorbs or

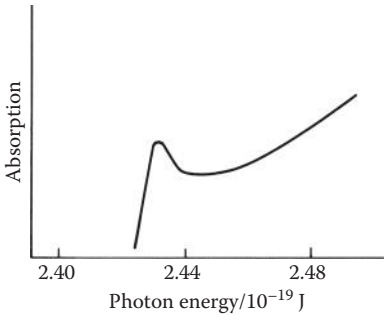


FIGURE 7.7 The absorption spectrum of GaAs.

emits radiation. Thus, an electron in the valence band with wave vector ($\mathbf{k}_i$) can only undergo allowed transitions to levels in the conduction band that also have the wave vector ($\mathbf{k}_i$). In some solids, for example, GaAs, the level at the top of the valence band and that at the bottom of the conduction band have the same wave vector. There is then an allowed transition at the band gap energy. Such solids are said to have a **direct band gap**.

For other semiconductors, for example silicon, the direct transition from the top of the valence band to the bottom of the conduction band is forbidden. These solids are said to have an **indirect band gap**. Illustrations of band structures for solids with direct and indirect band gaps are given in Figure 7.8. Note that in these diagrams, the energy for a band in *one direction* is plotted against the wave number (k). Similar diagrams were given in Chapter 4, but with the energy plotted against the density of states.

The simple free electron model might suggest that the lowest energy orbital in any band is that with $k = 0$. Figure 7.9, however, illustrates two combinations of orbitals that will have $k = 0$ for a chain because all the atomic orbitals are combined in phase. The combination of p orbitals is obviously antibonding and would be expected to have the highest energy in its band; the combination of s orbitals is bonding and would have the lowest energy in its band. If the p band lay below the s band, a transition between these levels would be allowed and would correspond to a direct transition across the band gap. In real solids, the highest and the lowest levels in the bands will contain contributions from different types of atomic orbital and it becomes difficult to predict whether a band gap will be direct or indirect.

One consequence of an indirect band gap is that an electron in the bottom level of the conduction band has only a small probability of emitting a photon and returning

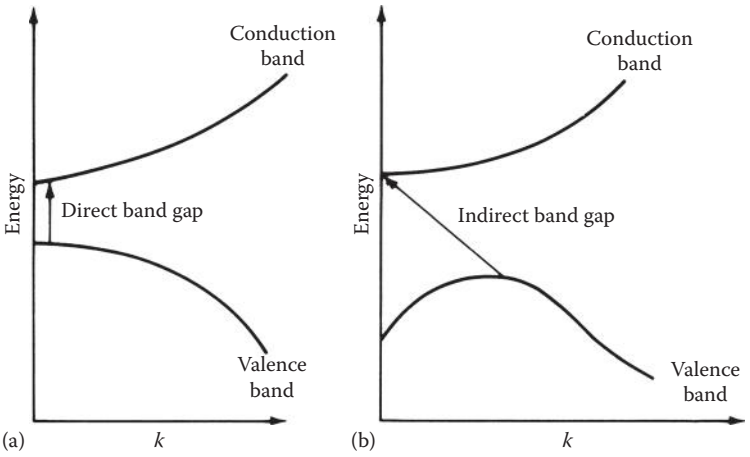


FIGURE 7.8 A sketch of the energy bands for (a) a solid with a direct band gap and (b) a solid with an indirect band gap. Note that in this diagram the horizontal axis is k, not the density of states. In this representation, a band is shown as a line going from 0 to the maximum value of k occurring for that band.

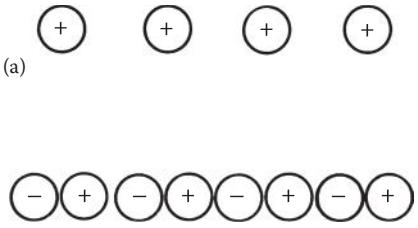


FIGURE 7.9 A row of (a) s orbitals and (b) p orbitals, both with $k = 0$.

to the top of the valence band. This is of importance when selecting materials for some of the applications we are going to consider.

Transitions across the band gap are also responsible for the appearance of many solids. Because a solid is very concentrated, the probability that a photon with energy that corresponds to an allowed transition, is absorbed, is very high. Many such photons will therefore be absorbed at or near the surface of the solid. These photons will then be re-emitted in random directions so that some will be reflected back towards the source of the radiation and some will travel further into the solid. Those travelling into the solid stand a very good chance of being reabsorbed and then re-emitted, again in random directions. The net effect is that the radiation does not penetrate the solid, but is reflected by its surface. If the surface is sufficiently regular, then solids that reflect visible radiation appear shiny. Thus, silicon, whose band gap is at the lower end of the visible region and has allowed transitions covering most of the visible wavelengths, appears shiny and metallic. Many metals have strong transitions between the conduction band and a higher energy band, which lead to their characteristic metallic sheen. Some metals, such as tungsten and zinc, have a band gap in the infrared region, and transitions in the visible region are not as strong. These metals appear relatively dull. Gold and copper have strong absorption bands due to the excitation of d band electrons to the s/p conduction band. In these elements, the d band is full and lies some way below the Fermi level (Figure 7.10).

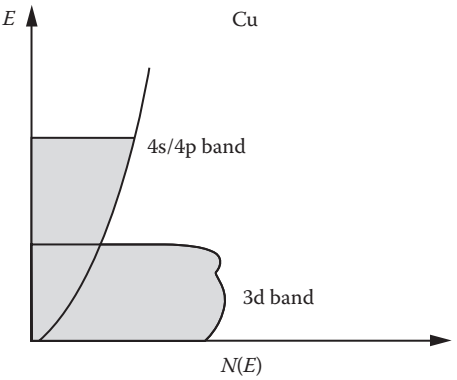


FIGURE 7.10 The band structure of copper.

The reflectivity peaks in the yellow part of the spectrum, and blue light and green light are less strongly absorbed; hence, the metals appear golden. Very thin films of gold appear green because the yellow light and red light are absorbed and only the blue and green are transmitted. Insulators typically have band gaps in the ultraviolet region and, unless there is a localised transition in the visible region, appear colourless.

The devices that we consider involving band gap transitions are, however, concerned with the emission of light rather than absorption or reflection; the electrons being initially excited by electrical energy.

7.3.1 Light-Emitting Diodes

LEDs are widely used for displays. Like transistors, they are based on the *p*–*n* junction, but the voltage applied across the *p*–*n* junction in this case leads to the emission of light. Figure 7.11 shows a *p*–*n* junction in a semiconductor such as GaAs. The band structure shown in Figure 7.11 is for the junction in the dark and with no electric field applied. Now suppose that an electrical field is applied so that the *n*-type semiconductor is made negatively charged relative to the *p*-type (i.e., in the reverse direction to the applied voltage in transistors, Chapter 4). Electrons will then flow from the *n*-type to the *p*-type. An electron in the conduction band moving to the *p*-type side can drop down into one of the vacancies in the valence band on the *p*-type side, emitting a photon in the process. This is more likely to happen if the transition is allowed, so that semiconductors with direct band gaps are usually used in such devices. To use the LED as a display, for example, it is then wired into a circuit so that an electric field is applied across the parts making up the required letters or numbers. Different colours can be produced by using semiconductors of differing band gap. GaP produces red light, but by mixing in various proportions of aluminium to form $Ga_{1-x}Al_xP$, green or orange light can be produced. Blue light is produced using InGaN sandwiched between GaN layers.

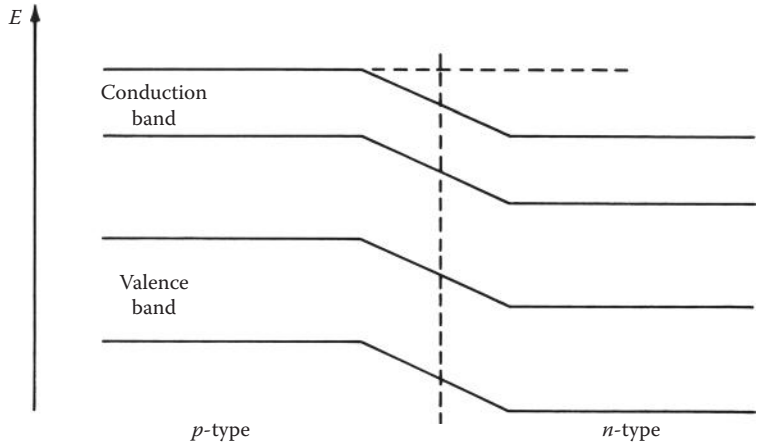


FIGURE 7.11 The energy bands near the junction in a *p*–*n* junction.

It should be noted that semiconductors with indirect band gaps are used for LEDs, but in these cases, the impurity levels play an important role. Thus, GaP is used although it has an indirect band gap. Silicon, however, is not suitable because there is a nonradiative transition available to electrons at the bottom of the conduction band and these electrons donate thermal energy to the crystal lattice rather than emitting light when they return to the valence band.

When the electric field causes conduction electrons to move across the *p*–*n* junction, the resulting situation is one in which the population in the conduction band is greater than the thermal equilibrium population. An excess of electrons in an excited state is an essential feature of lasers, and there are several semiconductor lasers based on the *p*–*n* junction. The best known of these is the gallium arsenide laser.

7.3.2 Gallium Arsenide Laser

The **gallium arsenide laser** actually contains a layer of GaAs sandwiched between layers of *p*-type and *n*-type gallium aluminium arsenide ($Ga_{1-x}Al_xAs$). As shown in Figure 7.12, the band gap of gallium aluminium arsenide is larger than that of gallium arsenide.

An electric field applied across the *p*–*n* junction as in LEDs produces an excess of electrons in the conduction band of the gallium arsenide. These electrons do not drift across into the gallium aluminium arsenide layer because the bottom of the conduction band in this layer is higher in energy, and the electrons would therefore need to gain energy in order to move across. The excess conduction band electrons are therefore constrained to remain within the GaAs layer. Eventually, one of these

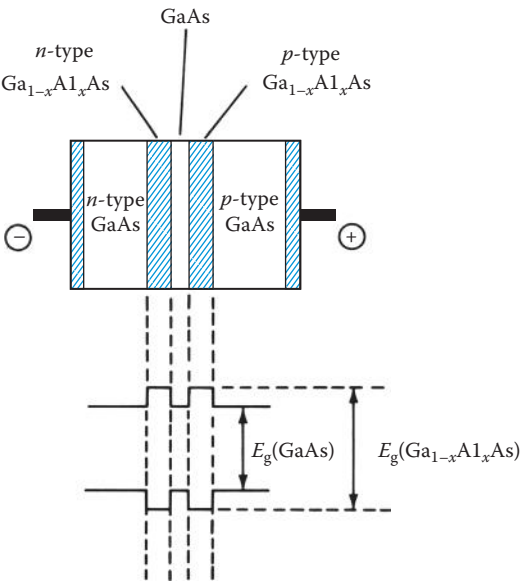


FIGURE 7.12 The arrangement of the different semiconductor regions in a GaAs laser. The band gap profile is shown below.

electrons drops down into the valence band, emitting a photon as it does so. This photon induces other conduction band electrons to return to the valence band and thus a coherent beam of light begins to build up. As in the ruby laser, the initial burst of photons is reflected back by mirrors placed at the ends, thereby inducing more emission. Eventually, a beam of infrared radiation is emitted.

Several such lasers have been developed, most of them based on III–V compounds—compounds of In, Ga and/or Al with As, P or Sb. It is possible to manufacture materials with band gaps over the range 400–1300 nm by carefully controlling the ratios of the different elements.

Infrared semiconductor lasers are used, among many other things, for reading compact discs. A compact disc consists of a plastic disc coated with a highly reflective aluminium film and is protected against mechanical damage by a layer of polymer. The original sound recording is split up into a number of frequency channels and the frequency of each channel is given as a binary code (i.e., a series of 0s and 1s). This is converted to a series of pits in tracks on the disk, spaced approximately 1.6 μm apart. A laser is focussed on the disk and reflected onto a photodetector. The pits cause some of the light to be scattered, thereby reducing the intensity of the reflected beam. The signal read by the photodetector is read as 1 when there is a high intensity of light and as 0 when the intensity is reduced by scattering. The binary code is thus recovered and can be converted back to sound or pictures. DVDs are similar, but they use a laser of shorter wavelength that can read more closely spaced pits, which means that more information can be packed into a given area.

7.3.3 Quantum Wells: Blue Lasers

Blue lasers allow a higher resolution, and hence a higher density of optical storage of information on devices such as DVDs, than the infrared GaAs lasers allow. The earliest blue lasers were based on ZnSe, but their lifetime proved too short for commercial applications. Lasers based on gallium nitride (GaN), first demonstrated in 1995, have proved to have greater lifetimes. In these lasers, the photons are produced in quantum wells rather than in a bulk semiconductor.

The active region of GaN lasers consists of GaN containing several thin layers (3–4 nm thick) of indium-doped GaN, $In_xGa_{1-x}N$. The addition of indium reduces the band gap within the thin layers, so that the bottom of the conduction band is at a lower energy than that of the bulk GaN. The electrons in this conduction band are effectively trapped because they need to gain energy from an external source to pass into the conduction band of the bulk GaN. Figure 7.13 schematically shows the conduction band for a series of thin layers of $In_xGa_{1-x}N$ in GaN.

The trapped electrons behave like particles in a box (Chapter 4), but with finite energy walls to the box. Such boxes are **quantum wells**. Within the well, the electron energy is quantised and the spacing of the energy levels depends on the (energy) depth and the (spatial) width of the well. The depth of the wells is controlled by the extent of the doping, that is, the value of x. Although the electrons in the wells have insufficient energy to surmount the energy barrier to reach the next well, there is a probability that they will move to a similar energy level in the next well through **quantum mechanical tunnelling** (see Chapter 2). The electrons travelling from the

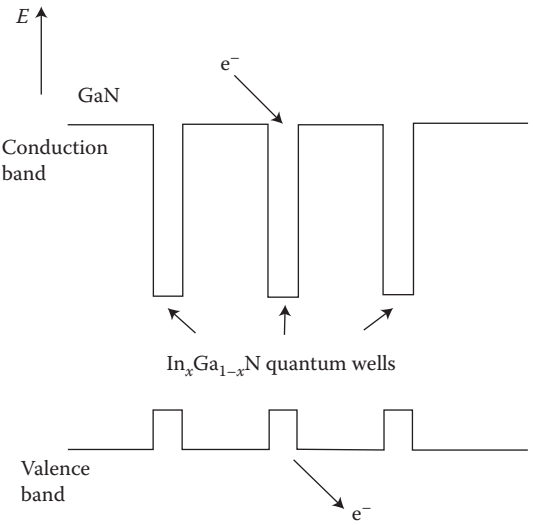


FIGURE 7.13 An energy level diagram for the conduction band in a series of thin layers of $In_xGa_{1-x}N$ in GaN.

bulk GaN enter a high level in the first well (Figure 7.13). From this level, the electrons can emit a photon and go to a lower level or can tunnel through to the next well. Tunnelling is faster than photon emission, so that the population of the higher levels builds up. There are now more electrons in the higher levels than in the lowest levels, that is, there is a population inversion—the requirement for laser action.

The active region containing the quantum wells is sandwiched between layers of *n*-doped and *p*-doped GaN and aluminium-doped GaN, $Al_yGa_{1-y}N$, which provide the electrons entering the quantum well and keep them confined to the active region. All these layers are built up on a substrate, for example, sapphire. The ends of the whole device are etched or cleaned to form a partial mirror that reflects the emitted photons, allowing a coherent beam to build up.

Blu-ray disks, introduced in 2006, are DVD-like devices using blue lasers to read the disk. However, the blue laser used is generally a diode-pumped solid-state laser in which light from an LED is used to excite ions in a solid-state laser, such as Nd/YAG (the crystal is an yttrium aluminium garnet doped with neodymium), and the frequency of the radiation emitted by the laser is doubled by passing it through a solid such as lithium borate.

7.4 REFRACTION

When light travels from a different medium through a solid, its velocity changes. A ray of light travelling at right angles to the surface of the solid will pass straight through it, but all other rays change direction. The size of the change in the velocity and hence the angle through which the radiation bends depend both on the material and the wavelength of the radiation. When electromagnetic radiation passes from a

medium with one refractive index, say n_1, to a medium with a different refractive index, say n_2, the radiation bends according to Snell's law, $n_1 \sin \theta_1 = n_2 \sin \theta_2$, where θ_1 and θ_2 are the angles between the radiation and the normal to the interface. This is illustrated in Figure 7.14.

To see why the refractive index is altered by composition, the atomic origin of the refractive index is considered briefly. Electromagnetic radiation has associated with it an oscillating electric field. Even when the radiation is not absorbed, this field has an effect on the electrons in the solid. If we think of an electrical field applied to an atom, we can imagine the electrons pulled by the field so that the atom is no longer spherical. The applied field produces a separation of the centres of positive and negative charges, that is, it induces an electric dipole moment. (A molecule in a solid may also have a permanent electric dipole moment produced by an unequal distribution of bonding electrons between the nuclei, but this is present in the absence of an applied field.) The oscillating field of the radiation can be thought of as pulling the electrons alternately one way and then the other way. The amount of response from the electrons depends on how tightly bound the electrons are to the nucleus. This property is called the **polarisability** and is higher for large ions with low charge, for example, Cs^+ and I^-, than for small, highly charged ions such as Al^{3+}. If there is a high concentration of polarisable ions in a solid, then the radiation will be slowed down more and the refractive index of the material increases. Adjusting the refractive index by adding carefully selected impurities is useful in a number of applications. For example, controlling the refractive index of glass is very important when making lenses for telescopes, binoculars and cameras. Lead ions (Pb^{2+}) are highly polarisable and are used to produce glass of a high refractive index.

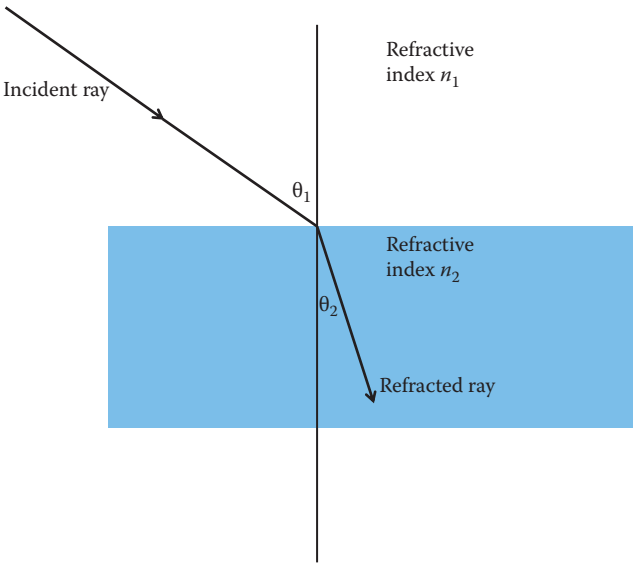


FIGURE 7.14 Diffraction of a ray of electromagnetic radiation when passing from one medium to another of a different refractive index.

7.4.1 Calcite

Calcite is the most stable polymorph of calcium carbonate ($CaCO_3$). Single crystals are transparent and display the interesting optical property of **birefringence**. Birefringent materials such as calcite have different polarisabilities in the directions of different crystal axes and hence different refractive indices for light polarised perpendicular to these axes. Calcite has a particular refractive index along a unique axis and a differing refractive index along the directions perpendicular to this axis. The unique axis is known as the **optical axis**. When light is passed through the material, it splits into two beams travelling at different speeds due to the different refractive indices. One beam obeys Snell's law and is known as the **ordinary ray**. The other is the **extraordinary ray**. The effect is illustrated in Figure 7.15, which shows the crystal structure of calcite; the optical axis is perpendicular to the planes of carbonate ions (CO_3^{2-}). The other two axes contain the planes. Thus, for calcite, the polarisability is not equal in all directions due to the asymmetry of the crystal structure; components in two directions are equal, but differ from that in the third direction, the optical axis. The ordinary rays are polarised in the plane perpendicular to the optical axis. The extraordinary rays are polarised in the plane parallel to the optical axis. If the beam enters at right angles to the surface, the ordinary ray is not deflected, but the extraordinary ray is deflected.

The interaction of the extraordinary ray with anisotropic oscillations in the crystal causes the ray to be propagated in a direction that is not at right angles to the wavefront and this causes it to deviate from the ordinary ray. Birefringence can only occur for crystals displaying asymmetry. It is not observed with cubic crystals, which are identical in all three directions, unless an asymmetrical stress is applied. Interestingly, the effect of stress is employed by engineers in testing structures. A model of the planned structure is made from clear plastic and then viewed through crossed polaroids. Where there are stresses in the structure, colours are seen.

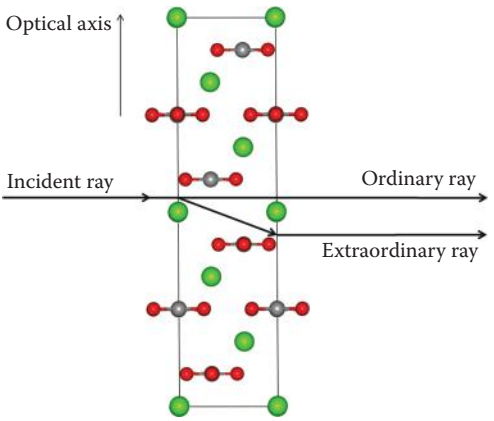


FIGURE 7.15 A ray of electromagnetic radiation passing into a birefringent crystal. The ray is entering perpendicular to the surface, which contains the optical axis, and the ordinary ray travels straight through.

One way of producing polarised light is to use a Nicol prism. The Nicol prism exploits the splitting by calcite of a beam of light into two rays to produce polarised light. Two crystals of calcite are glued together using Canada balsam cement that has a refractive index between that of the ordinary ray and that of the extraordinary ray. When light hits the prism at certain angles, the ordinary ray is totally internally reflected at the boundary, but the extraordinary ray continues into the second crystal. Thus, the light emerging from the second crystal is polarised.

At the end of 2010, the birefringence of calcite was used independently by two groups to produce a **cloak of invisibility** that worked at optical wavelengths; an object behind such a 'cloak' cannot be seen. George Barbastathis's group at the Singapore-MIT Alliance for Research and Technology Centre produced a 2-D cloak that worked under water, and a group led by John Pendry at Imperial College, London, and Shuang Zhang at the University of Birmingham, UK, has produced one that hides an object in air. The former optical cloak was made of two prisms of calcite glued together such that the optical axes of both are at 30° to the interface, as shown in Figure 7.16.

Polarised light enters one prism is refracted and then reflected by the object. The prisms are joined together at such an angle that the ray reflected from the object is refracted so that it appears to the observer that the light has simply been reflected off the base. Thus, both the object *and* the cloak are invisible.

7.4.2 Optical Fibres

Optical fibres are used to transmit light in the way that metal wires are used to transmit electricity. For example, a telephone call can be sent along an optical fibre in the form of a series of light pulses from a laser. The intensity, the time between pulses and the length of the pulse can be modified to convey the contents of the call in coded form. In order to transmit the information over useful distances (of the order of kilometres), the intensity of the light must be maintained so that there is

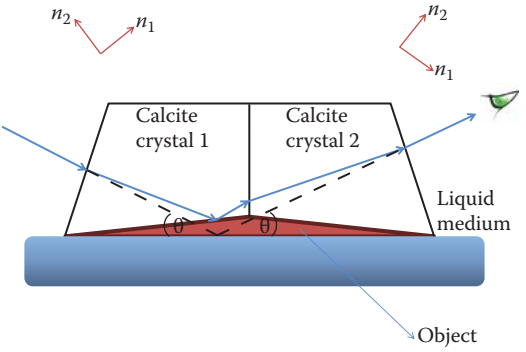


FIGURE 7.16 An optical cloak made from two crystals of calcite. The path of the light is shown by the blue trace. The ray exits in the direction it would have if reflected off the surface in the absence of object and cloak, as shown by the dashed trace. (Adapted from Zhang, BL., Luo, YA., Liu, XG., and Barbastathis, G., *Phys. Rev. Lett.*, **106**, 033901, 2011.)

still a detectable signal at the other end of the fibre. Thus, much of the art of making commercial optical fibres lies in finding ways of reducing energy loss as the beam travels along the fibre.

The first requirement is that the laser beam keeps within the fibre. Laser beams diverge less than conventional light beams so that using laser light is of help, but even so, there is a small tendency for the beam to stray outside the fibre. Therefore, fibres are usually constructed with a variable refractive index across the fibre. The beam is sent down a central core. The surrounding region has a lower refractive index than the core so that light deviating from a straight path is totally internally reflected and hence remains in the core. This is illustrated in Figure 7.17.

The refractive index can be varied by adding selected impurities. In the case illustrated in Figure 7.17, the totally internally reflected rays travel a longer path than those that travel straight along the core. This will lead to a pulse being spread out in time. One way to keep the pulses together is to use very narrow cores, so that essentially all the light travels the straight-ray path. An alternative is the variable refractive index core: the lower the refractive index, the faster is the speed of light. Thus, if the outer parts of the core have a lower refractive index, then the reflected light moves faster and this compensates for the longer path length.

There are bound to be some imperfections in the fibre and these are another source of energy loss. The imperfections cause scattering of the light of a type known as Rayleigh scattering. Rayleigh scattering does not cause any change in the wavelength of the light, only in its direction. The amount of scattering depends on $(1/\lambda^4)$, where λ is the wavelength, so there is much less scattering for longer wavelengths. Even the reduction in going from blue to red light is significant and is responsible for the colour of the sky. To reduce Rayleigh scattering, the lasers employed for optical fibre systems usually emit long-wavelength infrared radiation.

A third source of energy loss is absorption of light by the fibre. In a fibre several kilometres long, a very small amount of impurity can give rise to substantial absorption. We can get an idea of this by looking at a sheet of window glass edge-on. Instead of being clear, the glass appears green. This is due to absorption by the Fe^{2+} ions in the glass. A windowpane is only about half a metre across, so that we can see that in a fibre of a few kilometres in length, there would be considerable loss due to such absorption. In a glass, the spectrum of the impurity ions is similar to that in a crystal, but because the ions occupy several different types of site in a glass, the absorption bands are wider; each site giving rise to a band at a slightly different wavelength. In a 3 km long optical fibre operating at 1300 nm in the near infrared,

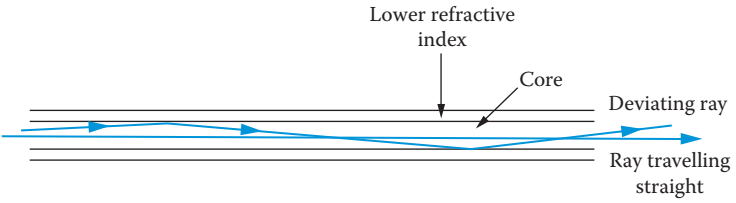


FIGURE 7.17 Rays of light travelling along an optical fibre.

the intensity of the Fe^{2+} absorption is still such that a concentration of two parts in 10^{10} would reduce the amount of radiation by half. The materials for optical fibres must therefore be very pure. One reason why silica has been widely used is because high purity silicon tetrachloride, developed for the semiconductor industry, is commercially available as a starting material to make impurity-free SiO_2.

Metal ions are not the only source of absorption, however. Using infrared radiation means that there is likely to be loss due to absorption by molecular vibrations. In silica glass, the structure may contain dangling SiO bonds that easily react with water to form –OH bonds. The vibrational frequencies of –OH bonds are high and close to the frequencies used for transmission. It is important, therefore, to exclude water when manufacturing silica optical fibres. Even when water is excluded and there are no –OH bonds, the absorption by vibrational modes cannot be neglected. Si–O bonds vibrate at lower frequencies than O–H bonds, so the maximum in the absorption does not interfere. However, the SiO absorption is very strong and the peak tails into the region of the transmission frequency. There has been some research into substances with lower vibrational frequencies than silica, particularly fluorides, but as yet such substances are not economically viable, being difficult to manufacture and more expensive than silica.

There are still losses in the fibres developed for commercial use; nonetheless, these have been reduced to a point where transmission over kilometres is possible. As well as the transmission of information in telephone systems and similar applications, it has been suggested that optical devices may replace conventional electronics in more advanced applications such as computers. For such applications, it will be necessary to develop optical switches, amplifiers and so on. In the next section, we look at new materials that have been developed, which may form the basis of integrated optical circuits. These materials are photonic crystals.

7.5 PHOTONIC CRYSTALS

Photonic crystals have been hailed as the optical equivalent of semiconductors. The idea of such crystals was first developed in the 1980s by Eli Yablonovitch at Bell Communications Research.

A photonic crystal consists of a periodic arrangement of two materials of different refractive indexes. At each boundary between the two materials, light, or other electromagnetic radiation, will refract and partly reflect. The beams from the different interfaces will reinforce or cancel each other out, depending on their relative phases. Whether two beams will be in phase is determined by the wavelength of the radiation, its direction of travel, the refractive index of the photonic crystal materials and the particular periodic arrangement. For certain wavelengths of radiation, refractive indices, and spacings of the materials, complete cancellation in all directions can occur so that such wavelengths are not transmitted by the crystal. The range of forbidden wavelengths is known by analogy with semiconductors as the **photonic band gap**. Figure 7.18 shows how a forbidden wavelength can occur for a one-dimensional arrangement—a row of slabs of dielectric material.

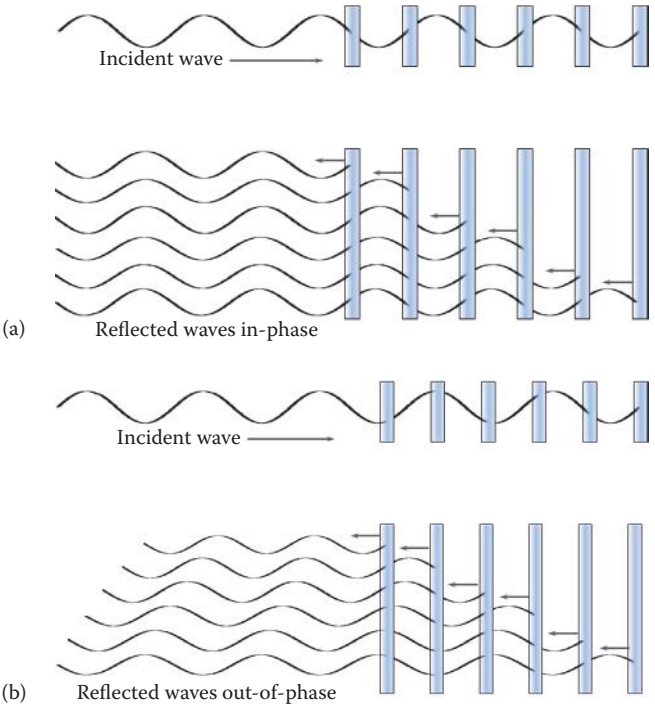


FIGURE 7.18 A row of dielectric slabs. Light travelling through can interfere destructively with the reflected light, giving rise to forbidden wavelengths. In (a) the reflected waves are all in phase with one another and out of phase with the incident wave and destructive interference occurs. In (b) the waves are reflected with slightly different phases and the incident light travels through.

As the light reaches each slab, some is reflected due to the change in the refractive index. For the correct spacing of the slabs, the reflected rays from each slab are in phase with each other but out of phase with the incident light. For such wavelengths, the incident and the reflected rays cancel each other (Figure 7.18a).

The original photonic crystal was produced by accurately drilling holes a millimetre in diameter in a block of material with a refractive index of 3.6. This crystal had a photonic band gap in the microwave region. Similar structures with band gaps in the microwave and radio regions are being used to make antennae that direct radiation away from the heads of mobile phone users. Producing photonic crystals with band gaps at shorter wavelengths—infrared and visible—is less straightforward. The hole size or lattice spacing needed for photonic crystals is roughly equal to the wavelength of the radiation divided by the refractive index. A GaAs laser produces radiation of wavelength 904 nm. To use this radiation with a photonic crystal composed of silica with a refractive index of 1.45 would require a spacing of 623 nm or 0.623 μm. Machining holes of this size is impractical, and the production of regular periodic structures with spacings of this order is a technical problem that needs to be overcome before integrated

optical circuits and other devices can be manufactured. One approach to this problem is to suspend spheres (usually of silica) with a diameter of less than 1 μm in a colloidal suspension or hydrogel. The spheres arrange themselves into a close-packed structure.

Two-dimensional photonic materials—materials in which light is blocked within a plane but transmitted perpendicular to the plane—are useful as optical fibres. An ingenious method of constructing such fibres is to pack a series of hollow capillary tubes around a central glass core. The structure is then heated and drawn until it is only a few micrometres thick. The central core is now surrounded by a periodic array of tubes of the right diameter to have a photonic band gap in the near infrared. It is also possible to replace the central glass core by air and this enables very high power laser signals to be transmitted along the fibre without damage to the fibre material.

Other potential applications of photonic crystals include crystals with rows of holes removed to guide radiation around sharp bends (something that cannot be attained with conventional optical fibres), nanoscopic lasers formed from thin films, ultrawhite pigment formed from a regular array of submicron titanium dioxide particles, radiofrequency reflectors for magnetic resonance imaging (MRI) and LEDs.

Photonic crystals have only been studied in the laboratory for a few decades, but there are naturally occurring examples, the best-known being the gemstone opal. Opals consist of tiny spheres of silica arranged in a face-centred cubic structure. These are thought to have formed from colloidal silica solutions and the colour depends on the size of the spheres.

7.6 METAMATERIALS: CLOAKS OF INVISIBILITY

Metamaterials like the photonic crystals just discussed, owe their properties of interest to their structure rather than their composition. They are composite materials, engineered to have properties beyond those of naturally occurring materials. They are often periodic, but you should note an important difference from the photonic materials described in the previous section. Metamaterials behave optically like a homogeneous material; they do not have a photonic band gap but they do have a refractive index that describes the bulk metamaterial. In photonic solids, the refractive index is characteristic of the materials from which the structure is made.

Great interest in these materials has centred around their unusual optical, electric and magnetic properties, in particular their negative refractive index. In 1968, Victor Veselago put forward the possibility of a negative refractive index, but no natural materials have been found to possess this property. In general, the refractive index (n) is given by $n = \pm (|\varepsilon_r||\mu_r|)^{1/2}$, where ε_r and μ_r are the relative permittivity and permeability of the material, respectively, at the wavelength of interest. For the majority of materials, ε_r and μ_r are positive and n takes the positive square root. Some materials, such as metals and ferroelectrics, have negative permittivity. Some ferrites have negative permeability. These two groups also have positive refractive indices. However, if both ε_r and μ_r are negative, then it can be shown that the negative square

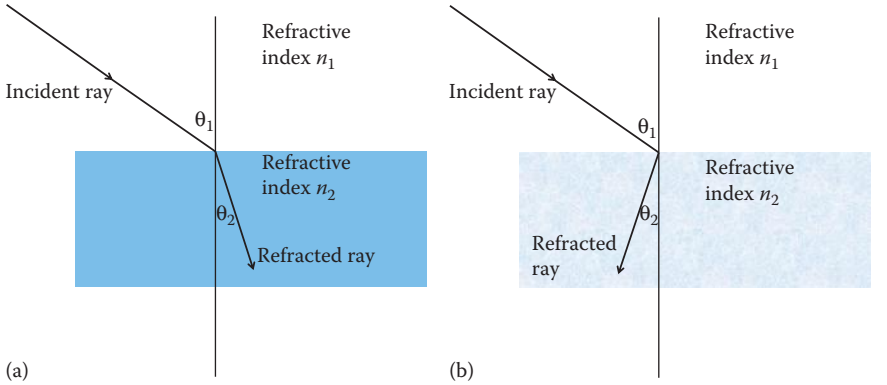


FIGURE 7.19 (a) Radiation passing from a medium with a refractive index n_1 to one with a refractive index n_2. (b) Radiation passing from a medium with a refractive index n_1 to one with a refractive index $-n_2$.

root has to be taken, meaning that the refractive index is negative. The first report of a negative refractive index material was not until 2000. This was a metamaterial with a negative refractive index in the microwave region.

The difference between materials with positive and negative refractive indices is illustrated in Figure 7.19. In Figure 7.19a, radiation passes from a medium, for example, air with a refractive index n_1 to a medium with a refractive index n_2, where, $n_2 > n_1$. The radiation approaches and is bent through an angle θ_2. In Figure 7.19b, the second medium has a refractive index $-n_2$ and is bent through the angle $-\theta_2$. That is, the ray in Figure 7.19b is bent through an angle of the same magnitude as in Figure 7.19a, but in the opposite direction.

The refractive index is related to the velocity of radiation in a medium. One of the consequences of a negative index of refraction is that the phase velocity of the radiation is negative, that is, the change of phase of the wave travels in the opposite direction to the direction of propagation (and of the direction of energy transfer). This leads to some unusual properties such that the Doppler effect is reversed with radiation travelling towards the observer being shifted to longer wavelengths (red shift). A useful consequence of the negative refractive index is that a lens made of such a material is not subject to the diffraction limit of ordinary lenses, so a higher resolution can be achieved. The limit of resolution of a conventional lens was first calculated by Abbé in 1873. Light from objects whose distance apart is less than half the wavelength of the light contains components that decay exponentially: **evanescent waves**. Lenses made of negative refractive index materials increase the amplitude of the evanescent waves. When they emerge from the lens, the waves decay again such that the amplitude at the image plane is equal to the original amplitude when the wave leaves the object. Thus, the evanescent wave component is not lost and all the light from the object is collected. Such lenses have thus been dubbed superlenses. In 2005, a superlens effect at optical frequencies was observed for a thin slab of silver.

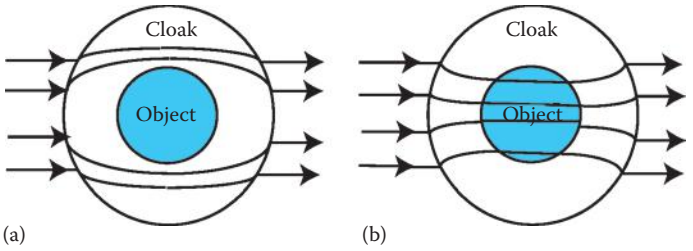


FIGURE 7.20 Two ways of cloaking. (a) Coordinate transformation and (b) cancellation of scattering.

Another proposed use for metamaterials is in MRI. Ricardo Marques and colleagues from the University of Seville, Spain, found that using a negative refractive index metamaterial as a magnetic flux guide enabled them to 'see' deeper into the body using MRI.

The application that made metamaterials famous, however, was their use in 'cloaks of invisibility'. The apparent invisibility rests on the presumption that electromagnetic radiation travels in straight lines. There are a number of methods of achieving a situation in which radiation is deviated in the vicinity of an object and led to continue parallel to the original beam on the other side of the object. The radiation then appears to have come straight through the object. We shall mention two methods, illustrated in Figure 7.20. In the first, coordinate transformation (Figure 7.20a), the radiation is diverted around the object, by varying the refractive index of the cloak. The skill in designing such a cloak lies in calculating the exact transformation needed at every point.

An alternative method is to select a cloak with a varying refractive index so that any scattering from the object is cancelled (Figure 7.20b). In this case, the radiation passes through the object but both the object and the cloak are invisible.

Both these methods depend on engineering a material with a refractive index varying in a precise way. Metamaterials have the advantage that they can be tuned to do this by varying the physical structure of a material. The first example of a cloak of invisibility, reported in 2006, used an array of metallic 'split ring' resonators, shown in Figure 7.21. This device shielded a copper cylinder at a range of microwave frequencies.

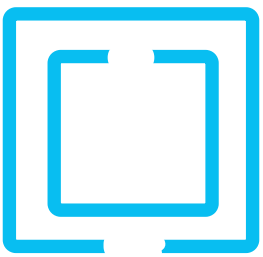


FIGURE 7.21 A split ring resonator.

A problem with producing metamaterial cloaks is that in order for the metamaterial to act as a single material, the structures it is composed of must be much smaller than the wavelength of the radiation used. Hence, producing a metamaterial that operates at visible wavelengths proved difficult. Recently (2009), cloaking with a silicon-based metamaterial has been demonstrated. A more recent surprise was the announcement of a cloak of invisibility operating at visible wavelengths that did not involve a metamaterial but, as we saw in Section 7.4.1, crystals of calcite.

QUESTIONS

1. In the oxide MnO, the Mn^{2+} ions occupy octahedral holes in an oxide lattice. The degeneracy of the 3d levels of the manganese ion are split into two, as for Ti^{3+}. The five d electrons of the Mn^{2+} ions occupy separate d orbitals and have parallel spins. Explain why the absorption lines due to transitions between the two 3d levels are very weak for Mn^{2+}.
2. Figure 7.22 shows the energy levels of Nd^{3+} in yttrium aluminium garnet ($Y_3Al_5O_{12}$), which are involved in the laser action of this crystal. Describe the processes that occur when the laser is working.
3. A phosphor commonly used on television screens is ZnS doped with Cu^{+}. This is much more efficient at transferring energy to the impurity sites for emission, than are the phosphors based on phosphates as host. ZnS is a semiconductor. Suggest a reason for the efficiency of transfer in this solid.
4. Figure 7.23 shows two bands of a semiconductor. Is the band gap of this solid direct or indirect?
5. Explain why silicon is used for solar cells but not for LEDs.

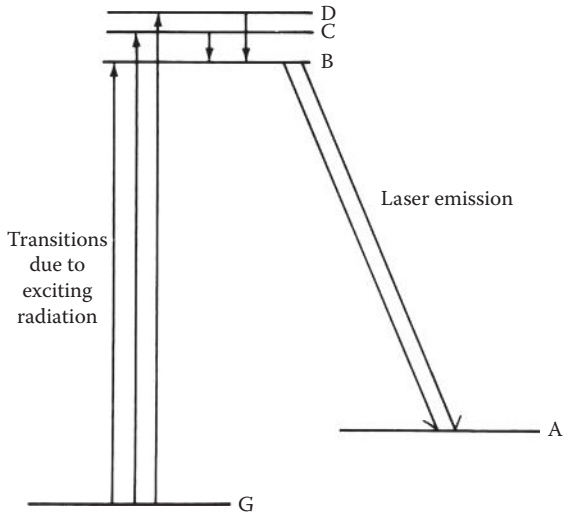


FIGURE 7.22 Energy levels of Nd^{3+} in yttrium aluminium garnet.

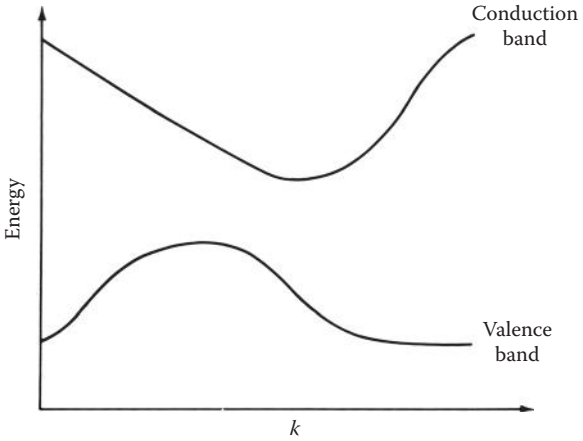


FIGURE 7.23 A plot of energy against k for two bands of a solid.

6. The core of silica optical fibres contains some B_2O_3, GeO_2 and P_2O_5. How do these impurities increase the refractive index?
7. Red-orange opals contain larger spheres than blue-green opals. How do the wavelengths of the photonic band gap vary with the colour?
8. Tetragonal calomel, whose empirical formula is HgCl, is birefringent, whereas alkali metal halides such as NaCl are not birefringent. The refractive index of the ordinary ray of calomel is higher than that of NaCl. From a consideration of the molecular and crystal structures of calomel and sodium chloride, explain these observations.

8 Magnetic and Electrical Properties

8.1 INTRODUCTION

One consequence of the closeness of atoms in a solid is that the properties of the individual atoms or molecules can interact cooperatively to produce effects not found in fluids. A well-known example of this is ferromagnetism. In a piece of iron used as a magnet, for example, the magnetism of the iron atoms aligns to produce a strong magnetic effect. Other cooperative magnetic effects lead to a cancelling (antiferromagnetism) or a partial cancelling (ferrimagnetism) of the magnetism of different atoms. Ferromagnets and ferrimagnets have many commercial applications, from compass needles and watch magnets, to hard-disk read heads and computer memory devices.

Cooperative effects are not confined to magnetism; similar effects can occur in the response of a crystal to mechanical stress and to electric fields. The electrical analogue of ferromagnetism is the ferroelectric effect, in which the material develops an overall electrical polarisation, that is, a separation of charge. Ferroelectric materials are important in the electronics industry as capacitors (for storing charge) and transducers (e.g., in the conversion of ultrasound to electrical energy). Ferroelectric crystals are a subclass of piezoelectric crystals, which have commercial uses of their own. For example, quartz watches use piezoelectric quartz crystals as oscillators.

This chapter looks at the types of materials that display cooperative magnetic and dielectric properties, and in the final section, we discuss multiferroics—materials that display more than one type of cooperative property, usually electric and magnetic.

To start, we consider the weaker magnetic effects that can be found in all types of matter. Then, we discuss the origin of the cooperative magnetic effects and their applications. Following the magnetic effects, we look at the dielectric effects, starting with the piezoelectric effect and its applications, and then we consider ferroelectric materials and, in particular, the oxide barium titanate ($BaTiO_3$), which is widely used as a capacitor. Finally, we consider multiferroics and their possible applications.

8.2 MAGNETIC SUSCEPTIBILITY

A magnetic field produces lines of force that penetrate the medium to which the field is applied. These lines of force show up, for example, when we scatter iron filings on a piece of paper covering a bar magnet. The density of these lines of force is known

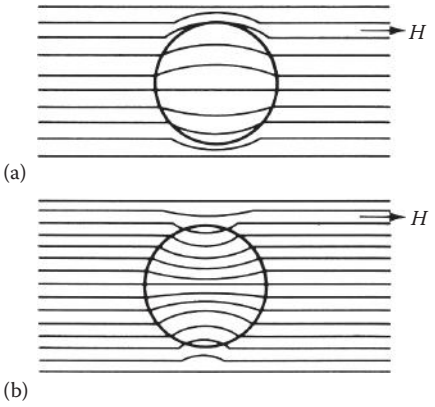


FIGURE 8.1 The flux density in (a) a diamagnetic and (b) a paramagnetic sample.

as the **magnetic flux density**. In a vacuum, the magnetic field and the magnetic flux density are related by the permeability of free space, μ_0

$$B = \mu_0 H \tag{8.1}$$

If a magnetic material is placed in the field, however, it can increase or decrease the flux density. Diamagnetic materials reduce the density of the lines of force, as shown in Figure 8.1, whereas paramagnetic materials increase the flux density.

The field of the sample in the applied field is known as its **magnetisation** (**M**). The magnetic flux density is now given by

$$B = \mu_0 (H + M) \tag{8.2}$$

The magnetisation is usually discussed in terms of the **magnetic susceptibility** (χ), where $\chi = M/H$, or the permeability (μ_r), where $\mu_r = 1 + \chi$.

Diamagnetism is present in all substances, but is very weak, so it is not normally observed if other effects are present. Diamagnetism is produced by the circulation of the electrons in an atom or a molecule. Atoms or molecules with closed shells of electrons are diamagnetic. Unpaired electrons, however, give rise to **paramagnetism**. Simple paramagnetic behaviour is found for substances such as liquid oxygen or transition metal complexes, in which the unpaired electrons on different centres are isolated from each other. In a magnetic field, the magnetic moments on different centres tend to align with the field and hence with each other, but this is opposed by the randomising effect of thermal energy, and in the absence of a field, the unpaired electrons on different centres are aligned randomly. The interplay of the applied field and thermal randomisation leads to the temperature dependence described by the **Curie law**:

$$\chi = \frac{C}{T} \tag{8.3}$$

where χ is the magnetic susceptibility, C is a constant known as the Curie constant and T is the temperature in kelvin. Different temperature dependence is observed when there is cooperative behaviour. The changeover from independent to cooperative behaviour is associated with a characteristic temperature. For **ferromagnetism**, the Curie law becomes

$$\chi = \frac{C}{(T - T_C)} \tag{8.4}$$

where T_C is the **Curie temperature** and for **antiferromagnetism**, the temperature dependence is of the form:

$$\chi = \frac{C}{(T + T_N)} \tag{8.5}$$

where T_N is the **Néel temperature**. These two behaviours are illustrated in Figure 8.2.

Ferrimagnetism has a more complicated form of temperature dependence, with ions on different sites having different characteristic temperatures.

The characteristics of the various types of magnetism that we shall be concerned with are given in Table 8.1.

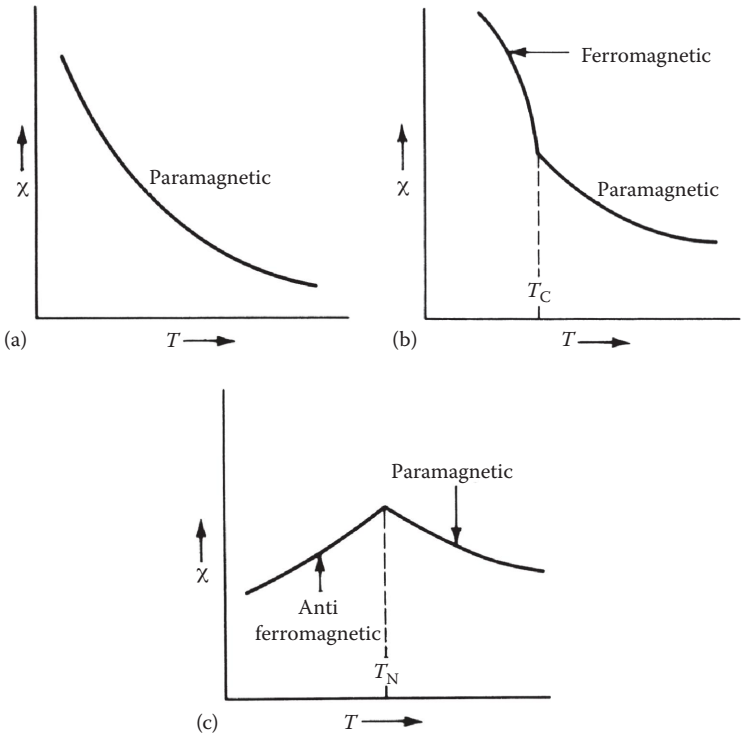


FIGURE 8.2 The variation of magnetic susceptibility with temperature for (a) a paramagnetic substance, (b) a ferromagnetic substance and (c) an antiferromagnetic substance.

TABLE 8.1
Characteristics of the Types of Magnetism

Type	Sign of χ	Typical χ Value (Calculated Using SI Units)	Dependence of χ on H	Change of χ with Increasing Temperature	Origin
Diamagnetism	–	$-(1–600) \times 10^{-5}$	Independent	None	Electron charge
Paramagnetism	+	0 – 0.1	Independent	Decreases	Spin and orbital motion of electrons on individual atoms
Ferromagnetism	+	$0.1–10^{-7}$	Dependent	Decreases	Cooperative interaction between magnetic moments of individual atoms
Antiferromagnetism	+	0–0.1	May be dependent	Increases	
Pauli paramagnetism	+	10^{-5}	Independent	None	Spin and orbital motion of delocalised electrons

8.3 PARAMAGNETISM IN METAL COMPLEXES

In solids containing metal complexes such that the unpaired electrons on the different metal atoms are effectively isolated, the susceptibility can be discussed in terms of magnetic moments. The isolated metal complex can be thought of as a small magnet. Each complex in a solid will produce its own magnetic field due to the unpaired electrons. If the solid consists of one type of complex, then each complex produces a magnetic field of the same magnitude. However, thermal motion causes the orientation of these fields to be random. In the Curie law (Equation 8.3), the temperature dependence is the result of this thermal motion, but the constant (C) gives us information on the value of the magnetic field, known as the **magnetic moment** (μ) of the complex. The magnetic moment, unlike the susceptibility, does not normally vary with temperature. The dimensionless quantity χ in Equation 8.2 is the susceptibility per unit volume. To obtain the size of the magnetic field due to an individual complex, χ is divided by the specific gravity to give the susceptibility per unit mass of the sample and then multiplied by the relative molecular mass to obtain the molar susceptibility χ_m. Assuming that each complex has a fixed magnetic moment (μ) and that the orientation is randomised by thermal motion, it can be shown that χ_m is proportional to μ^2, as in Equation 8.6.

$$\chi_m = \frac{N_A \mu_0}{3kT} \mu^2 \tag{8.6}$$

where N_A is Avogadro's number, k is Boltzman's constant, μ_0 is the permeability of the free space and T is the temperature in kelvin. In SI units, χ_m is in cubic metres per mole ($m^3\ mol^{-1}$) and the magnetic moment is in joules per tesla ($J\ T^{-1}$). It is usual to quote μ in Bohr magneton (BM or μ_B), where one **Bohr magneton** has a value of $9.274 \times 10^{-24}\ J\ T^{-1}$.

The magnetic moment (μ) is a consequence of the angular momentum of the unpaired electrons. The electrons possess both spin and orbital angular momenta. For the first-row transition elements, the contribution from the orbital angular momentum is greatly reduced or 'quenched' as a consequence of the lifting of the fivefold degeneracy of the 3d orbitals. In complexes of these elements, the magnetic moment is often close to that predicted for spin angular moment only:

$$\mu_s = g\sqrt{S(S+1)} \tag{8.7}$$

where g is a constant with a value of 2.00023 for a free electron and μ_s is in Bohr magnetons.

The value of S depends on the number of unpaired electrons. Table 8.2 gives the values of S and μ_s for the possible numbers of unpaired 3d electrons.

The contributions from the orbital angular momentum cause deviations from these values. For complexes containing heavier metal ions, the interaction of the spin and the orbital angular momenta is greater. For the lanthanoids, the magnetic moment depends on the total angular momentum of the electrons (**J**), not just or mainly on the spin angular momentum. The total angular momentum (**J**) is the vector sum of the orbital (**L**) and the spin (**S**) angular momenta.

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \tag{8.8}$$

If **J** is the quantum number for the total electronic angular momentum, then the magnetic moment is given by

TABLE 8.2
Values of S and μ_S for Unpaired 3d Electrons

No. Unpaired Electrons	Spin Quantum Number (S)	Magnetic Moment (μ_s) in Bohr Magneton
1	1/2	1.73
2	1	2.83
3	3/2	3.87
4	2	4.90
5	5/2	5.92

$$\mu = g\sqrt{J(J+1)} \tag{8.9}$$

where $g = 1 + \dfrac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}$

This can give rise to large magnetic moments, especially for shells that are more than half full. For example, Tb^{3+} with an f^8 configuration has a magnetic moment of 9.72 μ_B from Equation 8.9.

8.4 FERROMAGNETIC METALS

When discussing the electrical conductivity of metals, we describe them in terms of ionic cores and delocalised valence electrons. The core electrons contribute a diamagnetic term to the magnetic susceptibility, but the valence electrons can give rise to paramagnetism or one of the cooperative effects that we have described.

In filling the conduction band, we have implicitly put electrons into energy levels with paired spins. Even in the ground state of simple molecules, such as O_2, however, it can be more favourable to have the electrons in different orbitals with parallel spins than in the same orbital with paired spins. This occurs when there are degenerate or nearly degenerate levels. In an energy band, there are many degenerate levels and many levels very close in energy to the highest-occupied level. It might well be favourable to reduce electron repulsion by having the electrons with a parallel spin singly occupying the levels near the Fermi level. To obtain a measurable effect, however, the number of parallel spins would have to be comparable with the number of atoms; 10^3 unpaired spins would not be noticed in a sample of 10^{23} atoms. Unless the density of states is very high near the Fermi level, a large number of electrons would have to be promoted to high-energy levels in the band in order to achieve a measurable number of unpaired spins. The resulting promotion energy would be too great to be compensated for by the loss in electron repulsion. In the wide bands of the simple metals, the density of states is comparatively low, so that in the absence of a magnetic field, few electrons are promoted.

When a magnetic field is applied, the electrons acquire an extra energy term due to the interaction of their spins with the field. If the spin is parallel to the field, then its magnetic energy is negative; that is, the electrons are at lower energy than they were in the absence of a field. For an electron with a spin antiparallel to the field, it is now worthwhile to go to a higher-energy state and change the spin, as long as the promotion energy is not more than the gain in magnetic energy. This will produce a measurable imbalance of electron spins aligned with and against the field and hence the solid will exhibit paramagnetism. This type of paramagnetism is known as **Pauli paramagnetism**, and it is a very weak effect, giving a magnetic susceptibility much less than that due to isolated spins and comparable in magnitude to diamagnetism.

For very few metals, however, the unpaired electrons in the conduction band can lead to ferromagnetism. In the whole of the periodic table, only iron, cobalt, nickel and a few of the lanthanoids (Gd, Tb) possess this property. So, what is it about these elements that confers this uniqueness on them? It is not their crystal structure; they

each have different structures and their structures are similar to those of other non-ferromagnetic metals. Iron, cobalt and nickel, however, all have a nearly full, narrow 3d band.

The 3d orbitals are less diffuse than the 4s and 4p orbitals; that is, they are concentrated nearer the atomic nuclei. This leads to less overlap so that the 3d band is a lot narrower than the 4s/4p band. Furthermore, there are five 3d orbitals, so that for a crystal of N atoms, $5N$ levels must be accommodated. With more electrons and a narrower band, the average density of states must be much higher than in the ns/np band. In particular, the density of states near the Fermi level is high. In this case, it is energetically favourable to have substantial numbers of unpaired electrons at the cost of populating higher-energy levels. Thus, these elements have large numbers of unpaired electrons even in the absence of a magnetic field. For iron, for example, in a crystal of N atoms, there are up to $2.2N$ unpaired electrons, all with their spins aligned parallel. Note the contrast with a paramagnetic solid containing transition metal complex ions, where each ion may have as many as five unpaired electrons, but in the absence of a magnetic field, electrons on different ions are aligned randomly.

Ferromagnetism thus arises from the alignment of the electron spins throughout the solid, and this occurs for partially filled bands with a high density of states near the Fermi level. The 4d and 5d orbitals are more diffuse than the 3d orbitals and produce wider bands, so that ferromagnetism is not observed in the second-row and third-row transition elements. The 3d orbitals themselves become less diffuse across the transition series and lower in energy. In titanium, the valence electrons are in the 4s/4p band with a low density of states and, at the other end of the row in copper, the 3d band has dropped in energy so that the Fermi level is also in the 4s/4p band. Thus, it is only at the middle of the series that the Fermi level is in a region of high density of states. Schematic band diagrams for Ti, Ni and Cu are given in Figure 8.3. The occupied levels are indicated by shading.

The high density of states found in the 3d bands of Fe, Co and Ni also leads to a reduction of the mean free path of the electrons in this band. This causes a decrease in their mobility and hence in the electrical conductivity of these elements, compared with simple metals and copper, where the conduction electrons are in the s/p band.

The pure elements are not always suitable for applications requiring a metallic ferromagnet and many ferromagnetic alloys have been produced. Some of these contain one or more ferromagnetic elements and among these, alloys of iron, cobalt and nickel with the lanthanoids, for example, $SmCo_5$ and $Nd_2Fe_{14}B$, have produced some of the most powerful permanent magnets known. In the lanthanoid alloys, f electrons contribute to the magnetism. Potentially, this could lead to a very high magnetisation because there are seven f orbitals, and so a maximum possible magnetisation corresponding to seven electrons per atom. The theoretical maximum magnetisation for the transition metals is five electrons per atom, as there are only five d orbitals. In practice, this maximum is never reached. In the pure lanthanoid metals, the overlap of f orbitals is so small that they can be regarded as localised. In the ferromagnetic lanthanoids, the magnetism is produced by delocalised d electrons. The interaction between these d electrons and the localised f electrons causes the alignment of the d and f electrons in order to reduce the electron repulsion. Thus, the f electrons on

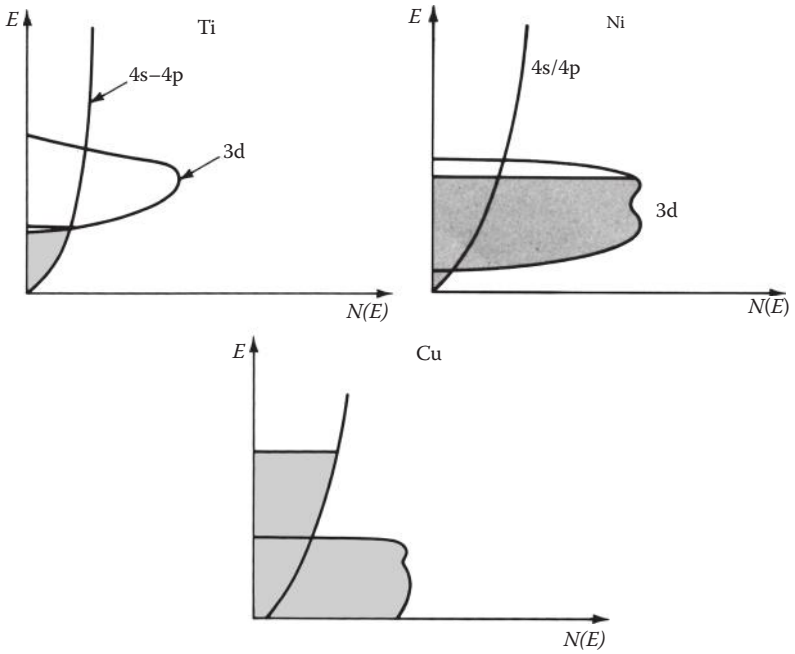


FIGURE 8.3 Schematic energy level diagrams for Ti, Ni and Cu. The shading represents the occupied energy levels.

different atoms are aligned through the intermediary of the d electrons. In alloys, the f electrons can align via the transition metal d electrons, and although not all of the d and f electrons are aligned, it can be seen that high values of the magnetisation could be achieved. It is not surprising, then, that it is these transition metal/lanthanoid alloys that are the most powerful magnets. Other alloys can be made from nonmagnetic elements, such as manganese, and in these, the overlap of the d orbitals is brought into the range necessary for ferromagnetism, by altering the interatomic distance from that in the element.

The usefulness of a particular ferromagnetic substance depends on factors such as the size of the magnetisation produced, how easily the solid can be magnetised and demagnetised and how readily it responds to an applied field. The number of unpaired electrons determines the maximum field, but the other factors depend on the structure of the solid and the impurities it contains, as discussed in the next section.

8.4.1 Ferromagnetic Domains

A drawback to the previous explanation may have occurred to us. If all $2.2N$ electrons are aligned in any sample of iron, why are all pieces of iron not magnetic? The reason for this is that our picture only holds for small volumes (typically $10^{-24} - 10^{-18}$ m^3) of metal within a crystal called **domains**. Within each domain, all the spins are aligned, but the different domains are aligned randomly with respect to each other.

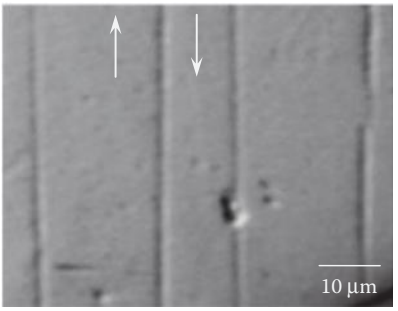


FIGURE 8.4 Domains in iron. (Reprinted with permission from Bathany, C., Le Romancer, M., Armstrong, J.N., and Choppra, H.D., *Physical Review B*, **82**, 184411, 2010. Copyright 2010 by the American Physical Society. http://link.aps.org/abstract/PRB/v82/e184411.)

It is actually possible to see these domains through a microscope on the polished surface of a crystal (Figure 8.4). What then causes these domains to form?

The spins tend to align parallel because of short-range **exchange interactions** stemming from electron–electron repulsion, but there is also a longer-range **magnetic dipole interaction** that tends to align the spins antiparallel. If you consider building up a domain starting with just a few spins, initially the exchange interactions dominate and, so the spins all lie parallel. As more spins are added, an individual spin will be subjected to an increasing magnetic dipole interaction. Eventually, the magnetic dipole interaction overcomes the exchange interaction and the adjacent piece of crystal has its spins aligned antiparallel to the original domain. Thus, within domains, exchange forces keep the spins parallel, whereas the magnetic dipole interaction keeps the spins of different domains aligned in different directions.

When a magnetic field is applied to a ferromagnetic sample, all the domains tend to line up with the field. This alignment can be accomplished in two ways. First, a domain of correct alignment can grow at the expense of a neighbouring domain. Between the two domains is an area of finite thickness, known as the domain wall. The changeover from the alignment of one domain to that of the next is gradual within the wall. When the magnetic field is applied, the spins in the wall nearest the aligned domain alter their spins to line up with the bulk of the domain. This causes the next spins to alter their alignment. The net effect is to move the wall of the domain further out, as shown in Figure 8.5. This process is reversible; the spins return to their former state after the magnetic field is removed.

If impurities or defects are present, it becomes harder for a domain to grow; there is an activation energy to aligning the spins through the defect, therefore a larger magnetic field is required. Once the domain has grown past the defect, however, it cannot shrink back once the magnetic field is removed because this will also need an energy input. In this case, the solid retains its magnetisation. The amount retained depends on the number and the types of defects. Thus, steel (which is iron with a high impurity content) remains magnetic after the field is removed, whereas soft iron, which is much purer, retains hardly any magnetisation.

The second mechanism of alignment, which only occurs in strong magnetic fields, is when the interaction of the spins with the applied field becomes large enough to

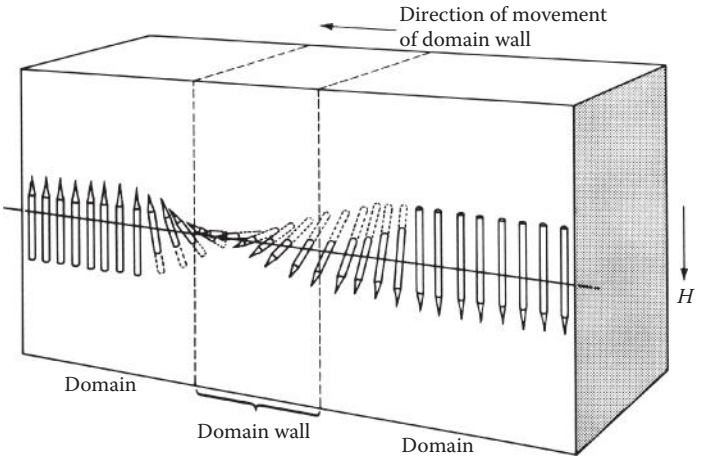


FIGURE 8.5 The movement of a domain wall. The broken lines show the domain wall, which separates two domains with the magnetic moments lined up in opposite directions. The moments twist to align with the applied field (H), and the wall moves in the direction of the arrow shown.

overcome the dipole interaction and entire domains of spins change their alignment simultaneously. The two mechanisms are compared in Figure 8.6.

The magnetic behaviour of different ferromagnetic substances is shown by their **hysteresis curves**. This is a plot of the magnetic flux density (B) against applied magnetic field (H). If we start with a nonmagnetic sample in which all the domains are randomly aligned, then in the absence of a magnetic field, B and H are zero. As the field is increased, the flux density also increases. The plot of B against H is shown in Figure 8.7. Initially, the curve is like *'oa'*, which is not simply a straight line because the magnetisation is increasing with the field. At point *'a'*, the magnetisation has reached its maximum value—all the spins in the sample are aligned. When the applied field is reduced, the flux density does not follow the initial curve. This is due to the difficulty of reversing processes where domains have grown through crystal imperfections. A sufficiently large field in the reverse direction to provide the activation energy for realignment through the imperfection must be applied before the magnetisation process can be reversed. At the point *'b'*, therefore, where H is zero, B is not zero because there is still a contribution from M. The magnetisation at this point is known as the **remanent magnetisation**. The field that needs to be applied

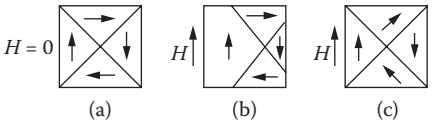


FIGURE 8.6 Magnetisation processes according to the domain model: (a) unmagnetised, (b) magnetised by domain growth and (c) magnetised by domain rotation (spin alignment).

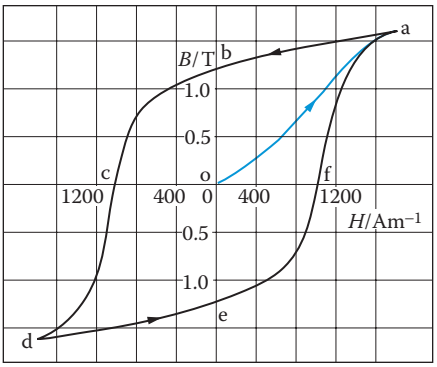


FIGURE 8.7 A *B*–*H* curve for a typical hard steel.

in the reverse direction to reduce the magnetisation to zero is the **coercive force** and is equal to the distance *'oc'*. If the coercive force is large, then the material is said to have high **coercivity**.

8.4.2 Permanent Magnets

Substances used as permanent magnets need a large coercive force, so that they are not easily demagnetised, and preferably should have a large remanent magnetisation. These substances have fat hysteresis curves. They are often made from alloys of iron, cobalt or nickel that form with small crystals and include nonmagnetic areas so that domain growth and shrinkage are difficult. The magnets for electronic watches, for example, are made from samarium/cobalt alloys. The best known of these alloys is $SmCo_5$, which has a coercive force of 6×10^5 A m^{-1} compared with 50 A m^{-1} for pure iron.

8.5 FERROMAGNETIC COMPOUNDS: CHROMIUM DIOXIDE

Chromium dioxide crystallises with a rutile structure (see Chapter 1) and is ferromagnetic with a Curie temperature of 392 K. CrO_2 has metal 3d orbitals that can overlap to form a band. In chromium dioxide, however, this band is very narrow, and, like Fe, Co and Ni, chromium dioxide displays ferromagnetism. The dioxides later in the row have localised 3d electrons (e.g., MnO_2) and are insulators or semiconductors. TiO_2 has no 3d electrons and is an insulator. VO_2 has a different structure at room temperature and is a semiconductor. However, it does undergo a phase transition to a metal at 340 K, when it becomes Pauli paramagnetic. Therefore, chromium dioxide occupies a unique position among the dioxides, similar to that of iron, cobalt and nickel among the first-row transition metals, in which the dioxides of the elements to the left have wide bands of delocalised electrons and the elements to the right have dioxides with localised 3d electrons. Because the metal atoms are further apart in the dioxides than in the elemental metals, the narrow bands that give rise to ferromagnetism occur earlier in the row than for the metallic elements.

8.6 ANTIFERROMAGNETISM: TRANSITION METAL MONOXIDES

These oxides have already been discussed in Chapter 5, and we may remember that they all had the sodium chloride structure, but had varying electrical properties. In this section, we shall see that their magnetic properties are equally varied. In TiO and VO, the 3d orbitals are diffuse and form delocalised bands. These oxides are metallic conductors. The delocalised nature of the 3d electrons also determines the magnetic nature of these salts and, like the simple metals, they are Pauli paramagnetic. MnO, FeO, CoO and NiO have localised 3d electrons and are paramagnetic at high temperatures. On cooling, however, the oxides become antiferromagnetic. The Néel temperatures for this transition are 122, 198, 293 and 523 K, respectively.

In antiferromagnetism, the spins on the different nuclei interact cooperatively but in such a way as to cancel out the magnetic moments. Antiferromagnetic materials therefore show a drop in magnetic susceptibility at the onset of cooperative behaviour, as we saw in Figure 8.2. The temperature that characterises this process is known as the Néel temperature (T_N).

The appearance of cooperative behaviour suggests that the d electrons on different ions interact, but the electronic properties were explained by assuming that the d electrons are localised. So how do we reconcile these two sets of properties?

The magnetic interaction in these compounds is thought to arise indirectly through the oxide ions; a mechanism known as **superexchange**. In a crystal of, say, NiO, there is a linear Ni–O–Ni arrangement. The d_{z^2} orbital on the nickel can overlap with the $2p_z$ on oxygen, leading to partial covalency. The incipient NiO bond will have the d_{z^2} electron and a $2p_z$ electron paired. The oxide ion has a closed shell, so there is another $2p_z$ electron, which must have the opposite spin. This electron forms a partial bond with the next nickel, so the d_{z^2} on this nickel pairs with the 2p electron of opposite spin. As shown in Figure 8.8, the net result is that adjacent nickel ions have opposed spins.

The alternating spin magnetic moments in antiferromagnets such as NiO can be observed experimentally using neutron diffraction. Because neutrons have a magnetic moment, a neutron beam used for diffraction responds not only to the nuclear positions but also to the magnetic moments of the atoms. X-rays, on the other hand, have no magnetic moment and respond only to the electron density and hence to the atomic positions. The structure of NiO as determined by X-ray diffraction is a simple

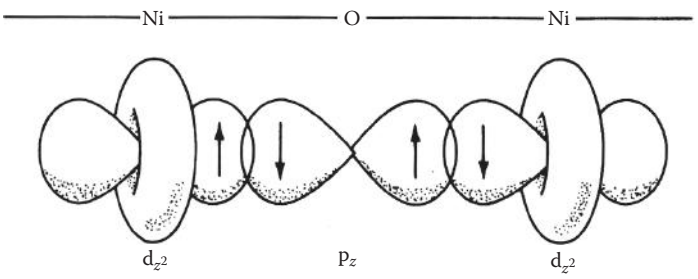


FIGURE 8.8 The overlap between the Ni d_{z^2} orbitals and the O p_z orbitals in NiO.

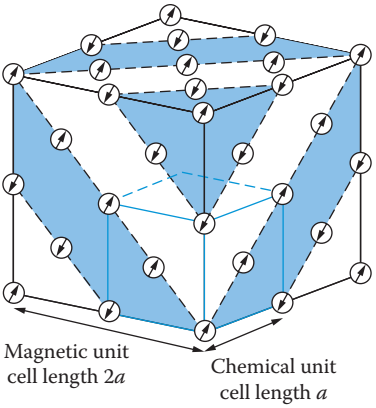


FIGURE 8.9 The magnetic unit cell of NiO with the crystallographic or chemical unit cell indicated in colour.

NaCl structure. When the structure is determined by neutron diffraction, however, extra peaks appear that can be interpreted in terms of a magnetic unit cell, which is twice the size of the unit cell determined by X-ray. The positions of the nickel ions in this cell are shown in Figure 8.9. The normal crystallographic unit cell is bounded by identical atoms. The magnetic unit cell is bounded by identical atoms with an identical spin alignment. The shading indicates layers of nickel ions parallel to the body diagonal of the cube. The spins of all the nickel ions in a given layer are aligned parallel, but antiparallel to the next layer.

8.7 FERRIMAGNETISM: FERRITES

The name ferrite was originally given to a class of mixed oxides having an inverse spinel structure (see Chapter 1) and the formula MFe_2O_4, where M is a divalent metal ion. The term has been extended to include other oxides, not necessarily containing iron, which have similar magnetic properties.

The spinel (AB_2O_4) structure is a common mixed oxide structure, typified by the spinel itself ($MgAl_2O_4$), in which the oxide ions form a face-centred cubic close-packed array. For an array of N oxide ions, there are N octahedral holes and $2N$ tetrahedral holes; the divalent A ions (Mg^{2+}) occupy one-eighth of the $2N$ tetrahedral sites and the trivalent B ions (Al^{3+}) occupy half of the octahedral sites, $A^{II}{}_{tet}(B^{III}{}_{oct}B^{III}{}_{oct})O_4$.

Figure 8.10 breaks this complex structure down into eight octants of two kinds, A and B, shown on the left of the diagram. In the inverse spinel structure, the oxide ions have the same cubic close-packed arrangement, but the divalent metal ions now occupy octahedral sites, and the trivalent ions are equally divided among tetrahedral and octahedral sites, $A^{III}{}_{tet}(B^{II}{}_{oct}B^{III}{}_{oct})O_4$.

Using Figure 8.10 to describe the inverse spinel ferrite structure (MFe_2O_4): half of the Fe^{3+} ions occupy tetrahedral positions in the A-type octants, together with the corners and face-centres of the unit cell, and the other half of the Fe^{3+} ions, together with the M^{2+} ions, occupy the octahedral sites in the B octants.

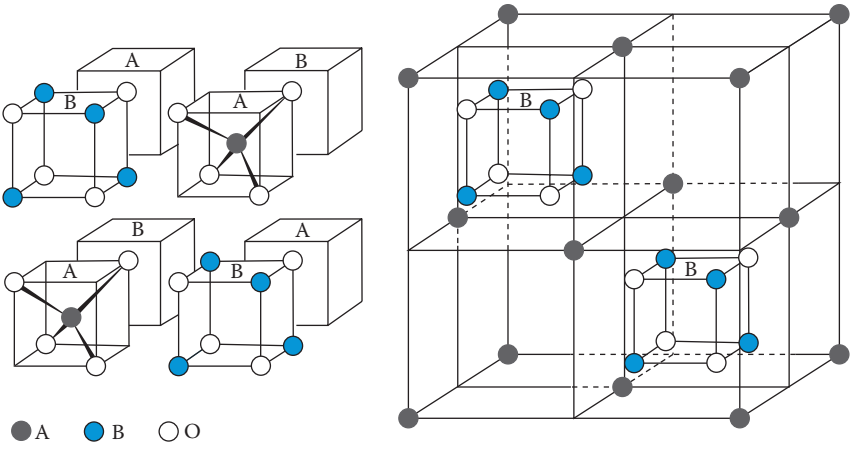


FIGURE 8.10 The spinel structure AB_2O_4.

The ions on octahedral sites interact directly with each other and their spins align parallel. The ions on octahedral sites also interact with those on tetrahedral sites, but in this case, they interact through the oxide ions and the spins align antiparallel, as in NiO.

In ferrites (MFe_2O_4), the Fe^{3+} ions on tetrahedral sites are therefore aligned antiparallel to those Fe^{3+} ions on octahedral sites, so that there is no net magnetisation from these ions. The divalent M ions, however, if they have unpaired electrons, tend to align their spins parallel to those of Fe^{3+} on the adjacent octahedral sites and hence with those of other M^{2+} ions. This produces a net ferromagnetic interaction for ferrites in which M^{2+} has unpaired electrons. The magnetic structure of such a ferrimagnetic ferrite is shown in Figure 8.11.

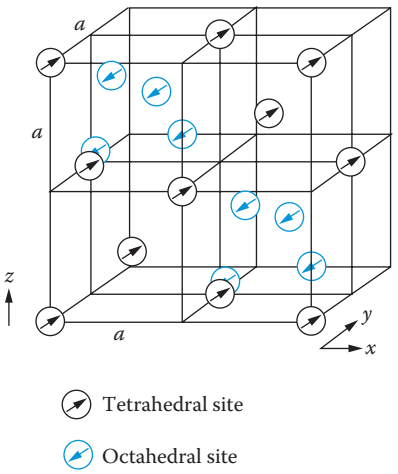


FIGURE 8.11 The magnetic structure of a ferrimagnetic inverse spinel.

In magnetite (Fe_3O_4), the divalent ions are also iron (Fe^{2+}), and the interaction between the ions on the adjacent octahedral sites is particularly strong. One way of looking at the electronic structure of this oxide is to consider it as an array of O^{2-} ions and Fe^{3+} ions with the electrons that would have made half of the Fe ions divalent, delocalised over all the ions on octahedral sites. The Fe^{3+} ions have five 3d electrons, all with parallel spins. Since there can only be five 3d electrons of one spin on any atom, the delocalised spin must have the opposite spin. Being delocalised, it must also have the opposite spin to the 3d electrons on the next Fe ion. Hence, the two ions must have their spins aligned, and these spins must be aligned with those of all the other Fe ions on octahedral sites. Delocalisation will be less for other ferrites. The delocalisation of the electrons means that Fe_3O_4 is a good conductor.

Magnetite is the ancient lodestone used as an early compass. Ferrites have also found use as memory devices in computers, as magnetic particles on recording tapes and as transformer cores.

8.7.1 Magnetic Strips on Swipe Cards

Magnetic strips are the brown or black strips found on many plastic cards, such as debit and credit cards. They were first used on the London Underground in the 1960s and were introduced into credit cards in the 1970s. They consist of small crystals of a magnetic material in a resin. A material commonly used for this is barium ferrite ($BaFe_2O_4$). Data such as your name and account number are encoded in the strip by altering the direction of the magnetisation of some crystals. Each character is represented by a set of crystals, some magnetised in one direction to represent 1 and others magnetised in the opposite direction to represent 0. When the card is swiped, the direction of the magnetisation is detected. High-coercivity materials, such as barium ferrite, are used as they retain their magnetisation so that the data are not accidentally erased.

8.8 SPIRAL MAGNETISM

When discussing ferromagnetism, ferrimagnetism and antiferromagnetism, we have assumed that the electron spin angular momentum can only point in two directions (spin up and spin down).

In many solids, however, the spins are at an angle to the direction of the net magnetisation, they are canted with each spin pointing in a direction slightly different from that of its neighbour. The spin direction changes by the same angle between each pair of spins to give a helical or a spiral effect. This tilting arises from the spin–orbit coupling interaction of the spin angular momentum with the orbital angular momentum of the ion and gives rise to **spiral magnetism**. There are various types of spiral arrangements (Figure 8.12). In the screw arrangement (a), successive spins point in different directions in a plane perpendicular to the wave vector. The cycloidal arrangement (b) is similar, except that the spin direction changes in a plane including the wave vector, rather like the change in direction of a particular spoke of a bicycle wheel as the wheel travels along the road. In the longitudinal and transverse conical

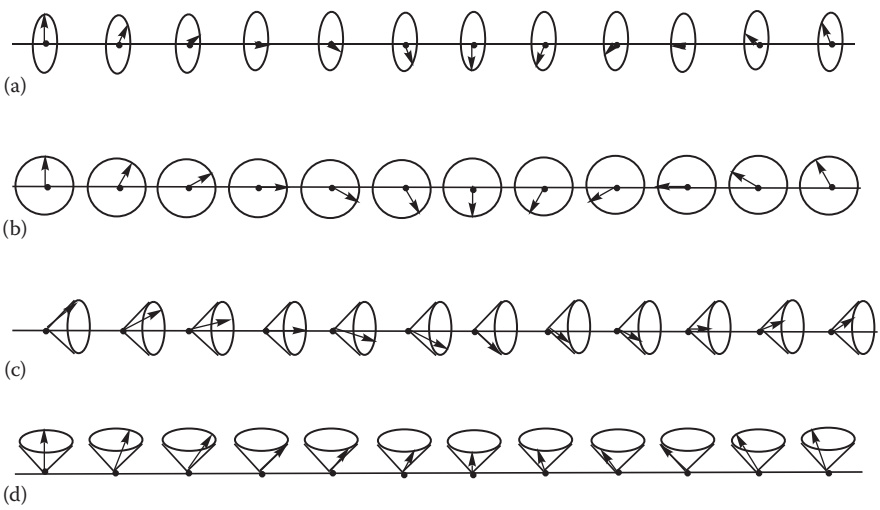


FIGURE 8.12 The spiral spin arrangements: (a) screw, (b) cycloidal, (c) longitudinal conical and (d) transverse conical.

(c and d) arrangements, the spins behave as though they were precessing from one ion to another. In longitudinal conical arrangements, the precession is around an axis coincident with the wave vector. In transverse conical arrangements, the precession is around an axis perpendicular to the wave vector.

Note that the net magnetisation cancels out for the screw and cycloidal arrangements, making solids with such arrangements effectively antiferromagnetic. For the conical arrangements, there is a net ferromagnetic effect. Recently, there has been much interest in spiral magnetism, as it can lead to the coupling of the magnetism with polarisation, producing materials that are both ferroelectric and ferromagnetic or antiferromagnetic. Such materials are examples of multiferroics and are discussed further in Section 8.12.

8.9 GIANT, TUNNELLING AND COLOSSAL MAGNETORESISTANCE

8.9.1 Giant Magnetoresistance

Magnetoresistance is the change in the electrical resistance of a material as a result of applying a magnetic field to the sample. In 1988, it was found that for certain metallic multilayer materials, the application of a magnetic field of strength 1–3 mT caused the resistivity to drop to half or less of its zero-field value. This phenomenon was termed **giant magnetoresistance (GMR)**. Within a decade, commercial devices such as computer hard-disk read heads based on GMR were available.

GMR is observed in metallic magnetic multilayers, which are stacks of nanometre-thick layers of different metals, in particular, alternate layers of a ferromagnetic metal and a nonmagnetic metal, for example, Fe/Cr/Fe… or Co/Cu/Co…. Within a single

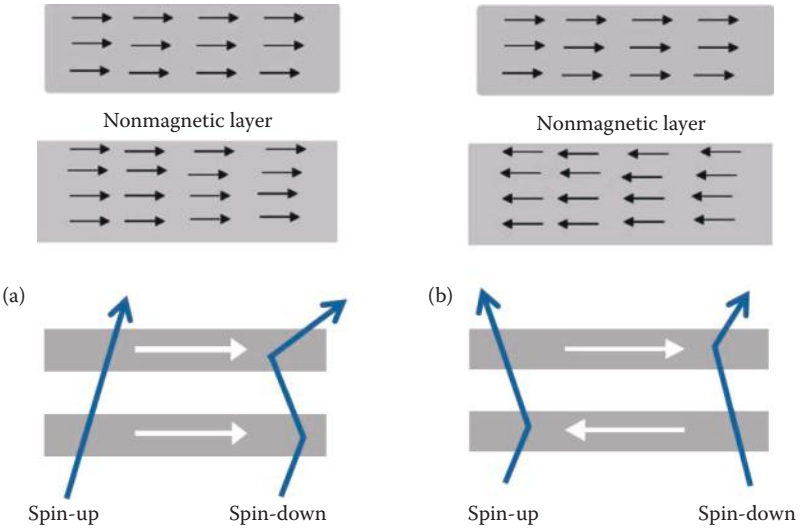


FIGURE 8.13 Two magnetic layers separated by a nonmagnetic layer showing the electron path, with (a) the spins in the two magnetic layers aligned ferromagnetically and (b) the spins in the two magnetic layers aligned antiferromagnetically.

ferromagnetic layer, all the spins are aligned, but the coupling between the adjacent ferromagnetic layers depends on the thickness of the intervening nonmagnetic layer. For certain thicknesses, the adjacent ferromagnetic layers are coupled antiferromagnetically (Figure 8.13b) and this increases the resistivity. However, when a magnetic field is applied, this aligns the layers in a ferromagnetic manner (Figure 8.13a) and there is a dramatic drop in electrical resistance.

GMR can be understood if we realise that in metals, spin-up and spin-down electrons conduct electricity independently and that electrical resistance occurs when the electrons are scattered by the material. The scattering is very different for spin-up and spin-down electrons. This means that the resistivities are different for the two spin states: the scattering is strong for spin-down electrons, which are antiparallel to the magnetisation direction, but is low for spin-up electrons, which are parallel to the magnetisation direction.

In a normal metal, there are the same number of spin-up and spin-down electrons, but in a ferromagnet, such as Fe, there are more electrons in which the spin is parallel to the direction of the magnetisation; these are the spin-up electrons (majority-spin electrons) and they are scattered only weakly. The minority, spin-down electrons are antiparallel to the direction of the magnetisation and they are scattered strongly. The resistance of a single layer of a ferromagnet thus depends on both resistances; we can think of the two conduction channels as resistances in parallel and the overall resistance is dominated by the low resistance of the majority-spin electrons.

Now look at the situation where we have two layers of ferromagnetic material separated by a nonmagnetic conducting layer. In (a), where there is an applied field and the layers are aligned parallel, the majority, spin-up, electrons pass through both layers with only weak scattering. The spin-down electrons are strongly scattered by both layers. The resistance will be dominated by the majority-spin electrons and is therefore low. In (b), the ferromagnetic layers are antiparallel. The spin-up electrons will be strongly scattered by the first antiparallel layer, but will then pass easily through the second layer. The minority, spin-down, electrons will pass easily through the first layer but will be strongly scattered by the second layer, so both categories of electrons will experience strong scattering in one or other of the layers. Hence, the overall resistance of this multilayer structure is high.

We now see why the application of a magnetic field, which switches the coupling from antiferromagnetic to ferromagnetic, leads to a large drop in resistance.

To understand why the electrons in one state conduct differently from the other state, we need to return to band theory. Consider the separate energy bands for spin-up and spin-down electrons. For d^8 metallic iron electrons, the majority-spin species, spin-up, the 3d bands are full, but for the spin-down electrons, the 3d band is only partially full. Because the spin-up 3d band is full, the spin-up current carriers occupy a higher energy s/p band, whereas the spin-down current carriers are in the 3d band. The mean free path of electrons (the average distance that electrons travel through the solid without being scattered) in the s/p band is greater than in the 3d band, so the spin-up electrons carry more current than the spin-down electrons.

8.9.2 Tunnelling Magnetoresistance

A related phenomenon, **tunnelling magnetoresistance (TMR)**, has superseded GMR as the basis of the action of current (2009) hard-disk read heads. TMR read heads can produce a larger signal and are easier to make into smaller devices.

TMR, like GMR, occurs for layered structures, but here the two ferromagnetic layers are separated by a very thin insulating layer—of a few nanometres—that allows electrons to flow from one ferromagnetic layer to the next by quantum mechanical tunnelling through the insulating oxide layer. When the magnetisations are parallel, it is more likely that the electrons can tunnel through than when they are antiparallel. A junction made of this material can thus be switched between high- and low-resistance states.

TMR was discovered in 1975 using a sandwich of Co/Ge/Fe, which showed a 14% difference in the resistance between the two states; however, subsequently, amorphous alumina (Al_2O_3) came to be the favoured material for the insulating layer; Al_2O_3 forms a very dense thin layer with few vacancies, so no gaps in the insulating layer, and has shown differences in resistance of up to 70%.

More recently, MgO has been used as the oxide layer. In the case of Fe/MgO/Fe, only wave functions of a particular symmetry can tunnel through efficiently. If the ferromagnetic material is chosen so that the bands at the Fermi level are of this symmetry for the majority carrier spins, but not for the minority carrier spins, then there is a sharp increase in resistance when the two ferromagnetic layers are aligned

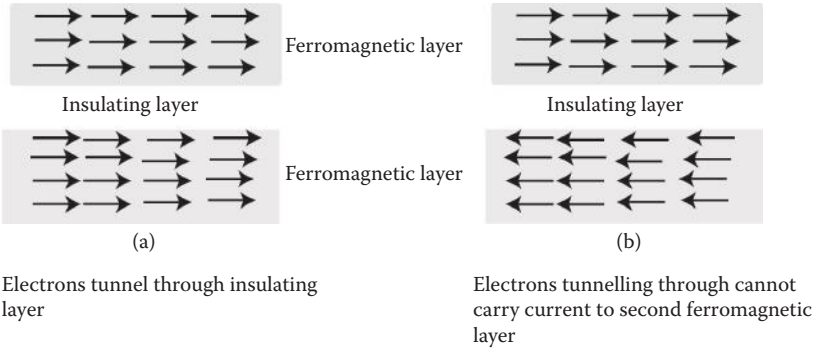


FIGURE 8.14 (a, b) Tunnelling magnetoresistance.

antiparallel (Figure 8.14). Theory has predicted that there could be a difference of several thousand percent between the two states, and at room temperature, ~600% has been observed and even higher at very low temperatures.

8.9.3 Hard-Disk Read Heads

In hard-disk read heads, one ferromagnetic layer has its spin orientation fixed by coupling to an antiferromagnetic layer (Figure 8.15). A second ferromagnetic layer separated from the first ferromagnetic layer by a nonmagnetic metal in the case of GMR, or an insulator in the case of TMR, is free to change its spin orientation when a field is applied.

As the read head moves over the hard disk, the magnetic fields on the disk cause the spins in the second layer to align either parallel or antiparallel to those in the first layer. The information on the hard disk is coded as a series of 0s and 1s corresponding to the different orientations of the magnetic field on the disk, and these give rise to a high or low current in the read head. A similar principle is used for magnetic random access memory chips (MRAM).

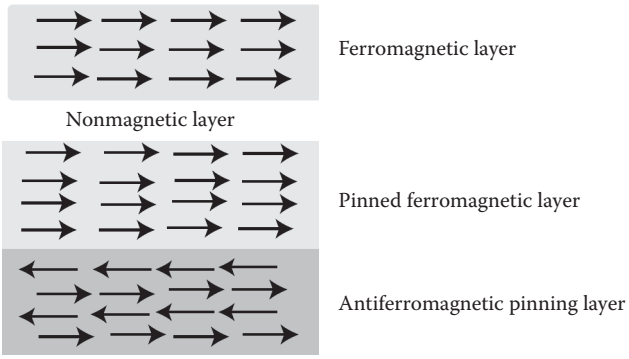


FIGURE 8.15 Two ferromagnetic layers as in a magnetic hard-disk read head showing pinning by an antiferromagnetic layer.

8.9.4 Colossal Magnetoresistance: Manganites

In 1993, **colossal magnetoresistance (CMR)** was observed for certain compounds such as doped manganite perovskites (e.g., $La_{1-x}Ca_xMnO_3$). In these compounds, a change in electrical resistance of orders of magnitude is observed, but large magnetic fields of the order of several tenths of a tesla (i.e., a hundred times stronger than those that produce GMR) or larger are needed.

Doped manganite perovskites exhibiting CMR have the general formula $Ln_{1-x}M_xMnO_3$, where Ln represents a lanthanoid element and M is a divalent metal such as Cu, Cr, Ba or Pb. The trivalent Ln ions and divalent M ions occupy the A sites in the perovskite structure (Figures 1.43 and 8.16) and they have 12-fold coordination to oxygen. The Mn ions occupy the octahedral B sites. Of the manganese ions, $(1–x)$ are Mn^{3+} and x are Mn^{4+}. The environment of the $Mn^{3+}(d^4)$ ions is distorted due to the Jahn–Teller effect and for small values of x, a cooperative distortion occurs.

As in other transition metal oxides such as NiO, the spins on the transition metal ions are coupled via the oxide ions. The Mn^{4+} ions couple antiferromagnetically with each other via superexchange as in NiO. The coupling of the Mn^{3+} ions varies and can be ferromagnetic or antiferromagnetic. The Mn^{3+} ions couple to the Mn^{4+} ions in a process known as **double exchange**. In this, a simultaneous hop of an electron from an Mn^{3+} ion to an O 2p orbital and from an O 2p orbital to an Mn^{4+} ion takes place.

$$Mn \uparrow\downarrow O \uparrow Mn \rightarrow Mn \uparrow O \downarrow\uparrow Mn$$

This produces ferromagnetic coupling.

Double exchange is strong in manganites with $x \approx \frac{1}{3}$, and these manganites form ferromagnetic phases at low temperatures. These phases can be described in terms of band theory. The e_g orbitals on manganese and the O 2p orbitals combine to form

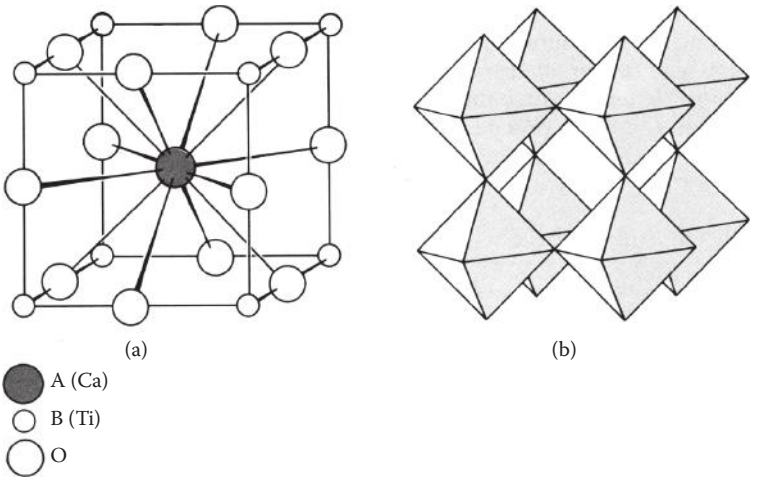


FIGURE 8.16 (a, b) The perovskite structure.

two bands, one for spin-up electrons and one for spin-down electrons. Unusually, there is a large band gap between the two bands, so that at low temperatures, one band is empty and the other is partly full. (Such solids are referred to as half metals because only one type of spin is free to carry an electrical current.) The ferromagnetism disappears at the Curie temperature. Above this temperature, the e_g electrons on manganese are better thought of as localised and the solid is paramagnetic and has a much higher electrical resistance. As the ferromagnetic phase approaches the Curie point, the electrical resistance rises as the thermal energy starts to overcome the double exchange. It is in this region that the manganites exhibit CMR. A strong magnetic field applied to the manganite realigns the spins, restoring the half-metallic state and thus decreasing the resistivity.

8.10 ELECTRICAL POLARISATION

Although solids consist of charged particles (nuclei and electrons), a solid has no overall charge. For most solids, there is also no net separation of positive and negative charges; that is, there is no net dipole moment. Even if a solid is composed of molecules with permanent dipole moments, for example ice, the molecules are generally arranged in such a way that the unit cell of the crystal has no net dipole moment, so the solid has none. If such a solid is placed in an electric field, then a field is induced in the solid which opposes the applied field. This field arises from two sources, a distortion of the electron cloud of the atoms or molecules and the slight movement of the atoms themselves. The average dipole moment per unit volume that is induced in the solid is the electrical polarisation (P), and is proportional to the field (E).

$$P = \varepsilon_0 \chi_e E \tag{8.10}$$

where ε_0 is the permittivity of free space ($= 8.85 \times 10^{-12}$ F m^{-1}) and χ_e is the (dimensionless) dielectric susceptibility. For most solids, the electric susceptibility lies between 0 and 10. The susceptibility is often determined experimentally by determining the capacitance of an electric circuit with and without the solid present. The ratio of these two capacitances is the relative permittivity or dielectric constant of the solid, ε_r.

$$\frac{C}{C_0} = \varepsilon_r \tag{8.11}$$

where C is the capacitance in farads in the presence of the solid, and C_0 is that in the absence of the solid. The dielectric constant (ε_r) is related to the dielectric susceptibility by

$$\varepsilon_r = 1 + \chi_e \tag{8.12}$$

If the experiment is performed using a high-frequency alternating electric field, then the atoms cannot follow the changes in the field and only the effect due to

electron displacement is measured. Electromagnetic radiation in the visible and the ultraviolet regions provides such a field, and the refractive index of a material is a measure of the electron contribution to the dielectric constant. Substances with high dielectric constants also tend to have high refractive indices, and therefore we will find that similar types of materials will be considered in the following sections as were discussed at the end of Chapter 7 in connection with optical devices.

Although most solids do not have a dipole moment in the absence of an electric field, the classes of solids that do have a dipole moment are commercially important and therefore form the subject matter of the next two sections.

8.11 PIEZOELECTRIC CRYSTALS: α-QUARTZ

A **piezoelectric crystal** is one that develops an electrical voltage when subjected to mechanical stress for example if pressure is applied to it, and conversely develops strain when an electric field is applied across it. The application of an electric field causes a slight movement of atoms in the crystal so that a dipole moment develops in the crystal. Most piezoelectric crystals must be made up from units that are non-centrosymmetric; that is, they do not possess a centre of symmetry. (There are a few crystals of high symmetry that cannot be piezoelectric because of other symmetry elements that they possess.)

α-quartz is based on SiO_4 tetrahedra. Tetrahedra do not have centres of symmetry, and in α-quartz the tetrahedra are distorted so that each unit has a net dipole moment. However, these tetrahedra are arranged in such a way (Figure 8.17) that normally the crystal does not have an overall polarisation.

External stress changes the Si–O–Si bond angles between the tetrahedra so that the dipole moments no longer cancel and the crystal has a net electrical polarisation. The effect in α-quartz is small; the output electrical energy is only 1% of the input strain energy, whereas for Rochelle salt (another commercially used piezoelectric crystal), the ratio of output energy to input energy is 81%. α-quartz, however, is useful in applications where an oscillator of stable frequency is needed, such as in quartz watches. An electric field causes distortion of quartz and if an alternating electric field is applied, the crystal vibrates. When the frequency of the electric field matches the natural vibration frequency of the crystal, then resonance occurs and a steady oscillation is set up with the vibrating crystal feeding energy back to the

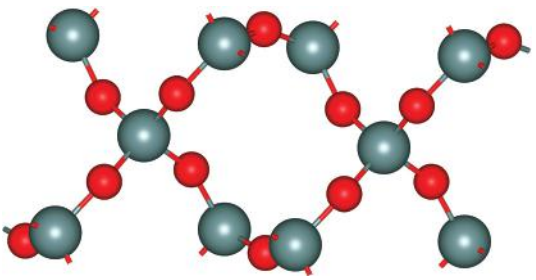


FIGURE 8.17 The structure of α-quartz.

electrical circuit. The importance of α-quartz in devices such as watches is due to the fact that for cuts along some crystal planes, the natural frequency of the crystal is independent of the temperature, so the crystal will oscillate at the same frequency, and the watch will keep time, however hot or cold the day is.

For other applications, such as ultrasonic imaging, it is more important that the conversion of mechanical energy to electrical energy is high.

Some piezoelectric crystals are electrically polarised in the absence of mechanical stress; one example is gem-quality tourmaline crystals. Normally, this effect is unnoticed because the crystal does not act as the source of an electric field. Although there should be a surface charge, this is rapidly neutralised by the charged particles from the environment and from the crystal itself. However, the polarisation decreases with increasing temperature, and this can be used to reveal the polar nature of the crystal. If tourmaline is heated, its polarisation decreases and it loses some of its surface charges. On rapid cooling, it has a net polarisation and will attract small electrically charged particles such as ash. Such crystals are known as **pyroelectric**, and ferroelectric crystals are a special subclass of pyroelectric crystals.

8.12 FERROELECTRIC EFFECT

Ferroelectric crystals possess domains of different orientation of electrical polarisation that can be reorientated and brought into alignment by an electric field. Among the most numerous ferroelectrics are the perovskites (Chapter 1), of which a classic example is barium titanate ($BaTiO_3$). This substance has a very large dielectric constant (around 1000) and is widely used in capacitors. Above 393 K, $BaTiO_3$ has a cubic structure, as in Figures 1.43 and 8.16, with Ba^{2+} ions in the centre, Ti^{4+} ions at the cube corners and an octahedron of O^{2-} ions around each titanium ion. The Ti^{4+} ion is small (75 pm radius), so there is room for it to move inside the O_6 cage. At 393 K, the structure changes to a tetragonal structure in which the Ti atom moves off-centre along a Ti–O bond. At 278 K, further change occurs in which the Ti atom moves off-centre along a diagonal between two TiO bonds, and at 183 K a rhombohedral phase is formed in which there is distortion along a cube diagonal. The three distortions are shown in Figure 8.18. This figure magnifies the effect: the Ti atom is moved about 15 pm off-centre.

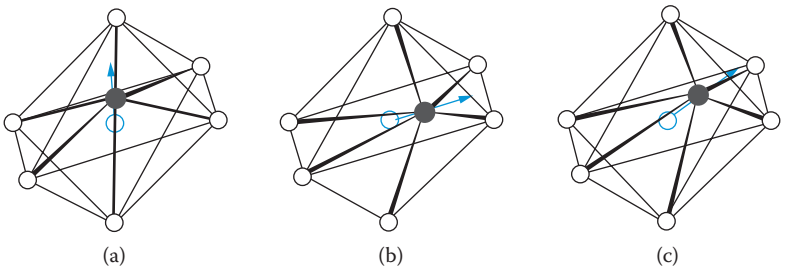


FIGURE 8.18 Distortions of TiO_6 octahedra in (a) the tetragonal structure, (b) orthorhombic structure and (c) rhombohedral structure of barium titanate.

In these three phases, the TiO_6 octahedra have a net dipole moment. To illustrate how ferroelectricity arises, we use the tetragonal structure (Figure 18.8a) as an example. If all the Ti atoms were slightly off-centre in the same direction, then the crystal would have a net polarisation. Like ferromagnets, however, ferroelectrics have domains within which there is a net polarisation, but different domains have their polarisation in different directions thus giving a net zero polarisation. In the tetragonal phase of $BaTiO_3$, the Ti atom can be off-centre in six directions along any one of the Ti–O bonds. As a result, neighbouring domains have polarisations that are either at 90° or at 180° to each other (Figure 8.19).

A domain can be of the order of 10^{-5} m or even more. Several methods can be used to obtain pictures of the domains. Figure 8.20 shows a thin slice of barium titanate under a polarising microscope in which different domains can clearly be seen. Note the sharpness of the domain boundaries.

When an external electric field is applied, the domains that are favourably aligned grow at the expense of others. As with ferromagnetics, the response to the field exhibits hysteresis; the polarisation grows until the whole crystal has its dipoles aligned, and this polarisation remains while the field is reduced to zero, only declining as a field of opposite polarity is applied. Figure 8.21 shows a hysteresis curve for barium titanate.

The dielectric susceptibility and dielectric constant of ferroelectric substances obey the Curie law dependence on temperature:

$$\varepsilon_r = \varepsilon_\infty + \frac{C}{(T - T_c)} \tag{8.13}$$

where ε_∞ is the permittivity at optical frequencies and T_C is the Curie temperature for ferroelectricity. The origin of the Curie temperature in barium titanate is quite easy

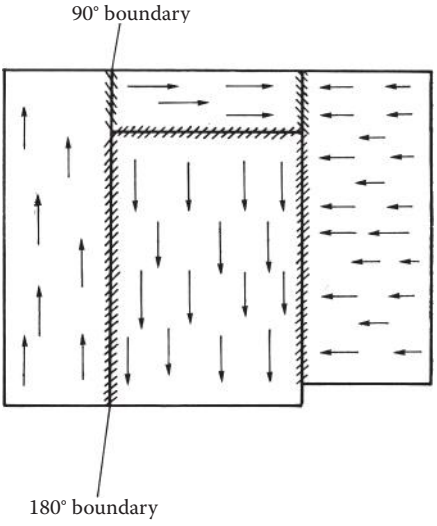


FIGURE 8.19 A sketch of the domains in barium titanate showing 90° and 180° boundaries.

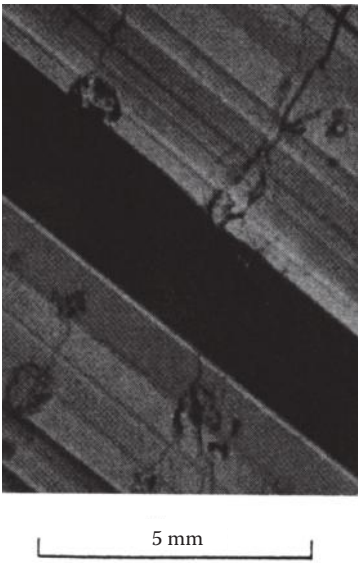


FIGURE 8.20 A photograph of a thin slice of barium titanate taken under a polarising microscope, showing domains of different polarisation. (From Guinier, A. and Julien, R., *The Solid State from Superconductors to Superalloys*, 3rd edn., Oxford University Press/ International Union of Crystallography, Oxford, Figure 2.9, p. 67, 1989. Reproduced by permission of Oxford University Press.)

to see because the Curie temperature is the temperature (393 K) at which barium titanate undergoes a phase transition to a cubic structure. Above the Curie temperature, the structure has a centre of symmetry and no net dipole moment. The dielectric constant is still high because the atoms can be moved off-centre by an applied electric field, but the polarisation is lost as soon as the field is removed.

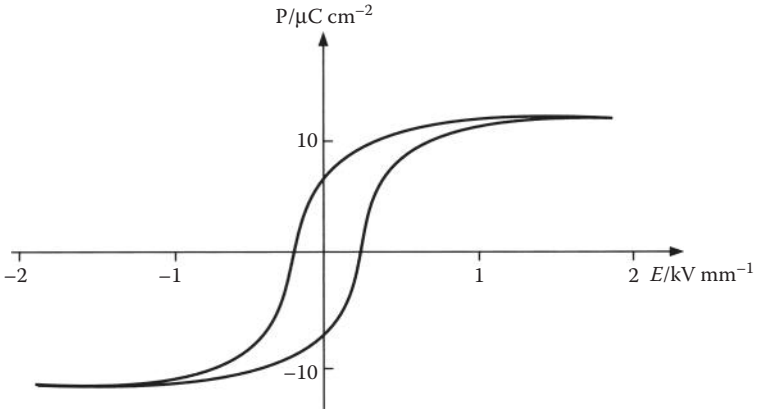


FIGURE 8.21 A plot of the polarisation versus the applied electrical field for barium titanate of grain size $2 \times 10^{-5} - 1 \times 10^{-4}$ m.

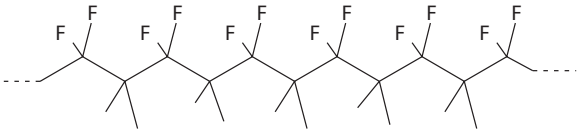


FIGURE 8.22 The alignment of $-CF_2$ groups in β-polyvinylidene fluoride.

In $PbZrO_3$, which also has a perovskite structure, the offset atoms are arranged alternately in opposite directions. This produces an antiferroelectric state. $PbZrO_3$ with some zirconium replaced by titanium gives the widely used ferroelectric material PZT ($PbZr_{1-x}Ti_xO_3$).

Remarkably, some polymers are ferroelectric. Polyvinylidene fluoride ($(-CH_2-CF_2-)_n$) as a thin film (approximately 25 μm thick) is used in shockwave experiments to measure stress. The pressure range over which this polymer operates is an order of magnitude larger than that of quartz or lithium niobate.

In the β-phase of polyvinylidene fluoride (PVDF), the $-CF_2$ groups in a polymer chain all point in the same direction (Figure 8.22), so that a dipole moment is produced. Within a domain, the $-CF_2$ groups on different chains are aligned. When an electric field is applied, the chains rotate through 60° increments to align with an adjacent domain.

8.12.1 Multilayer Ceramic Capacitors

Capacitors are used to store charge. An electric field is applied to induce charge in the capacitor. The capacitor then remains charged until a current is required. To be useful in modern electronic circuits, for computers, spacecraft, television sets and many other applications, a capacitor must be small. In order to retain a high capacitance, that is, to store a large amount of electrical energy whilst remaining small, a material needs a high permittivity. Hence, barium titanate with its very high permittivity has proved invaluable for this purpose. Pure barium titanate ($BaTiO_3$) has a high permittivity (of about 7000) close to the Curie temperature, but this rapidly drops with temperature to the room temperature value of 1–2000. While this is still high, for electronic circuits it would be useful to retain the higher value so that the size of the capacitor could be reduced. For some applications, it is also necessary for the permittivity to be constant with temperature over a range of 180 K, from 218 to 398 K. Barium titanate with some of the titanium substituted by zirconium or tin has a Curie temperature closer to room temperature and a flatter permittivity versus temperature curve. Further improvements can be made by partially substituting the barium ions. Materials made in this way consist of several phases mixed together, each with a different Curie temperature, and it is this that gives rise to the flatter permittivity versus temperature curve. Another factor that affects the dielectric properties of barium titanate is the grain size. On the surface of the tetragonal crystals, the structure is cubic, so that for small particles with a large surface-to-volume ratio, there is a high proportion of cubic material. This leads to a higher room temperature permittivity, but a smaller permittivity at the Curie temperature. For very small

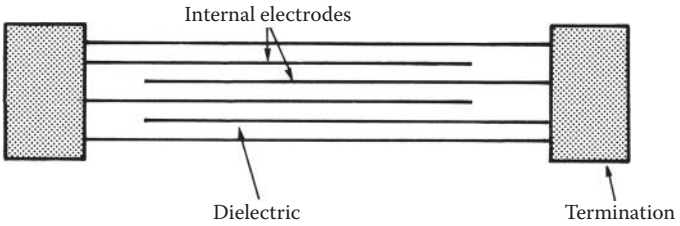


FIGURE 8.23 A section through a multilayer capacitor.

particles, there is no ferroelectric effect; therefore, it is important to produce grains of a suitable size to give the properties needed. To manufacture multilayer capacitors, barium titanate of a suitable grain size and appropriately doped, is interleaved with conducting plates (see Figure 8.23). This enables one device to be used in place of several single disc capacitors in parallel.

Barium titanate is one example of a ferroelectric material. Other oxides with the perovskite structure are also ferroelectric, for example, lead titanate and lithium niobate. One important set of such compounds, used in many transducer applications, are the mixed oxides PZT ($PbZr_{1-x}Ti_xO_3$). These, like barium titanate, have small ions in O_6 cages which are easily displaced. Other ferroelectric solids include hydrogen-bonded solids such as KH_2PO_4 and Rochelle salt ($NaKC_4H_4O_6.4H_2O$), salts with anions that possess dipole moments such as $NaNO_2$ and copolymers of PVDF. It has even been proposed that ferroelectric mechanisms are involved in some biological processes such as brain memory and voltage-dependent ion channels concerned with impulse conduction in the nerve and muscle cells.

8.13 MULTIFERROICS

Multiferroics are materials that possess two or more cooperative properties, such as ferromagnetism, ferroelectricity and ferroelasticity. In practice, it is generally used to refer to materials that are ferroelectric and ferromagnetic or antiferromagnetic. The interest in these materials stems from the possibility of applying a magnetic field to alter the charge or applying a voltage to change the spin. Possible applications include computer memory that can be laid down electrically but read magnetically, and electronic devices with 4-state logic. When writing to memory using ferroelectrics, charged plates are placed on either side of the ferroelectric material, making the atoms move up or down and giving a 0 or 1 as with magnetic memory. Writing using an electric field rather than a magnetic field has advantages, including using less energy. Reading back using electric fields results in the memory being wiped. However, if changing the polarisation changes the magnetic state, then the memory can be read by detecting the magnetism, a process that is nondestructive.

Multiferroics can be divided into two broad classes: **Type I multiferroics** possess both magnetic and electric cooperative properties, but the two arise from different features of the structure and are weakly coupled; **Type II multiferroics** have strongly coupled magnetic and electrical cooperative effects.

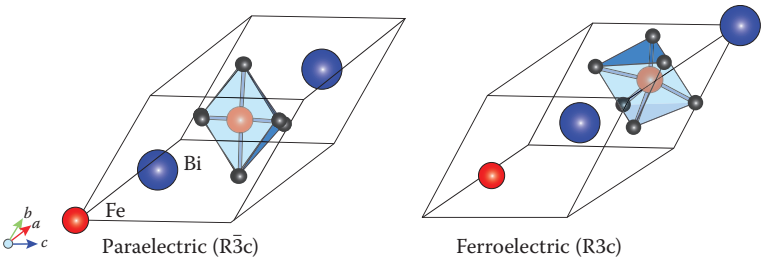


FIGURE 8.24 The structure of $BiFeO_3$. (From Stroppa, A. and Pirozzi, S., *PCCP*, **12**, 5407, 2010. Reproduced by permission of the PCCP Owner Societies.)

8.13.1 Type I Multiferroics: Bismuth Ferrite

One of the most studied examples of Type I multiferroics is $BiFeO_3$, bismuth ferrite. The interest in this compound was sparked by the publication of a paper in *Science* in 2003, reporting the results from Ramamoorthy Ramesh's group at the University of Maryland. The measurements showed that thin films of $BiFeO_3$ had a large remanent polarisation. Bismuth ferrite has a Néel temperature of 643 K and a ferroelectric transition temperature of 1100 K, so it is multiferroic at room temperature. It adopts a perovskite structure (Figure 8.24) with Bi–O and Fe–O layers. The spins on the Fe^{3+} ions are coupled antiferromagnetically. The ferroelectric effect is due to the alignment of the lone pairs of the Bi^{3+} ions, giving a cooperative offset arrangement.

However, because the polarisation and magnetisation are due to different ions, there is very little interaction between the two and the effect of a magnetic field on polarisation or of an electric field on magnetisation is very small.

Although the magnetoelectric coupling is small, bismuth ferrite does have potential applications. It has been shown that it can act as the insulating thin film in a tunnelling magneto-resistance sandwich when placed between ferromagnetic (La, Sr) MnO_3 and Co metal. A possibility is that it could be used to modify the TMR through the application of an electric field.

8.13.2 Type II Multiferroics: Terbium Manganite

The first Type II multiferroics to be studied were the terbium manganites, $TbMnO_3$ and $TbMn_2O_5$. $TbMnO_3$ adopts a perovskite structure. Polarisation is induced by spiral magnetism. For $TbMnO_3$, the spins on the manganese ions are in the cycloidal arrangement below 28 K. $TbMnO_3$ is an example of a lanthanoid manganite, which is the parent compound of colossal magneto-resistance materials. The Mn^{3+} ions are subjected to Jahn–Teller distortion in their surroundings. This encourages the ordering of the e_g orbitals in the *ab* plane and leads to spins being ferromagnetically coupled along the *a* and *b* axes and antiferromagnetically coupled along the *c* axis (Figure 8.25).

As the radius of the lanthanoid ion decreases, the distortion increases and the distance between Mn^{3+} ions becomes smaller. This enables ferromagnetic coupling of Mn^{3+} ions through space to compete with the antiferromagnetic Mn–O–Mn coupling.

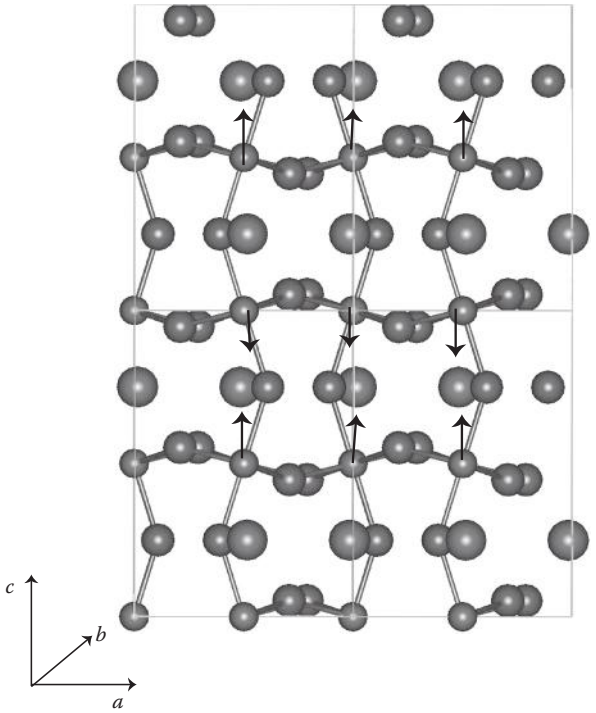


FIGURE 8.25 The structure of $TbMnO_3$ showing *a*, *b* and *c* axes.

The spins become frustrated and adopt a modulated spin ordering. Below 28 K, this modulation takes the form of a cycloidal arrangement of spins in the *bc* plane, and this, in turn, leads to the development of polarisation along the *c* axis. A magnetic field of about 5 T applied along the *b* axis was found to cause the electrical polarisation of $TbMnO_3$ to change direction from along the *c* axis to along the *a* axis.

There are a number of other solids in which cycloidal spin arrangements produce a magnetoelectric effect. Examples are CuO, $MnWO_4$ and $Ni_3V_2O_8$.

Solids with cycloidal spin arrangements are not the only solids to show Type II multiferroicity. It can be demonstrated that in a cubic crystal, the screw and longitudinal conical arrangements do not give rise to electrical polarisation, but the transverse conical arrangement does. The transverse conical arrangement is interesting because it gives rise to a solid that is both ferromagnetic and ferroelectric. An example of such a solid is $CoCr_2O_4$, which has a spinel structure. The inverse Dzyaloshinskii–Moriya or spin-current model predicts the value of the polarisation to be

$$P = a\sum_{ij} \mathrm{e}_{ij} \times \left(S_i \times S_j\right) \tag{8.14}$$

where e_{ij} is the unit vector connecting spins S_i and S_j and a is a constant dependent on the spin–orbit coupling constant, the spin exchange and the spin–lattice interaction.

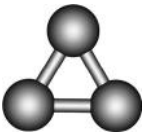


FIGURE 8.26 The structure of $MFeO_2$ showing a triangle of Fe ions.

Although the screw and longitudinal conical arrangements do not give rise to polarisation, it is possible for an applied magnetic field to alter the spin arrangement such that the solid becomes ferroelectric. One such example is the hexaferrite ($Ba_2Mg_2Fe_{12}O_{22}$; BMFO). This has a longitudinal conical spin arrangement below 50 K. The spins are arranged as though precessing around the *c* axis. By applying a magnetic field at an angle to the *c* axis, the cones are tilted so that a component perpendicular to the *c* axis is produced. The solid then becomes polarised.

A screw arrangement of spiral magnetism can lead to a magnetoelectric interaction if the symmetry of the crystal is lowered. The oxides AMO_2, where A = Cu, Ag and M = Cr, Fe have a structure in which each ion forms a triangular sublattice. The Cr and Fe ions prefer to couple antiferromagnetically to the adjacent ion of that type. Taking a triangle of Fe ions, as in Figure 8.26, there is a problem with this. If one ion has an opposite spin to both its neighbours, then the two neighbouring ions must have parallel spins, but there is a preference for them to have opposite spins. The spins are said to be frustrated because there is no way each spin can be an opposite spin to both its neighbours.

For $CuFeO_2$, the adoption of a screw arrangement of spins to avoid this frustration has been shown to lead to the appearance of ferroelectricity. The triangles of the Fe ions lie perpendicular to the *c* axis and polarisation develops parallel to this axis. Recently (2012) giant polarisation persisting up to 90 K has been reported for a solid with a screw arrangement of spiral magnetism, Ca Mn_7 O_{12}. The polarisation is induced parallel to the wave vector.

This chapter has been mainly concerned with cooperative properties with interesting magnetic and electric properties arising from all the atoms in a crystal or domain acting in concert. In the next chapter, we cover a property that arises from electrons cooperating throughout a crystal, that of superconductivity.

QUESTIONS

1. Although manganese is not ferromagnetic, certain alloys such as Cu_2MnAl are ferromagnetic. The Mn–Mn distance in these alloys is greater than in manganese metal. What effect would this have on the 3d band of manganese? Why would this cause the alloy to be ferromagnetic?
2. The compound EuO has an NaCl structure and is paramagnetic above 70 K, but magnetically ordered below it. Its neutron diffraction patterns at high and low temperatures are identical. What is the nature of the magnetic ordering?
3. $ZnFe_2O_4$ has an inverse spinel structure at low temperatures. What type of magnetism would you expect it to exhibit?

4. In transition metal pyrite disulfides (MS_2), the M^{2+} ions occupy octahedral sites. If a d band is formed, it will split into two as in the monoxides. Consider the information on some sulfides given below and decide whether the 3d electrons are localised or delocalised, which band the electrons are in if delocalised and, in the case of semiconductors, between which two bands the band gap of interest lies.

   MnS_2 antiferromagnetic ($T_N = 78$ K), insulator, above T_N paramagnetism fits five unpaired electrons per manganese
   FeS_2 diamagnetic, semiconductor
   CoS_2 ferromagnetic ($T_C = 115$ K), metal

5. In hydrogen-bonded ferroelectrics, the Curie temperature and permittivity alter when deuterium is substituted for hydrogen. What does this suggest about the origin of the ferroelectric transition in these compounds?
6. Pure $KTaO_3$ has a perovskite structure but is not ferroelectric or antiferroelectric. Replacing some K ions with Li, however, produces a ferroelectric material. Explain why the substitution of Li might have this effect.
7. $PbVO_3$ (in which Pb is present as Pb^{2+} ions) is a Type I multiferroic. What is the origin of the ferroelectricity of this solid? Why would you expect this to be decoupled from the magnetism?

9 Superconductivity

9.1 INTRODUCTION

Superconductors have two unique properties that have prompted commercial interest in them. First, they have zero electrical resistance, so they carry current with no energy loss: this could revolutionise power transmission in cities, for instance, and is already being exploited in the windings of superconducting magnets as used in NMR experiments and in the Large Hadron Collider. Second, they expel all magnetic flux from their interior, so they are forced out of a magnetic field. The superconductors can float or 'levitate' above a magnetic field. Until 1986, the highest temperature a superconductor operated at was 23 K, so they all had to be cooled by liquid helium (boiling temperature (b.t.) ~4 K). This, of course, made any use of superconductors extremely expensive.

In the late 1980s, the amount of research effort in the field of superconductors increased dramatically due to the discovery of the so-called 'high-temperature' superconductors by Bednorz and Muller in 1986. Their findings were thought to be so important that only a year later they were awarded the Nobel Prize in Physics by the Royal Swedish Academy of Sciences.

The discovery of a barium-doped lanthanum copper oxide that became superconducting at 35 K led to a flood of new high-temperature superconductors, some of which were superconducting above the boiling temperature of nitrogen, 77 K.

The past two decades have also seen the discovery of other unexpected types of superconductors. In 2001, a record temperature of 40 K for the onset of conventional superconductivity was observed in MgB_2. Although this temperature is not high enough for liquid nitrogen cooling, it is technologically important because mains-operated cryocoolers are now available that can reduce temperatures to 30 K, so that devices made from this material would not require liquid helium cooling. More recently, considerable interest has been raised by the announcement in 2008 of a high-temperature superconductor that was iron-based rather than copper-based.

Another unexpected discovery was of superconductors that are also ferromagnets, a combination of properties that is incompatible with the accepted theory of conventional superconductors.

We shall look at all these exciting developments; however, we start by looking at the properties of the low-temperature superconductors that were discovered nearly a century ago.

9.2 CONVENTIONAL SUPERCONDUCTORS

9.2.1 Discovery of Superconductors

In 1908, Kamerlingh Onnes succeeded in liquefying helium, and this paved the way for many new experiments to be performed on the behaviour of materials at low

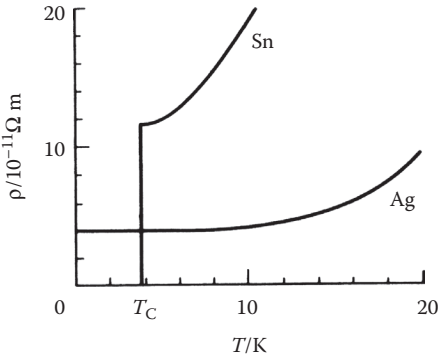


FIGURE 9.1 A plot of resistivity (ρ) versus temperature (T) showing the drop to zero at the critical temperature (T_C) for a superconductor and the finite resistance of a normal metal at absolute zero.

temperatures. For a long time, it had been known from conductivity experiments that the electrical resistance of a metal decreased with temperature. In 1911, Onnes was measuring the variation of the electrical resistance of mercury with temperature when he was amazed to find that at 4.2 K the resistance suddenly dropped to zero. He called this effect **superconductivity** and the temperature at which it occurs is known as the **(superconducting) critical temperature (T_C)**. This effect is illustrated for tin in Figure 9.1. One effect of the zero resistance is that there is no power loss in an electrical circuit made from a superconductor. Once an electrical current is established, it shows no discernible decay for as long as the experimenters have been able to watch.

For more than 20 years, little progress was made in the understanding of superconductors and only more substances exhibiting the effect were found. More than 20 metallic elements can be made superconducting under suitable conditions (see Figure 9.2) as can thousands of alloys. It was not until 1933 that Meissner observed a new effect.

9.2.2 Magnetic Properties of Superconductors

Meissner and Ochsenfeld found that when a superconducting material is cooled below its critical temperature (T_C), it expels all magnetic flux from within its interior (Figure 9.3a): the magnetic flux (B) is thus zero inside a superconductor. Since $B = \mu_0 H(1 + \chi)$, when $B = 0$, χ must equal -1, that is, superconductors are perfect diamagnets. If a magnetic field is applied to a superconductor, the magnetic flux is excluded (Figure 9.3b) and the superconductor repels a magnet. This is shown in Figure 9.4, where a magnet is seen floating in mid-air above a superconductor.

It is also found that the critical temperature (T_C) changes in the presence of a magnetic field. A typical plot of T_C against an increasing magnetic field is shown in Figure 9.5, where we can see that as the applied field increases, the critical temperature drops.

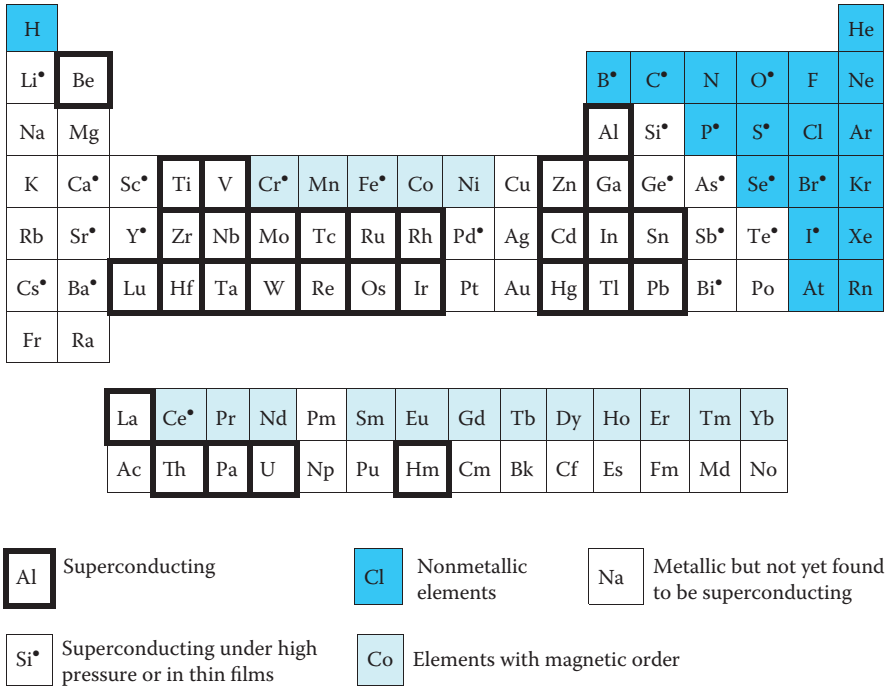


FIGURE 9.2 The superconducting elements.

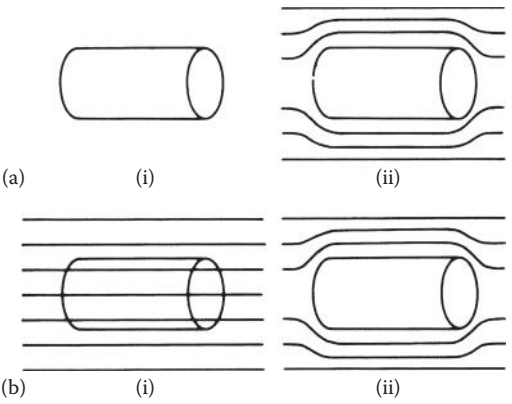


FIGURE 9.3 (a) (i) A superconductor with no magnetic field. When a field is applied in (ii), the magnetic flux is excluded. (b) (i) A superconducting substance above the critical temperature (T_C) in a magnetic field. When the temperature drops below the critical temperature (ii), the magnetic flux is expelled from the interior. Both are called Meissner effects.

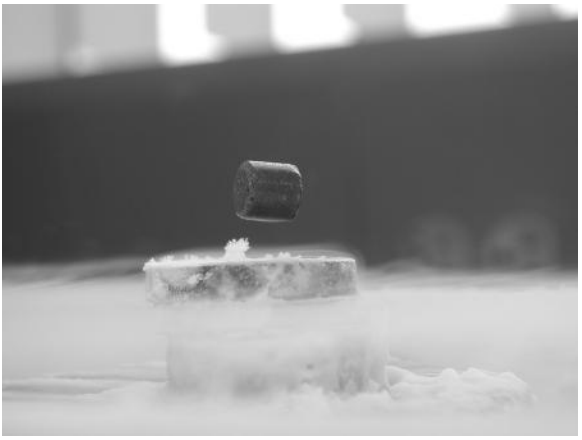


FIGURE 9.4 A permanent magnet floating over a superconducting surface. (From Darren Peets, UBC Superconductivity Group. With permission.)

It follows that a superconducting material can be made nonsuperconducting by the application of a large enough magnetic field. The minimum value of the field strength required to bring about this change is called the **critical field strength (H_C)**, its value depending on the material in question and on the temperature. Type I superconductors, which include most of the pure metal superconductors, have a single critical field. Type II superconductors, which include alloys and the high-T_C superconductors, allow some penetration of the magnetic field into the surface above a critical field, H_{C1}, but do not return to their nonsuperconducting state until a higher field, H_{C2}, is reached. Type II superconductors tend to have higher critical temperatures than Type I superconductors.

Similarly, if the current in the superconductor exceeds a **critical current**, the superconductivity is destroyed. This is known as the **Silsbee effect**. The size of the critical current is dependent on the nature and geometry of the particular sample.

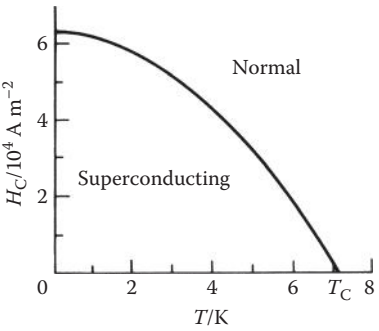


FIGURE 9.5 The variation of the critical field strength (H_C) for lead. Note that H_C is zero when the temperature (T) equals the critical temperature (T_C).

9.2.3 BCS Theory of Superconductivity

This section attempts to give a qualitative picture of the ideas involved and also to give some familiarity with the terminology.

Physicists worked for many years to find a theory that explained superconductivity. To begin with, it looked as though the lattice played no part in the superconducting mechanism because X-ray studies showed that there was no change in either the symmetry or the spacing of the lattice when superconductivity occurred. However, in 1950, an **isotope effect** was first observed: for a particular metal, the critical temperature was found to depend on the isotopic mass (M) such that

$$T_C \alpha \frac{1}{\sqrt{M}}. \tag{9.1}$$

It is well known that the frequency (ν) of vibration of a diatomic molecule is given by

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}, \tag{9.2}$$

where μ is the reduced mass of the molecule and k is the force constant of the bond. We can see that a vibration also changes frequency on isotopic substitution such that the frequency (ν) is proportional to $1/\sqrt{\text{mass}}$, suggesting to physicists that superconductivity was in some way related to the vibrational modes of the lattice and not just to the conduction electrons. The vibrational modes of a lattice are quantised, as are the modes of an isolated molecule: the quanta of the lattice vibrations being called **phonons** (see Chapter 4).

Frohlich suggested that there could be a strong phonon–electron interaction in a superconductor, leading to an *attractive* force between two electrons that is strong enough to overcome the Coulomb repulsion between them. Very simply, the mechanism works like this: as a conduction electron passes through the lattice, it can disturb some of the positively charged ions from their equilibrium positions, pushing them together and giving a region of increased positive charge density. As these oscillate back and forth, a second electron passing this moving region of increased positive charge density, is attracted to it. The net effect is that the two electrons have interacted with one another, using the lattice vibration as an intermediary. Furthermore, the interaction between the electrons is *attractive* because each of the two separate steps involved an attractive Coulomb interaction.

It is the scattering of conduction electrons by the phonons that produces electrical resistance at room temperature. At low temperatures, it is predominantly the scattering by lattice defects that gives electrical resistance. Contrary to what we might have expected intuitively, a superconductor will have *high* resistance at room temperature, because it has strong electron–phonon interactions. Indeed, the best room-temperature electronic conductors—silver and copper—do not superconduct at all.

Superconductors do not have low electrical resistance above the superconducting critical temperature, T_C.

In 1957, Bardeen, Cooper and Schrieffer published their theory of superconductivity, known as the **BCS theory**. It predicts that under certain conditions, the attraction between two conduction electrons due to a succession of phonon interactions can slightly *exceed* the repulsion that they exert directly on one another due to the Coulomb interaction of their like charges. The two electrons are thus weakly bound together, forming the so-called **Cooper pair**. It is these Cooper pairs that are responsible for superconductivity. In conventional superconductors, these electrons are paired so that their spin and orbital angular momenta cancel; they are described by a wave function, known as an **order parameter**, which has symmetry similar to the wave function of s electrons, that is, it is s-like and represents a singlet state.

The BCS theory shows that there are several conditions that have to be met for a sufficient number of Cooper pairs to be formed and superconductivity to be achieved. It is beyond the scope of this book to go into this in any depth: suffice it to say that the electron–phonon interaction must be strong and that low temperature favours pair formation—hence, high-temperature superconductors were not predicted by the BCS theory. The relatively high critical temperature of MgB_2 is thought to be due to the high vibrational frequencies associated with the light boron atoms and the strong interaction between the electrons and lattice vibrations. Evidence for the involvement of B atom vibrations comes from the observation that T_C is increased by about 1 K when ^{11}B is replaced by ^{10}B.

Cooper pairs are weakly bound, with typical separations of 10^6 pm for the two electrons. They are also constantly breaking up and reforming (usually with other partners). Thus, there is an enormous overlap between different pairs, and the pairing is a complicated dynamic process. The ground state of a superconductor, therefore, is a 'collective' state, describing the ordered motion of large numbers of Cooper pairs. When an external electrical field is applied, the Cooper pairs move through the lattice under its influence. However, they do so in such a way that the ordering of the pairs is maintained: the motion of each pair is locked to the motion of all the others, and none can be individually scattered by the lattice. Because the pairs cannot be scattered by the lattice, the resistance is zero and the system is a superconductor.

9.3 HIGH-TEMPERATURE SUPERCONDUCTORS

By 1973, the highest temperature found for the onset of superconductivity was 23.3 K—this was for a compound of niobium and germanium (Nb_3Ge)—and here it stayed until 1986 when Georg Bednorz and Alex Müller reported their findings. Their Nobel Prize citation states: 'Last year, 1986, Bednorz and Müller reported finding superconductivity in an oxide material at a temperature 12°C higher than previously known'. The compound that prompted their initial paper has been shown to be $La_{2-x}Ba_xCuO_4$, where $x = 0.2$, with a structure based on that of K_2NiF_4, a perovskite-related layer compound. They observed the onset of superconductivity at 35 K. The insight that they brought to this field was to move away from the investigation of metals and their alloys and to systematically study the solid-state physics and chemistry of metallic oxides.

Soon, the idea was born that it might be possible to raise the temperature even further by substitution with different metals. Using this technique, it was Chu's group in Houston, Texas, that finally broke through the liquid–nitrogen temperature barrier, with the superconductor that is now known as '1-2-3': this superconductor replaces lanthanum with yttrium and has the formula $YBa_2Cu_3O_{7-x}$. The onset of superconductivity for 1-2-3 occurs at 93 K. The highest critical temperature at ambient pressure discovered so far (2009) is 134 K for a doped $HgBa_2Ca_2Cu_3O_{8+\delta}$.

9.3.1 Cuprate Superconductors

The first high T_C superconductors discovered contained weakly coupled copper oxide (CuO_2) planes. Since then, many other high T_C superconductors with such planes have been reported. These are referred to as cuprate superconductors. The highest critical temperatures are found for cuprates containing a Group 2 metal (Ca, Ba, Sr) and a heavy metal, such as Tl, Bi or Hg. The structures of all the cuprate superconductors are based on or related to the perovskite structure. The perovskite structure is named after the mineral $CaTiO_3$ (see Chapter 1): many oxides of the general formula ABO_3 adopt this structure (also fluorides [ABF_3] and sulfides [ABS_3]). The so-called perovskite A-type unit cell (with the A-type atom in the centre of the cell) is shown in Figure 9.6.

The central A atom (Ca) is coordinated by 8 titaniums at the corners of the cube and by 12 oxygens at the midpoints of the edges. The perovskite structure can be equally well represented by moving the origin of the unit cell to the body centre: this has the effect of putting Ca (A) atoms at each corner, Ti (B) atoms at the body centre and an O atom in the centre of each face (Figure 9.7).

The crystal structure of the 1-2-3 superconductor, $YBa_2Cu_3O_{7-x}$, is shown in Figure 9.8. Figure 9.8a shows only the positions of the metal atoms: we can see the strong similarity between this and the structure of $La_{2-x}Ba_xCuO_4$. If, as before, we discuss it in terms of the perovskite structure ABO_3, where B = Cu, the central section is now an A-type perovskite unit cell and above and below it are also A-type perovskite unit cells with their bottom and top layers missing. This gives copper

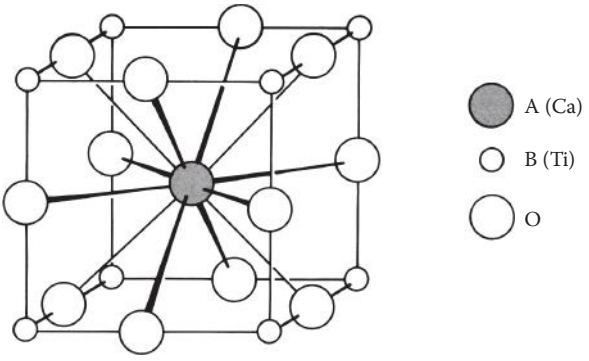


FIGURE 9.6 The A-type unit cell of the perovskite structure for compounds ABO_3, such as $CaTiO_3$.

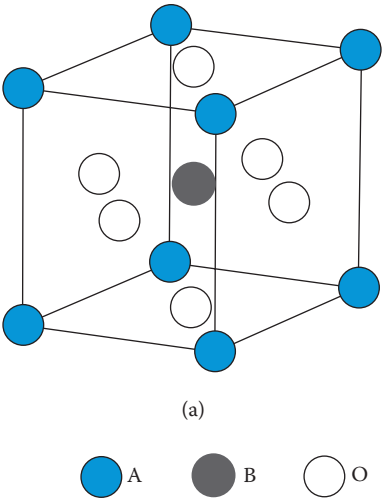


FIGURE 9.7 A B-type cell for the ABO_3 perovskite structure.

atoms at the unit cell corners and on the unit cell edges at fractional coordinates $\frac{1}{3}$ and $\frac{2}{3}$. The atom at the body centre of the cell (i.e., at the centre of the middle section) is now **yttrium**. The atoms at the centres of the top and bottom cubes are **barium**.

If, in this structure, all three sections were based exactly on perovskite unit cells, we would expect to find the oxygen atoms in the middle of each cube edge (Figure 9.8b), giving an overall formula of $YBa_2Cu_3O_9$. This formula is improbable because it gives an average oxidation state for the three copper atoms of $\frac{11}{3}$, implying that the unit cell contains both Cu(III) and Cu(IV), which is unlikely because Cu(IV) complexes are extremely rare. The unit cell, in fact, contains only approximately seven oxygen atoms ($YBa_2Cu_3O_{7-x}$): when $x = 0$, the oxygen atoms on the vertical edges of the central cube are not there and there are also two missing from both the top and the bottom faces (Figure 9.8c). A unit cell containing seven oxygen atoms has an average copper oxidation state of 2.33, indicating the presence of Cu(II) and Cu(III) in the unit cell (but no longer Cu(IV)).

In the 1-2-3 structure (when $x = 0$), the yttrium atom is coordinated by 8 oxygens and the barium atoms by 10 oxygens. The oxygen vacancies in the 1-2-3 superconductor create sheets and chains of linked copper and oxygen atoms running through the structure (this is shown slightly idealised in the diagrams as, in practice, the copper atoms lie slightly out of the plane of the oxygens): the copper is in fourfold square–planar or fivefold square–pyramidal coordination (Figure 9.8d). The superconductivity is found in directions parallel to the copper planes, which are created by the bases of the Cu/O pyramids and are separated by layers of yttrium atoms. If this superconductor is made more deficient in oxygen, at $YBa_2Cu_3O_{6.5}$ ($x = 0.5$), the superconducting critical temperature (T_C) drops to 60 K and at $YBa_2Cu_3O_6$ the superconductivity disappears. The oxygen is not lost at random, but goes from specific sites, gradually changing the square–planar coordination of the Cu along the *c*

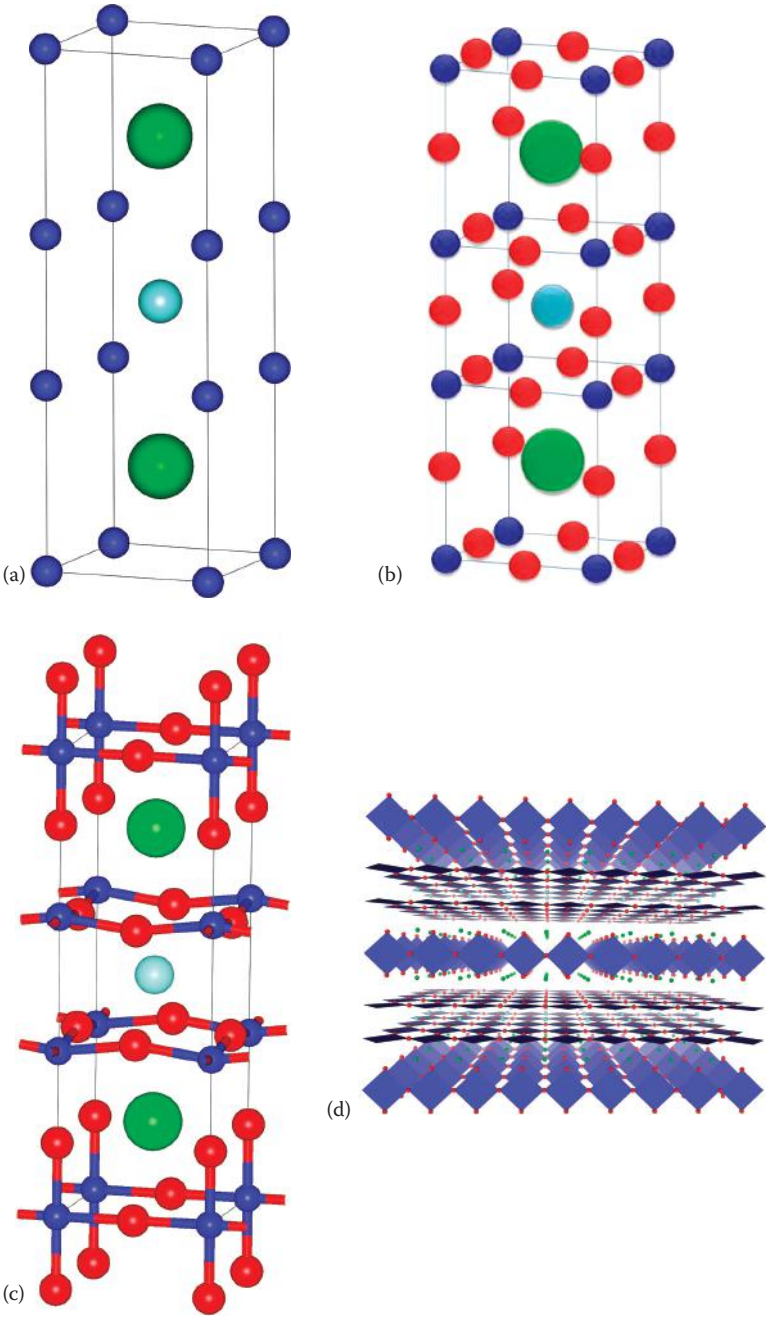


FIGURE 9.8 The structure of 1-2-3; (a) the metal positions; (b) an idealised unit cell of the hypothetical $YBa_2Cu_3O_9$, based on three perovskite A-type unit cells; (c) an idealised structure of $YBa_2Cu_3O_{7-x}$; and (d) the structure of $YBa_2Cu_3O_7$, showing copper–oxygen planes formed by the bases of the pyramids, with the copper–oxygen diamonds in between.

direction into the twofold linear coordination characteristic of Cu^+ and the arrangement of the copper and oxygen atoms in the base of the pyramids is not affected. However, when the formula is $YBa_2Cu_3O_6$, all the square–planar units along the *c* direction have become chains containing Cu^+ and the pyramid bases contain only Cu^{2+}; the unpaired spins of the Cu^{2+} are aligned antiparallel and the compound is antiferromagnetic. It is not until the oxygen content is increased to $YBa_2Cu_3O_{6.5}$ that the antiferromagnetic properties are destroyed and the compound becomes a superconductor. It is thought that this compound contains copper in all three oxidation states—I, II and III. $YBa_2Cu_3O_7$ contains Cu(II) and Cu(III) both in the sheets and in the chains. Clearly, the oxidation state of the copper atoms in the structure (and thus their bonding connections and bond lengths) is extremely important in determining both whether superconductivity occurs at all and the temperature below which it occurs (T_C).

The Hg-containing cuprate superconductors, of which more than 40 are known and include the material with the highest known (in 2009) T_C at ambient pressure, are also perovskite-based structures. The structures of these solids fall into one of three classes, 1201 based on $HgBa_2CuO_4$, 1212 based on $HgBa_2CaCu_2O_6$ and 1223 based on $HgBa_2Ca_2Cu_3O_8$. The Hg layers in the parent compounds contain no oxygen and are poor superconductors. In 1993, it was found that annealing these compounds in oxygen resulted in oxygen atoms being inserted into the ($\frac{1}{2}$, $\frac{1}{2}$, 0) position in the Hg layer and that $HgBa_2CuO_{4+\delta}$ was superconducting. Under pressures above 10 GPa, $HgBa_2Ca_2Cu_3O_{8+\delta}$ has a critical temperature greater than 150 K. By replacing some of the Hg atoms with metal atoms in higher oxidation states, compounds with extra oxygen in the Hg layer are formed, which are more stable at normal pressures. The highest critical temperature at the time of writing is for the 1223 structure with one-fifth of the Hg atoms replaced by Tl: $Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8+0.33}$. The structure of this compound is shown in Figure 9.9.

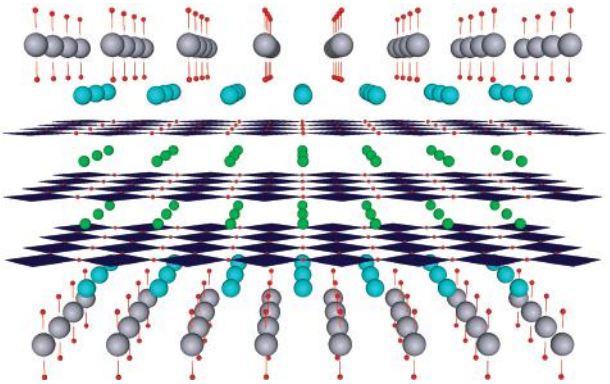


FIGURE 9.9 The structure of $HgBa_2Ca_2Cu_3O_8$ showing copper oxygen diamonds between the Ca layers and the copper oxygen layers forming the bases of the pyramids. The apices of the pyramids are in the barium oxygen layers. In the superconductor $Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8+0.33}$, one-fifth of the Hg^{2+} ions are replaced by Tl^{3+} ions and there are additional oxygen ions in the mercury layer.

The unit cell contains four CuO_2 layers with Cu in square–pyramidal coordination and with the apices of the pyramids alternately pointing up and down. Ba occupies sites close to the apices of these pyramids and the Ca positions are just above or below the bases. At the top and bottom of the cell, between two layers of Ca atoms are CuO layers with the Cu in square–planar coordination. In the centre of the cell is the Hg layer. Overdoping with Tl gives a metallic compound with poor superconducting properties.

In 1-2-3, $La_{2-x}Ba_xCuO_4$ and the mercury cuprates, the average oxidation state of the Cu is greater than 2 and, as a result, positive holes are formed in the valence bands. The charge is carried by the positive holes and, as a consequence, such materials are known as ***p*-type superconductors**. In 1-2-3 and the mercury cuprates, positive holes are formed by adding oxygen, but there are other methods. In $SrCuO_2F_{2+\delta}$, superconductivity is induced by the insertion of fluorine, and in the collapsed oxycarbonates, such as $TlBa_2Sr_2Cu_2(CO_3)O_7$, superconductivity can be produced by shearing so that a shift appears along one plane every n octahedra (where n is typically 3–5), which leaves the CuO_2 layers unchanged and mixes the TlO and CO layers. T_C for the oxyhalides can reach 80 K and for the collapsed carbonates, T_C is in the range of 60–77 K.

Until 1988, all the high-temperature superconductors that had been found were *p*-type and it was assumed by many that this would be a feature of high-temperature superconductors. However, some ***n*-type superconductors** have also been discovered, where the charge carriers are electrons: the first to be found was based on the compound Nd_2CuO_4 with small amounts of the three-valent neodymium substituted by four-valent cerium—$Nd_{2-x}Ce_xCuO_{4-y}$, where $x \sim 0.17$ (samarium, europium or praseodymium can also be substituted for the neodymium). Other similar compounds have since been found based on this structure where the three-valent lanthanide is substituted by, for example, four-valent thorium in $Nd_{2-x}Th_xCuO_{4-y}$. The superconductivity occurs at $T_C \leq 25$ K for these compounds.

It seems clear that in cuprate superconductors, the superconductivity takes place in the CuO_2 planes and that the other elements present and the spacings between the planes change the superconducting transition temperature—exactly how is not yet understood.

An unusual superconductor that is structurally related to the cuprates was identified in 1994: Sr_2RuO_4 has a crystal structure almost identical to La_2CuO_4, the parent compound of the high T_C superconductor 1-2-3. However, whereas La_2CuO_4 is an antiferromagnetic insulator, Sr_2RuO_4 is metallic and, below about 1.5 K, it is superconducting. Interest in this compound, despite its low T_C, is due to the presence of ferromagnetism as well as superconductivity. Conventional superconductors are diamagnetic and, previously, diamagnetism had been thought of as a necessary condition for superconductivity. The ruthenium in Sr_2RuO_4 is formally present as Ru(IV), giving an electronic configuration of $4d^4$. The Ru is in the centre of an octahedron of oxygen and so the d bands are split into t_{2g} and e_g. The four 4d electrons partly occupy the t_{2g} band. It is the electrons close to the highest-occupied levels of this band that are responsible for the superconductivity.

9.3.2 Iron Superconductors

Due to the idea that magnetism and superconductivity were believed to be mutually exclusive, the use of a magnetic ion to form a superconducting layer was generally avoided. Then, in 2008, Hideo Hosono reported superconductivity with $T_C = 26$ K in fluorine-doped LaFeAsO. This prompted researchers around the world to search for iron-based superconductors. At the time of writing (2011), the record T_C for such compounds was 55 K in $SmFeAsO_{1-x}F_x$, under an applied pressure. The structure of the parent compound, SmFeAsO, is shown in Figure 9.10.

Like the cuprates, the iron-based superconductors form a layer structure. The characteristic layer consists of Fe arranged on a square lattice with P, As, Se or Te tetrahedrally coordinated to the iron above and below the plane. These layers are either simply stacked or interspersed with layers of Group 1, Group 2 or lanthanoid oxyfluorides, or with a perovskite-type structure. Five structural classes have been identified. The Fe ions form the conducting layer, and the highest critical temperatures so far are for solids containing As.

The parent compounds, such as LaFeAsO and Fe_2Se_2, contain iron in a formal oxidation state of +2. It is perhaps surprising, therefore, that most of the superconductors found contain dopants that result in the addition of electrons to the Fe layer. The electrons are supplied in a variety of ways—substitution of fluoride for oxide, oxygen vacancies, insertion of Group 1 or 2 metals in between layers to give M^{2+} or M^+ ions, or substitution of cobalt on iron sites. The addition of electrons to the iron layer will, of course, reduce the oxidation state of Fe. This together with the high coordination number of the Fe, which includes coordination to adjacent Fe as well as to P, As, Se and Te, suggests that the superconducting layer is more like an alloy than an ionic structure.

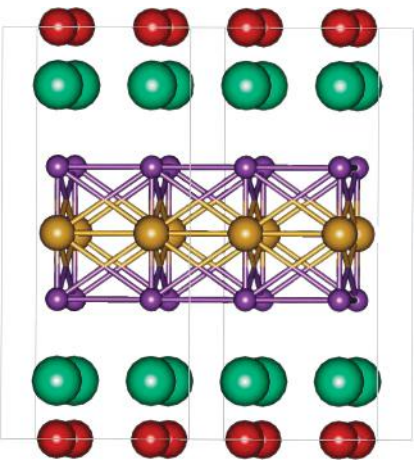


FIGURE 9.10 The structure of SmFeAsO. The Fe and As atoms form a closely packed layer with the Fe atoms at the centre. Above and below this layer are Sm ions and above and below these are oxide ions.

There are also some Fe-based superconductors in which doping withdraws electrons from the iron layer, for example, by the substitution of Group 2 metals on lanthanoid sites.

9.3.3 Theory of High T_C Superconductors

The general consensus is that, in common with the conventional superconductors, high-temperature superconductors contain Cooper pairs (Section 9.2.3). However, there is a difference in the pairs formed. In conventional superconductors, the Cooper pairs couple to give a total angular momentum of zero and are described by an order parameter of s-wave symmetry. In high-temperature cuprate superconductors, the Cooper pairs couple to give nonzero angular momentum and are described by an order parameter of d-wave symmetry. The pairing state in the BCS theory does not need to be caused by electron–phonon interaction, and it is thought that pairing in the d state is due to antiferromagnetic spin fluctuations or electron–magnon coupling. Such a possibility had been examined by Kohn and Luttinger, who found that a weak collective residual attraction could be generated from the Coulomb repulsion between electrons, but only if the electrons in the Cooper pairs were prevented from close encounters. This could be achieved if the pairs had angular momentum, that is, if they were described by p- or d-waves. Their initial estimate of T_C in such systems was very low, but it was hoped that higher temperatures might be found in metals with strong spin fluctuations. As in the BCS theory for conventional superconductors, the coupling is a collective property and the pairing is a dynamic process. Spin-mediated coupling is far more effective in bringing about superconductivity in two dimensions than in three, and this may explain why high T_C superconductivity is found in quasi–two-dimensional solids such as the cuprates. Experiment suggests that the quasi–two-dimensional organic superconducting polymers are also d-wave superconductors.

The presence of d-wave symmetry has been demonstrated in an elegant experiment using scanning tunnelling microscopy (STM, see Chapter 2). Collaborators at Berkeley and Tokyo, led by J.C. Séamus Davis and Shin-ichi Uchida, used STM to measure differential tunnelling conductance on a specimen of bismuth strontium calcium copper oxide with zinc atoms replacing a small number of copper atoms. Differential tunnelling conductance is proportional to the density of states available to electrons tunnelling from the sample into the microscope's tip. This should be zero within a defined energy above the highest-occupied levels, known as the superconducting symmetry gap. For d-wave symmetry, the energy gap is zero in certain directions resembling the nodes of d orbitals. The Zn atoms broke the superconducting pairs at their positions so that a conductance peak appeared at the positions of the Zn atoms. Emanating from the Zn positions were lines of relatively high conductance in the direction of the nodes expected for the d-wave function.

The iron-based superconductors are also thought to contain Cooper pairs and there is some evidence from isotope studies that phonons are involved. As for the cuprate semiconductors, the solids are antiferromagnetic above T_C and the onset

of superconductivity may be linked to the quashing of magnetic ordering. NMR has shown that these superconductors are singlet, $S = 0$. The wave function of the Cooper pair must be antisymmetric with respect to the exchange of electrons. For singlet Cooper pairs, exchange leads to a change in spin, so the order parameter must be even (s-, d-, etc.). Measurements on some solids have found the superconducting symmetry gap to be nodeless, suggesting an s-wave symmetry. However, the type of s-wave symmetry is not that of the conventional superconductors because there is evidence of phase change of the s-wave. The nature of the mechanism, however, is still a matter of debate.

Ferromagnetic superconductors are thought to have Cooper pairs that are formed by magnetic interactions as in high-T_C superconductors, but with their spins aligned, giving a triplet state. For a triplet pair, both electrons have the same spin, so the order parameter must be odd. It is generally accepted that the ferromagnetic superconductors have p-wave order parameters.

9.4 USES OF HIGH-TEMPERATURE SUPERCONDUCTORS

Conventional superconductors are used to form magnets, for example, for use in NMR spectrometers. Other uses include superconducting quantum interference devices (SQUIDS), which are loops of wire containing Josephson junctions (constructed from superconducting layers) and can measure very small voltages, currents and magnetic fields. Cooling to liquid helium temperatures is expensive, therefore using high-temperature superconductors would reduce costs in these and other applications. However, the cuprates are brittle ceramic materials that are not easily formed into wires and the superconductivity in these materials is anisotropic. Progress has been made on overcoming these problems.

The bismuth strontium calcium copper oxides, $Bi_2Sr_2Ca_2Cu_3O_{10}$ and $Bi_2Sr_2CaCu_2O_8$, have been processed into long wires by packing a powdered precursor into a silver tube, heating it to form the superconductor and drawing the silver tubes to form a wire. The bismuth-based cuprates were chosen because they are more flexible than other high T_C superconductors due to the weak bonds across the double Bi/O layers that allow the layers to slip. Wires that can carry 100 A per strand at 77 K are available in lengths of 100 m. Second-generation, high-temperature superconductor wires use yttrium barium copper oxide (YBCO) and are produced as ribbons by depositing thin flexible films of YBCO on a metal substrate. This wire can carry 100 times the current of copper wires.

Electric motors using high-temperature superconductors, and magnets containing such materials have been developed. In 1997, a high-T_C superconducting magnet was installed in the beamline of a carbon-dating van der Graaf accelerator at the Institute for Geological and Nuclear Sciences in Wellington, New Zealand. In 1998, Oxford Magnet Technology and Siemens Corporate Technology built a prototype MRI body scanner with two coils made from high T_C wires, and Hitachi have demonstrated a 23.4 Tesla superconducting magnet close to the 23.5 Tesla needed for NMR spectrometers operating at 1 GHz.

The absence of electrical resistance could be extremely useful in the distribution of electricity throughout the country, as the copper and aluminium wires now used lose 5%–8% of power due to the resistance of these elements. Using superconductors would not eliminate power loss entirely. Although superconductors carry direct current (dc) with no loss of power and without generating heat, there is a small loss of power when they carry alternating current (ac) due to the production of radio waves. However, this loss is much less than that with copper and aluminium wires. The National Grid is unlikely to be replaced by superconducting cables in the near future, but such cables are capable of providing a solution to the problem of supplying extra power to cities. Power is run into cities via cables in conduits. Many conduits running power into large cities and towns are full and to provide extra power would mean buying up land and digging new ones. By replacing the existing cables with high T_C cables, more power could be supplied using the existing conduits. In May 2001, 150,000 residents of Copenhagen received electricity through a 30 m high-temperature superconductor cable. In the same year, Pirelli replaced nine copper cables at a substation in Detroit by a 120 m cable consisting of three high-temperature superconducting cables. The total mass of this high-temperature cable was 70 times less than the copper it replaced. For the first time, power was delivered on a commercial basis to US customers via superconducting wires.

QUESTIONS

1. A party of late-night revellers, returning home, decide to take a shortcut across a field. Unfortunately, the night is dark and moonless and the field is known to contain deep potholes. Someone has the bright idea that if everyone links arms and advances together across the field, then if one of the company does encounter a pothole, she will be lifted clear by dint of collective support! Taking this story as an analogy (not a perfect one, of course) with the BCS theory, try to identify as many components as possible with corresponding components of the BCS theory.
2. Draw a packing diagram (see Chapter 1) of the perovskite A-type cell ($CaTiO_3/ABO_3$), determine the number of ABO_3 formula units and describe the coordination geometry around each type of atom. Repeat this procedure for the perovskite B-type unit cell.
3. Complete the following table, summarising the types of superconductivity.

Material	Symmetry of order parameter	Spin state
Metals and alloys		
Cuprates		
Iron-based superconductors		
Ferromagnetic superconductors		

4. Calculate the average oxidation state of Cu in the mercury cuprate $Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8.33}$. Assume Tl is present as Tl^{3+} and Hg as Hg^{2+}.
5. Calculate the average oxidation state of Fe in the superconductor $K_{0.75}Fe_2Se_2$ and comment on the value you obtain. Assume Se is present as Se^{2-}.
6. Using Figures 9.9 and 9.10, compare the structures of cuprate and iron-based superconductors, in terms of broad features of similarity and difference between the two classes.

10 Nanostructures and Solids with Low-Dimensional Properties

10.1 NANOSCIENCE

In 1905, as part of his doctoral thesis, Einstein measured the size of a sugar molecule using diffusion techniques and found it to be about 1 nm in diameter. A hydrogen atom is about 0.1 nm in diameter, so for chemists, thinking on the nanometre scale is thinking on the atomic or molecular scale, which is what we are used to doing. Therefore, what is different about nanoscience from what chemists have always done? Current research on the nanoscale has two crucial strands. The first is to investigate and utilise the very different properties displayed by very small particles compared with the properties of bulk solids. Chemists have roles here in synthesising these particles, investigating their properties and developing new uses for them; nanoscience covers such topics as nanotubes, coatings, new alloys, composites, particles for sunscreens, catalysts, colloids and quantum dots, to name but a few. The second is to be able to manipulate or manufacture things very precisely on the nanometre scale. Some techniques, such as photolithography for the printing of silicon chips, are the so-called top-down processes because they are carving out or etching nanometre-size structures; these are clearly based on technology, although often use chemical processes. The building up of structures, self-assembly processes and the direct manipulation of atoms into nanostructures, the so-called bottom-up processes, once again fall into the realm of chemistry (see Chapter 3).

Nanoparticles are usually considered to have at least one dimension less than 100 nm, although this definition is not rigid. Such particles have a larger surface area–to-volume ratio than larger particles, affecting the way that they react with other particles and with each other. A 10 nm diameter nanoparticle has about 15% of the atoms on their surface, and by comparison, this drops to <1% for a bulk solid. Because a particle may only consist of a few atoms, the energy levels associated with extended solids such as bands in metals no longer apply, and the electronic energy levels are more similar to the quantised levels found in individual atoms. This affects their conductivity and the way that they interact with light and other forms of energy.

The Royal Society has come up with two working definitions:

- **Nanoscience** is the study of phenomena and manipulation of materials at atomic, molecular and macromolecular scales, where properties differ significantly from those at larger scales.
- **Nanotechnology** is the production and application of structures, devices and systems for controlling shape and size at nanometre scale.

Much of the research centres around the organic/bioorganic area and is involved with using self-assembly methods. We touched on this area in Chapter 3, but we will not discuss it any further here. In the following sections, we will concentrate only on areas that relate in some way to solid-state chemistry. First, we consider the physical and electronic effects on the nanoscale.

10.2 CONSEQUENCES OF THE NANOSCALE

10.2.1 Nanoparticle Morphology

The high surface-to-bulk ratio in nanostructures also means that the crystal structure is determined by a balance between bulk terms such as lattice energy, surface energy terms and terms due to faults such as dislocations, as all these terms are now significant. This can lead to unusual crystal structures, such as thin films of body-centred cubic (*bcc*) copper, compared with the normal bulk structure, which is cubic closed packed (*ccp*).

Varying the condition of a film in chemical vapour deposition (CVD) can alter the morphology of the nanocrystals formed, and the techniques for producing specific morphologies could be very important in the production of catalysts, because different crystal faces have been found to catalyse very specific reactions.

Many nanomaterials can be made in different forms. In Section 10.3, we look at carbon, which can be found as diamond films, single layers of graphite, carbon black, fullerenes, and multi- and single-walled nanotubes. Inorganic materials can also form a variety of nanostructures, for example, MoS_2 can be made as nanotubes, 'onions' (multi-walled fullerene-type structures) and thin films.

The methods and conditions of nanostructure manufacture are crucial because, as we shall see, the properties depend critically on the size and the shape of the particles produced. The synthetic techniques used are often so particular to each system that they make a subject in their own right. Some of the techniques were covered in Chapter 3.

10.2.2 Electronic Structure

In Chapter 4, we discussed how the energy levels of a crystal could be obtained by thinking of a crystal as a very large molecule. For crystals of, say, micrometre dimensions, the number of energy levels is so large and the gap between them so small that we could treat them as essentially infinite solids with continuous bands of allowed energy. At the nanometre scale, we can still think of the particles as giant molecules, but a typical nanoparticle contains 10^2–10^4 atoms, very large for a molecule but not large enough to make an infinite solid a good approximation. The result is that in nanoparticles, we can still distinguish bands of energy, but the gaps between the bands may differ from those found in larger crystals and, within the bands, the energy levels do not quite form a continuum so that we can observe the effects due to the quantised nature of the levels within the bands. This is illustrated in Figure 10.1 for the nano-sized crystals of a semiconductor (quantum dots).

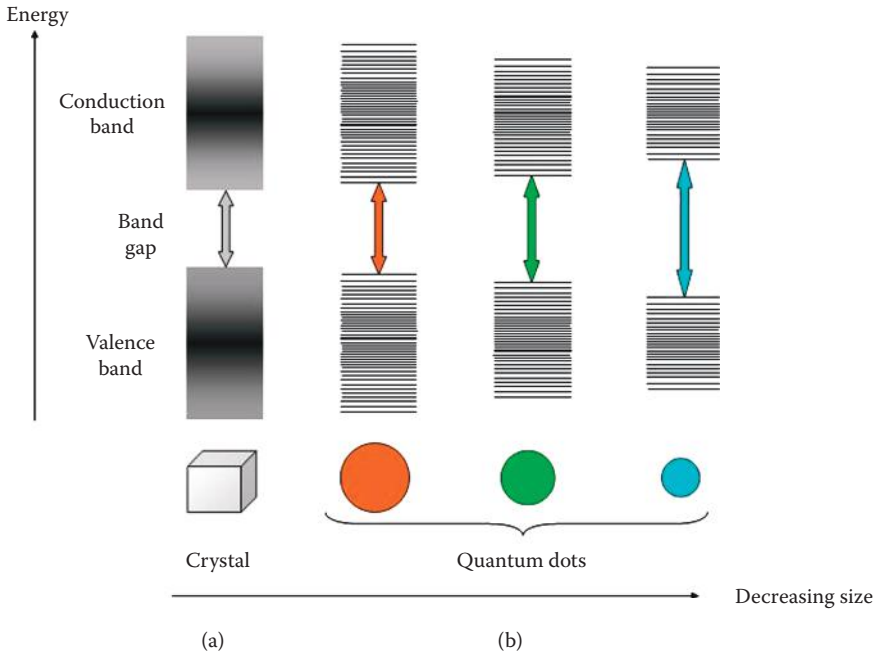


FIGURE 10.1 (a, b) The band gap of a semiconductor depends on its size.

The band diagram for a semiconductor is shown in Figure 10.1a. The bonding electrons are held in the lower valence band consisting of a continuum of many energy levels, and two electrons can occupy each energy level. The orbital density of states diagrams show that, in general, there is a lower density of states at the top and the bottom of the band and a higher density of states in the middle (indicated by shading). Above the valence band there is a conduction band, with a band gap separating the two bands. If the electrons are promoted from the valence band to the empty conduction band by supplying sufficient energy for them to jump the band gap, then the solid conducts.

As a crystal of a semiconductor becomes smaller in size, there are fewer atomic orbitals available to contribute to the bands. The orbitals are removed from each of the band edges until a point when the crystal is very small—a 'dot'—the bands are no longer a continuum of orbitals, but individual quantised orbital energy levels (Figure 10.1b). At the same time, this has the effect of increasing the band gap: as the size of the nanoparticles of most semiconductors decreases, so does the number of orbital energy levels, but the band gap increases. The band gap in CdSe crystals, for example, is approximately 1.8 eV for crystals of diameter 11.5 nm, but it is approximately 3 eV for crystals of diameter 1.2 nm.

Quantum dots are nanometre scale in three dimensions, but structures that are only nanometre scale in two dimensions (quantum wires) or one dimension (quantum wells or films) also display interesting properties. The quantised nature of the bands in nanostructures can be seen in the density of states. Schematic, theoretical

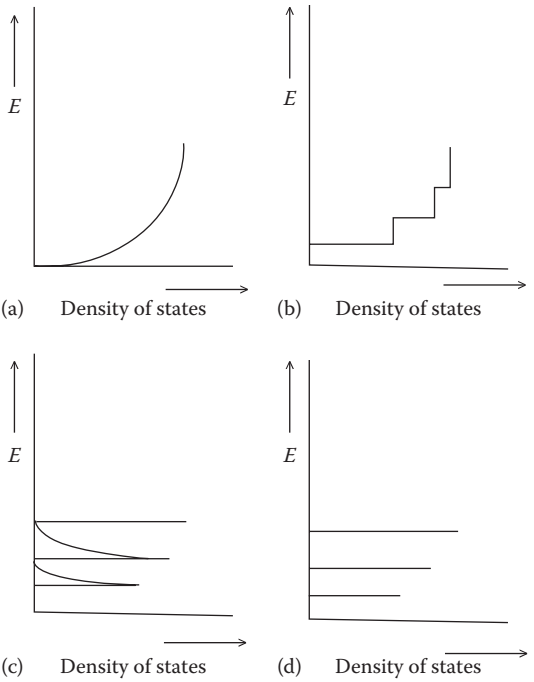


FIGURE 10.2 The theoretical density of states diagrams for (a) bulk material, (b) a quantum well, (c) a quantum wire and (d) a quantum dot.

density of states diagrams for bulk material, quantum wells, quantum wires and quantum dots are shown in Figure 10.2.

Figure 10.3 shows the density of states for a specific example, a semiconducting nanotube. The density of states for carbon nanotubes is predicted to show sharp peaks (known as van Hove singularities) corresponding to specific energy levels. These can be seen in Figure 10.3 and have been confirmed by scanning tunnel microscopy (STM) experiments. In Figure 10.3, there is a gap with zero density about the Fermi energy ($E = 0$ on the figure). This is the band gap that we associate with semiconductors. For semiconducting carbon nanotubes, the band gap generally increases with decreasing diameter, but for particular nanotube structures, the band gap becomes zero and the nanotubes are then metallic conductors similar to graphite.

Electrical conductance in solids (other than ionic conductors) depends on the availability of delocalised orbitals close enough together in energy to form bands. When we have structures on the nanoscale where the levels at the top of occupied bands are not as close in energy as we have seen, then this can affect the conductance. Under certain conditions, electrical conductance through a nanostructure is quantised and increases in a stepwise fashion with increasing voltage. The electrons tunnel into the structure and fill the lowest empty quantised levels until all the levels below the highest-filled level providing electrons is full (Figure 10.4). If the thermal energy is insufficient to raise the electrons to the next discrete energy level, then

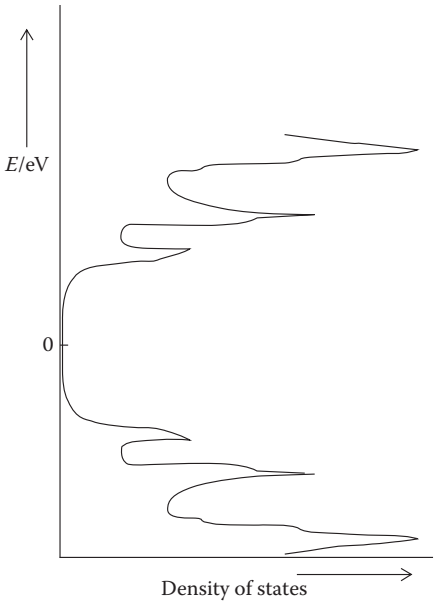


FIGURE 10.3 The density of states for a semiconducting nanotube.

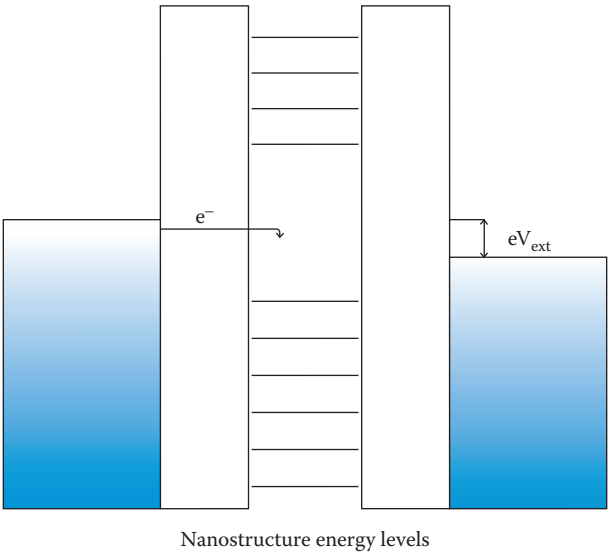


FIGURE 10.4 Electrons fill the lowest empty quantised levels until all the levels below the highest-filled level providing electrons are full.

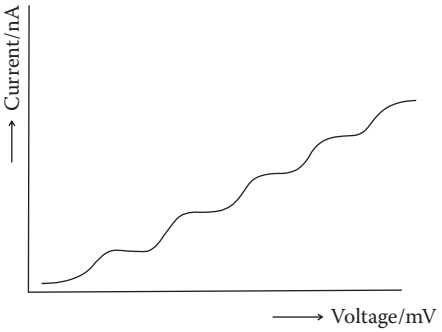


FIGURE 10.5 The increase of current with gate voltage along a quantum wire connecting two GaAs/AlGaAS interfaces in a transistor.

no more electrons can tunnel in. The conductance thus drops until the voltage is increased to a value V, such that the energy $e \times V$ is sufficient to raise the energy of the electrons in the adjacent solid so that they can reach the next energy level, when the current increases again. Figure 10.5, for example, shows the increase in conductance with gate voltage along a quantum wire connecting two GaAs/AlGaAs interfaces in a transistor.

10.2.3 Optical Properties

In this section, we consider two aspects of the interaction of light with nanostructures—the differences in the optical properties of bulk solids that arise because of the different energy levels of nanostructures, and the changes in the scattering properties.

The last section revealed that semiconductor band gaps for nanostructures vary with the size of the structure. The wavelength of light emitted when an electron in the conductance band returns to the valence band will therefore also vary. Thus, different colour fluorescence emission can be obtained from different-sized particles of the same substance. To produce fluorescence, light of greater photon energy than the band gap is shone onto the nanostructure. An electron is excited to a level in the conduction band, from where it reaches the lowest energy level in the conduction band through a series of steps by losing energy as heat. The electron then returns to the valence band, emitting light as it does so (Figure 10.6).

Figure 10.6a depicts the irradiation of a large quantum dot (smaller band gap). It then decays into the valence band, emitting a photon of light, which is the coloured fluorescence seen. If the smaller quantum dot (larger band gap) (Figure 10.6b) undergoes the same process, we can see that the photon emitted as it decays back into the valence band has more energy. From the Planck–Einstein equation, $E = hv$, the higher energy photon will have the higher frequency and thus be closer to the blue end of the spectrum, accounting for the difference in colour. The smaller the nanostructure, the larger is the band gap hence the shorter is the wavelength of the emitted

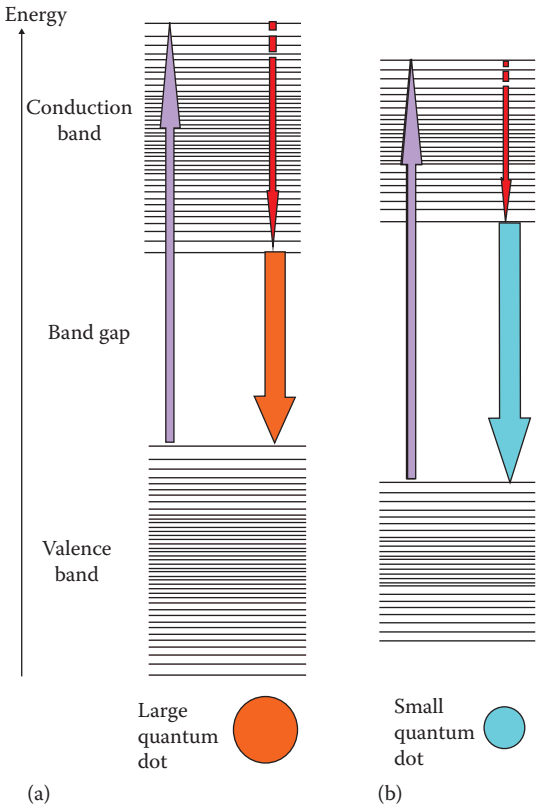


FIGURE 10.6 (a, b) The colour of the fluorescence from a nano-sized particle depends on its dimensions.

light. Thus, 5.5 nm diameter particles of CdSe emit orange light, whereas 2.3 nm diameter particles emit turquoise light.

In the absorption spectra of nanoparticles of CdSe and other semiconductors, not only can the shift in wavelength be observed, but there are also bands corresponding to absorption to discrete energy levels in the conduction band. For example, 11.5 nm diameter particles of CdSe have an absorption spectrum showing an almost featureless edge, but particles of diameter 1.5 nm show features resembling molecular absorption bands shifted about 200 nm to shorter wavelengths, as shown in Figure 10.7.

The colours produced by nanoparticles of gold (colloidal gold) have been used to colour glass since Roman times and have been studied since Faraday in the mid-nineteenth century. In metals, light interacts with the surface electrons and is reflected. The surface electrons in nanoparticles are induced by the light to oscillate at a particular frequency; absorption at this frequency gives rise to the colour. The oscillation frequency and hence the colour depend on the size of the particle. For example, whereas bulk gold is yellow, thin films of gold appear blue when

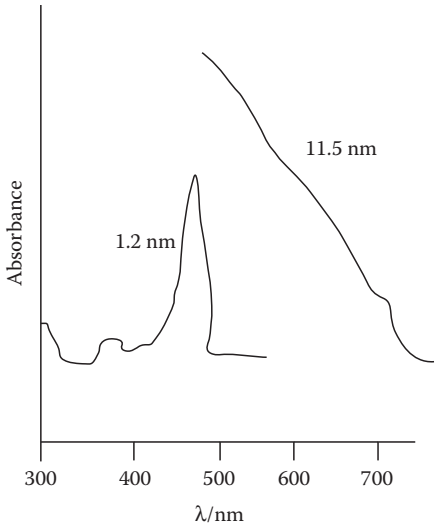


FIGURE 10.7 Schematic spectra of nanoparticles of CdSe. The absorption spectrum of 11.5 nm diameter particles has an almost featureless edge, but particles of diameter 1.2 nm show features resembling molecular absorption bands of about 200 nm shifted to shorter wavelengths.

light passes through them, but as the particle size is reduced, the wavelength of the absorbed light decreases and the film appears first red then, for 3 nm diameter particles, orange. Oscillation frequencies in the visible region are only observed for Ag, Au, Cu and the Group 1 metals.

Particles are known to scatter light as well as absorb it, and this produces the white or pale appearance of fine powders. The even-smaller nano-sized particles, however, are transparent because the scattering efficiency is reduced. This effect has led to the use of nanoparticles in sunscreens and cosmetics. These will still absorb ultraviolet light but will scatter less visible light.

10.2.4 Magnetic Properties

In Chapter 8, the idea of ferromagnetic domains was introduced. Domains typically have dimensions of the order of 10–1000 nm. In nanocrystals, therefore, we can reach a situation where the domain size and the crystal dimensions are comparable. Such single-domain crystals have all the electron spins in the crystal aligned. Single-domain crystals of magnetite are found in certain bacteria (magnetotactic bacteria), which use the magnets to locate the sediment/water interface in muddy ponds.

In larger crystals, the main mechanism for magnetisation and demagnetisation is the rotation of the domain walls (Section 8.3.1). If the crystal size is reduced, as the single-domain region is approached, it becomes harder to demagnetise the crystal by applying a magnetic field. This is because the only possible mechanism is now the disruption of spin–spin coupling within a domain. If the particle size is decreased

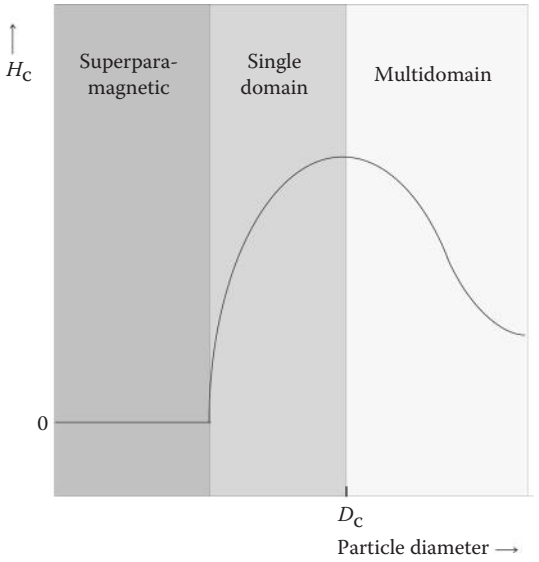


FIGURE 10.8 A plot of the magnetic field needed to demagnetise ferromagnetic particles (coercivity, H_C) as a function of the particle size. The particle becomes single domain at radius D_C.

still further, however, the number of spins decreases and the force aligning them becomes weaker. Eventually, as this force gets weaker, and under the influence of temperature and in the absence of an applied magnetic field, the spins are able to flip over to their other orientation. Then, the crystal is no longer ferromagnetic but **superparamagnetic**. Superparamagnetism is different from normal paramagnetism because it occurs *below* the Curie temperature, and at this temperature the spins are able to flip to the opposite orientation, thereby reversing the direction of magnetisation. If the time between flips is short, when the magnetisation is measured it appears to be zero. Figure 10.8 is a plot of the magnetic field needed to demagnetise ferromagnetic particles (coercive force) as a function of the particle size. The radius D_C is the radius at which the particle becomes single domain. The values for D_C for the ferromagnetic transition metals and the ferrimagnetic Fe_3O_4 (magnetite) are given in Table 10.1.

TABLE 10.1
Single-Domain Radii for Selected Ferromagnetic Solids

Solid	D_C/nm
Fe	14
Co	70
Ni	55
Fe_3O_4	128

In the superparamagnetic state, if a magnetic field is applied, above a critical temperature (T_B) known as the blocking temperature, the particle can have its spins aligned by a magnetic field and becomes magnetised. It behaves in a similar fashion to paramagnetic materials, but with a much larger susceptibility and magnetic moment than those observed for paramagnetism. Below the blocking temperature, there is insufficient kinetic energy to overcome the energy barrier to reorientation of the spins. The blocking temperature varies with the strength of the applied field. In the superparamagnetic state, the particles, like paramagnetic solids, do not show hysteresis.

The existence of superparamagnetism in ferromagnetic nanoparticles sets a limit to the minimum size of the particle that can be used in hard-disk drives and therefore on the possible storage density.

So far, we have considered only isolated magnetic nanoparticles. When nanosized grains are in close contact, the magnetic effects can differ. This is because spin exchange coupling can occur between the grains. In Chapter 8, we saw that thin layers of magnetic material could have the spin state pinned by an adjacent antiferromagnetic layer. Exchange coupling between grains of hard and soft magnetic materials can produce nanocomposites with high remanence and high coercivity, that is, they remain strongly magnetic when the applied magnetic field is reduced to zero and need a large magnetic field to demagnetise them.

10.2.5 Mechanical Properties

The properties of solids, such as hardness and plasticity, have long been known to vary with the grain size. For example, down to micrometre-sized grains, the resistance to plastic deformation increases with decreasing grain size. However, as the grain size is reduced still further, the resistance levels off or even decreases. The effect has been attributed to a different mechanism for plastic deformation in nanosized grains. In larger crystals, deformation is mainly governed by the movement of dislocations (Chapter 5) in the crystal. Such movements are inhibited by grain boundaries. As the grain size is reduced, the ratio of grain boundary to bulk grain increases, so deformation becomes harder. Eventually, as the grain size reaches 5–30 nm, the movement of the dislocations becomes negligible, but deformation through atoms sliding along the grain boundaries becomes favourable. This latter mechanism is aided by a large ratio of the grain boundary to the bulk grain, so softening increases as the grain size is reduced.

10.2.6 Melting

Nanocrystals show a melting temperature depression with decreasing size. Melting of nanocrystals can be observed in a transmission electron microscope, and gold nanoclusters have been found to melt at 300 K compared with a melting temperature of 1338 K for elemental gold in bulk, and 3 nm CdS crystals melt at about 700 K in a vacuum compared with 1678 K in bulk. The surface energy becomes an increasing factor as the size of a crystal becomes smaller; thermodynamics predicts a lowering of the melting temperature if the surface energy of the solid is higher than that of the liquid.

10.3 LOW-DIMENSIONAL AND NANOSTRUCTURAL CARBON

10.3.1 Carbon Black

Carbon black is made by the vapour-phase incomplete pyrolysis of hydrocarbons to produce a fluffy fine powder consisting of nanoparticles of amorphous carbon. The method of preparation produces particles with few impurities in contrast to powders of similar appearance such as soot. Worldwide, about 8 million tonnes a year is produced. It is used as a reinforcing agent in rubber products such as tyres, as a black pigment in printing inks, paints, and plastics, in photocopier toner, and in electrodes for batteries, and brushes in motors. Adding carbon black to tyres improves the tensile strength and resistance to wear of the tyres, with smaller particles (around 20 nm) having a greater effect than larger ones.

10.3.2 Graphite

Graphite is, of course, a very familiar substance, with many uses. The lead in lead pencils is graphite, and finely divided forms of graphite are used to absorb gases and solutes. Its absorption properties find a wide range of applications from gas masks to decolouring food. Graphite formed by the pyrolysis of oriented organic polymer fibres is the basis of carbon fibres. Graphite is also used as a support for several industrially important catalysts. Its electronic conductivity is exploited in several industrial electrolysis processes where it is used as an electrode. Crystals of graphite are, however, only good conductors in two dimensions, and it is this two-dimensional aspect that we are concerned with here.

Crystals of graphite contain layers of interlocking hexagons, as shown in Chapter 1 and reproduced in Figure 10.9.

Figure 10.9 shows the most stable form in which the layers are stacked ABAB, with the atoms in every other layer directly above each other. There is another

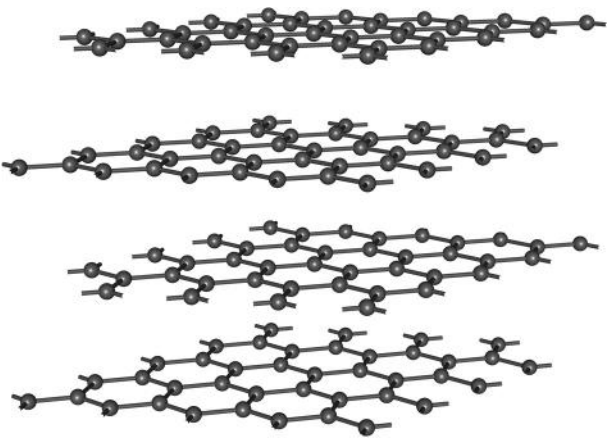


FIGURE 10.9 Hexagonal structure of graphite layers.

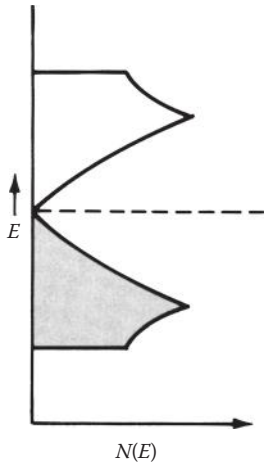


FIGURE 10.10 The band structure of graphite.

form that also contains layers of interlocking hexagons, but the layers are stacked ABCABC.

Each carbon can be thought of as having three single bonds to the neighbouring carbons. This leaves one valence electron per carbon in a p orbital at right angles to the plane of the layer. These p orbitals combine to form delocalised orbitals that extend over the whole layer. If the layer contains n carbon atoms, then n orbitals are formed, and there are n electrons to fit in them. Thus, half of the delocalised orbitals are filled. Were the orbitals to form one band, this would explain the conductivity of graphite very nicely, since there would be a half-filled band confined to the layers. The situation is close to this, but not quite as simple. The delocalised orbitals, in fact, form two bands, one bonding and one antibonding. (This is reminiscent of diamond, where, as shown in Figure 4.8b, the s/p band splits with a gap where nonbonding orbitals would lie.) The lower band is full and the upper band empty. Graphite is a conductor because the band gap is zero, and thus the electrons are readily promoted to the upper band. The band structure for graphite is shown in Figure 10.10. Because the density of states is low at the Fermi level, the conductivity is not as high as that for a typical metal. It can, however, be increased, as we will now see.

10.3.3 Intercalation Compounds of Graphite

Because the bonding between the layers in graphite is weak, it is easy to insert molecules or ions into the spaces between the layers. The solids produced by reversible insertion of such guest molecules into lattices are known as **intercalation compounds**, and although originally applied to layered solids, the term is also being applied to other solids with similar host–guest interactions. Since the 1960s, attention has been paid to the intercalation compounds as of possible importance as catalysts and as electrodes for high-energy-density batteries.

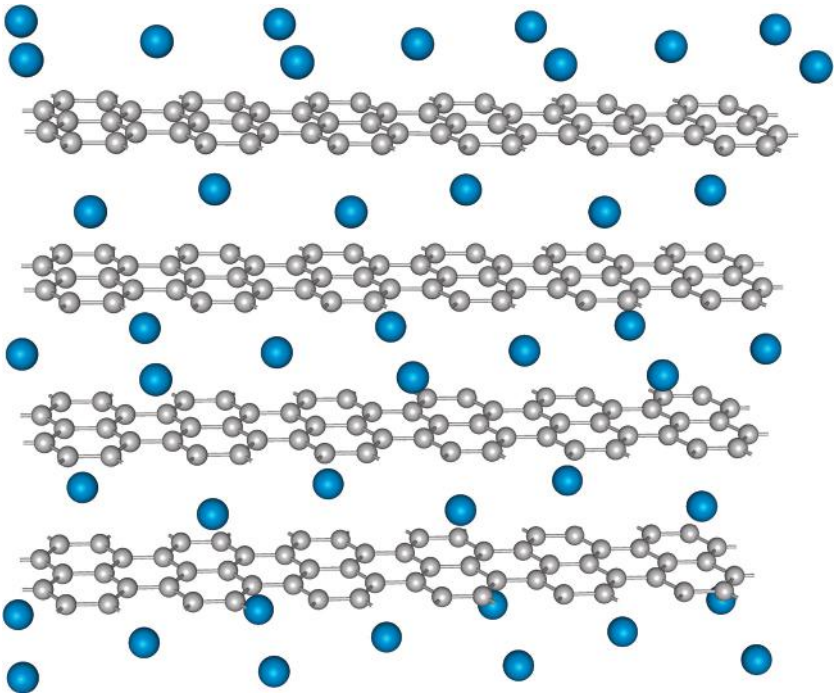


FIGURE 10.11 The structure of KC_8. K, blue; C, grey.

Many layered solids form intercalation compounds, but graphite is particularly interesting because it forms compounds with both the electron donors and the electron acceptors. Among the electron donors, the most extensively studied are the alkali metals. The alkali metals enter the graphite between the layers and produce strongly coloured solids in which the layers of carbon atoms have moved further apart. For example, potassium forms a golden compound (KC_8) in which the interlayer spacing is increased by 200 pm (Figure 10.11). The potassium donates an electron to the graphite (forming K^+) and the conductivity of the graphite now increases because it has a partially full antibonding band.

In 1841, the first intercalation compound was made, containing sulfate, an electron acceptor. Since then, many other electron acceptor intercalation compounds have been made with, for example, NO_3^-, CrO_3, Br_2, $FeCl_3$ and AsF_5. In these compounds, the graphite layers donate electrons to the inserted molecules or ions, thus producing a partially filled bonding band. This increases the conductivity, and some of these compounds have electrical conductivity approaching that of aluminium.

In graphite, the current is carried through the layers by delocalised p electrons. Layered structures in which the current is carried by d electrons are common in transition metal compounds, such as the disulfides of Ti, Zr, Hf, V, Nb, Ta, Mo and W, and mixed lithium metal oxides. At one time, there was considerable interest in the disulfides because some of their intercalates were found to be superconductors at very low

temperatures. It was hoped that by altering the interlayer spacing, a compound that was superconducting at higher temperatures would be found. Unfortunately, it soon transpired that altering the spacing by inserting different molecules had very little effect on the temperature at which superconductivity appeared. It was concluded that the superconductivity was confined to the layers. When high-temperature superconductors were discovered, these were also layer structures, as we saw in Chapter 9.

10.3.4 Graphene

The 2010 Nobel Prize in Physics was awarded to Andre K. Geim and Konstantin S. Novoselov for producing, isolating, analysing and identifying graphene. Graphene consists of a single layer of carbon atoms, essentially a single layer of graphite; the first truly two-dimensional solid (Figure 10.12).

The electronic structure of graphene had been predicted theoretically in 1947, but it had proved difficult to be isolated and identified. In 2004, Geim and Novoselov extracted layers of graphene from graphite using Scotch tape and transferred them to a silicon substrate. The thinnest layers were characterised using atomic force microscopy (AFM). Some of these proved to be the single-layer graphene, and they were able to measure its electrical properties.

Since then, it has become possible to manufacture large sheets of graphene. Graphene is stronger than steel (its breaking strength is 200 times that of steel), but it is very flexible. Its electronic properties are also unusual. It is, for example, a transparent electrical conductor and could replace indium tin oxide (ITO) in touch screens and solar panels. In addition, it has a high thermal conductivity.

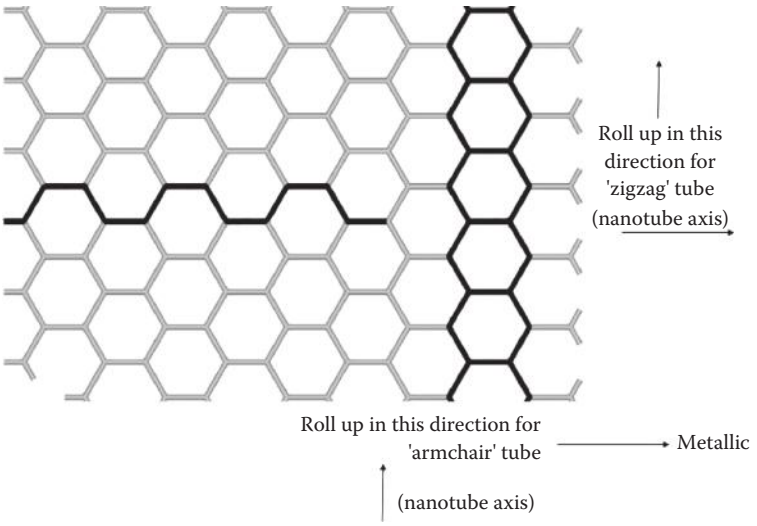


FIGURE 10.12 The structure of graphene. The sheet can be rolled up in the directions of the two arrows, forming distinct types of nanotubes (see Section 10.3.6).

Buckminster fullerene and carbon nanotubes, which we go on to discuss in the next sections, were discovered earlier and are related to graphene in that they can be thought of as being formed by rolling up sheets of graphene.

10.3.5 Buckminster Fullerene

This polymorph of carbon was only discovered in 1985 by Sir Harry Kroto at the University of Sussex while he was looking for carbon chains. It is made by passing an electric arc between two carbon rods in a partial atmosphere of helium. Kroto was awarded the 1996 Nobel Prize in Chemistry along with two American researchers Robert F. Curl Jr. and Richard E. Smalley.

The crystals of buckminster fullerene are not necessarily nanoparticles, but we have included fullerenes here because they are related to carbon nanotubes, which we discuss in the next section.

The molecule has the formula C_{60} and has the same shape as a soccer ball—a truncated icosahedron; it takes its name from the engineer and philosopher Buckminster Fuller, who discovered the architectural principle of the hollow geodesic dome that this molecule resembles (a geodesic dome was built for EXPO 67 in Montreal). The structure is shown in Figure 10.13.

The molecule is extremely symmetrical with every carbon atom in an identical environment; it consists of 12 pentagons of carbon atoms joined to 20 hexagons. The carbon–carbon distances between the adjacent hexagons is 139 pm, and where a hexagon joins a pentagon, the carbon–carbon distance is 143 pm—similar to that found in the graphite layers. Indeed, we can think of this structure as a carbon sheet with graphite-type delocalised bonding, which bends back on itself to form a polyhedron. Other fullerenes are also known, such as C_{70}, and they all have 12 pentagons of carbon atoms linked to different numbers of hexagons. The C_{60}

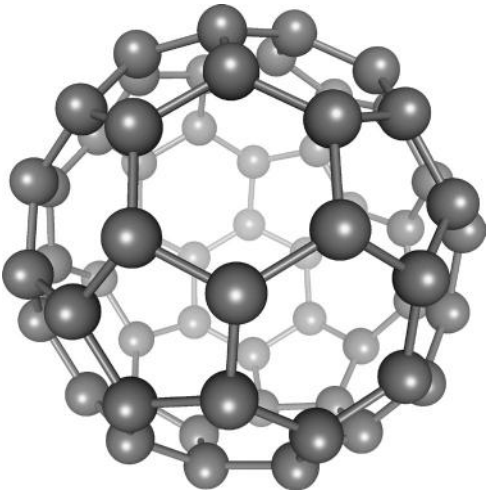


FIGURE 10.13 The structure of buckminsterfullerene (C_{60}).

molecules, commonly called buckyballs, pack together in a cubic close-packed array in the crystals. Fullerene chemistry is currently a very active research area. The use of fullerenes has been suggested for electronic devices and even in targeted cancer therapy. One of the interesting features is the formation of salts with the alkali metals, known as buckides. For instance, potassium buckide (K_3C_{60}) has a *ccp* array of buckyballs, with all the octahedral holes and tetrahedral holes filled by potassium atoms: this is a metallic substance, which becomes superconducting below 18 K. Other alkali metal buckides become superconducting at even higher temperatures.

10.3.6 Carbon Nanotubes

In 1991, Sumio Iijima discovered nanotubes. Similar in structure to buckyballs, they consist of sheets of graphene that roll up to form cylinders. Initially multilayered (multi-walled nanotubes; MWNTs), they can now be made single-walled. Single-walled nanotubes (SWNTs) (Figure 10.14) have diameters of about 1 nm and lengths of the order of 10^{-6} m.

The tubes are capped at each end by half of a fullerene-type structure. They have remarkable properties and are being used for a variety of high-performance applications. They can adsorb 100 times their volume of hydrogen, so they could be developed as a safe storage medium for hydrogen for fuel cells (Chapter 5). They have a remarkable tensile strength, 100 times greater than steel, but are very light, with half the density of aluminium, and are being used in high-performance sports equipment such as tennis rackets. Although strong, they are also very flexible. Using AFM, they can be bent into a variety of shapes, including springs. They can also be made as conductors and semiconductors, and one use has been found in car bumpers, where they not only provide strength but also prevent the build-up of static electricity.

The electronic properties of SWNTs depend on the direction in which the graphene sheet rolls up. In the structure of graphene shown in Figure 10.12, notice that the line joining the rows of hexagons in the vertical direction is a simple zigzag, but

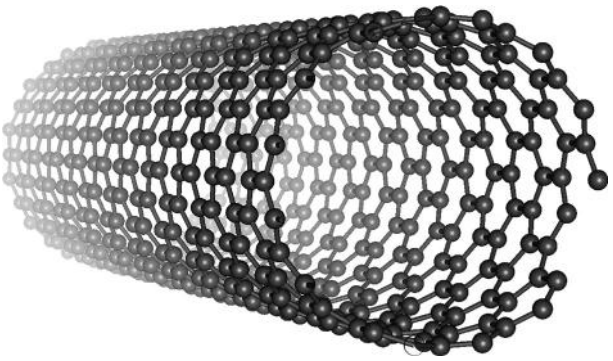


FIGURE 10.14 A computer simulation of a single-walled carbon nanotube.

those at right angles to this, the so-called 'armchair lines', join the rows. Tubes that roll up along the armchair direction—the so-called *armchair* nanotubes—always exhibit metallic levels of conductivity. If graphite sheets roll up along the zigzag lines, with the armchairs along the axis of the tube (*zigzag* nanotubes) or if the sheets roll up along any other direction except the zigzag or the armchair lines, forming *helical* nanotubes, a band gap is introduced, and the tubes can be semiconducting. It is this property that has allowed the formation of field-effect transistors (FETs), single electron transistors, and diodes from carbon nanotubes.

One chemically interesting use of nanotubes is as 'test-tubes'. Solids can be prepared inside nanotubes. Because of the small dimensions of the tube, unusual crystalline structures can be obtained.

10.4 CARBON-BASED CONDUCTING POLYMERS

10.4.1 Discovery of Polyacetylene

The extension of the ideas of delocalised electrons in conjugated organic molecules led to a suggestion that a conjugated polymer might be an electrical conductor or semiconductor. The initial attempts to make polyacetylene and similar solids, however, resulted in short-chain molecules or amorphous, unmeltable powders. Then, in 1961, Hatano and co-workers in Tokyo managed to produce thin films of polyacetylene, by polymerising ethyne (acetylene). Polyacetylene exists in two forms, *cis* and *trans*, as shown in Figure 10.15, of which the *trans* form is the more stable. Ten years after Hatano's work, Shirakawa and Ikeda made films of *cis*-polyacetylene, which could be converted into the *trans* form. They achieved this by directing a stream of ethyne gas onto the surface of a Ziegler–Natta catalyst (a mixture of triethyl

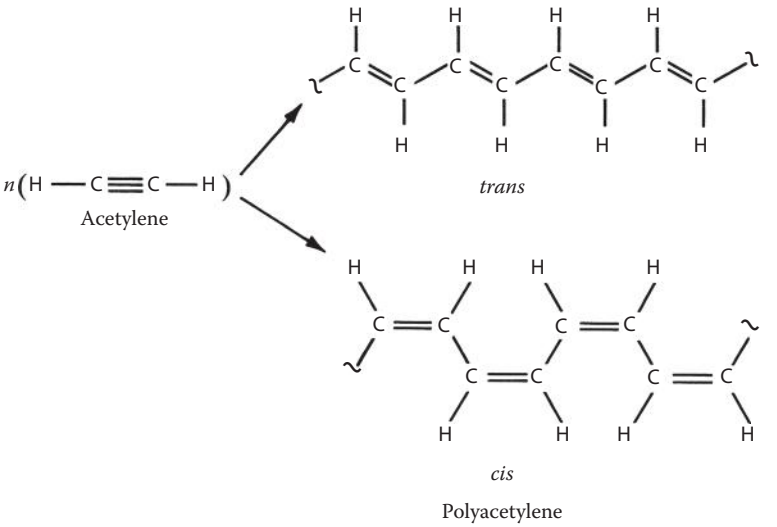


FIGURE 10.15 *cis*- and *trans*-Polyacetylene.

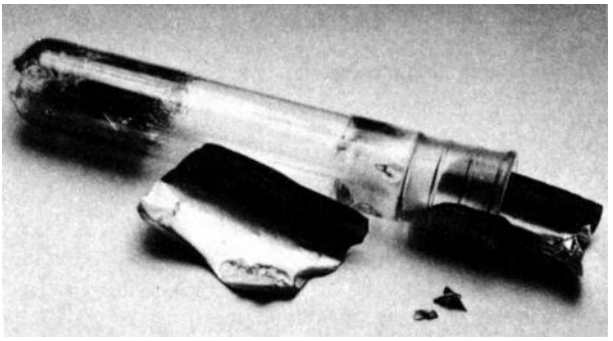


FIGURE 10.16 A film of polyacetylene forms on the inner surface of the reaction vessel, after ethyne gas passes over the catalyst solution of the walls. The paper-thin flexible sheet of polyacetylene is then stripped from the walls prior to doping. (Photograph by James Kilkelly. Originally published in *Scientific American*.)

aluminium and titanium tetrabutoxide). To make a large film, the catalyst solution can be spread in a thin layer over the walls of a reaction vessel (Figure 10.16) and then ethyne gas is allowed to enter. The polyacetylene produced in this way has a smooth shiny surface on one side and is a sponge-like structure. It can be converted to the thermodynamically more stable *trans* form by heating. The conversion is quite rapid above 370 K. After conversion, the smooth side of the film is silvery in appearance, becoming blue when the film is very thin.

A way of improving the conductivity was found when Shirakawa visited McDiarmid and Heeger in Pennsylvania later in the 1970s. This led to these three scientists being awarded the Nobel prize for chemistry in 2000. The Americans had been working on smaller conjugated molecules to which they added an electron acceptor in order to make them conducting. It was a natural step to try this approach with polyacetylene. If an electron acceptor such as bromine is added to polyacetylene forming $[(CH)^{\delta+}\ Br^{-}_{\delta}]_n$, its conductivity is greater than that of the undoped material. Other examples of dopants that can oxidise polyacetylene are I_2, AsF_5 and $HClO_4^-$. The effect of these dopants is shown in Figure 10.17, where it can be seen that the conductivity rises from 10^{-3} S m^{-1} to as much as 10^5 S m^{-1} using only small quantities of dopant.

The conductivity of polyacetylene is also increased by dopants that are electron donors. For example, the polymer can be doped with alkali metals to give, for example, $[Li^{+}_{\delta}(CH)^{\delta-}]_n$. The wide range of conductivities produced by these two forms of doping is illustrated in Figure 10.17.

Polyacetylene is very susceptible to attack by oxygen. The polymer loses its metallic lustre and becomes brittle when exposed to air. However, other conjugated polymers were found. Polypyrrole, polythiophene, polyaniline, polyphenylvinylene and others (Figure 10.18) are conjugated polymers whose bonding and conductivity are similar to those of polyacetylene. These polymers are, however, less sensitive to oxygen and by attaching suitable side chains, can be made soluble in nonpolar organic solvents and are therefore easier to process.

As for polyacetylene, the conductivity of these polymers is sensitive to doping. This is exploited in polypyrrole gas sensors, which are based on the variation of

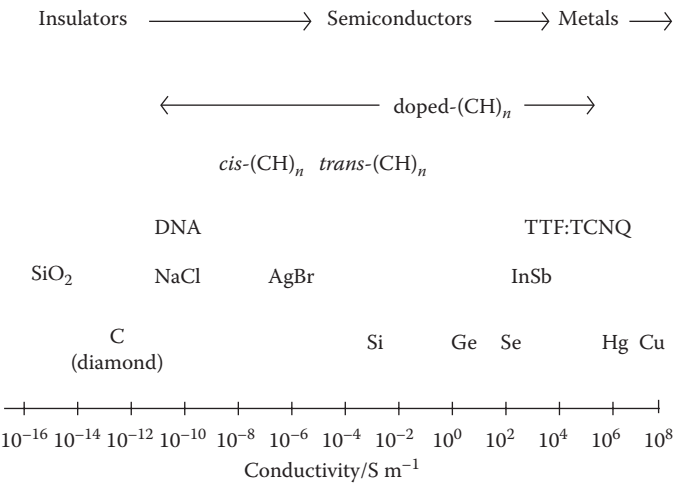


FIGURE 10.17 The conductivites of undoped and doped polyacetylene $(CH)_n$ compared with the values for some of the better-known insulators, semiconductors and metals.

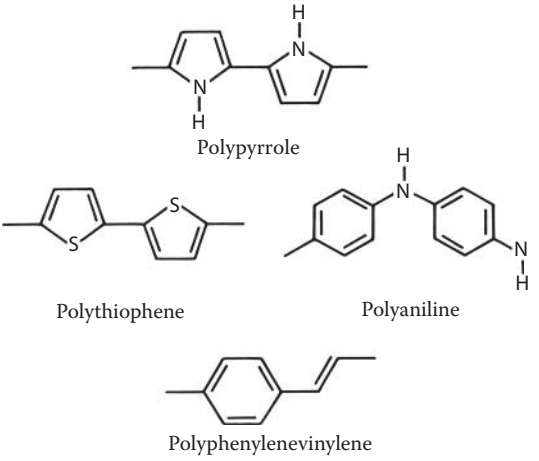


FIGURE 10.18 Repeating units of some conducting polymers.

conductivity of a thin polymer film when exposed to gases such as NH_3 and H_2S. Doped conducting polymers can also be used as a metallic contact in organic electronic devices. Promising commercial applications of conjugated polymers include light-emitting diodes (LEDs), transistors, photovoltaic cells and diodes.

10.4.2 Bonding in Polyacetylene and Related Polymers

In small conjugated alkenes such as butadiene with alternate double and single bonds, the π electrons are delocalised over the molecule. If we take a very long conjugated

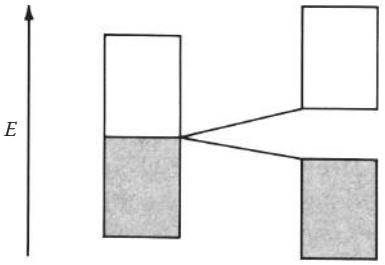


FIGURE 10.19 The band gap in polyacetylene produced by the alternation of long and short bonds.

olefin, we might expect to obtain a band of π levels, and if this band were partly occupied, we would expect to have a one-dimensional conductor. Polyacetylene is just such a conjugated long-chain polymer. Now, if polyacetylene consisted of a regular, evenly spaced chain of carbon atoms, the highest-occupied energy band, the π band, would be half full and polyacetylene would be an electrical conductor. In practice, polyacetylene shows only modest electrical conductivity, comparable with semiconductors such as silicon: the *cis* form has a conductivity of the order of 10^{-7} S m^{-1} and the *trans* form, 10^{-3} S m^{-1}. The crystal structure is difficult to determine accurately, but diffraction measurements indicate that there is an alternation in the bond lengths of about 6 pm. This is much less than would be expected for truly alternating single and double bonds (C–C, 154 pm in ethane; C=C, 134 pm in ethene). Nonetheless, this does indicate that the electrons tend to localise in double bonds rather than be equally distributed over the whole chain.

What, in fact, is happening is that two bands, bonding and antibonding, form with a band gap where nonbonding levels would be expected. There are just enough electrons to fill the lower band. As shown in Figure 10.19, this leads to a lower energy than the half-full single band.

This splitting of the band is an example of **Peierls' theorem**, which asserts that a one-dimensional metal is always electronically unstable with respect to a non-metallic state—there is always some way of opening an energy gap and creating a semiconductor. Thus, although a very simple bonding picture of such solids would suggest a half-filled band and metallic conductivity, the best that can be expected is a semiconducting polymer.

If an electron acceptor is added, it takes the electrons from the lower π bonding band. The doped polyacetylene now has holes in its valence band and, like *p*-type semiconductors, has a higher conductivity than the undoped material. Electron donor dopants add electrons to the upper π band, making this partly full, thereby producing an *n*-type semiconductor.

10.4.3 Organic LEDs and Photovoltaic Cells

At the end of 2007, Sony released the first organic LED screen television. The Sony XEL-1 had a screen size of only 11 inches, but was notable for being only 3 mm

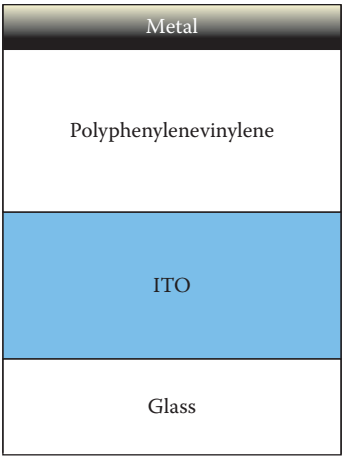


FIGURE 10.20 A section through a polymer LED.

thick. The XEL-1 used small molecule LEDs. Polymer LEDs are being developed for flat panel displays in mobile phones, laptops and televisions. A typical polymer LED is built up on a glass substrate. At the bottom is a layer of a semitransparent metallic conductor, usually ITO, which acts as an electrode (Figure 10.20). On top of this is a layer of an undoped conjugated polymer, such as polyphenylenevinylene, and on top of that is an easily ionised metal, such as Ca, Mg/Ag or Al. The metal forms the second electrode. To act as an LED, a voltage is applied across the two electrodes such that the ITO layer is positively charged. At the ITO/polymer interface, the electrons from the polymer move into the ITO layer, attracted by the positive charge. This leaves gaps in the lower energy band. At the same time, the electrons from the negatively charged second electrode move into the polymer and travel towards the ITO layer. When such an electron reaches a region of the polymer where there are vacancies in the lower energy band, it can jump down to this band, emitting light as it does so. The wavelength of the light, as in semiconductor LEDs, depends on the band gap. Polyphenylenevinylene LEDs emit green/yellow light, polyphenylene-based LEDs emit blue light and some polythiophenes emit red light.

For photovoltaic cells, small band-gap conjugated polymers can be mixed with C_{60} or C_{70}. Light excites the electrons across the band gap of the polymer. The excited electrons then transfer to the fullerenes, thereby preventing recombination in the polymer. The fullerenes then donate electrons to one electrode, while the other electrode passes electrons to the valence band of the polymer.

10.4.4 Polymers and Ionic Conduction: Rechargeable Lithium Batteries

In the 1970s, Whittingham developed a battery that operated at room temperature based on the intercalation of Li in TiS_2. In the **lithium/titanium disulfide battery**, one electrode is lithium metal and the other electrode is titanium disulfide bonded

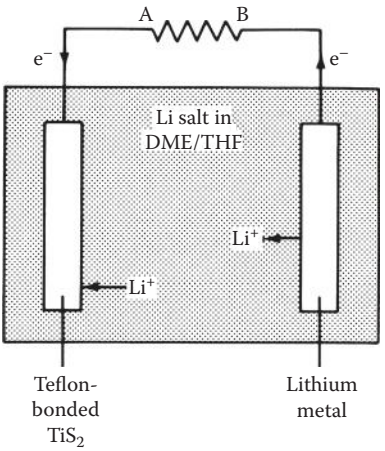


FIGURE 10.21 The Li – TiS_2 battery during a discharge phase.

to a polymer such as Teflon. The electrolyte is a lithium salt dissolved in an organic solvent. Typically, the solvent is a mixture of dimethoxyethane (DME) and tetrahydrofuran (THF). The setup is shown in Figure 10.21.

When the circuit is complete, lithium metal from the lithium electrode dissolves giving solvated ions, and the solvated ions in the solution are deposited in the titanium disulfide. These ions intercalate into the disulfide, and the charge is balanced by the electrons from the external circuit. Thus, the two electrode reactions are

$$\mathrm{Li(s) = Li^+(solv) + e^-} \tag{10.1}$$

and

$$x\mathrm{Li^+(solv) + TiS_2(s) + }x\mathrm{e^- = Li}_x\mathrm{TiS_2(s)}, \tag{10.2}$$

giving an overall reaction

$$x\mathrm{Li(s) + TiS_2(s) = Li}_x\mathrm{TiS_2(s)}. \tag{10.3}$$

The intercalation compound formed has layers of a variable number of lithium ions between the sulfide layers, as shown in Figure 10.22.

When the titanium disulfide has completely discharged, the battery can be recharged by applying a voltage across AB. A voltage high enough to overcome the free energy of the cell reaction will cause the lithium ions to return to the solvent and the lithium to be deposited on the metal electrode, thus restoring the battery to its original condition. The usefulness of the lithium/titanium disulfide electrode

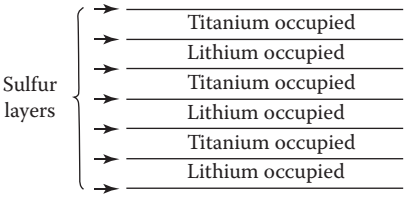
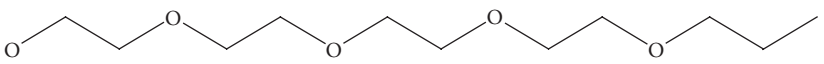


FIGURE 10.22 The occupation pattern of octahedral holes between close-packed layers of sulfur atoms in $LiTiS_2$.

depends on the electrode material being a good conductor and on the electrode reaction being reversible. Li_xTiS_2 is a particularly good electrical conductor, because not only is the electronic conductivity increased by the donation of the electrons to the conduction band, but also the lithium ions themselves act as current carriers, so the compound is both an electronic and an ionic conductor. Compounds other than TiS_2 have proved more useful for commercial batteries (ionic conductors and lithium batteries are described in Chapter 5).

Lithium polymer batteries are similar in principle to the lithium batteries previously described, but the electrolyte is a polymer. The advantage of these batteries is the absence of liquid in the cell, so the batteries do not leak. The polymer electrolyte is a polymer–alkali metal salt complex. The best-known such electrolytes are complexes of poly(ethylene)oxide (PEO):



Poly(propylene)oxide, poly(ethylene)succinate, poly(epichlorohydrin) and poly(ethylene imine) have also been investigated as possible polymer bases for polymer electrolytes. These are all polymers containing electronegative elements, oxygen or nitrogen, in the chain, which will coordinate to metal cations. The electrical current is carried by the ions of the salts moving through the polymer. The conductivity depends on the polymer molecular weight and the polymer/salt ratio, and experiments suggest that it is the amorphous part of the polymer that produces high ionic conductivity. Some values for Li^+ conductivity in various polymer–lithium salt complexes are given in Table 10.2. A conductivity of the order of 10^{-5} S m^{-1} is suitable for batteries. In commercial batteries, the conductivity is improved by combining the polymer electrolyte with an electrolyte gel.

In lithium-ion polymer batteries, one electrode is lithium foil or, in some cases, another electrically conducting material such as graphite, and the other electrode is a reversible intercalation compound, as in liquid electrolyte lithium batteries. The compounds used as intercalation electrodes include $LiCoO_2$ and V_6O_{13}.

Lithium-ion polymer batteries became available in 2000 and are used in, for example, mobile phones and notebook computers.

TABLE 10.2
Conductivities of Polymer–Metal Salt Complexes

Polymer	Molecular Mass	Salt	Conductivity/S m^{-1}	Temperature Range/K
PEO	400	$LiBF_4$	3.1×10^{-4}–3.1×10^{-5}	298–450
PEO	5×10^6	$LiBF_4$	10^{-5}–10^{-7}	298–450
PEO	4×10^6	CH_3COOLi (O/Li = 4)	3.16×10^{-4}–10^{-9}	298–420
PEO	4×10^6	CH_3COOLi (O/Li = 9)	10^{-4}–10^{-9}	298–420
PEO	4×10^6	CH_3COOLi (O/Li = 18)	10^{-4}–10^{-8}	194–420
PEO (cross-linked)	3000	$LiClO_4$	10^{-2}–10^{-9}	253–273
PEO (linear)	3000	$LiClO_4$	10^{-8}–10^{-11}	253–273

10.5 NONCARBON NANOPARTICLES

The large surface area of nanoparticles lends increasing dominance to the behaviour of the atoms on the surface of the particles; in catalysis, this is exploited to improve the rate of production in commercial processes and in the structure of electrodes, to improve the performance of batteries and fuel cells (see Sections 5.5.4, 7.5 and 7.6). The interactions between these surface atoms and a surrounding matrix determine the properties of high-performance nanocomposites. Because the dimensions of the nanoparticles are less than the wavelength of visible light (400–700 nm), visible light is no longer scattered by them, rendering them transparent; this is a useful property for the manufacture of cosmetics and coatings.

Nanoparticles have been known and used for centuries; think, for instance, of the pigments used to colour stained glass and ceramic glazes and of the colloidal gold particles used for making 'ruby' glass, known since Roman times. Within the modern era, fumed silica—small particles of silica—has been added to solids and liquids to improve their flow properties since the 1940s.

At the simplest level, nanoparticles of hard substances are useful as polishing powders which are able to give very smooth, defect-free surfaces. Indeed, 50 nm nanoparticles of cobalt tungsten carbide are found to be much harder than the bulk material and therefore can be used to make cutting and drilling tools that will last longer.

10.5.1 Fumed Silica

This very pure form of silica is made by reacting $SiCl_4$ with an oxyhydrogen flame. The resulting SiO_2 particles, with dimensions 7–50 nm, have an amorphous structure. The particles have silanol (Si–OH) groups on their surface, and these hold the particles together by hydrogen bonding, forming chain-like structures. When added to liquids, this three-dimensional network traps the liquid and

increases the viscosity, but when the thickened liquid is subsequently brushed out or sprayed, the liquid and any trapped air are released. When the shear force is removed, the liquid thickens again. This property is known as thixotropy and is very useful in paints for preventing the settling of pigments and for improving the flow properties of paints, coatings and resins. It is also used to improve the flow properties of powdered solids, such as pharmaceuticals, cosmetics, cement, inks and abrasives, helping to prevent caking.

10.5.2 Quantum Dots

A quantum dot or **nanocrystal** is defined as a crystal of a semiconductor which is a few nanometres in diameter, typically containing only 10^2–10^4 atoms. As we saw in Section 10.2, quantum dots exhibit quantum size effects in their physical properties, having interesting electronic, magnetic and optical properties that are a consequence of their size rather than their chemical composition. For instance, a cadmium selenide (CdSe) dot of diameter 2.3 nm gives out a turquoise fluorescent light when irradiated with UV light, but a CdSe dot of 5.5 nm glows orange (Figure 10.6).

There are considerable practical difficulties in making quantum dots of a regular size and shape and also in preventing the dots, once made, from coalescing into a larger crystal. CdSe quantum dots are currently made by dissolving selenium and dimethylcadmium ($(CH_3)_2Cd$) in hot trioctylphosphine oxide (TOPO), from where large quantities of the CdSe quantum dots precipitate. The size of the dots is controlled by the time they remain in the hot liquid. The shape of the dots is improved by the growth of a thin layer of zinc sulfide (ZnS) on the surface, and they are then coated in TOPO to make them hydrophobic and prevent them from coalescing. Now we must ask how the unusual properties of quantum dots are being exploited.

In Chapter 7, we saw how in LEDs, a voltage is applied across the *p*–*n* junction of a semiconductor, causing the electrons to move across the junction from the *n*-type side and drop into vacancies in the valence band on the *p*-type side, emitting a photon of light. As we have seen, quantum dots of different sizes emit light of different wavelengths and if made reliably to a particular size, can be used to make LEDs of a very pure colour. This purity of colour is what is needed for use as pixels in colour displays.

The purity of the colour produced by quantum dots is also leading to research into their use in special dyes and polymers, which could be used, for instance, to combat counterfeiting, as they will be extremely difficult to reproduce.

The light-emitting properties of quantum dots mean that they can be used as fluorescent probes in biological systems, where they can have many advantages in replacing conventional organic fluorescent probes. Quantum dots have been used both in vivo and in vitro; if they can be attached to a biological molecule of interest, such as an antibody or a protein, they can be used to follow the reactions of that molecule. They have the advantages of being stable, can be excited by broadband excitation, but emit a narrow band of frequencies of high intensity, and are available in many colours; because the colour emitted by each size of dot is pure, several different colour dots can be used at the same time to track different processes.

In an area of interest that we have already heard about (Section 6.8), quantum dots of CdS have been grown inside zeolite cages, thus controlling their size. This gives the prospect of forming new semiconducting materials useful for data storage. A similar idea under investigation is the growth of quantum dots inside silica diatom skeletons where the thousands of different species could provide the templates for complex arrays of dots.

Research is also looking at the possibility of a quantum dot laser that would allow rapid light pulses from a minute source to further increase the speed of circuitry and communications. Such lasers should be tunable, as the frequency of the light emitted will depend on the size of the quantum dot.

10.5.3 Explosives

Aluminium particles are made with diameters in the range of 20–200 nm with a protective shell of aluminium oxide of about 4 nm thick, and when mixed with a suitable oxidising agent, are used for rocket propulsion fuels.

Aluminium powders are also used for the thermite reaction (Figure 10.23); this is the reaction traditionally used to weld rail tracks together in situ:

$$2Al(s) + Fe_2O_3(s) = 2Fe(1) + Al_2O_3(s)$$

as it produces so much heat that molten iron is produced, but it can also be used to produce other less easily reduced metals, such as Mo, from their oxides. Using nanometre particles dramatically improves performance, and these reactions are also used as priming reactions for munitions and in fireworks.

FIGURE 10.23 The thermite reaction. (From Adrian Guy, Blundell's School, Devon. With permission.)

10.5.4 Magnetic Nanoparticles

As discussed in Section 10.2.4, nanoparticles of iron oxides (Fe_2O_3 or Fe_3O_4) are superparamagnetic, having zero remanence, that is, the magnetism induced by a magnetic field disappears when the field is removed. This property can be used as a simple switch to separate the components of chemical or biological solutions. The particles are coated with a functional group that will attach to the required component in the solution. If a magnetic field is then applied to one side of the container, the particles become magnetised and drift towards the magnet, taking the attached component with them.

Iron oxide nanoparticles are used to enhance magnetic resonance images (MRI scans) acting as contrast agents. Ferromagnetic Fe or Ni nanoparticles can also be used for this purpose, with the advantages of a larger magnetic moment and therefore a larger signal; unfortunately, they are much more difficult to produce.

There are other promising medical uses for these magnetic nanoparticles. If they can be functionalised appropriately, say, with an antibody, then they could also be injected into the bloodstream and used to seek out markers of disease or infection such as cancer sites; this would then be detectable by MRI. If a drug molecule could be attached to the surface of a magnetic nanoparticle, then a magnetic field could be used to draw the drug towards the site of the infection, thereby targeting the drug more efficiently.

The superparamagnetic properties of iron oxide nanoparticles when suspended in hydrocarbons are used in **ferrofluids**; these were invented by NASA as a way to control liquid rocket fuels in space, as the flow can be controlled by magnetic fields. With the advent of solid rocket fuels, they are now used mainly as bearing seals in the hard drives of computers and as dampers in speakers. A magnetic fluid containing nanoparticles of iron nitride (Fe_3N) in kerosene has been produced by reacting iron carbonyl ($Fe(CO)_5$) and ammonia. Similarly, thermal or sonochemical decomposition of iron carbonyl in decalin also produces a ferrofluid containing nanoparticles of α-Fe.

Cobalt clusters embedded in silver, display giant magnetoresistance (GMR) with an increase in resistance of up to 20%, and these are used for magnetic recording and data storage.

10.5.5 Medical and Cosmetic Use

Silver nanoparticles have been found to have very good antibacterial action and are being used to impregnate bandages (Figure 10.24); they are also used to impregnate socks and are added to underarm deodorants, as the antibacterial action kills the bacteria responsible for the unpleasant smells.

TiO_2 (and ZnO) has been used in sunscreens for many years because of its ability to absorb the UV radiation that is harmful to the skin. But TiO_2 also has a very high refractive index and therefore scatters light very efficiently; so efficiently, in fact, that it makes a very good white pigment and consequently, it is used in most white paint because of its high covering power. For many years, the use of TiO_2 in sunscreens and sunblocks with the highest factor numbers has made them appear white on the skin and not very attractive. Nanoparticles of TiO_2 of about 50 nm diameter

FIGURE 10.24 Silver-impregnated plasters. (From Elastoplast.)

are now being manufactured. They are transparent as they are too small to scatter the visible light, but they still absorb the harmful shorter-wavelength UV radiation, and therefore are now being used in sunscreens to get around this problem. Larger nanoparticles of TiO_2 are being used to impregnate fabrics, to make clothing with a built-in sun protection.

Nanoparticles of zirconium oxide (ZrO_2) are being used in UV-cured dental fillings. They give strength and are transparent to visible light, but are opaque to X-rays. Rare-earth oxides can also be used.

10.6 NONCARBON NANOFILMS AND NANOLAYERS

Nanofilms can be used simply as very thin coatings. For example, nanoscale coatings to protect and enhance modern plastic spectacle lenses have been developed, including self-assembling top coatings for nonreflective lenses to protect the antireflective layer from dirt, dust and skin oils, and a super-hard coating of carbides to protect from scratching.

A spray-on coating containing nanometre-sized particles of TiO_2 combined with a photocatalyst can be used on surfaces such as glass to give a dirt-repellent surface: the photocatalyst uses UV light to break down organic debris, but also the coating stops rainwater from forming droplets, so that it simply runs off, collecting dirt as it goes (Figure 10.25). (Similarly, nanoparticles are being used to impregnate fabrics, making them stain-resistant.)

The development of electronics using organic compounds has led to nanofilm electronic devices. Such films can be used in flat-screen displays for computer monitors. There are two types of such displays, thin-film transistor liquid-crystal

FIGURE 10.25 The difference in performance of ordinary glass and self-cleaning glass. (From Saint Gobain Glass Group.)

displays (TFT-LCDs) and organic LEDs (OLEDs). In TFT-LCDs, one transistor is connected to each of the three LCD elements (red, green and blue) for each pixel. By making the bank of transistors a film of organic material 20 nm thick, rather than a 2 mm thick layer of silicon, thinner, lighter and more flexible screens can be produced. In OLEDs, the colours are produced by an array of LEDs in which the light-emitting layer is a nanofilm of organic polymer, as we described earlier in this chapter.

The flexibility of thin-film organic electronic devices means that they can be printed onto paper and other materials. An alternative approach is to develop single-fibre transistors that could then be woven into cloth. The incorporation of electronics into cloth has inspired some interesting proposals for their use. One suggestion is that military uniforms could be printed with organic films containing electronic circuits that could be used to send commands to the soldiers in the field. Another is for LEDs to be printed onto curtains to provide light when the curtains are closed. One problem to be overcome is that the organic materials are moisture-sensitive, so the cloth tends to attract moisture.

The electronic and magnetic properties of nanolayers are important in devices formed from more conventional electronic materials. We have already discussed quantum-well lasers (Chapter 7) and colossal magnetoresistance (CMR) devices that are used for hard-disk read heads (Chapter 8). Quantum-well lasers may be an important component of light-based computers. Other possibilities include magnets with unusual properties.

10.7 NONCARBON NANOTUBES, NANORODS AND NANOWIRES

MWNTs have now been synthesised from inorganic compounds, and some, such as WS_2 nanotubes, are routinely synthesised in bulk quantities. Nanotubes are often free of bulk defects such as dislocations and grain boundaries, and this is responsible for their efficiency in potential applications.

TiO_2 nanotubes have been investigated because of their potential application in dye-sensitised solar cells. The band gap of TiO_2 nanotubes is close to that of nanosheets, but is higher than that of bulk anatase. TiO_2 nanotubes are more efficient at transporting charge than nanoparticles, and this is attributed to a reduction of the resistance due to the grain boundaries. WS_2 and MoS_2 nanotubes have potential applications in composite materials of ultrahigh strength. Observation of their behaviour under strain and stress shows that the WS_2 tubes are very flexible and exhibit elastic behaviour almost to failure point. At the point of failure, it is thought that a bond at the centre of the tube breaks, and this initiates a succession of bond breakages. This indicates that failure due to macroscopic defects, such as dislocations, is absent and that the tubes are free of such defects.

To prevent nanotubes clumping or to make them soluble, groups have been attached to the surface of the nanotubes. For example, WS_2 nanotubes have been made water-soluble by interaction with Ni^{2+} coordinated to 2,2′,2″-nitrilotriacetic acid ($N(CH_2COOH)_3$).

Nanowires have been prepared from many materials, for instance, from semiconductors such as Si and Ge, metals, and oxides particularly ZnO. In Section 10.2, we described the electrical properties of nanowires and we saw that conductance along such wires as a function of voltage could be stepped. In nanowires of less than 100 nm diameter, scattering at the surface is the dominant mechanism for electrical resistance.

Stepped magnetisation hysteresis loops have been observed for iron nanowires. As for nanorods, the mechanical properties of nanowires indicate that they are free of bulk defects.

Many applications for nanowires have been proposed. Boron-doped and phosphorus-doped silicon nanowires have been used to build rectifiers and transistors. The rectifier consists of crossed *p*- and *n*-type nanowires, and the transistors consist of heavily doped and lightly doped *n*-type wires crossing a common *p*-type wire. There is interest in using Si nanowires for solar photovoltaic cells. One advantage of these nanowires is that the reflectance of light from arrays of Si nanowires is much less than that for polished Si wafers. Nickel nanowires have high coercivity and are potentially useful in magnetic recording. Laser action has been observed for ZnO nanowires pumped with radiation from a Nd:YAG laser. The output from the Nd:YAG laser was frequency tripled to give a radiation of 355 nm wavelength. The ZnO laser radiation appears in the wavelength region of 385–390 nm.

QUESTIONS

1. Magnetotactic bacteria build rows of nanoparticles of magnetite that consist of a single domain which they use to help them locate the bottom of muddy ponds. What advantage might forming particles of this size have for the bacteria?
2. Which orbitals would you expect to combine to form a delocalised band in polyphenylenevinylene?
3. Would doping of polyacetylene with (a) Rb or (b) H_2SO_4 give an *n*-type or a *p*-type conductor?

4. When polyacetylene is doped with chloric(VII) acid ($HClO_4$), part of the acid is used to oxidise the polyacetylene and part is used to provide a counter anion. The oxidation reaction for the acid is given in Equation 10.4. Write a balanced equation for the overall reaction.

$$ClO_4^- + 8H^+ + 8e^- = Cl^- + 4H_2O \tag{10.4}$$

5. Undoped polythiophene is red. What does this suggest about the size of the conduction band/valence band energy gap in this polymer? Would you expect a polythiophene LED to emit a red light?
6. Sketch the band structure of graphite intercalated with (a) an electron donor and (b) an electron acceptor.
7. Fullerenes consist of 12 pentagons and a number of hexagons; 20 in the case of C_{60} and 25 in the case of C_{70}. What would be the formula of the smallest possible fullerene?

Further Reading

This is a fast-moving field, and to keep up with current developments the reader is referred to review magazines and journals such as ***Chemistry World*** (http://www.rsc.org/chemistryworld/), ***Education in Chemistry*** (http://www.rsc.org/Education/EiC/), ***New Scientist, Scientific American, Angewandte Chemie, Science, Physics Today, Materials Today*** and ***Physics World*** (http://physicsweb.org).

BIBLIOGRAPHY

General

Bruce, D.W., O'Hare, D. and Walton, R.I. (2010) *Molecular Materials*, John Wiley-Blackwell, Chichester.

Cox, P.A. (2010) *Transition Metal Oxides: An Introduction to Their Electronic Structure and Properties*, Oxford University Press, New York.

Ropp, R.C. (2003) *Solid State Chemistry*, Elsevier Science, Amsterdam.

Tilley, R. (2004) *Understanding Solids*, Wiley, Chichester.

West, A.R. (1985) *Solid State Chemistry and Its Applications*, Wiley, New York.

West, A.R. (1999) *Basic Solid State Chemistry*, 2nd edn, John Wiley and Sons Ltd, Chichester.

CHAPTER 1

Cotton, F.A. (1990) *Chemical Applications of Group Theory*, 3rd edn, Wiley, New York.

Hahn, Th. (Ed) (2005) *International Tables for Crystallography*, vol. A, 5th edn (print); and (2006) 1st edn online as a pdf, International Union of Crystallography (IUCr), Chester.

Li, W.-K., Zhou, G.-D. and Mak, T. (2008) *Advanced Structural Inorganic Chemistry*, International Union of Crystallography (IUCr), Chester.

CHAPTER 2

Aberration correction past and present (2009) *Philosophical Transactions of the Royal Society A*, **367**.

Bonnell, D.A. (2001) *Scanning Probe Microscopy and Spectroscopy: Theory, Techniques and Applications*, 2nd edn, Wiley-VCH, New York.

Chen, C. J. (2007) *Introduction to Scanning Tunnelling Microscopy*, 2 edn, Oxford University Press, Oxford.

De Graef, M. and McHenry, M.E. (2007) *Structure of Materials: An Introduction to Crystallography, Diffraction and Symmetry*, Cambridge University Press, New York.

Dinnebier, R.E. and Billinge, S.J.L. (2008) *Powder Diffraction: Theory and Practice,* Royal Society of Chemistry, Cambridge.

Duer, M.J. (2004) *Introduction to Solid-State NMR Spectroscopy*, Blackwell, Oxford.

Fultz, B. and Howe, J.M. (2008) *Transmission Electron Microscopy and Diffractometry of Materials*, 3rd edn, Springer.

Imaging at the picoscale (2004) *Materials Today*, **7**(12), 42–48.

Ladd, M. F. C. and Palmer, R. A. (2003) *Structure Determination by X-Ray Crystallography*, 4th edn, Plenum, New York.
Laing, M. (2001) *An Introduction to the Scope, Potential and Applications of X-Ray Analysis*, International Union of Crystallography (free download from their website).
Peng, L.M., Dudarev, S.L. and Whelan, M.J. (2011) *High Energy Electron Diffraction and Microscopy*, Oxford University Press, Oxford.
Special Issue on Microscopy (2008) *Materials Today*, **11**, Elsevier, UK.

CHAPTER 3

Altavilla, C. and Ciliberto, E. (Eds) (2010) *Inorganic Nanoparticles: Synthesis, Applications, and Perspectives*, CRC Press, Boca Raton, FL.
Lalena, J.N., Cleary, D.A, Carpenter, E. and Dean, N.F. (2008) *Inorganic Materials Synthesis and Fabrication*, Wiley-Blackwell, New York.
Patil, K.C., Hegde, M.S., Rattan, T. and Aruna, S.T. (2008) *Chemistry of Nanocrystalline Oxide Materials: Combustion Synthesis, Properties and Applications*, World Scientific Publishing, Singapore.
Rao, C.N.R. (2000) *Chemical Approaches to the Synthesis of Inorganic Materials*, John Wiley and Sons Inc, New York.
Tompsett, G.A., Conner, W.C. and K. Yngvesson, S. (2006) Microwave synthesis of nanoporous materials, *Chem Phys Chem*, **7**, 296–319.
Wright, J.D. and Sommerdijk, N.A.J.M. (2000) *Sol–Gel Materials: Chemistry and* Applications, CRC Press, Boca Raton, RL.

CHAPTER 4

Kittel, C. (1996) *Introduction to Solid State Physics*, 7th edn, John Wiley, New York.
Solymar, L. and Walsh, D. (2010) *Electrical Properties of Materials*, 8th edn, Oxford University Press, New York.

CHAPTER 5

Dell, R.M. and Rand, D.A.J. (2001) *Understanding Batteries*, The Royal Society of Chemistry, Cambridge.
Fuel Cells (2000) see website: www.fuelcells.org.
Tilley, R.J.D. (2008) *Defects in Solids*, Wiley-Blackwell.

CHAPTER 6

Barrer, R.M. (1982) *Hydrothermal Chemistry of Zeolites*, Academic Press, New York.
Breck, D.W. (1974) *Zeolite Molecular Sieves*, John Wiley, New York.
Catlow, C.R.A., Smit, B. and van Santen, R.A. (Eds) (2004) *Computer Modelling of Microporous Materials*, Academic Press, London.
Cejka, J., Corma, A. and Zones, S. (Eds) (2010) *Zeolites and Catalysis: Synthesis, Reactions and Applications*, Wiley-VCH, Weinheim.
Kuznicki, S.M. and UOP (2012) *Zeolite Molecular Sieves: Structure Chemistry and Use*, 2nd edn, Wiley-Blackwell.
Long, J. and Yaghi, O. (Eds) (2009) Metal Organic Frameworks, *Chemical Society Reviews*, **5**, 1201–1508.

Xu, R., Pang, W., Yu, J., Huo, Q. and Chen, J. (2007) *Chemistry of Zeolites and Related Porous Materials: Synthesis and Structure*, Wiley-Blackwell.

CHAPTER 7

Duffy, J.A. (1990) *Bonding, Energy Levels and Bands in Inorganic Solids*, Longman, London.
Padilla, W.J., Basov, D.N., and Smith, D.R. (2006) Negative refractive index materials, *Materials Today*, **9**, 28–35.
Pendry, J. (2009) Taking the wraps off cloaking, *Physics*, **2**, 95.
Stafford, N. (2010) LEDs to light up the world, *Chemistry World*, **7**, 42–45.
Yablonovitch, E. (2001) Photonic crystals, *Scientific American*, **285**(6), 47–55.
Zhang, X. and Liu, Z. (2008) Superlenses to overcome the diffraction limit, *Nature Materials*, **7**, 435–441.

CHAPTER 8

Awschalom, D.W., Flatté, M.E. and Samarth, N. (2002) Spintronics, *Scientific American* (June), 68–73.
Coey, J. M. D. (2009) *Magnetism and Magnetic Materials*, Cambridge University Press, Cambridge, UK and New York.
Guinier, A. and Julien, R. (1989) *The Solid State from Superconductors to Superalloys*, Oxford University Press, New York.
Khomskii, D. (2009) Classifying multiferroics: Mechanisms and effects, *Physics*, **2**, 20.
Spaldin, N. (2011) *Magnetic Materials, Fundamentals and Applications*, 2nd edn, Cambridge University Press, Cambridge, UK and New York.
Tokura, Y. and Seki, S. (2010) Multiferroics with spiral spin orders, *Advanced Materials*, **22**, 1554–1565.

CHAPTER 9

Buchel, W. (2004) *Superconductivity – Fundamentals and Applications*, Wiley-VCH.
Flouquet, J. and Buzdin, A. (2002) Ferromagnetic superconductors, *Physics World*, January (http://physicsweb.org).
Gough, C. (2001) New metallic superconductor makes an immediate impact, *Physics World*, April (http://physicsweb.org).
Grant, P. (1987) Do-it-yourself superconductors, *New Scientist*, **115**, 36–39.
Hazen, R.M. (1990) Perovskites, *Scientific American*, June, 52–61.
Maeno, Y., Rice, T.M. and Sigrist, M. (2001) The intriguing superconductivity of strontium ruthenate, *Physics Today*, January, 42–47.
Norman, M.R. (2011) The challenge of unconventional superconductivity, *Science*, **332**, 196–200.
Ren, Z.-A. and Zhao, Z.-X. (2009) Research and prospects of iron-based superconductors, *Advanced Materials*, **21**, 4584–4592.

CHAPTER 10

Broadwith, P. (2010) Bend me, shape me anyway you want me, *Chemistry World*, **7**, 23. (Structures made from carbon nanotubes.)
Cox, J. (2003) Chemistry in Britain, *A Quantum Paintbox*, September, 21–25.

Heeger, A.J. (2010) Semiconducting polymers: The third generation, *Chemical Society Reviews*, **39**, 2354–2371.

Higgins, S.J., Eccleston, W., Sedgi, N. and Raja, M. (2003) Plastic electronics, *Education in Chemistry*, May, 70–73.

Ozin, G.A., Arsenault, A. and Cademartiri, L. (2009) *Nanochemistry: A Chemical Approach to Nanomaterials*, Royal Society of Chemistry, Cambridge.

The first graphene touchscreen (2010) *Chemistry World*, **7**, 22.

Answers to Questions

ANSWERS FOR CHAPTER 1

1. a. NF_3. Figure A.1(a) shows the threefold (C_3) axis and three planes of symmetry.
   b. SF_4. Figure A.1(b) shows the twofold (C_2) axis and two planes of symmetry.
   c. ClF_3. Figure A.1(c) shows the twofold (C_2) axis and two planes of symmetry.
2. No. Tetrahedral molecules do not have a centre of inversion. A twofold (C_2) axis of symmetry is coincident with the $\bar{4}$ (S_4).
3. There are four. Unit cell (2) contains two lattice points; (3a) and (3b) each contain three lattice points; (4) contains four lattice points.
4. $P\bar{1}$: the triclinic space group contains only a centre of symmetry; monoclinic $P2_1/c$ has a primitive unit cell with a 2_1 screw axis along the monoclinic direction, with a *c* glide at right angles to it; orthorhombic C2/m has a C-centred unit cell containing a twofold axis of symmetry with a mirror plane at right angles; orthorhombic Pbca has a primitive unit cell with three glide planes mutually at right angles; cubic $F\bar{4}3m$ (zinc blende) has an F-centred unit cell containing a fourfold inversion (or improper) axis, a threefold axis of symmetry and a mirror plane.
5. The indices are B—$1\bar{1}$; C—*01*; D—*21*; E—$2\bar{1}$. (By choosing a different line, you may have come up with the answers: $\bar{1}1$, $0\bar{1}$, $\bar{2}\,\bar{1}$ and $\bar{2}1$. These are equally valid answers.)
6. Figure 1.27(b) shows the *200* planes: the planes are parallel to *y* and *z* and divide *a* into two.
   Figure 1.27(c) shows the $11\bar{1}$ planes: they leave *a, b* and *c* undivided.
   Figure 1.27(d) shows the *110* planes: the planes are parallel to *z* and leave both *a* and *b* undivided.
7. Figure A.2 shows the three planes: (a) *100*, (b) *110* and (c) *111*. The area of the *100* plane shown is a^2 and contains $(1 + 4 \times \frac{1}{4}) = 2$ atoms. The *110* plane contributes an area $a^2\sqrt{2}$ and contains $(2 \times \frac{1}{2} + 4 \times \frac{1}{4}) = 2$ atoms. The *111* plane contributes an area of $a^2\sqrt{3}/2$ and contains $(3 \times \frac{1}{2} + 3 \times \frac{1}{6}) = 2$ atoms. The relative densities per unit area for these three planes are thus:

$$100 : 110 : 111 = 2 : 1.414 : 2.31.$$

   Notice that the *111* planes are the close-packed layers of the structure and so have the densest packing of atoms, as we would expect.
8. a. The unit cell projection for CsCl is shown in Figure A.3a.
   b. The unit cell projection for NaCl is shown in Figure A.3b. This is a view looking down *c* on the unit cell.
   c. The unit cell projection for ZnS (zinc blende) is shown in Figure A.3c.

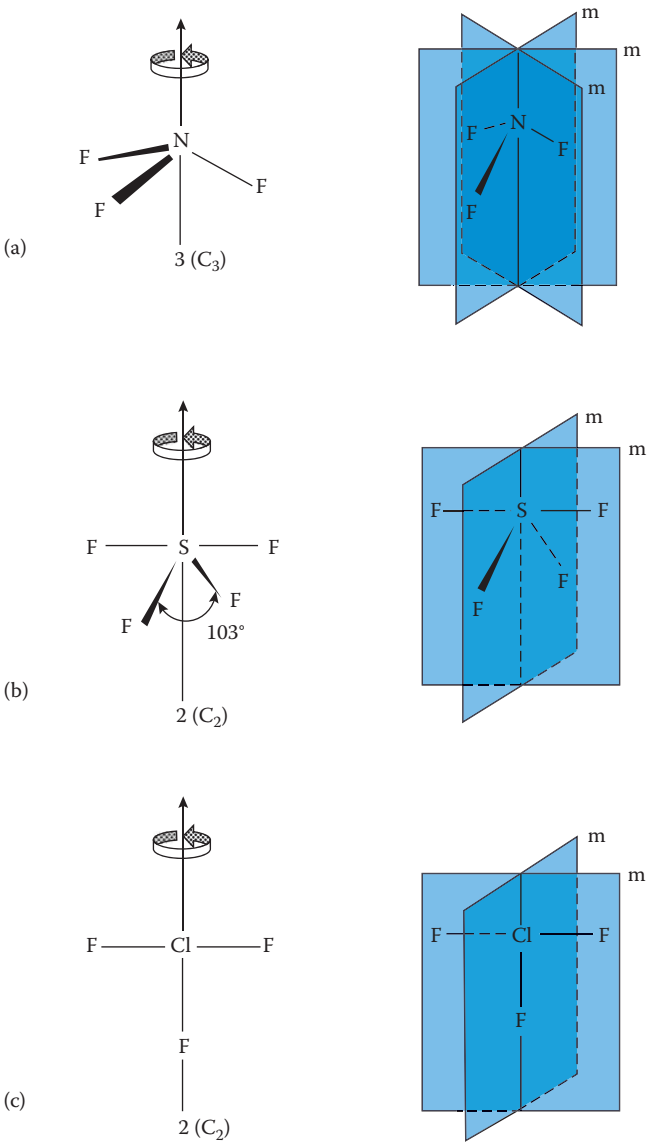


FIGURE A.1 (a) Threefold (C_3) axis in NF_3, and three planes of symmetry in NF_3, (b) twofold axis (C_2) in SF_4, and two planes of symmetry in SF_4, (c) twofold axis (C_2) in ClF_3, and two planes of symmetry in ClF_3.

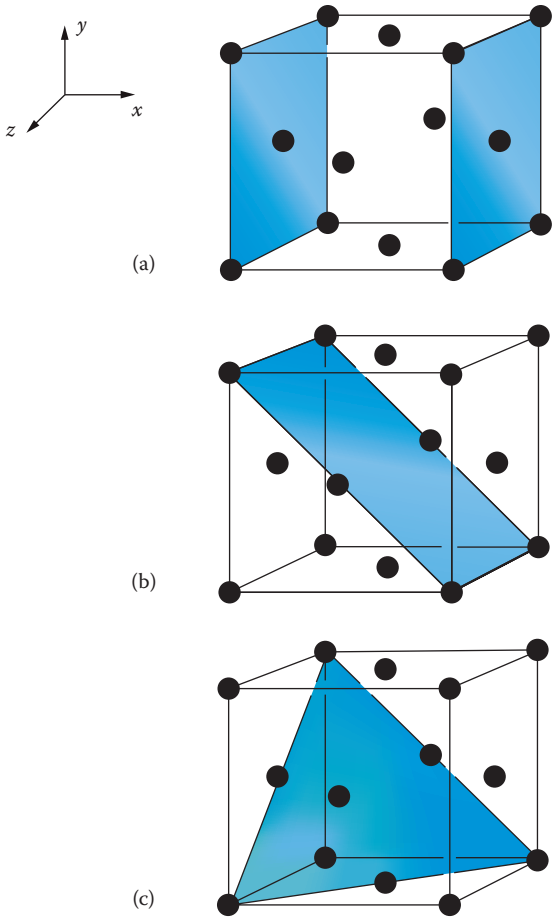


FIGURE A.2 The diagrams show different planes in the face-centred unit cell of the *ccp* structure: (a) *100*, (b) *110*, (c) *111*.

9. Four. There are $(6 \times \frac{1}{2}) = 3$ sulfur units at the centres of the faces and $(8 \times \frac{1}{8}) = 1$ at the corners. These four are matched by the four zinc units entirely enclosed in the cell.
10. a. Figure A.4a shows the hexagonal unit cell for *hcp*.
    b. Figure A.4b shows a possible unit cell for *ccp*, which has threefold symmetry. A more usual unit cell for this structure is a cubic one, which is more difficult to visualise because the close-packed planes lie parallel to a body-diagonal of the cell (see Figure 1.32).
11. The relative molecular mass of NaCl is $(22.9898 + 35.453) = 58.4428$. There are four molecules in the unit cell, so the relative mass of the unit cell in kg is:

$$\frac{4 \times 58.4428}{6.0220 \times 10^{26}} = 3.8819 \times 10^{-25}\,\text{kg}$$

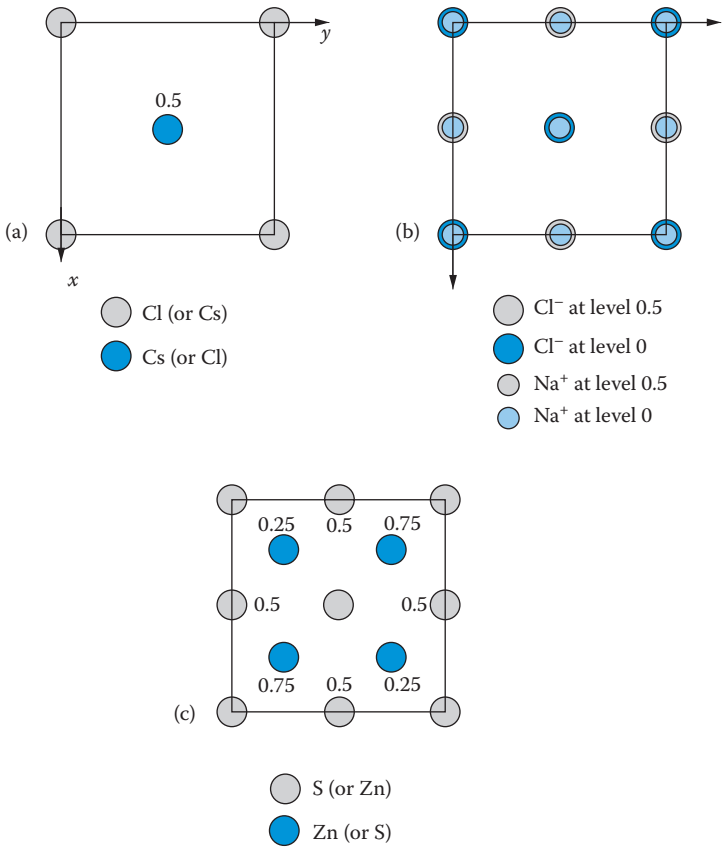


FIGURE A.3 (a) Plan of the CsCl unit cell, (b) plan of the NaCl unit cell, (c) plan of the zinc blende ZnS unit cell.

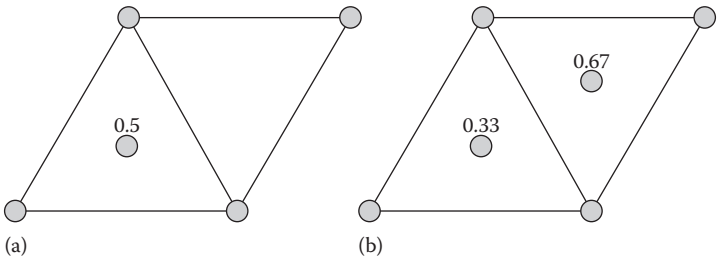


FIGURE A.4 (a) Hexagonal close-packing, (b) cubic close-packing.

The volume of the unit cell is:

$$(564 \times 10^{-12}\,\mathrm{m})^3 = 1.79 \times 10^{-28}\,\mathrm{m}^3$$

The density is therefore

$$\frac{3.8819 \times 10^{-25}}{1.79 \times 10^{-28}} \fallingdotseq 2.17 \times 10^3\,\mathrm{kg\ m^{-3}}$$

12. Let the atomic mass of X be x. The mass of the AgX unit cell in kg is thus:

$$\frac{(4 \times 107.868) + 4x}{6.0220 \times 10^{26}}\,\mathrm{kg}$$

The volume of the unit cell is:

$$(577.5 \times 10^{-12})^3\,\mathrm{m}^3 = 1.926 \times 10^{-28}\,\mathrm{m}^3$$

From the density and the volume, we calculate the mass of the unit cell to be:

$$6.477 \times 10^3\,\mathrm{kg\ m^{-3}} \times 1.926 \times 10^{-28}\,\mathrm{m}^3 = 1.247 \times 10^{-24}\,\mathrm{kg}$$

Equating the two values for the mass of the unit cell gives:

$$\frac{(4 \times 107.868) + 4x}{6.0220 \times 10^{26}}\,\mathrm{kg} = 1.247 \times 10^{-24}\,\mathrm{kg}$$

Solving for x, $431.5 + 4x = 751.2$
$x = 79.93$, and so X is bromine.

13. Assuming that anion–anion contact occurs as in Figure 1.46b, the iodide ion radius is $300/\sqrt{2}$ or 212 pm.
14. From the internuclear distance in NaI, $r(Na^+) = (323 - 212)\mathrm{pm} = 111$ pm. Then from the internuclear distance in NaF, $r(F^-) = (231 - 111)\mathrm{pm} = 120$ pm. The same procedure for RbI and RbF gives $r(Rb^+) = 154$ pm and $r(F^-) = 128$ pm.
15. The *100* and *111* planes are the same as the planes for the cubic close-packed structure in Question 7, Figure A.2. The only plane that has changed is the *110* (Figure A.5), which now has two additional atoms. Thus, the relative densities per unit area become:

$$100 : 110 : 111 = 2 : 2.83 : 2.31$$

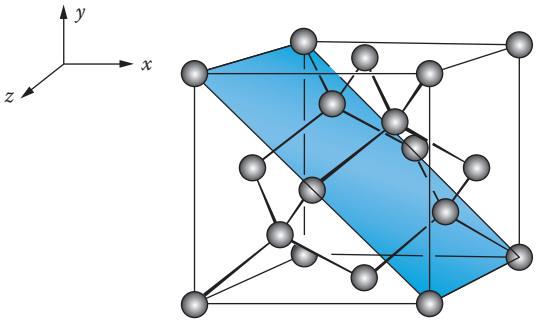


FIGURE A.5 The *110* plane of the diamond structure.

(Note that the structure is no longer truly close-packed, i.e., the carbon atoms are not touching, because the *110* planes are now more densely packed than the *111*. Remember that the radius of a tetrahedral hole is only 0.225*r*, and the carbon atoms here all have the same radius, so although we can think in terms of the carbon atoms occupying tetrahedral holes in a close-packed structure, they are actually too big to do so, and so this description is only useful in terms of the relative geometry, in the same way as we discussed for the fluorite structure.)

16. $L(CaCl_2, s) = -2255.8$ kJ mol^{-1}.
From Equation 1.19, we can write

$$\Delta H_f^{\ominus}\left(CaCl_2,s\right) = \Delta H_f^{\ominus}\left(Ca,s\right) + I_1\left(Ca\right) + I_2\left(Ca\right) + D\left(Cl\text{–}Cl\right)$$
$$-2E\left(Cl\right) + L\left(CaCl_2,s\right)$$

Rearranging and inserting the values from Table 1.19 gives the answer.
The lattice energy for calcium chloride has a large negative value. Note the large effect of the second ionisation energy of calcium.

17. The Madelung constant, $A = -3.99$.
There are seven ions in the structure in Figure 1.61: six cations and one anion. First calculate the contribution to the potential energy of interactions of the six cations with the central anion. Each cation is at a distance r_0 from the central ion.

$$E_1 = -\frac{6e^2}{4\pi\varepsilon_0 r_0}$$

Each cation also interacts with a diametrically opposite cation (distance $2r_0$). There are three such interactions, so

$$E_2 = +\frac{3e^2}{4\pi\varepsilon_0 2r_0}$$

Finally E_3 is calculated from the interactions between adjacent cations (distance $\sqrt{2}r_0$) of which there are twelve:

$$E_3 = +\frac{12e^2}{4\pi\varepsilon_0\sqrt{2}r_0}$$

$$E = E_1 + E_2 + E_3$$

$$= -\frac{e^2}{4\pi\varepsilon_0 r_0}\left(6 - \frac{3}{2} - \frac{12}{\sqrt{2}}\right)$$

$$= -\frac{Ae^2}{4\pi\varepsilon_0 r_0}$$

so,

$$A = 6 - \frac{3}{2} - \frac{12}{\sqrt{2}}$$

$$A = -3.99$$

18. The appropriate Born–Haber cycle is shown in Figure 1.56 and leads to the relationship in Equation 1.2:

$$\Delta H_f^{\ominus}(\mathrm{MCl,s}) = \Delta H_{atm}^{\ominus}(\mathrm{M,s}) + I_1(\mathrm{M}) + \tfrac{1}{2}D_m(\mathrm{Cl{-}Cl}) - E(\mathrm{Cl}) + L(\mathrm{MCl,s}) \qquad (1.2)$$

Using the data in Table 1.20, we have

$$L = (-436.7 - 89.1 - 418 - 122 + 349)\,\mathrm{kJ\,mol^{-1}}$$

$$= -717\,\mathrm{kJ\,mol^{-1}}$$

For KCl, $Z_+ = 1$, $Z_- = 1$, $r_{K^+} = 152$ pm, $r_{Cl^-} = 167$ pm, $v = 2$, and using Equation 1.15, this gives $L = -677$ kJ mol^{-1}.

There is a discrepancy of 40 kJ mol^{-1} between the calculated value and the thermodynamic value.

19. An appropriate cycle is shown in Figure A.6. Note that sulfur is a solid in its standard state, not a gas.

We wish to calculate $[E(\mathrm{S}) + E(\mathrm{S}^-)]$, which we can do if L(FeS,s) can be calculated. Values of all the other terms in the cycle are known. If the lattice energy relationship of equation 1.15 is used, substituting $v = 2$, $Z_+ = 2$, $Z_- = 2$, $r_+ = 92$ and $r_- = 170$, it gives

$$L = -3295\,\mathrm{kJ\,mol^{-1}}$$

From the cycle in Figure A.6

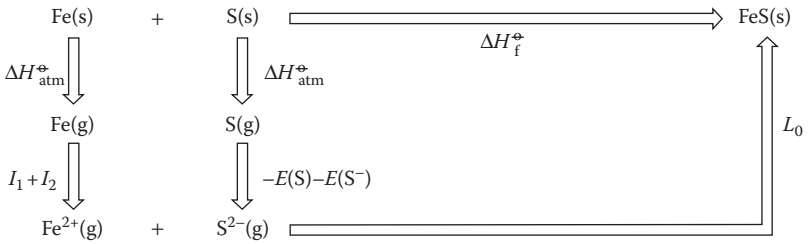


FIGURE A.6 Born–Haber cycle for the calculation of the electron affinity of sulfur.

$$E(S) + E(S^-) = -\Delta H_f^\ominus(FeS,s) + \Delta H_{atm}^\ominus(Fe,s) + (I_1 + I_2) + \Delta H_{atm}^\ominus(S,s) + L(FeS,s)$$

$$= (100.0 + 416.3 + 761 + 1561 + 278.8 - 3295)$$

$$= -177.9\ \text{kJ mol}^{-1}$$

$$E(S) = 200, \text{ so } E(S^-) = -177.9 - 200 = -377.9\ \text{kJ mol}^{-1}$$

The value of the electron affinity is the heat given out on the addition of an electron, so this implies that the enthalpy change for the addition of an electron to the $S^-(g)$ anion is endothermic:

$$S^-(g) + e^-(g) = S^{2-}(g); \quad \Delta H_m^\ominus = 377.9\ \text{kJ mol}^{-1}$$

Not surprisingly, it appears to be energetically unfavourable to add an electron to a negatively charged ion.

20. An appropriate cycle is shown in Figure A.7.

We wish to calculate $[E(O) + E(O^-)]$, which we can do if $L(MgO,s)$ can be calculated. Values of all the other terms in the cycle are known. The lattice energy relationship of Equation 1.15 is used:

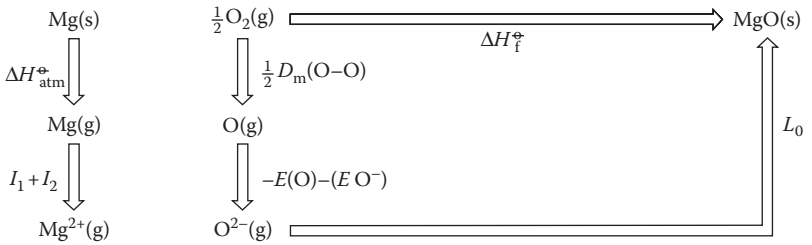


FIGURE A.7 Born–Haber cycle for magnesium oxide.

Substituting $v = 2$, $Z+ = 2$, $Z- = 2$, $r+ = 86$ and $r^- = 126$, it gives

$$L = -4072 \text{ kJ mol}^{-1}$$

From the cycle in Figure A.7

$$\begin{aligned} E(\text{O}) + E(\text{O}^-) &= -\Delta H_\text{f}^\ominus(\text{MgO,s}) + \Delta H_\text{atm}^\ominus(\text{Mg,s}) + (I_1 + I_2) \\ &\quad + \tfrac{1}{2} D_\text{m}(\text{O–O}) + L(\text{MgO,s}) \\ &= (601.7 + 147.7 + 736 + 1452 + 249 - 4072) \\ &= -885.6 \text{ kJ mol}^{-1} \end{aligned}$$

$E(\text{O}) = 141$, so $E(\text{O}^-) = -885.6 - 141 = -1026.6 \text{ kJ mol}^{-1}$

This implies that the enthalpy change for the addition of an electron to the O^- (g) anion is also endothermic:

$$\text{O}^-(\text{g}) + \text{e}^-(\text{g}) = \text{O}^{2-}(\text{g}); \quad \Delta H_\text{m}^\ominus = 1026.6 \text{ kJ mol}^{-1}$$

21. First we use Equation 1.15 to calculate a value for the lattice energy: For NH_4Cl, $v = 2$, $Z_+ = 1$, $Z_- = 1$, $r_+ = 151$ pm and $r_- = 167$ pm, giving

$$L = -679 \text{ kJ mol}^{-1}$$

From Figure 1.58:

$$\begin{aligned} P(\text{NH}_3,\text{g}) &= -\Delta H_\text{f}^\ominus(\text{NH}_4\text{Cl,s}) + \Delta H_\text{f}^\ominus(\text{NH}_3,\text{g}) + \tfrac{1}{2} D(\text{H–H}) + I(\text{H}) \\ &\quad + \tfrac{1}{2} D(\text{Cl–Cl}) - E(\text{Cl}) + L(\text{NH}_4\text{Cl,s}) \\ &= (314.4 - 46.0 + 218 + 1314 + 122 - 349 - 679) \text{ kJ mol}^{-1} \\ &= 894.4 \text{ kJ mol}^{-1} \end{aligned}$$

The addition of a proton to the ammonia molecule is an exothermic process

$$\text{NH}_3(\text{g}) + \text{H}^+(\text{g}) = \text{NH}_4^+(\text{g}); \quad \Delta H_\text{m}^\ominus = -894.4 \text{ kJ mol}^{-1}$$

(The experimentally determined value is 871 ± 15 kJ mol^{-1}, quite good agreement!)

22. The Born–Haber cycle for a chloride, MCl, is given in Figure 1.56 and leads to equation 1.2:

$$\Delta H_f^{\ominus}(\mathrm{MCl,s}) = \Delta H_{atm}^{\ominus}(\mathrm{M,s}) + I_1(\mathrm{M}) + \tfrac{1}{2} D_m(\mathrm{Cl\text{–}Cl}) - E(\mathrm{Cl}) + L(\mathrm{MCl,s}) \qquad (1.2)$$

giving a value for the lattice energy of NaCl of $L = -786$ kJ mol^{-1}. Assuming that this is also the value of L for MgCl and AlCl, we can calculate the following:

$$\Delta H_f^{\ominus}(\mathrm{MgCl,s}) = -129\ \mathrm{kJ\ mol^{-1}}$$

and

$$\Delta H_f^{\ominus}(\mathrm{AlCl,s}) = -110\ \mathrm{kJ\ mol^{-1}}$$

$\Delta S_f^{\ominus}(\mathrm{NaCl,s}) = -90.6\ \mathrm{J\ K^{-1}\ mol^{-1}}$, and using this value for MgCl and AlCl gives

$$\Delta G_f^{\ominus}(\mathrm{MgCl,s}) = -102\ \mathrm{kJ\ mol^{-1}} \text{ and } \Delta G_f^{\ominus}(\mathrm{AlCl,s}) = +83\ \mathrm{kJ\ mol^{-1}}$$

On the face of this may seem a surprising result because this tells us that both MgCl(s) and AlCl(s) are stable with respect to decomposition into their elements under standard conditions, and yet we know that they do not exist! The answer is that in each case, there is another thermodynamically favourable decomposition route—a **disproportionation**—that can take place and is more favourable:

$$2\mathrm{MgCl(s)} = \mathrm{Mg(s)} + \mathrm{MgCl_2(s)};\ \Delta G_f^{\ominus} = -388\ \mathrm{kJ\ mol^{-1}}$$

$$3\mathrm{AlCl(s)} = 2\mathrm{Al(s)} + \mathrm{AlCl_3(s)};\ \Delta G_f^{\ominus} = -380\ \mathrm{kJ\ mol^{-1}}$$

ANSWERS FOR CHAPTER 2

1. The spacings for these planes are $d_{100} = a$, $d_{110} = a/\sqrt{2}$ and $d_{111} = a/\sqrt{3}$ and so the reflections occur in the order *100*, *110* and *111*.
2. Since the separation of planes decreases as $h^2 + k^2 + l^2$ increases, it follows that the value of sinθ increases as $h^2 + k^2 + l^2$ increases. The $h^2 + k^2 + l^2$ values are therefore: for *220*, 8; for *300*, 9 and for *211*, 6. The sequence of reflections is *211*, *220* and *300*.
3. If the *100* and *110* reflections are absent, then the crystal is likely to be face-centred cubic.
4. In this case, the common factor of the first two lines is 0.05. Dividing all the $\sin^2\theta$ values by 0.05 gives the series 3, 4, 8, 11, 12, 16, 19, …, which is characteristic of face-centring.

The common factor 0.05 is equal to $\lambda^2/4a^2$, and so

$$a^2 = \frac{\lambda^2}{4 \times 0.05} = \frac{154.2^2}{0.2}$$

$$a = 345 \text{ pm}$$

5. The $\sin^2\theta$ values do not have the ratio 1 : 2 : 3 : 4 : 5 : 6 : 8, …, or 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8, …, so the cell is not primitive or body-centred. If it is face-centred, then the common factor A will be 0.0746 − 0.0560 = 0.0186. Dividing the $\sin^2\theta$ values by the common factor gives the $h^2 + k^2 + l^2$ values, and these are listed in below table rounded to the nearest integer. *hkl* only takes values where *hkl* are either all even or all odd, so the data fits a face-centred cubic structure.

Powder Diffraction Data for NaCl

θ_{hkl}	$\sin^2\theta$	$h^2 + k^2 + l^2$	*hkl*
13°41′	0.0560	3	*111*
15°51′	0.0746	4	*200*
22°44′	0.1492	8	*220*
26°56′	0.2052	11	*311*
28°14′	0.2239	12	*222*
33°7′	0.2984	16	*400*
36°32′	0.3544	19	*331*
37°39′	0.3731	20	*420*
42°0′	0.4477	24	*422*
45°13′	0.5036	27	*511, 333*
50°36′	0.5972	32	*440*
53°54′	0.6529	35	*531*
55°2′	0.6715	36	*600, 442*
59°45′	0.7462	40	*620*

6. The mean value of the common factor A is 0.01875, and this is equal to $\lambda^2/4a^2$.
This gives

$$a^2 = \frac{154.2^2}{4 \times 0.01875}$$

$$a = 563.1 \text{ pm}$$

7. If the mass of the unit cell contents is *M* and the unit cell volume is *V*, then the density, ρ, is given by

$$\rho = \frac{M}{V} = 2.17 \times 10^3 \text{ kg m}^{-3}$$

but

$$V = (563.1 \times 10^{-12})^3 \text{ m}^3$$

the mass of one mole of NaCl $= (22.99 + 35.45) \times 10^{-3}$ kg dividing by the Avogadro constant, we get that the mass of one formula unit of NaCl $= \dfrac{(22.99 + 35.45) \times 10^{-3}}{6.022 \times 10^{23}}$ kg, and if there are Z formula units in one unit cell, then the mass of the unit cell contents is

$$M = \frac{Z(22.99 + 35.45) \times 10^{-3}}{6.022 \times 10^{23}} \text{ kg}$$

so

$$\rho = 2.17 \times 10^3 = \frac{Z(22.99 + 35.45) \times 10^{-3}}{6.022 \times 10^{23} \times (563.1 \times 10^{-12})^3} \text{ kg m}^{-3}$$

and rearranging gives

$$Z = \frac{2.17 \times 10^3 \times 6.022 \times 10^{23} \times (563.1 \times 10^{-12})^3}{(22.99 + 35.45) \times 10^{-3}}$$

$Z = 3.99$, or $Z = 4$ (rounding to the nearest whole number)

8. $\theta = 19.1°$
$\sin\theta = 0.327$

From the Bragg equation

$$\lambda = 2d_{111}\sin\theta_{111}$$

so, $d_{111} = \dfrac{\lambda}{2\sin\theta_{111}} = \dfrac{154.2}{0.654} = 235.8$ pm

the d spacing between the *111* planes is $a/\sqrt{3}$

so, $a = (\sqrt{3} \times 235.8)$ pm

$= 408.4$ pm

density, $\rho = \dfrac{M}{V}$

$$V = \left(408.4 \times 10^{-12}\right)^3 \text{ m}^3$$

$$M = \frac{Z \times 107.9 \times 10^{-3}}{N_A}$$

where $Z = 4$, the number of atoms of silver in the unit cell, and N_A is the Avogadro constant.

So, $N_A = \dfrac{4 \times 107.9 \times 10^{-3}}{10.5 \times 10^3 \times \left(408.4 \times 10^{-12}\right)^3}$

$$= 6.03 \times 10^{23} \text{ mol}^{-1}$$

9. $$\rho = \frac{M}{V}$$

$$3.35 \times 10^3 \doteqdot \frac{Z(40.08 + 15.99) \times 10^{-3}}{6.022 \times 10^{23}} \div \left(481 \times 10^{-12}\right)^3 \text{ kg m}^{-3}$$

$$Z = \frac{3.35 \times 10^3 \times \left(481 \times 10^{-12}\right)^3}{9.311 \times 10^{-26}}$$

$$Z = 4$$

10. $$\rho = \frac{M}{V}$$

$$8.5 \times 10^3 \doteqdot \frac{Z\left(232.04 + 2(78.96)\right) \times 10^{-3}}{6.022 \times 10^{23}} \div \left(442.0 \times 761.0 \times 906.4 \times \left(10^{-12}\right)^3\right) \text{ kg m}^{-3}$$

$$Z = \frac{8.5 \times 10^3 \times 3.049 \times 10^{-28}}{6.476 \times 10^{-25}}$$

$$Z = \frac{2.591 \times 10^{-24}}{6.476 \times 10^{-25}}$$

$$Z = 4$$

11. From Equation 2.7, we know that $\sin^2\theta_{hkl} = \lambda^2/4a^2\,(h^2 + k^2 + l^2)$.

A cubic close-packed structure has a face-centred unit cell. The first two reflections observed will therefore be *111* and *200*, with $h^2 + k^2 + l^2$ values

of 3 and 4, respectively. Thus, $\sin^2\theta_{111} = 3\lambda^2/4a^2$ and $\sin^2\theta_{200} = 4\lambda^2/4a^2$, and $\lambda^2/4a^2 = \sin^2\theta_{200} - \sin^2\theta_{111} = 0.181 - 0.136 = 0.045$

$$a = 363.5 \text{ pm}$$

In a close-packed structure where the atoms are considered to be in contact, the radius of an atom, r, is $\frac{1}{4}$ of the length of the body-diagonal.

$$r = \tfrac{1}{4}\sqrt{3}a$$

$$r = \left(\tfrac{1}{4} \times 1.732 \times 363.5\right) \text{ pm}$$

$$r = 157.4 \text{ pm}$$

12. The scattering power is determined by the number of electrons possessed by each atom, that is, the atomic number. In increasing order, the atoms can be arranged as follows: H, O, F, Na, Cl, Co, Cd, Pt, Tl.
13. Both *111* and *222* are observed in P and F lattices, but *111* is not present for I. The *001* is not observed for F but would be present in a P unit cell. The Bravais lattice is thus F.
14. $Co(CO)_4$ is a tetrahedral molecule. The first coordination shell corresponds to four carbon atoms surrounding the Co at a distance of 177 pm. The second coordination shell refines to four oxygen atoms at a distance of 292 pm.
15. The peaks in the spectrum maximise at approximately −88, −93, −99 and −105 ppm. If you mark these values on the chart in Figure 2.34(a), then you will see that the best correspondence is to the four linkages: $Si(OAl)_3(OSi)$, $Si(OAl)_2(OSi)_2$, $Si(OAl)(OSi)_3$ and $Si(OSi)_4$.
16. There is only one peak in the new spectrum at approximately −108 ppm. Reference to the chart in Figure 2.34(a) suggests that this is due to a $Si(OSi)_4$ environment. Clearly, the treatment with $SiCl_4$ has removed all the tetrahedral Al from the framework. The intensity measurements on this peak indicate a Si/Al ratio of 55.
17. The starting sample (a) clearly shows the presence of tetrahedral Al in the framework (peak at 61 ppm). After treatment with $SiCl_4$ (b), the amount of Al in the framework has been reduced considerably, but there is a very strong peak due to $[AlCl_4]^-$ (at l00 ppm) and also a peak due to octahedral aluminium at 0 ppm. The first washings (c) remove $Na^+[AlCl_4]^-$ from the sample, and repeated washing (d) also removes some of the octahedrally coordinated Al.
18. The relative molecular mass (RMM) of MnC_2O_4 is $(54.9 + (2 \times 12.0) + (4 \times 16.0)) = 142.9$, and the RMM of H_2O is $((2 \times 1.0) + 16.0) = 18.0$.
25 mg of $MnC_2O_4 \cdot xH_2O$ gives 20 mg of MnC_2O_4, so 20/142.9 moles MnC_2O_4 are produced from 25/X moles of $MnC_2O_4 \cdot xH_2O$, if X is the RMM of $MnC_2O_4 \cdot xH_2O$. Therefore, $X = (25 \times 142.9)/20 = 178.6$. The molecular mass of water, $18x = 178.6 - 142.9 = 35.7$ and $x = 2$.

The oxalate is thus the dihydrate, $MnC_2O_4 \cdot 2H_2O$.

At higher temperatures, the oxalate decomposes to CO and CO_2 and MnO. Above 900°C, the MnO oxidises to Mn_3O_4.

19. The first exotherm at about 60°C coincides with a sharp weight loss and is due to dehydration of the ferrous sulfate. The second exotherm at 90°C does not coincide with any weight loss and must therefore be due to a phase change (this is the melting temperature). The third exotherm at about 600°C again coincides with weight loss and is due to decomposition.
20. ± 0.001 Å

ANSWERS FOR CHAPTER 3

1. This compound could be made from the elements in the correct stoichiometric proportions. The reactants would have to be very well mixed, and the reaction vessel would have to be closed to prevent the loss of the volatile sulfur (see SmS preparation).
2. Of the methods described in the chapter, the following are suitable: (a) sol–gel because the gel can be sliced thinly; CVD methods; (b) hydrothermal synthesis, vapour phase and molecular beam epitaxy; (c) vapour phase and molecular beam epitaxy; (d) sol–gel precursor method.
3. Advantages—more homogeneous product with a smaller range of grain sizes; thin layers can be formed for microcapacitors. Disadvantages—the process is more complex and would be expensive to set up on an industrial scale.
4. The compound given could be used as a precursor as it has a ratio of Cu:Cr of 1 : 2. Heating would pyrolyse the ammonia and ammonium, which would be driven off as nitrogen oxides and water. The product would be more homogeneous than a ceramic product because the components are mixed at a molecular level. The temperature needed to pyrolyse the ammonia and ammonium is less than that required to enable the solid-state reaction to proceed at a reasonable rate.

   The precursor compound contains ammonia of solvation and was probably therefore precipitated from liquid ammonia.
5. β-TeI is metastable and contains Te in an unusual oxidation state. A method that employs temperatures of this order and can produce compounds in unusual oxidation states is hydrothermal synthesis.
6. Alumina is insoluble in water even under hydrothermal conditions but needs to be dissolved for the synthesis. However, alumina is amphoteric and will dissolve in alkali. OH^- ions from the alkali coordinate to the aluminium, forming aluminate ions $Al(OH)_4^-$ which are soluble.
7. The tetrapropylammonium ion acts as a template and directs the formation of the network of channels around the propyl ligands. Heating in air oxidises the cation and leaves the zeolite framework intact.
8. *At least one component of the reaction mixture must absorb microwave* radiation strongly. Splitting the mixed oxides up into their component oxides

that would be used as starting materials, we get $CaO + TiO_2$; $BaO + PbO_2$; $ZnO + Fe_2O_3$; $ZrO_2 + CaO$; $K_2O + V_2O_5$. PbO_2, ZnO and V_2O_5 absorb microwaves strongly, and so $BaPbO_3$, $ZnFe_2O_4$, $Zr_{1-x}Ca_xO_{2-x}$ and KVO_3 are good candidates.

9. Since the reaction is exothermic, it is driven to the left by raising the temperature. Thus, the crystals grow at the hotter end of the tube.
10. Niobium is in a high oxidation state in lithium niobate; the addition of oxygen to the carrier would prevent any tendency for decomposition to a lower oxidation state plus oxygen. The hydrogen in the preparation of mercury telluride would act to reduce the tellurium (in HgTe, it is in an oxidation state of −2) and form hydrocarbons with the ethyl radicals.

ANSWERS FOR CHAPTER 4

1. For the Fermi level in sodium, $E = 2.8$ eV and the mass of an electron is 9.11×10^{-31} kg. With 1 eV $= 1.602 \times 10^{-19}$ J, this gives

$$2.8 \times 1.602 \times 10^{-19} = \tfrac{1}{2} \times 9.11 \times 10^{-31} \times v^2$$

$$v = \left(\frac{2 \times 2.8 \times 1.602 \times 10^{-19}}{9.11 \times 10^{-31}} \right)^{\frac{1}{2}}$$

$$= 9.9 \times 105 \text{ m s}^{-1}$$

2. 10^{-12} m^3 of metal contains 1740×10^{-12} kg of magnesium, but one atom of magnesium weighs $24/6.022 \times 10^{23}$ g $= 24 \times 10^{-3}/6.022 \times 10^{23}$ kg. So there are $1740 \times 10^{-12} \times 6.022 \times 10^{23}/24 \times 10^{-3}$ atoms of magnesium $= 4.4 \times 10^{16}$ atoms.
3. a. $N = 10^{-12} \times (2 \times 9.11 \times 10^{-31} \times 2.8 \times 1.602 \times 10^{-19})^{3/2}/3\pi^2 \times (1.055 \times 10^{-34})^3 = 2.125 \times 10^{16}$
   b. 2.125×10^{22}
   c. 0.2

   Each level can take two electrons, and a crystal of N atoms of magnesium has $2N$ electrons to fill the band. As you can see, the agreement between the number of filled levels predicted by this very simple theory and the number needed to accommodate the available electrons is very good for crystals. Note also that this question illustrates how the energy level spacing increases as the electrons are confined to a smaller and smaller volume.
4. The band gap lies in the energy range for visible photons, and so the photons of visible light can promote electrons from the valence band to the conduction band. Electrons in both bands can then conduct electricity.
5. Si, Ge
6. (a) *n*-type, (b) neither, (c) *p*-type, (d) *p*-type and (e) *p*-type.
7. Carborundum, like silicon and germanium, has $4N$ valence electrons for a crystal of N atoms. The tetrahedral diamond structure will be favoured

because all $4N$ electrons will then be in bonding orbitals, and the energy is lower than in the higher coordination structure.

8. In general, compounds with broader d bands will be metallic. Broad bands will tend to occur for elements at the beginning of the transition series. Thus going from Ti to V to Mn, the bands would be expected to narrow and the compounds become less metallic.
9. Examples include the following: limited computer resource available, materials with very large unit cells, problems involving very highly diluted defects, tracking of the motion of an individual ion through the solid.

ANSWERS FOR CHAPTER 5

1. $n_S \approx Ne^{-\Delta H_S/2RT}$ where $\Delta H_S = 200$ kJ mol^{-1} and R = 8.314 J K^{-1} mol^{-1}

Schottky Defect Concentration in MX Compound at Various Temperatures

Temperature/°C	Temperature/K	n_S/N	n_S/mol^{-1}
27	300	3.87×10^{-18}	2.33×10^{6}
227	500	3.57×10^{-11}	2.15×10^{13}
427	700	3.45×10^{-8}	2.08×10^{16}
627	900	1.57×10^{-6}	9.45×10^{17}

2. Rearrange Equation 5.4 and take logs:

$$\frac{n_S}{N} \approx e^{-\Delta H_S/2RT}$$

$$\ln \frac{n_S}{N} = -\frac{\Delta H_S}{2RT}, \text{ or}$$

$$\log \frac{n_S}{N} = -\frac{\Delta H_S}{2.303 \times 2RT}$$

A plot of $\log n_s/N$ against $1/T$ gives a straight line plot passing through the origin. The slope of this graph, $-[\Delta H_s/(2.303 \times 2R)]$, gives a value of 183.4 kJ mol^{-1} for ΔH_S. Dividing by Avogadro's number, 6.022×10^{23}, gives the enthalpy of formation of one Schottky defect, 3.045×10^{-19} J.

3. Increasing the impurity levels does not affect the intrinsic (left-hand side) of the graph. It does, however, increase the value of σ (and thus of lnσ) in the extrinsic region. As the activation energy, E_a, for cation movement stays the same, the slope of the graph is unchanged (see Figure A.8). The purer the crystal, the lower the transition temperature at which thermally generated defects take over.

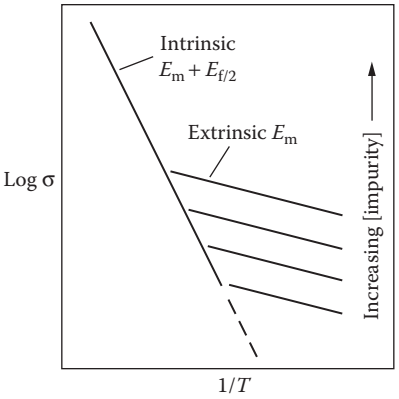


FIGURE A.8 Schematic ionic conductivity of doped NaCl crystals. Parallel lines in the extrinsic region correspond to different dopant concentrations. (After West, A.R. *Solid State Chemistry and Its Applications*, Wiley, 1984.)

4. (a) Almost none. (b) Conductivity increases as two Na^+ ions are replaced by each Mg^{2+} ion, thus creating a vacancy each time. (c) Almost none. (d) Two Cl^- ions are replaced by each O^{2-} ion, creating anion vacancies; this will result in a small increase in the conductivity because, although the major charge carrier is Na^+, some current can be carried by the chloride ions.
5. The oxide has more highly charged ions (M^{4+} and O^{2-}) than a fluoride. This means that the coulombic interactions will be stronger and thus the energy of defect formation will be higher (see Table 5.1), so reducing the defect concentration at a given temperature.
6. The path is shown in Figure A.9. The anion jumps from the tetrahedral site, through a trigonal position on the common triangular face and into the octahedral site at the body-centre.
7. The anion at the body-centre of the unit cell in Figure 5.3c is surrounded by six anions at distance $(a/2)$ and by four cations at a distance of $0.43a$. The interstitial site at the body-centre of the unit cell in Figure 5.3a is surrounded by *eight* anions at a distance of $0.43a$ and by *six* cations at a distance of $a/2$.

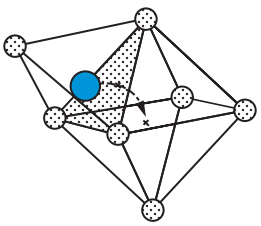


FIGURE A.9 Path taken by an anion moving from one tetrahedral site to an octahedral site.

for the normal anion site:

$r = 0.43 \times 537 \times 10^{-12}$ m and $Z = +2$, for interaction with four cations. $r = 0.5 \times 537 \times 10^{-12}$ m and $Z = -1$, for interaction with the six anions.

$$E = -\left(\frac{2.31 \times 10^{-28}\,\text{J m}}{537 \times 10^{-12}\,\text{m}}\right)\left(\frac{4 \times 2}{0.43} + \frac{6 \times (-1)}{0.5}\right)$$

$$\fallingdotseq -\left(4.302 \times 10^{-19}\,\text{J}\right)\left(6.605\right)$$

$$\fallingdotseq -2.84 \times 10^{-18}\,\text{J}$$

for the interstitial site:

$$E = -\left(4.302 \times 10^{-19}\,\text{J}\right)\left(\frac{8 \times (-1)}{0.43} + \frac{6 \times 2}{0.5}\right)$$

$$\fallingdotseq -\left(4.302 \times 10^{-19}\,\text{J}\right)\left(5.395\right)$$

$$= -2.32 \times 10^{-18}\,\text{J}$$

The energy of defect formation is the difference in energy between the two sites and so is given by:

$$\left(-2.32 \times 10^{-18}\,\text{J}\right) - \left(-2.84 \times 10^{-18}\,\text{J}\right) = 5.2 \times 10^{-19}\,\text{J}$$

The experimental value for fluorite is given in Table 5.1 as 4.49×10^{-19} J. This calculation gives a very good level of agreement considering that we have ignored all the more distant interactions, internuclear repulsion, lattice vibrations and lattice relaxation!

8. Both structures contain Ag^+ ions in tetrahedral sites, but in γ-AgI, half of the tetrahedral holes are occupied, so there is only *one* vacant equivalent site per Ag^+ ion, whereas in α-AgI, there are *five*. Both structures contain vacant octahedral and trigonal sites. In γ-AgI, the octahedral holes lie at the body-centre and at the mid-point of each edge. The trigonal sites lie on the triangular faces of the octahedra, where adjacent octahedra and tetrahedra join (Figure A.10). Both structures have the same monovalent ions and a polarisable anion.

   The jump from one tetrahedral site to another in γ-AgI could take either of the routes sketched out in Figure A.10, which first pass through a trigonal face and then through an octahedral hole. Alternatively (not shown), if the ion passed through only the top half of the octahedron, it would be five coordinate before passing through another three coordinate face. All these routes would be of higher energy than the pathways we looked at for α-AgI.

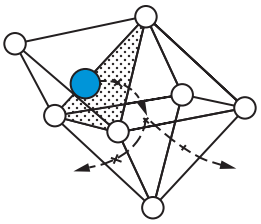


FIGURE A.10 Path taken by an anion moving from one tetrahedral site to another.

9. I^-, S^{2-}, Se^{2-} and Te^{2-} are all polarisable anions.
10. In undoped β-alumina, the excess Na^+ ions are balanced by additional oxygen in the conduction plane. Beyond a certain point, it seems likely that these will begin to block the motion of the Na^+ ions, as observed. By contrast, doping with Mg^{2+} suggests an alternative charge compensation mechanism: simple substitution for Al^{3+} ions in the spinel-like blocks allows extra Na^+ ions into the conduction planes without the need for oxygen interstitials. This is what happens in practice.
11. Unit cell volume is $(428.2\text{ pm})^3 = 7.8513 \times 10^{-29}\text{ m}^3$. Mass of contents for *iron vacancies:*

    $[(4 \times 55.86 \times 0.910) + (4 \times 16.00)]/(N_A \times 10^3)$ kg giving a density of $5.653 \times 10^{-3}\text{ kg m}^{-3}$

    Mass of contents for *oxygen interstitial*:

    $[(4 \times 55.85) + (4 \times 16.00 \times 1/0.910)]/(N_A \times 10^3)$ kg giving a density of $6.213 \times 10^3\text{ kg m}^{-3}$

    Comparing these theoretical values with the experimental value, we again see that the evidence supports an iron vacancy model.
12. From Table 5.6, we see that as the Fe:O ratio decreases, the unit cell volume also decreases; this is the trend we would expect to see as more vacancies are introduced. If the interstitial model were correct, as the Fe:O ratio decreases, the number of interstitial oxygen rises and we would expect to see a slight *increase* in lattice parameter.
13. A crystal containing F-centres contains anion vacancies. We would expect, therefore, that the density would be *lower* than that of the colourless crystal.
14. (a) The central section has *two* Fe^{3+} ions in tetrahedral sites and *seven* vacancies, so it is known as a 7 : 2 cluster. (b) There are 32 oxide anions. There are seven octahedral vacancies and two Fe^{3+}_{tet} interstitial ions, so there will be a total of 27 Fe cations. (There are *two* Fe^{3+} ions and *six* Fe_{oct} ions enclosed within the cluster. The outer layer will be the same as the Koch-Cohen cluster with $(8 \times \frac{1}{8}) = 1\ Fe_{oct}$ at the corners; $(12 \times \frac{1}{4}) = 3\ Fe_{oct}$ at the mid-points of the edges and $(30 \times \frac{1}{2}) = 15\ Fe_{oct}$ on the faces; this makes 27 Fe ions in total). The formula would be $Fe_{27}O_{32}$. (c) The 32 oxide ions provide 64 negative charges to be balanced. The two tetrahedral Fe^{3+} ions reduce this to 58 to be balanced by the ions in octahedral positions.

Setting up simultaneous equations, we know that if x is the number of Fe^{2+} ions and y the number of Fe^{3+}, then

$$x + y = 25$$

And adding up the charges, we get

$$2x + 3y = 58$$

Solving gives $x = 17$ and $y = 8$

15. **Titanium vacancies:** there are 8 at the corners $(8 \times \frac{1}{8}) = 1$ and $(2 \times \frac{1}{2}) = 1$ on cell faces.
Titanium ions: cell edges, $(4 \times \frac{1}{4}) = 1$. Cell faces, $(8 \times \frac{1}{2}) = 4$ on the top and bottom. There are 5 ions contained within the cell boundary, making 10 in total.

The titanium stoichiometry of the unit cell is obviously representative of the whole structure: of the 12 sites, 10 are occupied and 2 are vacant. This is also true for oxygen:
Oxygen vacancies: cell faces, $(4 \times \frac{1}{2}) = 2$.
Oxide ions: cell faces, $(8 \times \frac{1}{2}) = 4$. Cell edges, $(8 \times \frac{1}{4}) = 2$. There are 4 ions contained within the cell boundary, making 10 in all.
16. Taking the titanium positions first: there are vacancies at all eight corners $(8 \times \frac{1}{8}) = 1$ and 1 vacancy in the centre of the cell, making 2 vacancies in all. There are no Ti^{2+} ions on cell edges. There are Ti^{2+} ions on only two of the faces—the top and bottom, which have 4 each—contributing $(2 \times 4 \times \frac{1}{2}) = 4$. There are 4 Ti^{2+} ions enclosed within the cell. Four edges have O^{2-} ions giving $(4 \times 1 \times \frac{1}{4} = 1$; only the top and bottom faces have O^{2-} ions, and these have 5 each contributing $(2 \times 5 \times \frac{1}{2}) = 5$; there are 4 O^{2-} ions enclosed within the cell, making 10 in all. The overall content of this unit cell is thus, Ti_8O_{10}, which corresponds to $TiO_{1.25}$.
17. The structure has Ti vacancies—every fifth Ti is missing. For every absent Ti^{2+} ion, there must be two Ti^{3+} present or one Ti^{4+}.
18. The effects of sharing are shown in Figure A.11.
19. Figure A.12 shows the shear structure with a unit cell added. Within the boundary, there is one group of four edge sharing octahedra and four $[WO_6]$ octahedra. The formula is thus $W_4O_{11} + 4WO_3 = W_8O_{23}$.
20. Nonstoichiometric ZnO is an *n*-type semiconductor. Gallium ions entering the structure of ZnO have a charge of +3. If the Ga^{3+} substitutes for Zn^{2+} and the crystal maintains its stoichiometry, oxygen will be lost during the reaction. The electrons made available from the oxide ions becoming oxygen molecules will remain in the structure to effect the necessary charge compensation, thus enhancing the *n*-type semiconduction. An equation for the reaction is:

$$x\mathrm{Ga_2O_3} + 2(1 - x)\mathrm{ZnO} = 2\mathrm{Ga}_x\mathrm{Zn}_{1-x}\mathrm{O} + \tfrac{1}{2}x\mathrm{O_2}$$

where we have used 1 : 1 ZnO for simplicity.

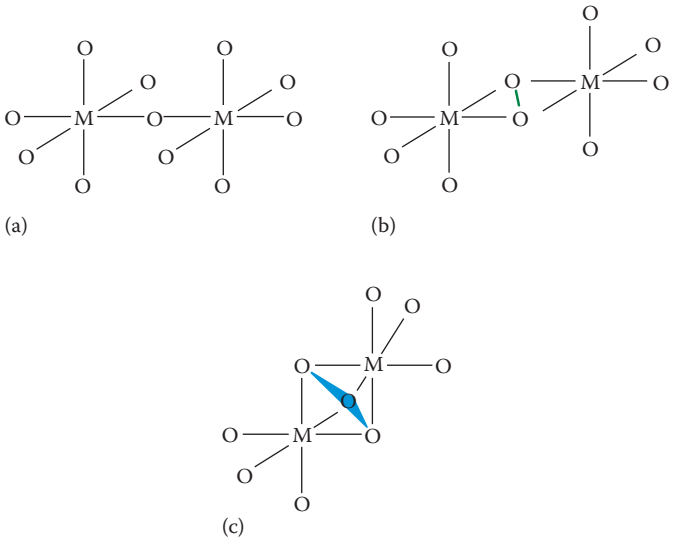


FIGURE A.11 (a) Sharing a corner M_2O_{11}. (b) Sharing an edge M_2O_{10}. (c) Sharing a face M_2O_9.

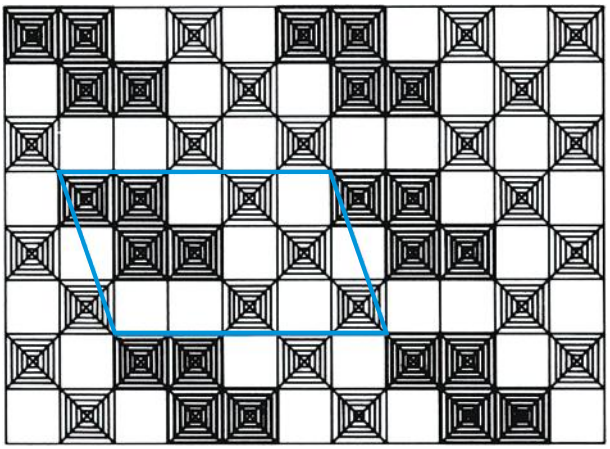


FIGURE A.12 Unit cell of W_8O_{23}.

ANSWERS FOR CHAPTER 6

1. The chart in Figure 2.34 shows that the most likely Si environment for zeolite A is $Si(OA1)_3(OSi)$. However, we know that the Si/Al ratio is 1, so this coordination is not possible without systematically breaking Loewenstein's rule. It was the spectroscopic work on this structure (and on zeolite ZK-4) that eventually confirmed the $Si(OAl)_4$ structure of zeolite A with strict alternation of Si and Al, and led to the extended ranges shown in the chart.

2. Catalytic activity tends to parallel the *increase in acidity* of the active sites as the Si ratio increases. Beyond a certain point, however, the *number* of acidic sites has decreased so much that catalytic activity also decreases.
3. This system demonstrates reactant shape-selective catalysis. The branched hydrocarbons are too bulky to pass through the pore openings in the catalyst.
4. We see here product shape-selective catalysis in operation: both reactant molecules are small enough to diffuse into the pores and be hydrogenated, but the slightly larger propane molecule cannot leave.
5. Zeolite A (Ca) shows reactant selectivity. The straight-chain *n*-hexane can pass through the windows and undergo reaction, but the branched-chain 3-methylpentane is excluded. The selective cracking of straight-chain hydrocarbons in the presence of branched chains is an important industrial process known as **selectoforming**, which improves the octane number of the fuel.
6. There are two possible reasons for the increased percentage of *para*-xylene. First, the increased crystal size increases the distance travelled by the diffusing molecules along the channels, giving more time for isomerisation to the *para*-form. Second, the shape-selective reactions take place in the internal pores of the zeolites, rather than on the surface. Larger crystals have a higher ratio of internal to external sites, and so we would expect selectivity to improve.
7. As the Si/Al ratio increases, the –OH stretching frequency falls. This indicates a decrease in the covalent bond strength, making ionisation of H^+ easier, that is, an increase in acid strength.

ANSWERS FOR CHAPTER 7

1. In Mn^{2+}, an electron can only go from one 3d level to another if it changes its spin. Transitions in which an electron changes its spin are forbidden and so give rise to weak spectral lines.
2. Radiation from a lamp excites electrons from the ground state, G, to states B, C and D. Electrons in states C and D undergo nonradiative transitions to state B. Radiative transitions from B to lower states are forbidden, and so a large population of electrons in state B builds up. Eventually, a photon is emitted and an electron goes to state A. This photon induces further emission from B to A, eventually producing a beam of laser light.

   Transitions from B to both A and G must be forbidden for a large population to build up in B, but the transition B↔A must be less forbidden than the transition B↔G as it is B↔A that gives rise to laser action, and hence it must be the more likely to occur.
3. ZnS is a semiconductor with a full valence band and empty conduction band. When the phosphor is illuminated, electrons are promoted to the conduction band. Since the orbitals in this band are delocalised, the energy can easily be transferred to other parts of the crystal, in particular to the dopant atoms.

4. The solid has an indirect band gap because the lowest energy interband transition (that is from the highest energy level in the valence band to the lowest energy level in the conduction band) does not correspond to $\Delta k = 0$ and is therefore forbidden.
5. Silicon is an indirect band gap solid with an available nonradiative pathway from the conduction band to the valence band. In photo-voltaic cells, electrons are promoted from the valence band to the conduction band and are then used to do electrical work. The promoted electrons do not return directly to the valence band either by emitting energy or by a nonradiative pathway. In LED's, it is the return of the electrons to the valence band by emitting light that is important. This return has low probability because of the indirect band gap and the electrons use the nonradiative pathway instead.

   Promotion to the conduction band in the solar cell will also be of low probability, but there is no competing nonradiative route.
6. These oxides (at least formally) contain small highly charged ions. These have low polarisability and hence impart high refractive indices.
7. The photonic band gap wavelength increases as the size of the spheres increases, so that it will be larger for the longer wavelength orange-red colours.
8. Sodium chloride has a cubic structure. Although the empirical formula of calomel is HgCl, its formula is usually written Hg_2Cl_2. It has a tetragonal crystal structure containing Hg_2 units. Thus, calomel crystals are asymmetric, a prerequisite for birefringence. Hg will be more polarisable than Na, giving a higher refractive index.

ANSWERS FOR CHAPTER 8

1. Because the Mn atoms are further apart, the overlap of the 3d orbitals will be less. The 3d band will therefore be narrower than in manganese metal. With a narrower band, there is a larger interelectronic repulsion and a state with a number of unpaired spins comparable to the number of atoms becomes favourable. The alloy is thus ferromagnetic.
2. The magnetically ordered unit cell is identical to the high-temperature (paramagnetic) unit cell, so that all the layers of europium ions must be aligned with their spins parallel, giving a ferromagnetic compound.
3. The Zn^{2+} and half the Fe^{3+} ions are on octahedral sites with spins aligned, and the remaining Fe^{3+} ions are on tetrahedral sites aligned antiparallel. The net moment of the Fe^{3+} ions is zero. As all the electron spins are paired in Zn^{2+} ions, there is no overall magnetic moment and the compound is antiferromagnetic.
4. In MnS_2, the 3d electrons are localised on the evidence given. Above the Néel temperature, the solid is an insulator with a paramagnetic susceptibility corresponding to 5 unpaired electrons per Mn and this could be explained on the basis of filled bands below the 3d level and localised 3d electrons. Below the Neél temperature, the localised electrons on different Mn ions interact to cancel out magnetic moments, possibly through a superexchange mechanism involving the disulfide ions.

The properties of FeS_2 suggest that all spins are paired either in a band or localised on the iron ions. There are six 3d electrons per Fe ion, and this is just enough to fill the lower t_{2g} band. The semiconducting properties suggest that there is an empty band not much higher in energy. This is probably the e_g band.

In CoS_2, there are enough electrons to partly occupy the e_g band, and thus this compound is metallic. The band is narrow and so gives rise to ferromagnetism.

5. The effect of deuterium substitution suggests that the hydrogen atoms are displaced when the ferroelectric phase is formed.
6. Li^+ is smaller than K^+ and so is easily displaced.
7. The ferroelectricity will derive from displacement of ions due to the lone pair on Pb^{2+}. The magnetism is due to the V^{4+} ions. Thus, the ferroelectricity and ferromagnetism arise from different ions and are unlikely to be strongly coupled (see $BiFeO_3$).

ANSWERS FOR CHAPTER 9

1. In the analogy, each person represents a Cooper pair, and the linking of arms denotes the overlap between the pairs, giving an ordered system. Because of the ordered collective motion, scattering from defects—tripping over potholes—cannot take place! The analogy breaks down in that it allows for overlap only between adjacent pairs rather than over large numbers; also, the loss of superconductivity leads to the scattering of individual electrons, not of individual Cooper pairs.
2. The packing diagram for the A-type unit cell is shown in Figure A.13.
The contents of the unit cell are as follows: 12 oxygen atoms midway on cell edges, each shared by four unit cells: one calcium atom at the body-centre, eight titanium atoms at the cube corners—each shared by eight unit cells: $[(12 \times \frac{1}{4})O + 1Ca + (8 \times \frac{1}{8})Ti] = 3O + 1Ca + 1Ti = 1$ formula unit. The Ca in the centre is surrounded by twelve O atoms, and eight Ti atoms at the corners of a cube. A Ti atom is surrounded octahedrally by six O atoms, and by eight Ca atoms at the corners of a cube. An oxygen atom is linearly coordinated by two Ti atoms and by four Ca atoms in a square-planar configuration.

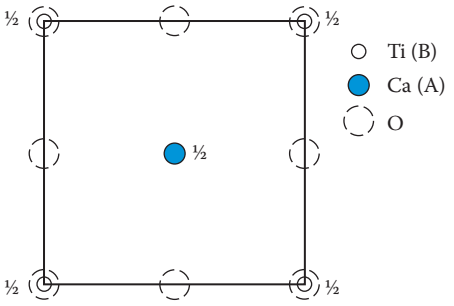


FIGURE A.13 A packing diagram for a unit cell of perovskite, $ABO_3/CaTiO_3$.

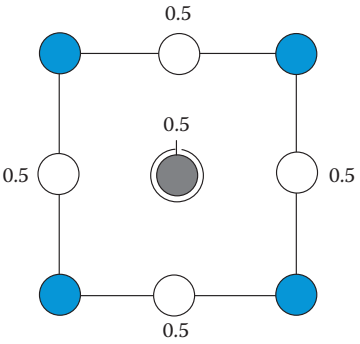


FIGURE A.14 A packing diagram for a perovskite B-type cell.

The packing diagram for the B-type cell is shown in Figure A.14. This has the Ti/B atom at the centre of the cell and the Ca/A atoms at the corners—each shared by eight other unit cells; there are six face-centred oxygen atoms, each shared by two unit cells. Counting up as before gives: $[(8 \times \frac{1}{8})\text{Ca} + 1\text{Ti} + (6 \times \frac{1}{2})\text{O}] = 1\text{Ca} + 1\text{Ti} + 3\text{O} = 1$ formula unit. The coordination geometry of course remains identical—we have only moved the origin of the unit cell.

3.

Material	Symmetry of Order Parameter	Spin State
Metals and alloys	s-wave	singlet
Superconducting cuprates	d-wave	singlet
Iron-based superconductors	s-wave	singlet
Ferromagnetic superconductors	p-wave	triplet

4. The net oxidation state on the cations other than Cu is $(2 \times 0.8) + (3 \times 0.2) + (2 \times 2) + (2 \times 2) = 10.2$. There are 8.33 oxygen atoms with oxidation state −2 giving −16.66. To achieve charge balance, the net oxidation state on the copper atoms is thus $16.66 - 10.2 = 6.46$. There are three copper atoms, and so the average oxidation state is $6.46/3 = 2.153$.
5. There are two Se in oxidation state −2 giving −4. K forms K^+ so $K_{0.75}$ gives 0.75+. To achieve charge balance, the net oxidation state on Fe is $(4 - 0.75)/2 = +1$. 625. Fe is usually found in oxidation states +2 and +3, so the oxidation state here is unusually low. It is possible the Fe_2Se_2 layer is more alloy-like than ionic.
6. Both types have layered structures. In the cuprates, the superconducting layer is Cu-O_2 in square coordination. In the iron-based superconductors, the conducting layer is a square mesh of Fe with the anions (As^{3+}, Se^{2-}, Te^{2-}) above and below the Fe plane. The superconducting layers in both are interspersed with nonsuperconducting layers.

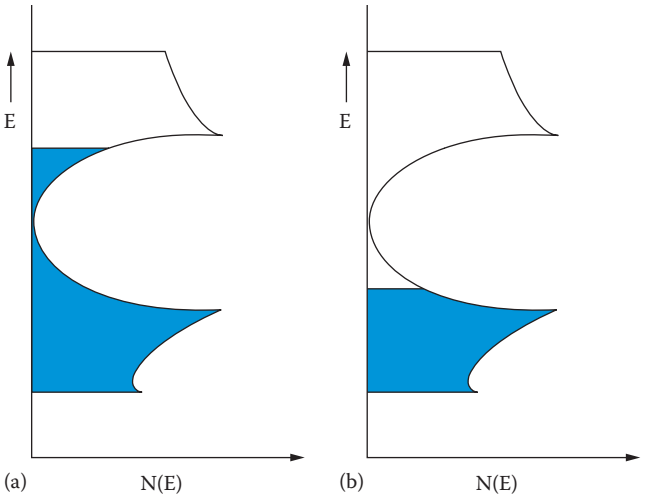


FIGURE A.15 Band structure of graphite doped with (a) an electron donor and (b) an electron acceptor.

ANSWERS FOR CHAPTER 10

1. Because the particles are single domain, they are resistant to demagnetisation and so the bacteria can rely on them always lining up to indicate in which direction the bottom of the pond lies.
2. Polyphenylenevinylene has a π system delocalised over the benzene ring. It is likely that this is also delocalised over the conjugated double bond and hence to the benzene ring of the next unit. Thus, like polyacetylene, this polymer will have a delocalised π system, but it will include the π ring orbitals.
3. a. Rb would lose an electron to the polyacetylene, forming Rb^+, and give *n*-type polymer with a partially full conduction band.
   b. H_2SO_4 would act as an electron acceptor and give *p*-type polyacetylene.
4. $8(CH)_n + 9\delta nHClO_4 = 8[(CH)^{\delta+}(ClO_4^-)_\delta]_n + \delta nHCl + 4\delta nH_2O$ (10.5)
5. If the polymer appears red, it is absorbing light in the blue/green region of the visible spectrum, suggesting that the band gap energy lies in the blue/green region. When acting as an LED, the energy of the emitted photons would correspond to the band gap energy or greater, thus the emitted light would be blue/green rather than red.
6. Figure A.15 shows the band structure of graphite intercalated with (a) an electron donor and (b) an electron acceptor.
7. The smallest would be that with no hexagons and just twelve pentagons. This is a regular dodecahedron that has 20 vertices and the fullerene would have the formula C_{20}.

   In practice, pentagons are more strained than hexagons, and some hexagons are needed for stability.

Chemistry

FOURTH EDITION

SOLID STATE CHEMISTRY
AN INTRODUCTION

Building a foundation with a thorough description of crystalline structures, **Solid State Chemistry: An Introduction, Fourth Edition** presents a wide range of the synthetic and physical techniques used to prepare and characterize solids. Going beyond basic science, the book explains and analyzes modern techniques and areas of research.

The book covers

- A range of synthetic and physical techniques used to prepare and characterize solids
- Bonding, superconductivity, and electrochemical, magnetic, optical, and conductive properties
- STEM, ionic conductivity, nanotubes and related structures such as graphene, metal organic frameworks, and FeAs superconductors
- Biological systems in synthesis, solid state modeling, and metamaterials

This largely nonmathematical introduction to solid state chemistry includes basic crystallography and structure determination, as well as practical examples of applications and modern developments to offer students the opportunity to apply their knowledge in real-life situations and serve them well throughout their degree course.

New in the Fourth Edition

- Coverage of multiferroics, graphene, and iron-based high temperature superconductors, the techniques available with synchrotron radiation, and metal organic frameworks (MOFs)
- More space devoted to electron microscopy and preparative methods
- New discussion of conducting polymers in the expanded section on carbon nanoscience

K12133

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

ISBN: 978-1-4398-4790-9
90000
9 781439 847909