

S279 Our dynamic planet: Earth and life
Science: Level 2

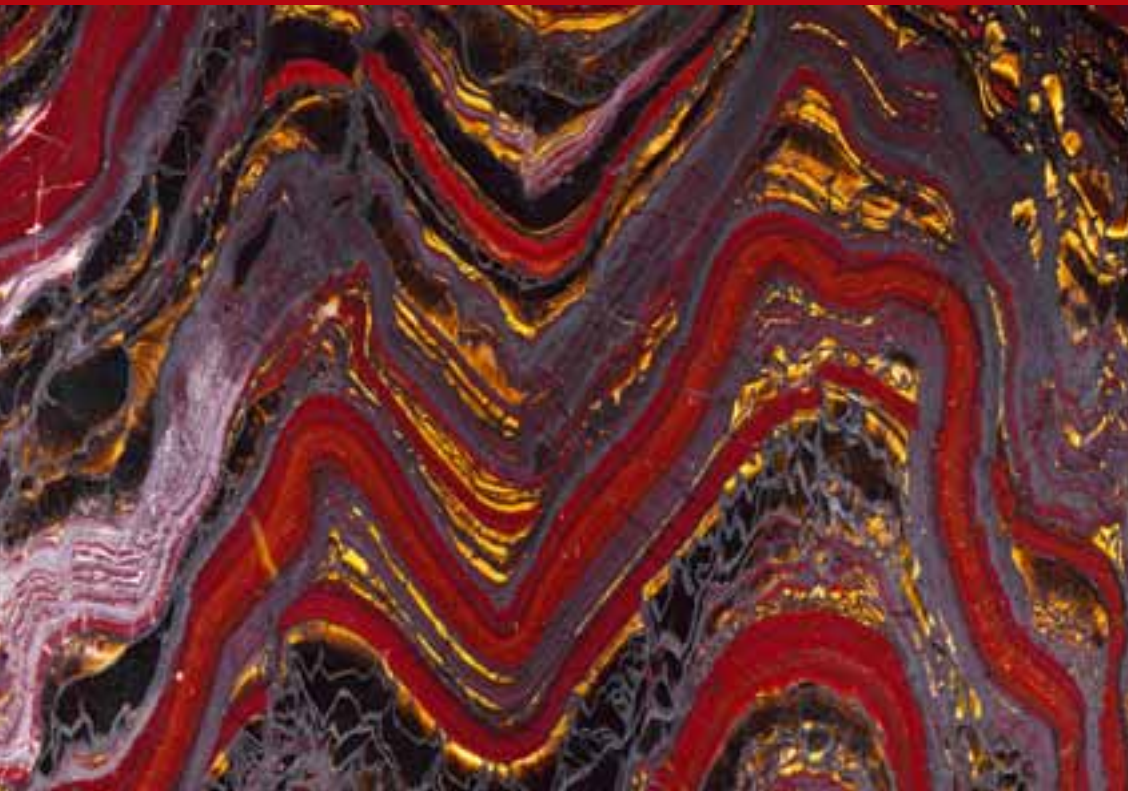


The Open University

Book 1 Part 1 Our Dynamic Planet

Edited by Nick Rogers

Authors: Stephen Blake, Kevin Burton, Nick Rogers and Mike Widdowson



Acknowledgement

The S279 Course Team gratefully acknowledges the contributions of members of the S267 *How The Earth Works: The Earth's Interior* (1993) Course Team.

Cover image: Folded layers of a banded iron formation (BIF or banded ironstone). Banded ironstone is a rock formed from alternating layers of the iron oxide mineral hematite (red) and tiger eye and jasper. At greater than 3 billion years of age these are among the oldest rocks on Earth. They were laid down in shallow seas where primitive bacteria caused iron to be oxidised and precipitated. Subsequent deep burial of the rocks, and tectonic movements, have caused the rock to alter and deform. This small sample is around 2 cm across. The rock has been cut open to expose the banding (strata), and also polished for display. Many BIFs come from Western Australia. (Dirk Wiersma/Science Photo Library)

This publication forms part of an Open University course S279 *Our dynamic planet: Earth and life*. Details of this and other Open University courses can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom: tel. +44 (0)845 300 60 90, email general-enquiries@open.ac.uk

Alternatively, you may visit the Open University website at <http://www.open.ac.uk> where you can learn more about the wide range of courses and packs offered at all levels by The Open University.

To purchase a selection of Open University course materials visit <http://www.ouw.co.uk>, or contact Open University Worldwide, Michael Young Building, Walton Hall, Milton Keynes MK7 6AA, United Kingdom for a brochure. tel. +44 (0)1908 858793; fax +44 (0)1908 858787; email ouw-customer-services@open.ac.uk

The Open University
Walton Hall, Milton Keynes
MK7 6AA

First published 2007.

Copyright © 2007 The Open University

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted or utilised in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without written permission from the publisher or a licence from the Copyright Licensing Agency Ltd. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS; website <http://www.cla.co.uk/>.

Open University course materials may also be made available in electronic formats for use by students of the University. All rights, including copyright and related rights and database rights, in electronic course materials and their contents are owned by or licensed to The Open University, or otherwise used by The Open University as permitted by applicable law.

In using electronic course materials and their contents you agree that your use will be solely for the purposes of following an Open University course of study or otherwise as licensed by The Open University or its assigns.

Except as permitted above you undertake not to copy, store in any medium (including electronic storage or use in a website), distribute, transmit or retransmit, broadcast, modify or show in public such electronic materials in whole or in part without the prior written consent of The Open University or in accordance with the Copyright, Designs and Patents Act 1988.

Edited, designed and typeset by The Open University.

Printed and bound in the United Kingdom at the University Press, Cambridge.

ISBN 978 0 7492 1902 4

1.1

Contents

CHAPTER 1 AN INTRODUCTION TO THE STRUCTURE AND COMPOSITION OF THE EARTH

5

1.1 Planet Earth is unique	7
1.2 The structure of the Earth	14
1.3 The Sun, meteorites and the bulk composition of the Earth	24
1.4 The composition of Earth layers	36
1.5 The Earth's crust: continents and oceans	44
Summary of Chapter 1	49
Learning outcomes for Chapter 1	49

CHAPTER 2 THE EARLY EARTH

51

2.1 Heating and differentiation of the Earth	51
2.2 The age of the Earth and its layers	61
2.3 The Moon	73
2.4 The origins of the atmosphere and hydrosphere	79
2.5 The earliest continents	84
Summary of Chapter 2	91
Learning outcomes for Chapter 2	93

CHAPTER 3 PLATE TECTONICS

95

3.1 From continental drift to plate tectonics	96
3.2 The theory of plate tectonics	105
3.3 Plate tectonic motion	125
3.4 Plate driving forces	132
3.5 Implications of plate tectonics	138
Summary of Chapter 3	142
Learning outcomes for Chapter 3	142

CHAPTER 4 PROCESSES AT CONSTRUCTIVE PLATE BOUNDARIES

143

4.1 The structure and composition of oceanic lithosphere	144
4.2 Rocks and minerals in the upper mantle	148
4.3 Experiments with peridotite	156
4.4 Why does partial melting happen?	162
4.5 Magma generation at constructive plate boundaries	171
4.6 From mantle to crust	177
4.7 Interactions with the oceans	190
Summary of Chapter 4	198
Learning outcomes for Chapter 4	198

ANSWERS TO QUESTIONS	A1
APPENDICES	A11
A The elements	A11
B SI fundamental and derived units	A14
C The Greek alphabet	A15
D Enlarged figures	A16
FURTHER READING	A18
ACKNOWLEDGEMENTS	A19
SOURCES OF FIGURES, DATA AND TABLES	A21

An introduction to the structure and composition of the Earth

Chapter I

The Earth is a special place, not least because it is our home planet. It is the only place in the Universe where we are sure we can survive and thrive and, as far as we know, the only place where life has evolved. Despite the efforts of numerous space probes and ground-based searches for both extraterrestrial life and intelligence, at the time of writing (2006), this is all there is.

Perhaps the failure of projects designed to discover intelligent life elsewhere, such as the Search for Extraterrestrial Intelligence (SETI), is not so surprising if you consider the scale of the Universe and the time it takes for signals to travel between stars. Our closest stellar neighbour, for example, is four light-years away and even the briefest text message would have to wait eight years for an answer, assuming that it was decipherable by the recipient. But distance is not the only factor in our apparent isolation. Life needs the right conditions to start and an environment that is not so extreme that it is annihilated before it has had time to evolve. Planet Earth provides such an environment; an environment that all too often is taken for granted and exploited at great risk to our future viability, both as a civilisation and as a species.

But how did that environment evolve? Where did the atmosphere, the oceans, and the rocks come from? Why are there oceans and continents? Has the Earth always looked like it does today or has it changed through time? These and many other similar questions are fundamental to our understanding of the evolution of Earth and life and to the foundations of modern Earth system sciences – the study of Planet Earth as a whole, functioning system.

Our everyday experience of the Earth is probably one of underlying constancy and unchanging stability. The mountains and hills appear unchanged, rivers continue to flow to the sea and the oceans are always in the same place. Only plants and animals respond to the seasons, but even those follow an apparently unchanging annual cycle controlled by the constant orbit of the Earth about the Sun. Geology, it seems, remains a solid foundation on which biological activity relies. But the simple scene of geological inactivity is not the same everywhere and at all times.

Consider, for example, cliff erosion and the landscape where the land meets the sea. There are many examples where buildings and land that were once thought to be secure are now precariously perched on cliff tops ready to slip into the sea during the next severe storm.

At other times, a tranquil landscape is disturbed not by weather and climate but by internal forces that change the Earth's surface for ever. Earthquakes lay waste to large areas, reducing buildings, towns and sometimes cities to piles of rubble at terrible cost to individuals and society at large. Dormant volcanoes come to life in paroxysmal explosions, inundating the surrounding area with thick volcanic ash and lava. One such event happened in May 1980 when Mt St Helens in Washington State, USA, came to life after a period of over 150 years of quiescence and within a very short period of time destroyed itself and large swathes of the surrounding forests. Spirit Lake, so beloved of postcard photographers and tourists alike, was emptied of its contents and filled with volcanic rock debris in a matter of minutes as the side of the volcano slipped away (Figure 1.1).

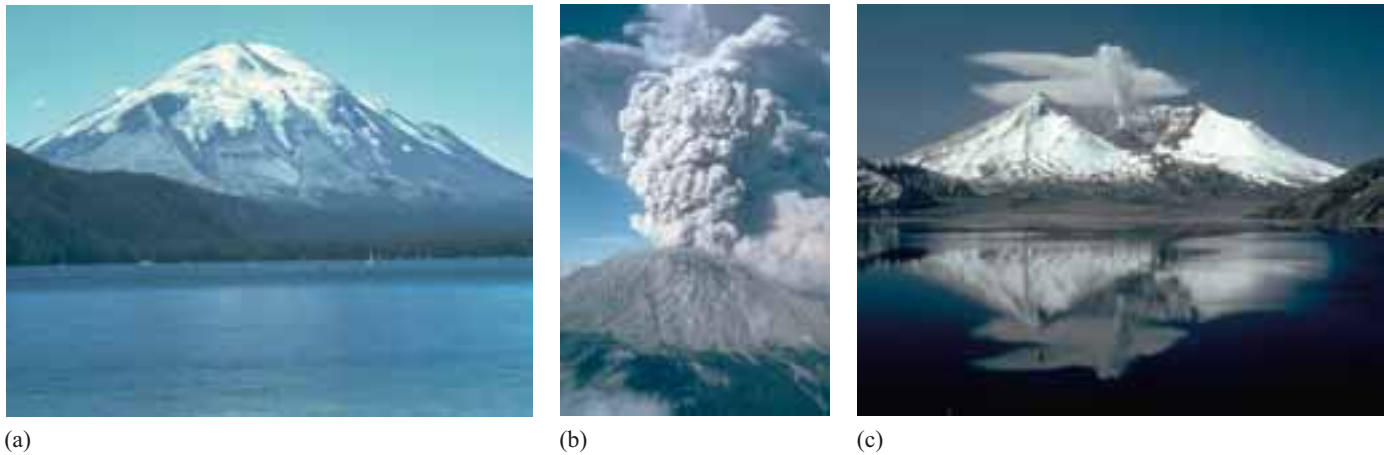


Figure 1.1 Mt St Helens in the USA: (a) before, (b) during and (c) after the eruption of May 1980. By volcanological standards, this was a modest event, erupting $\sim 1 \text{ km}^3$ of rock in an eruption column that reached about 10 km altitude. The geological record reveals numerous volcanic eruptions in the recent past (up to 1 Ma ago) that were 100 or even 1000 times larger than this one. (USGS)

Volcanic eruptions and earthquakes are probably the most dramatic of natural events that illustrate the powerful forces that shape the fabric of the landscape and the topography of the planet. Individually, they may prove catastrophic to the local environment, but they represent the surface expression of the processes that have produced the continents, atmosphere and oceans over the history of the Earth. If we are to understand these processes we need to expand our mental horizons beyond the everyday and explore the depths of the Earth and the duration of geological time. On these different levels and timescales we see a very different planet: one in a state of dynamic activity where the continents move across the surface, ocean basins are created and destroyed and mountain belts rise and fall in concert to the pulse of the Earth's internal energy.

Such, then, is the aim of this book – to stretch your mental horizons by:

- describing and analysing the way the solid Earth works
- investigating the composition of the Earth and its constituent layers
- revealing how the Earth has evolved from a cloud of gas and dust via an ocean of magma to a planet with a solid crust, oceans and a breathable atmosphere.

To do that you will need to call on physics and chemistry, mathematics and geology – these disciplines provide knowledge and techniques that should help you to understand our dynamic planet.

In this first chapter, you will be introduced to your home planet. The first part explores very briefly what is unique about the Earth when compared with the Moon and other terrestrial planets: Mercury, Venus and Mars. The second part of the chapter investigates the structure of the Earth: how it is divided into core, mantle and crust, and the evidence on which that division is based. Finally, the lines of evidence concerning the composition of the Earth and its constituent layers are drawn together from various sources, such as meteorites and samples of the mantle. This should give you the necessary background information to embark on your journey into the depths of the Earth and back through geological time.

► **Figure 1.2** Images of the surface/planetary discs of (a) the Earth; (b) the Moon; (c) Mars; (d) Venus; (e) Mercury. (NASA)

1.1 Planet Earth is unique

Many everyday aspects of our environment that we take for granted are most unusual in the Solar System – yet they contribute to making the Earth as amenable to life as we know it. Our knowledge of the surface features and conditions of all four terrestrial planets, i.e. Mercury, Venus, Mars and the Earth, together with the Moon, is becoming ever more detailed thanks to the spectacular results from numerous interplanetary probes (Figure 1.2).

Study the images in Figure 1.2, and then consider the following question.

- What are the characteristics of the Earth's surface that make it so favourable for life?
- The presence of liquid water and the availability of oxygen in the atmosphere.

None of the other terrestrial planets currently has liquid water on its surface and, although Mars has water locked away in its ice caps and reveals geological evidence for the presence of water in the past, its atmosphere is thin and dominated by carbon dioxide (CO_2) and nitrogen (N_2). Only Venus has a thick atmosphere, but it is dominated by CO_2 , sulfur dioxide (SO_2) and surface temperatures and pressures that would be intolerable for life. The other two bodies, Mercury and the Moon, are small, dry and have insignificant atmospheres.

These differences, although critical to life, appear superficial to the solid planets that lie beneath the atmospheres (see Box 1.1). Consequently, the question that arises is whether or not the solid planets share other common features and structures.

Box 1.1 Earth's interacting systems

In this book, the different components of the Earth are divided according to their physical state into the atmosphere, hydrosphere and geosphere. The term **atmosphere** should be familiar to you: it describes the envelope of gases that surrounds the Earth. The **hydrosphere** may be less familiar, but it refers to all of the water on the planet – in the oceans, seas, lakes and rivers. Part of the hydrosphere is locked up in the form of ice and is sometimes referred to separately as the **cryosphere**. The solid Earth, which makes up the bulk of the planet by mass, is known as the **geosphere**. The geosphere is the main subject of this book, but the atmosphere and hydrosphere also interact with the solid Earth through a variety of processes that will be explored later. Finally, all forms of life are collectively described as the **biosphere**.

It is useful to consider the relative sizes of the different Earth systems and their influence on our lives. The Earth has a radius of 6371 km, yet the atmosphere becomes too thin to breathe just above the top of the highest mountains, which is about 10 km above sea level. Even though atmospheric



(a)



(b)



(c)



(d)



(e)

physicists extend their range of study to the boundaries of space some 300 km above the surface, atmospheric pressures for most of this distance are less than 1% of that at sea level.

The oceans appear vast and the dark ocean floor is more alien to us than the face of the Moon, yet on average it is only 3 km below the surface.

- The Atlantic Ocean is about 5000 km across at its widest. If that width was reduced to a puddle 5 m across, how deep would the puddle be to maintain the true scale of the ocean?
- As 1 km = 1000 m and 1 m = 1000 mm, the scale reduction is 1:1 000 000, so our puddle should be

$$\frac{3 \text{ km}}{1\,000\,000} = 0.000\,003 \text{ km} = 3 \text{ mm deep.}$$

While the oceans appear very deep, on the scale of the whole Earth they represent a very thin film on the Earth's surface – which is about 0.05% of the Earth's radius deep.

And what of the biosphere? Life extends to about 10 km above the surface. It also appears that life can exist at the greatest depths yet explored in the oceans, again about 10 km below sea level, making the biosphere about 20 km thick. But, as with the oceans, on the scale of the whole Earth this represents just 0.3% of the Earth's radius. No wonder then that the biosphere has been described as nothing more than 'a thin green smear' on the surface of the Earth.

1.1.1 Surface features

The Earth's surface is characterised by continents and oceans, but is there evidence for similar features on the other planets?

- Study Figure 1.2 again. Are the surfaces of the different planets uniform or diverse?
- Diverse: with the possible exception of Venus, each image has dark and light regions and areas with many craters and other areas with relatively few.

While only the Earth has a permanent hydrosphere, there is clearly a contrast on the other planets between the highland regions, which may be equivalent to the continents on Earth, and the lowlands or basins, which may be analogous to terrestrial oceans. Some planetary surfaces are also pockmarked with craters, while on others such features are rare.

The Moon

The Moon orbits the Earth; some of the other planets have moons.

Our nearest neighbour, the Moon, is the Earth's constant companion. Results from the Apollo missions in the 1960s and 70s revealed just how old the Moon is – it is in excess of 4.4 Ga and is as old as the Earth itself. However, the Moon is exceptional, not because of its age – all planetary bodies orbiting the Sun are of

roughly the same age – but because of its size. In relation to the mass of the Earth, the Moon is the largest of any of the satellites of the other planets. Ganymede and Callisto, which are moons of Jupiter, and Titan – the largest moon of Saturn – may be slightly larger than the Moon, but they are tiny in relation to their parent bodies. Of the terrestrial planets, the only other planet with satellites is Mars. These small objects, named Phobos and Deimos, appear to have been captured from the **asteroid belt** lying between Mars and Jupiter, and are a relatively late addition to the Martian system.

The Moon has two different surfaces: the **highlands**, which are pale-coloured regions with a dense covering of impact craters, and **maria**, which are darker, flatter regions with fewer craters. This suggests a division into elevated regions that are analogous to continents and depressions that were once thought to be filled with water – hence *maria* (Latin for sea). The number of craters on a planet's surface is related to its age – more craters means an older surface – so, even before the Apollo missions in the 1960s and 70s, scientists realised that the highlands are older than the maria. What was not known was the absolute age of the rocks until samples were returned and subjected to radiometric dating – the determination of the age of a rock using naturally occurring radioactive isotopes.

Direct dating of samples from different areas of the Moon revealed that the highlands date from 4.4 Ga, whereas the maria are younger at 4.2–3.8 Ga, confirming and quantifying the conclusions from crater counting. The study of Moon samples also indicates that these two surfaces are made of different rock types:

- the maria are dominated by **basalts** that are similar to those found on Earth
- the highlands are made of a rock known as **anorthosite**, which is an uncommon rock type on Earth that is rich in **plagioclase feldspar**.

The craters and the surface ages show that the Moon experienced a period of intense meteorite bombardment early in its history, but that this had declined by the time the maria formed.

Mercury, Venus, Earth and Mars

Similar cratering studies of the other terrestrial planets have revealed an even greater diversity in surface ages. Mercury, for which there is only limited information, appears very similar to the Moon with a highly cratered ancient surface. The different surfaces of Mars also appear similar to those of the Moon: cratered highland regions surround basins with fewer impact craters. Although this division is less easy to see in Figure 1.2, and the absolute ages of these different surfaces are as yet unknown, ages of 3.8 Ga have been suggested for the highlands – assuming a similar impact history to the Moon.

Radar images of the surface of Venus reveal a somewhat different story, with a lower cratering density suggesting a much younger surface age. Impact craters on Earth are much less easily recognised. Certainly impact features similar to those on the surface of the Moon are rare.

- What do these observations suggest about the mean age of the surface of the Earth compared to that of the Moon?
- The Earth's surface must be considerably younger than that of the Moon.

This conclusion is confirmed by many observations on the ages of rocks at the Earth's surface today. Even though the oldest terrestrial rocks are almost 4 Ga old, most rocks at the surface of the Earth are much younger, with ages of tens to hundreds of millions of years. Most notably, no rocks recovered from the oceans are older than 200 Ma.

The cratering record on the Moon, Mercury and parts of Mars suggests an extraordinary violent birth and early life for these planets, involving innumerable impacts of asteroid-sized objects – a violent early history that must have been shared with the Earth and Venus. This period of time on Earth, between its formation at 4.55 Ga and the ages of the earliest rocks at ~ 3.9 Ga, is known as the **Hadean**. As there are no rocks on Earth older than ~ 3.9 Ga, much of our knowledge of the processes that affected Earth during this period is derived from comparison with other planetary bodies.

- Why do you think there is no direct record of the earliest surfaces of the Earth?

- They have been removed by the effects of surface processes.

In part, the lack of a terrestrial record of this time (i.e. the Hadean) is a consequence of surface processes. Unlike the Moon, which has no atmosphere, the Earth's surface, and to a lesser extent the surface of Mars, is being continually reworked by chemical and physical interactions with the atmosphere and hydrosphere by the processes of **weathering** and **erosion**. On Mars, the thinness of the atmosphere and the ephemeral nature of the hydrosphere allow older features such as impact craters to be preserved for longer than on the other planets.

- Do the same arguments concerning weathering and erosion apply to Venus?

- In part yes, but Venus does not have a hydrosphere, and as many terrestrial surface processes require a flowing liquid, the relative youth of the surface of Venus may have a different explanation.

There are other processes that relate to the deeper working of the planets that also renew planetary surfaces on geological timescales, for example volcanic activity. These are clearly apparent on the Earth and are the subject of scientific investigation on Venus, but have not been dominant in the evolution of Mercury or the Moon and are only occasionally active on Mars. Such internal processes, as you will discover later in this book, are driven by the internal heat of a planet and so are related to the planet's size. Given that Earth and Venus are of a similar size and are the largest of the terrestrial planets, the relatively young ages of their surfaces also relate to how much heat each holds deep in its interior.

- Assuming the Earth dissipates its heat through the processes of plate tectonics, which of the terrestrial planets might also be expected to show plate-like features?

- Venus, because of its similar size.

Figures 1.3 and 1.4 (overleaf) show the surfaces of the Earth, Venus and Mars in false colours to denote the topography of each planet. Shades of blue represent depressions or basins with elevations below the mean altitude of the surface while shades of red denote highlands. If the water from the oceans could be removed to reveal the topography of the ocean floors, you would see something like the image in Figure 1.3.

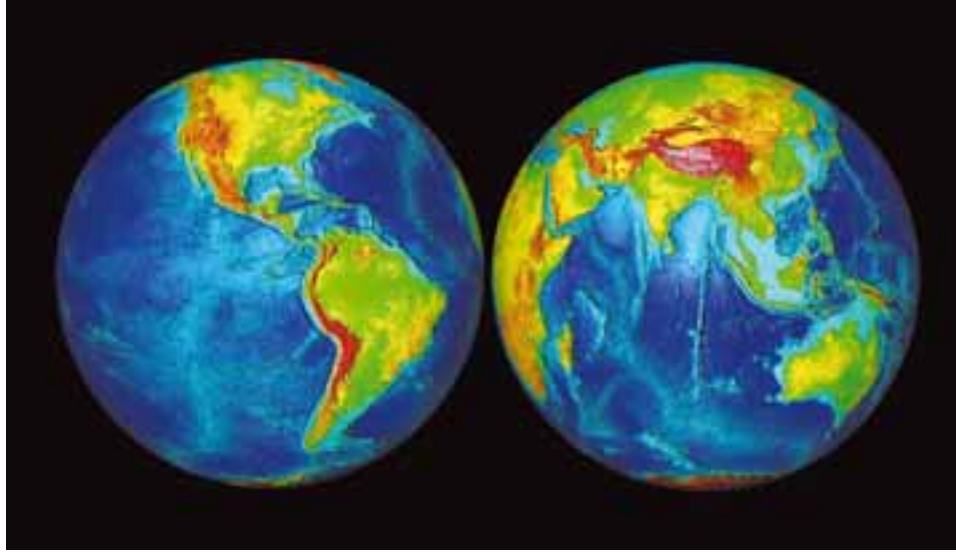


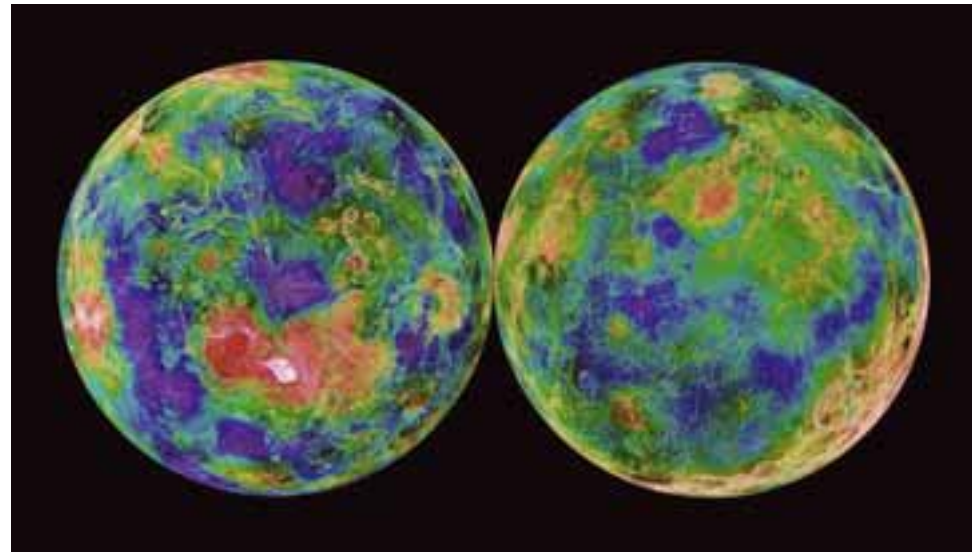
Figure 1.3 Topography of the Earth with the oceans removed. The blue areas corresponding to the Pacific and Indian Oceans indicate low elevations – the deeper the colour, the lower the elevation. Pale-blue areas close to shorelines are equivalent to the continental shelf; land is represented by shades of colour – green: 0–500 m (e.g. Amazon Basin); yellow: 500–1000 m (e.g. Western Australia); brown: 1000–2000 m (e.g. Iran and Afghanistan); red: 2000–5000 m (e.g. Andes); and grey: >5000 m (e.g. Himalaya). (NOAA/NASA)

The ocean basins are far from flat or uniform. Beneath the sea are hidden some of the most dramatic topographic features on Earth – huge volcanoes, deep trenches and the longest mountain chain on Earth, the ocean-ridge system, which links all the ocean basins together. The ocean ridges mark regions on the Earth's surface where new crust is being created. If new crust is being created then older crust must be being absorbed back into the Earth's interior (assuming that the Earth is not expanding), and the deep ocean trenches that surround the Pacific Ocean and lie south of Indonesia in the Indian Ocean mark the places where this is happening. So if similar recycling processes are happening on other planets, then similar surface features should also be visible.

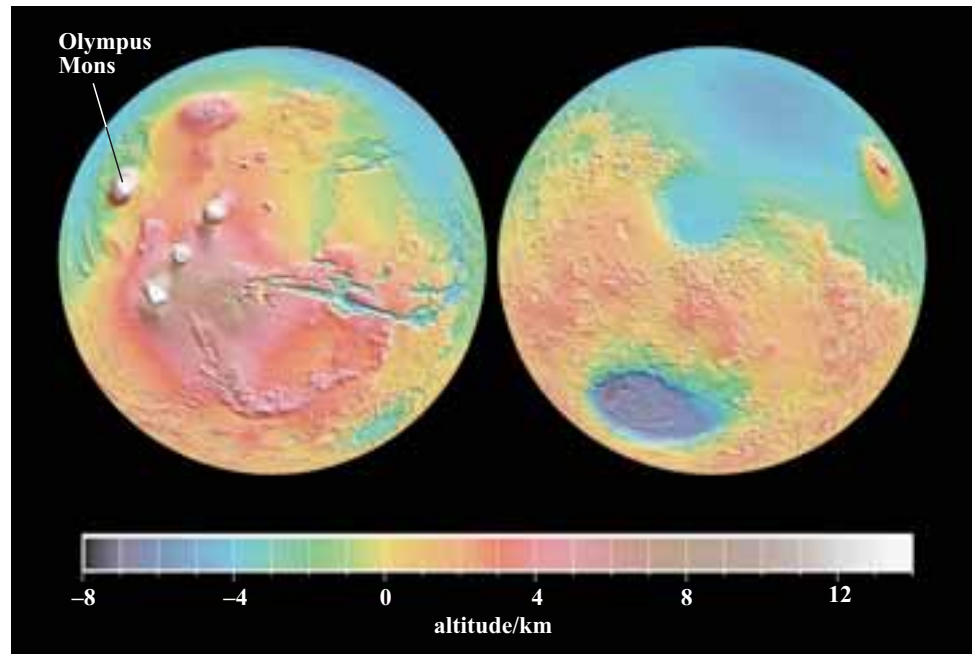
- Compare Figure 1.3 with Figure 1.4a. Do you see any structures or features on Venus that resemble ocean ridges or trenches?
- No. There are no such features on Venus.

Superficially, none of the major features on the surfaces of either Venus or Mars suggest plate tectonics is an active process on either planet. Both planets have volcanic features – Olympus Mons on Mars is the largest volcanic structure in the Solar System and such features are the external expressions of internal processes, either active now or in the past. But major features that might be related to plate tectonics, such as ocean ridges, are not apparent. Taken with the ages of the surfaces from the cratering record, the evidence indicates that plate tectonics is not an active process on either planet.

Figure 1.4 False-colour images of (a) Venus and (b) Mars, colour-coded to provide elevation information across the total topographic range of each planet. The blue (e.g. circular region in centre of left Venus image) denotes areas 1–2 km below the mean surface elevation (mse); green denotes 0–1 km below mse; yellow denotes 0–1 km above mse and red denotes areas 1–2 km above mse. Higher elevations are in shades of pink to white (small area below centre of left Venus image). The total topographic range on the image of Venus is much less than that of Mars. (NASA)



(a)



(b)

While the cratering record can be used to provide information about the relative ages of surfaces and the intensity of processes that modify the surface, it does not tell us anything about the internal structure, such as whether other planets have continents and oceans with Earth-like compositions and structures. Do these different surfaces reflect the differences between continents and oceans, as on Earth?

We can investigate the similarities and differences quantitatively, using a device known as the **hypso metric plot**. The hypso metric plot is a histogram of topographic height over a whole planet, with heights determined above or below the median surface. The plot for Earth is shown in Figure 1.5a.

- The hypsometric plot for Earth is obviously bimodal. By comparison with Figure 1.3, what do you think each of the two peaks in Figure 1.5a represents?
- The peak at relatively low height (i.e. the one on the left) represents the floor of the oceans (which are typically at 3–6 km below sea level) and the peak at greater relative height represents the surface of the continental crust (which is mostly less than 1 km above sea level).

The hypsometric plot clearly reveals the difference in elevations between the oceans and the continents on Earth, and we shall return to the causes of this difference later in the chapter.

- How does the hypsometric plot for Venus (Figure 1.5d) compare with that for the Earth?
- It is different. Although both have a peak between 0 and 1 km, this is the only peak in the Venusian data (Earth has two peaks) and the range of elevation is less.

Venus is often described as the Earth's sister planet because the two planets have similar sizes. Yet the contrasting hypsometric plots strongly suggest that their crustal structures are fundamentally different. This comparison may be flawed because of the different surface processes that operate on Venus compared with on Earth, largely because of the lack of a biosphere and a hydrosphere on Venus. Nevertheless, it is difficult to see how a unimodal distribution of heights could be transformed into a bimodal distribution, or vice versa, simply as a consequence of surface processes.

The hypsometric plot of Venus, therefore, suggests that the surface features are very different from those on Earth. The division into continents and oceans that is apparent on Earth is lacking on Venus and also on Mars. Continents and oceans (with or without water) appear to be another unique feature of Earth, and later in this book you will discover how they came into being and the processes responsible for their formation.

In summary, in addition to an oxygen-rich atmosphere and a permanent and active hydrosphere, the differences between the Earth and other terrestrial planets are more than just skin deep. Continental and oceanic crust, youthful surfaces, the presence of plate tectonics and the coupling of the Earth with the Moon are all features that are unique to our planet. In addition, the Earth has a strong magnetic field and a relatively rapid rate of rotation. In the following chapters you will explore how some of these unique features have come about, the processes that produced them, and the processes they control.

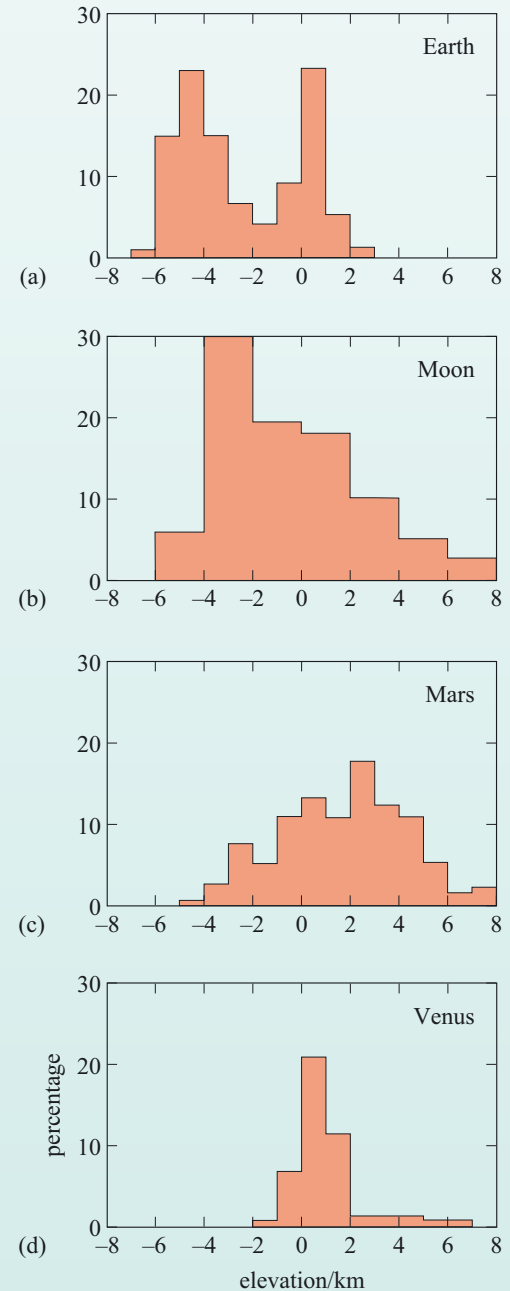


Figure 1.5 Hypsometric plots for (a) the Earth, (b) the Moon, (c) Mars and (d) Venus showing the percentage of the solid surface above and below sea level for the Earth, and mean surface elevation for the Moon, Mars and Venus. (Watts, 2001)

I.2 The structure of the Earth

Our knowledge of the structure of the Earth comes from a variety of sources, but the most fruitful line of discovery is through the study of earthquakes and seismic waves – the science of **seismology**. Earthquakes are familiar to most people, if not through first-hand experience, then via reports in the media when they lay waste to areas of population and result in the deaths of sometimes many thousands of people. It is perhaps ironic that these most violent of natural phenomena, that have killed so many people throughout history, are eagerly awaited by seismologists, providing yet more opportunities to probe the interior of our planet.

Earthquakes result from the release of strain energy that builds up in parts of the Earth where large bodies of rock are moving relative to each other. Such an example is the San Andreas Fault in California, where a slither of western California is moving northwards relative to the rest of North America. While the relative movement continues gradually, no earthquakes occur, but should the fault become locked for some reason, then the rocks become increasingly deformed or strained. When the fault unlocks, the energy stored as strain is released almost instantaneously and dissipates in the form of seismic waves. It is these seismic waves that both deliver the destruction and provide information on the structure of the Earth.

I.2.1 Seismic waves – Earth probes

Seismic waves are of four main types: P, S, Love and Rayleigh. The latter two types are confined largely to the surface of the Earth and, while they are the most destructive and are used to measure the magnitude of an earthquake, they do not provide information about the interior of the Earth, so they will not be considered further here. By contrast, P-waves and S-waves are both **body waves** and travel through the interior of the Earth.

P-waves involve the transmission of pulses of compression and dilation (expansion), in which the particles of the medium vibrate backwards and forwards in the direction the wave is travelling (propagating). P-waves are like sound waves, but at much lower frequencies than can be detected by the human ear.

S-waves involve shear displacements, meaning that the motion of particles in the medium is at right angles (perpendicular) to the direction of wave propagation. The motion can be vertical, horizontal, or a combination of the two, depending on the nature of the earthquake. S-waves are similar to the motion induced when a taught rope is shaken.

The speed of a seismic wave is related to the physical properties of the medium through which the wave travels. In general, the wave speed is a function of two properties – the **elastic modulus** and the density – and is of the form:

$$\text{speed} = \left(\frac{\text{elastic modulus}}{\text{density}} \right)^{\frac{1}{2}} \quad (1.1)$$

The elastic modulus is a measure of how much deformation or strain occurs in a material when it is subjected to a given amount of stress, and is defined as the ratio of stress to strain. Because materials can be strained in a number of ways, they can have more than one elastic modulus. For example, the *bulk modulus* relates to the compressibility of a medium whereas the *shear modulus* describes how rigid that medium is. For P-waves, the speed depends on both the bulk modulus and the shear modulus, whereas for S-waves it depends only on the shear modulus.

The equation for the speed of P-waves, v_P , is:

$$v_P = \left(\frac{K + \frac{4\mu}{3}}{\rho} \right)^{\frac{1}{2}} \quad (1.2)$$

where K is the bulk modulus, μ is the shear modulus and ρ is the density. The equation for the speed of S-waves, v_S , is:

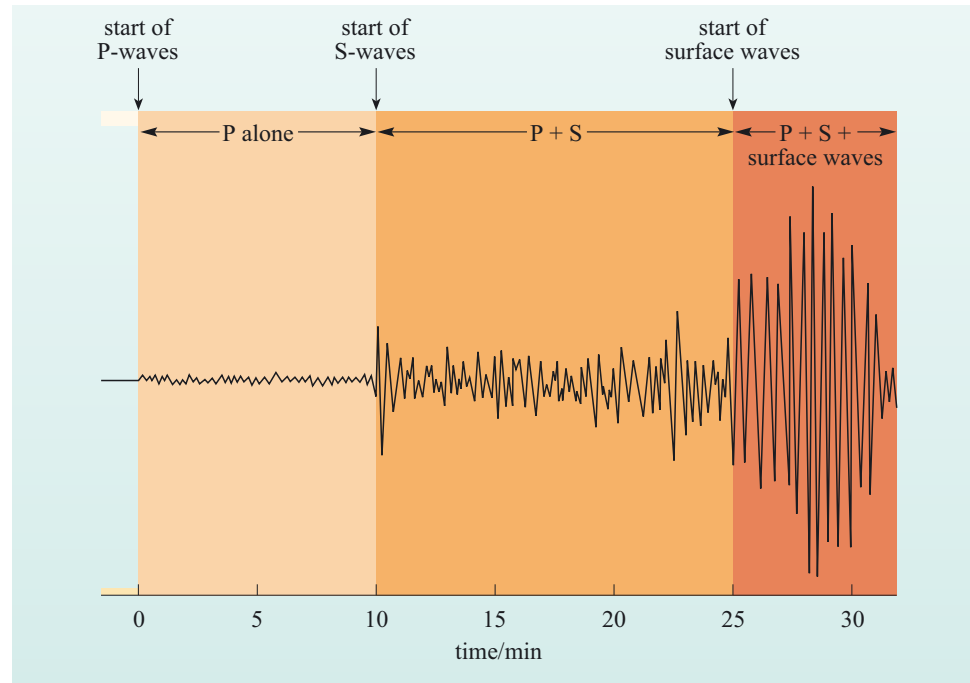
$$v_S = \left(\frac{\mu}{\rho} \right)^{\frac{1}{2}} \quad (1.3)$$

where μ is the shear modulus and ρ is the density.

- Can you see why v_P is greater than v_S ?
- Given that both K and μ are positive numbers, $K + \frac{4\mu}{3}$ is always greater than μ , so v_P will be greater than v_S .
- What do Equations 1.2 and 1.3 infer about P- and S-wave speeds in liquids of similar densities to solids?
- Given that a liquid has no rigidity (it occupies the shape of any containing vessel), it cannot be deformed. Hence μ must be zero. Therefore, liquids cannot transmit S-waves. Also, for a given density, the P-wave velocity is lower in a liquid than in a solid.

Seismic waves are detected using instruments known as seismometers. A simple example of the output from such an instrument, called a seismogram, is shown in Figure 1.6. Because P-waves always travel faster than S-waves, P-waves always arrive at a seismic station first. The speed of the seismic waves through the Earth can be measured in the laboratory, so the difference in their arrival times can be used to calculate the distance a seismogram is from the centre of an earthquake using the simple relationship between speed, distance and time.

Figure 1.6 A simple seismogram, in this case produced by an earthquake in Turkey and recorded in the USA (Cambridge, Massachusetts). Notice the difference in the displacements caused by the different types of waves – the first arrival of P-waves followed by S-waves and then the surface waves (Love and Rayleigh).



So far, all references have been made to the speed of seismic waves. However, objects and waves have two attributes to their motion – a magnitude of motion and a direction. These are the two attributes of velocity. In everyday speech the two terms are frequently used interchangeably, but in science the difference between speed and velocity is quite definite. Subsequently, all references to seismic waves will use the term ‘velocity’, as the propagation of seismic waves through the Earth involves changes in both magnitude and direction.

1.2.2 Seismic refraction

Seismic velocity is a function of both density and the physical state of the deep Earth, and determining the variation of seismic velocity with depth in the Earth has been a major factor in the determination of the structure of the Earth’s interior.

The rules that govern the propagation of seismic waves through the Earth are similar to those that govern the passage of light through transparent media, such as glass and water. You have probably noticed the way in which a stick appears to be bent when partially immersed in water. The light from the stick below the water is bent or **refracted** at the water’s surface because the velocity of light in water is different from that in air. So it is with seismic waves as they pass through the different layers of the Earth. This is illustrated in Figure 1.7. The angle of refraction (r) is directly related to the angle of incidence (i) and the ratio of the seismic velocities of the two layers, such that:

$$\frac{\sin i}{\sin r} = \frac{v_1}{v_2} \quad (1.4)$$

This equation summarises what is known as **Snell’s law** and applies to all forms of wave energy, including light and sound as well as seismic waves. Boundaries also reflect some of the incident energy; a possible reflected wave path is shown in Figure 1.7.

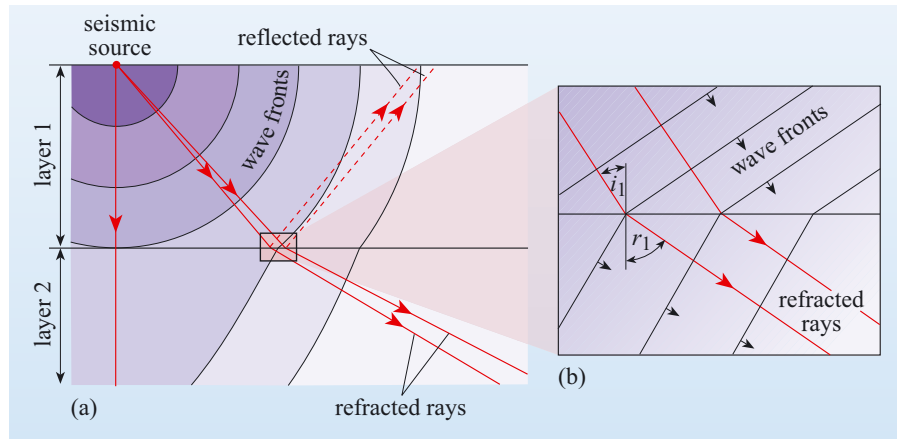


Figure 1.7 Refraction of seismic waves. In (a), evenly spaced wavefronts spread out from a seismic source. When they enter the lower layer with a higher velocity, they change orientation and some of the energy is also reflected. In (b), two wave paths (in red) are also shown tracing two different points on the wave front (in black).

Within the Earth, most changes in seismic velocity take place continuously with depth, such that refraction occurs as a continuous process.

- Imagine the Earth consists of a series of thin concentric shells each with a seismic velocity slightly greater than the one above. What will be the shape of the ray path?
- The ray path will be refracted to a shallower angle at each boundary and curve back towards the surface.

The trace of a ray path will be similar to that in Figure 1.8. At each boundary between successive layers, the ray path is refracted to increasingly shallow angles. Eventually, as seismic velocities increase and the angle of incidence increases, a particular angle is reached at which

$$\frac{v_1}{v_2} = \sin i, \text{ i.e. } \sin r = 1, \text{ meaning that } r = 90^\circ.$$

- What do you think happens to the wave path when it strikes a boundary at the critical angle?
- It is reflected rather than refracted.

The angle at which this occurs is known as the **critical angle**, and the wave then traces a new path back towards the Earth's surface, some distance from the original seismic source.

The distance of a seismic receiver from a seismic source can be measured in kilometres but is usually expressed in terms of the angle subtended at the centre of the Earth (as shown in Figure 1.8) and is denoted by Δ (delta). A plot of arrival times for different seismic waves against Δ (the **epicentral angle**), known as a **travel-time curve**, is shown in Figure 1.9.

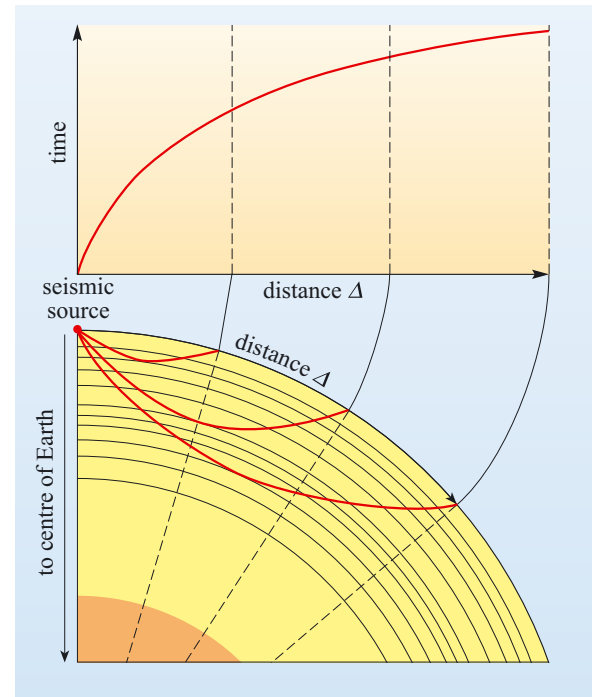


Figure 1.8 Schematic paths of seismic waves in a multiple-layered Earth. Rays are refracted and reflected at each boundary and the overall ray path curves towards the surface. The distance from the earthquake is measured as the angle subtended at the centre of the Earth between the receiving station and the earthquake source.

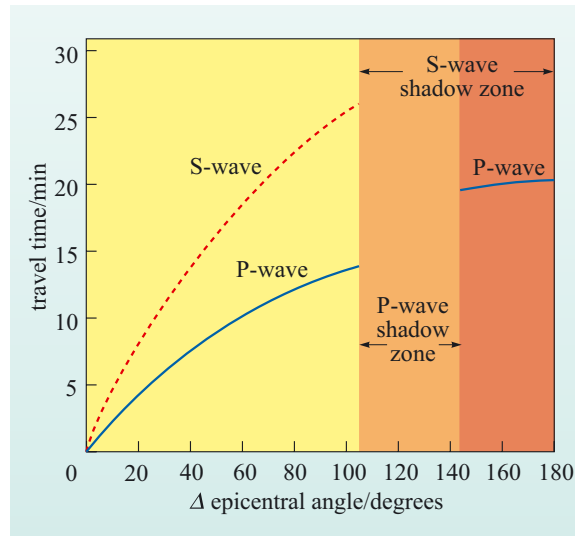


Figure 1.9 P-wave and S-wave travel times for the whole Earth. There is a P-wave shadow between epicentral angles 105° and 142° and an S-wave shadow at all angles greater than 105° .

- Study Figure 1.9. Do the arrival times of S- and P-waves vary continuously for all epicentral angles?
- No. Arrival times of P-waves increase up to an epicentral angle of $\sim 105^\circ$, after which there is a gap before they appear again at about 142° , but at much later times. Arrival times of S-waves also increase up to $\sim 105^\circ$ but there are no S-waves at epicentral angles greater than 105° .

The areas of the Earth's surface at these epicentral angles where there are no seismic waves are known as **shadow zones** and they are more or less the same for any earthquake, no matter where on the Earth it originates. This simple observation shows that the internal structure of the Earth must be radially symmetric.

An everyday analogy for the structure of the Earth is that of a peach, with a core (the peach stone) surrounded by what is known as the mantle (the peach flesh). (One might pursue this analogy further and suggest that the Earth's crust is represented by the peach skin.)

You have seen from Figure 1.8 that waves which emerge at the greatest epicentral angles have penetrated to the greatest depths. In Figure 1.9 you can also see that at epicentral angles of less than 105° the gradient of the travel-time curve decreases as Δ increases.

- What does this tell you about seismic velocity changes with depth in the Earth?
- The gradient of the curve is equal to time divided by distance, which is the inverse of speed. As the gradient decreases with epicentral angle this must mean that, up to a certain depth represented by an epicentral angle of 105° , seismic P- and S-wave speeds increase.
- What does the S-wave shadow suggest about a part of the Earth's interior?
- You will recall from earlier in this chapter that S-waves do not travel through liquids. The S-wave shadow, therefore, shows that a significant part of the Earth's deep interior is liquid.

This important break in the seismic signal at 105° is known as a **discontinuity** and this one, by definition, marks the boundary between the Earth's core and the overlying mantle. Figure 1.10 shows how this discontinuity comes about by tracing the paths of a number of seismic waves through the Earth. The abrupt velocity decrease at the core–mantle boundary causes rays entering the core to be refracted towards the vertical and so producing the P-wave shadow starting at an epicentral angle of $\sim 105^\circ$. The refracted wave passes through the outer core and emerges, much later than would be expected if the core were not there, at a much greater value of Δ . Other wave paths that strike the core–mantle boundary at lower incident angles pass through the outer and inner core and the minimum epicentral angle at which they can emerge is 142° .

Figure 1.10 also shows other wave paths through the Earth and labels them according to a particular code that is related to the different Earth layers through which they have passed, and whether or not they were reflected or refracted. For example PKP represents a P-wave that passed through the mantle, followed by the outer core and then back to the surface through the mantle as a P-wave. SKP denotes a shear wave that converted to a P-wave at the core–mantle boundary, was refracted through the core and was then refracted back through the mantle to the surface.

- What do you think PKIKP denotes?
- A P-wave that passed through the mantle, through the outer core and the inner core, and then back through the outer core and mantle to the surface.

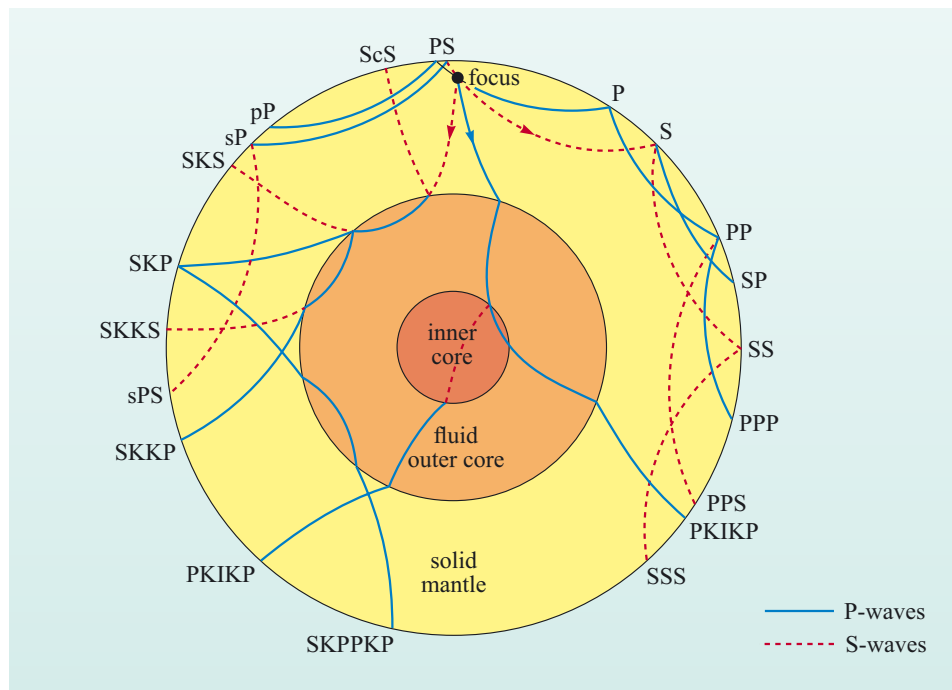
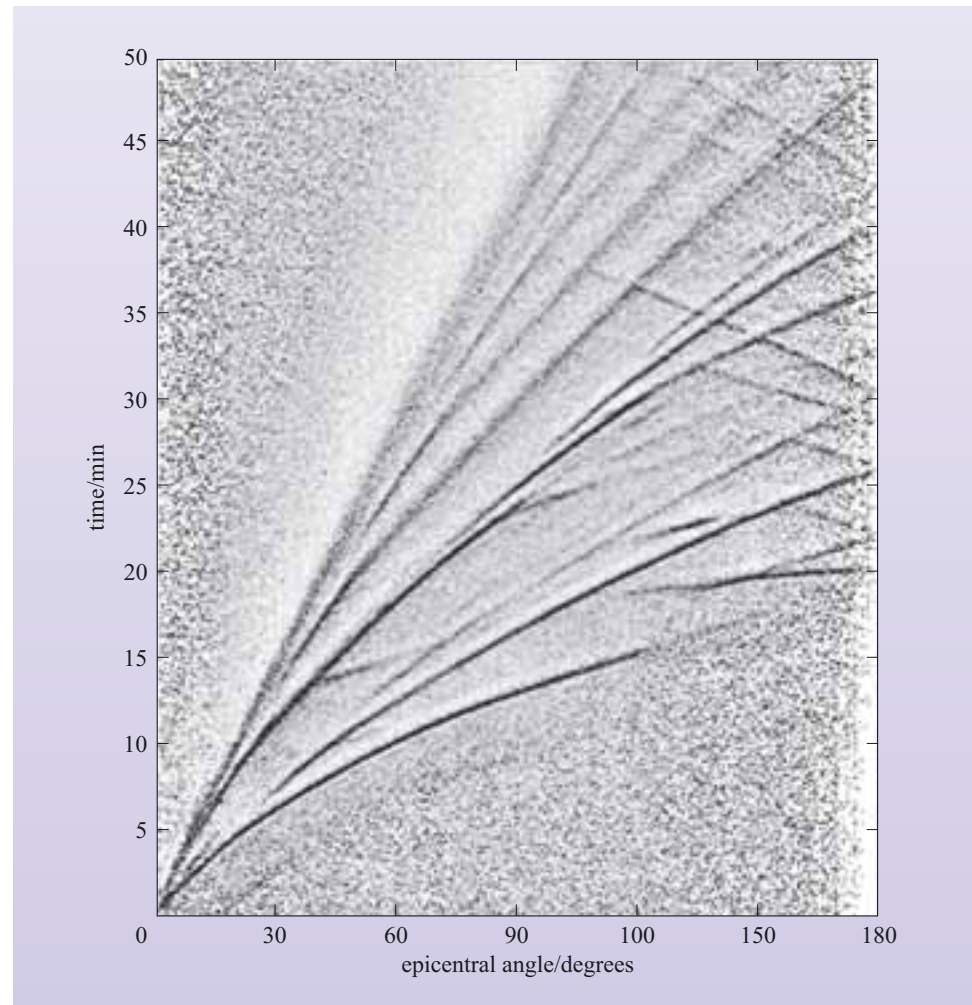


Figure 1.10 A selection of P- and S-wave paths through the Earth. The waves are emitted from the earthquake focus near the top of the diagram.

You should now realise that Figure 1.9 is a gross over-simplification of the real data from seismology and, to give you a flavour of the complexity of what seismologists deal with, a compilation of over 60 000 arrival time records relating to 2995 earthquakes are shown in Figure 1.11. In addition to simple P- and S-wave arrivals, this diagram shows arrival times for waves variously refracted and reflected within the Earth.

Figure 1.11 Travel-time curves from over 60 000 arrival records relating to 2995 seismic events recorded digitally between 1980 and 1987. (Earle and Shearer, 1994)



The above paragraphs describe how seismic waves pass through the Earth and how arrival times are determined by the Earth's structure. In real life, however, all the seismologist has to go on are the arrival times of P- and S- waves and a reasonable knowledge of the seismic velocities of various rock types as they occur at the Earth's surface. The process of converting these data into information on the internal structure of the Earth is complex and was originally achieved using a laborious mathematical procedure known as **inversion**. (The details of this procedure are beyond the scope of this book, but see Fowler (2005), Appendix 3, for a detailed explanation.) The application of inversion techniques to numerous earthquakes gave results that were so good that they formed the basis of a more modern approach, based on the use of fast computers to **forward-model** earthquake events. In forward modelling, an initial estimate of the depth–velocity profile from earlier inversion results is used to predict the

arrival times of a given seismic event. These predicted arrival times are compared with the actual arrival data and adjustments are then made to the velocity profile until the model matches the observed arrival times. This approach allows seismologists to identify regions within the Earth with anomalous seismic velocities with a good degree of accuracy.

1.2.3 The density of the Earth and Earth layers

While seismic refraction data can inform us of both the depth of seismic discontinuities in the Earth and the variation in seismic velocities with depth to quite remarkable accuracy, it is difficult to extract the density of the different layers within the Earth with the same degree of certainty from seismic velocities alone.

- What do you think is the reason for this?
- The two equations that relate P- and S-wave velocities to density (Equations 1.2 and 1.3) also contain other variables: the two elastic moduli, K and μ . Hence there is no unique solution.

It is possible to eliminate μ from Equations 1.2 and 1.3 by combining them:

$$v_p^2 = \frac{K + \frac{4\mu}{3}}{\rho} \text{ and } v_s^2 = \frac{\mu}{\rho}$$

So:

$$\mu = \frac{3}{4}(\rho v_p^2 - K) = v_s^2 \rho$$

Rearranging gives:

$$\frac{K}{\rho} = v_p^2 - \frac{4v_s^2}{3} \quad (1.5)$$

But Equation 1.5 still requires a knowledge of K , and how it varies with depth, to give any information about density, ρ . In 1923, two geophysicists, Adams and Williamson, formulated another equation that eliminated K . They considered how gravity and density would change through the Earth if density changed only as a result of the mass of the overlying layers, by what is known as **self-compression**. The so-called Adams–Williamson equation takes the form:

$$\frac{\Delta\rho}{\Delta r} = \frac{\text{change in density}}{\text{change in depth}} = \frac{\rho G m}{r^2 \left(v_p^2 - \frac{4v_s^2}{3} \right)} \quad (1.6)$$

Equation 1.6 is a differential equation that describes the change of density ($\Delta\rho$) with the change of depth (Δr) as a function of distance from the Earth's centre (r), the seismic velocities (v_p and v_s), density (ρ) and total mass (m) at that distance. G is the universal gravitational constant. It is applied by evaluating the right-hand

side of the equation at the surface, followed by successive layers at increasing depths. In practice, the integration begins at the top of the mantle, based on a density of 3200 kg m^{-3} – taken from occasional mantle samples found at the Earth's surface – and the process is repeated to the base of the mantle, where there is a change from solid to liquid, and a new density is chosen. Choosing a density for the core is more hit-and-miss than it is for the mantle but, with the added constraint that the total mass assumed in the calculation cannot exceed the known mass of the Earth, density variations through the core can be determined with a degree of certainty. The results of these calculations are summarised in Figure 1.12, which shows how density, gravity and pressure vary with depth in the Earth.

1.2.4 Earth reference models

The information summarised in Figure 1.12 is known as an **Earth reference model**. It describes the important physical properties within the Earth from the surface to the core. Such reference models, and there are more than one, provide the background structure against which anomalies can be compared and recognised. The agreement between different Earth reference models is now so good that seismic velocities throughout most of the mantle and core are known to $\pm 0.01 \text{ km s}^{-1}$, while density, pressure and gravity are defined equally well.

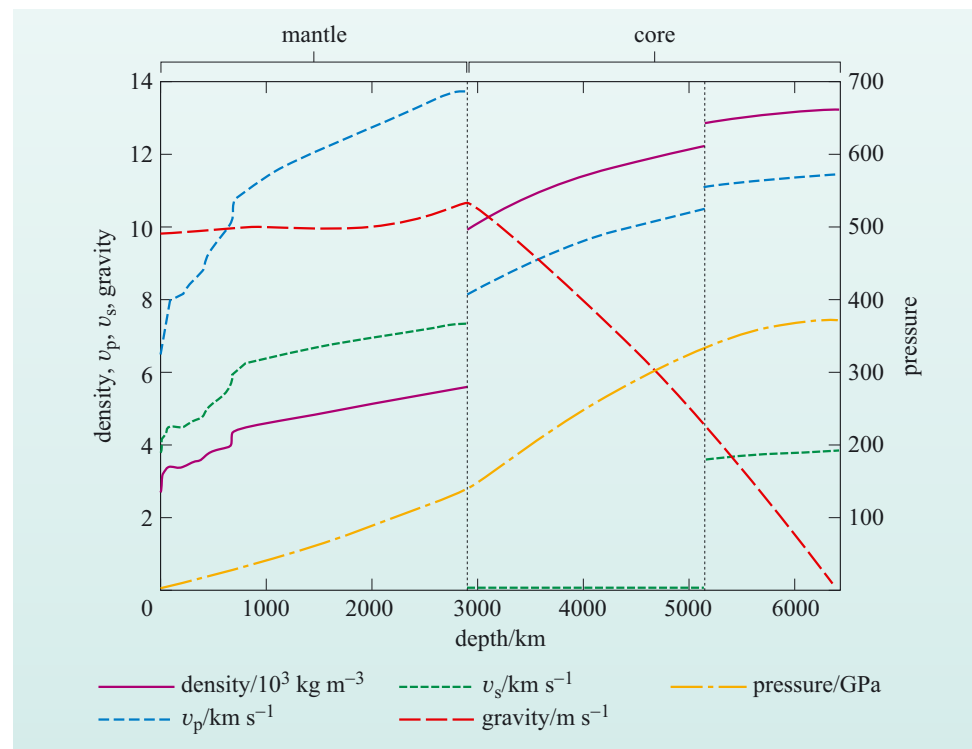


Figure 1.12 Variations of seismic (P- and S-waves) velocities, density, pressure and gravity in the Earth. (Based on the PREM reference model, Dziewonski and Anderson, 1981)

A rapid tour from the surface through the mantle reveals just how thin the Earth's crust is. The crust beneath the continents varies, with an average thickness of about 35 km, but ranging between 15 km and 75 km. Beneath the oceans the crust is much more uniform and thinner (7–8 km in thickness). Overall, the density of the continents is lower than that of the oceanic crust, with typical continental crustal densities of $2600\text{--}2800\text{ kg m}^{-3}$ compared with $2800\text{--}3000\text{ kg m}^{-3}$ for oceanic crust.

The crust is separated from the mantle by a major seismic discontinuity, the **Mohorovičić discontinuity**, which is almost invariably abbreviated to the **Moho**. The change in seismic velocity across the Moho is greater than in any other region of the Earth, except for the boundary between the core and mantle.

Below the crust is the mantle, which extends down to its boundary with the core at 2900 km. It is far from uniform, especially the uppermost part of the mantle, which in places shows low seismic velocities. The region beneath the Moho and down to 220 km is sometimes referred to as the **low velocity zone**. Below 220 km, seismic velocities increase smoothly to a depth of about 400 km where there is the first of two small discontinuities, the first at 410 km followed by a second at 670 km. Both involve increases in P- and S-wave velocities of 5–7%, and this region is sometimes referred to as the **mantle transition zone**. This is a region within the mantle where the crystal structures of the constituent minerals, olivine and pyroxene, are transformed into a dense mineral known as **perovskite**. The 670 km discontinuity marks the conventional boundary between the upper and lower mantle, which is also divided into two layers, denoted D' and D'' (D-prime and D-double prime). D' extends from 670 km down to 2700 km and D'' extends from 2700 km down to 2900 km. Although relatively thin, the D'' layer is marked by reduced velocity gradients compared with the monotonic increase in seismic velocity through D'. Seismic velocities in the D'' layer also vary by up to 4% and there is much debate as to whether these relate to differences in composition, temperature, or a combination of the two.

Invariably, increases in seismic velocity mirror increases in density because, despite the inverse relationship between velocity and density in Equations 1.2 and 1.3, the bulk moduli increase with depth more rapidly than density. Near the Moho a mantle density of around 3200 kg m^{-3} is typical, and this increases either smoothly, as in the lower mantle, or in a series of discrete jumps, as in the upper mantle, to a density of almost 5000 kg m^{-3} at the core–mantle boundary.

The core is distinguished from the lower mantle by the **Gutenberg discontinuity**, named after its discoverer. This and the Moho are the two major seismic discontinuities in the Earth. It is marked by a drop in P-wave velocity from 13.7 km s^{-1} to about 8.1 km s^{-1} , while S-waves drop from about 7.3 km s^{-1} to zero. The curves in Figure 1.12 show that there is a marked density change at the core–mantle boundary, the density almost doubling from 5000 kg m^{-3} to $10\,000\text{ kg m}^{-3}$, in addition to a phase change from solid to liquid. Seismic variations within the outer core are smooth because it is liquid, and the outer core extends to a depth of 5150 km. Finally, the innermost layer of the Earth is the inner core. This is known to be solid, with P-wave velocities of about 11 km s^{-1} and a density of $12\,000\text{--}13\,000\text{ kg m}^{-3}$.

1.3 The Sun, meteorites and the bulk composition of the Earth

1.3.1 Starting conditions

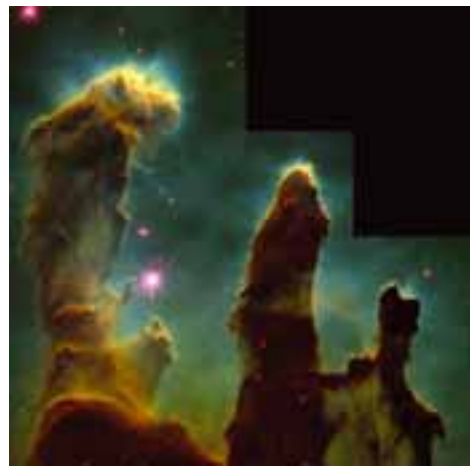
In order to understand how the Earth works today, it is important to know its bulk composition. To gain that knowledge you need to know how the planet formed initially: both the process of formation and the raw materials used. The dramatic changes at the core–mantle boundary, in seismic velocity, density and phase (solid to liquid), almost certainly mean that the composition changes with depth. To determine just what that change is requires an investigation of the original material that made up the Earth, and for that we need to change our direction of observation outwards from the Earth into the Solar System – to the Sun and meteorites.

The Earth is one of only eight planets (Pluto was downgraded to a ‘dwarf planet’ in 2006) in the Solar System, which also includes innumerable smaller bodies such as **asteroids**, **comets** and **planetary satellites**, all of which formed, as far as we know, about 4.6 Ga ago out of roughly the same primordial material. Today, there are two types of planet:

- the so-called terrestrial planets, which are small rocky bodies orbiting close to the Sun (e.g. Earth)
- the gas giants that follow more distant orbits in the colder outer reaches of the Solar System.

Planetary formation is linked to the formation of stars themselves. Stars are now widely considered to form from clouds of gas, mostly hydrogen and helium, which are found throughout the Galaxy. Such clouds also contain small amounts of ices and dust composed of elements heavier than helium. Figure 1.13 shows one such example of a stellar nursery. It is one of the best-known images from the Hubble Space Telescope and shows pillars of gas and dust in the Eagle Nebula in the constellation Serpens.

Figure 1.13 (a) Columns of cold gas and dust in the Eagle Nebula. The columns protrude from the wall of a vast cloud of molecular hydrogen and are up to four light-years long. In places the interstellar gas is dense enough to collapse under its own weight, forming young stars that continue to grow as they accumulate more and more mass from their surroundings. (b) A close-up of the so-called evaporating gaseous globules (EGGs) at the tips of the gas ‘fingers’. (NASA)



(a)



(b)

To give you some idea of the scale of the gas clouds shown in Figure 1.13, the individual columns are up to four light-years long and emerge from an even larger and more diffuse cloud of gas and dust (shown towards the bottom of Figure 1.13a). Of importance to stellar and planetary formation are the small globules protruding from the ends of the gas ‘fingers’ (Figure 1.13b) that have been dubbed evaporating gaseous globules (EGGs) within which the density of the gas has increased to such an extent that the cloud is locally collapsing to form a star. These EGGs are roughly the size of our Solar System. Eventually, when the embryonic star in the centre of the EGG starts to shine by its own nuclear energy, the remnant gas and dust is blown away and the system stops growing.

What goes in an EGG? How does a star and its planetary system form from a collapsing and rotating cloud of dust and gas? One possible sequence of events and processes in the formation of our Solar System is outlined in Box 1.2.

Box 1.2 Formation of the Solar System

The evolution from condensation of grains in the solar nebula through planetary embryos and into planets occurred about 4.6 Ga ago and is summarised in Figure 1.14.

Condensation of the solar nebula

The first phase of the formation of the Solar System involved the gravitational collapse of an interstellar dust cloud dominated by hydrogen and helium, but also containing traces of metallic and silicate dust and ices, such as water (H_2O), methane (CH_4) and ammonia (NH_3) (Figure 1.14a). Once the Sun started to shine, the heat vaporised most of the dust and ices and the vapour was transported further away from the Sun by the early intense solar wind to the cooler regions of the forming Solar System where it cooled and condensed (Figure 1.14b).

The start of accretion

Condensed particles of dust and ice collided with each other and tended to stick together (Figure 1.14c). These random collisions increased the particle size to about 10 mm over a period estimated to have been about 10 000 years. In the inner orbits, these clumps would have been dominated by silicates and metal; their composition is preserved in the most primitive of the chondritic meteorites.

Formation of planetesimals

Particle collisions continued increasing the size of clumps and decreasing the number of objects in orbit around the newly formed Sun, eventually producing a profusion of bodies ranging in size from 0.1–10 km in diameter, termed **planetesimals** (‘tiny planets’) (Figure 1.14d).

Accretion and the development of planetary embryos

Once formed, the larger planetesimals would become the focus of accretion because of their larger gravitational attraction. At this time, the larger

bodies would begin to heat up because of the release of kinetic energy as smaller bodies impacted. Above 10 km this focusing of accretion on the larger planetesimals would have led to the relatively rapid growth of larger objects, known as planetary embryos, with diameters of up to a few thousand kilometres (Figure 1.14f). It is estimated that the planetary embryos would have swept up any remaining planetesimals within a few thousand years.

Planetary embryos and giant impacts

The next stage of growth would have been slower, involving fewer, chance collisions between planetary embryos. Such giant impacts probably fragmented both of the impacting bodies, with the debris subsequently recombining to form a new, larger body (Figure 1.14g). The heat released was also probably capable of melting the newly combined mass, creating a molten mantle of silicate material known as a magma ocean. Metallic material sank through the magma ocean and formed a dense metallic core, thus producing a differentiated planetary embryo.

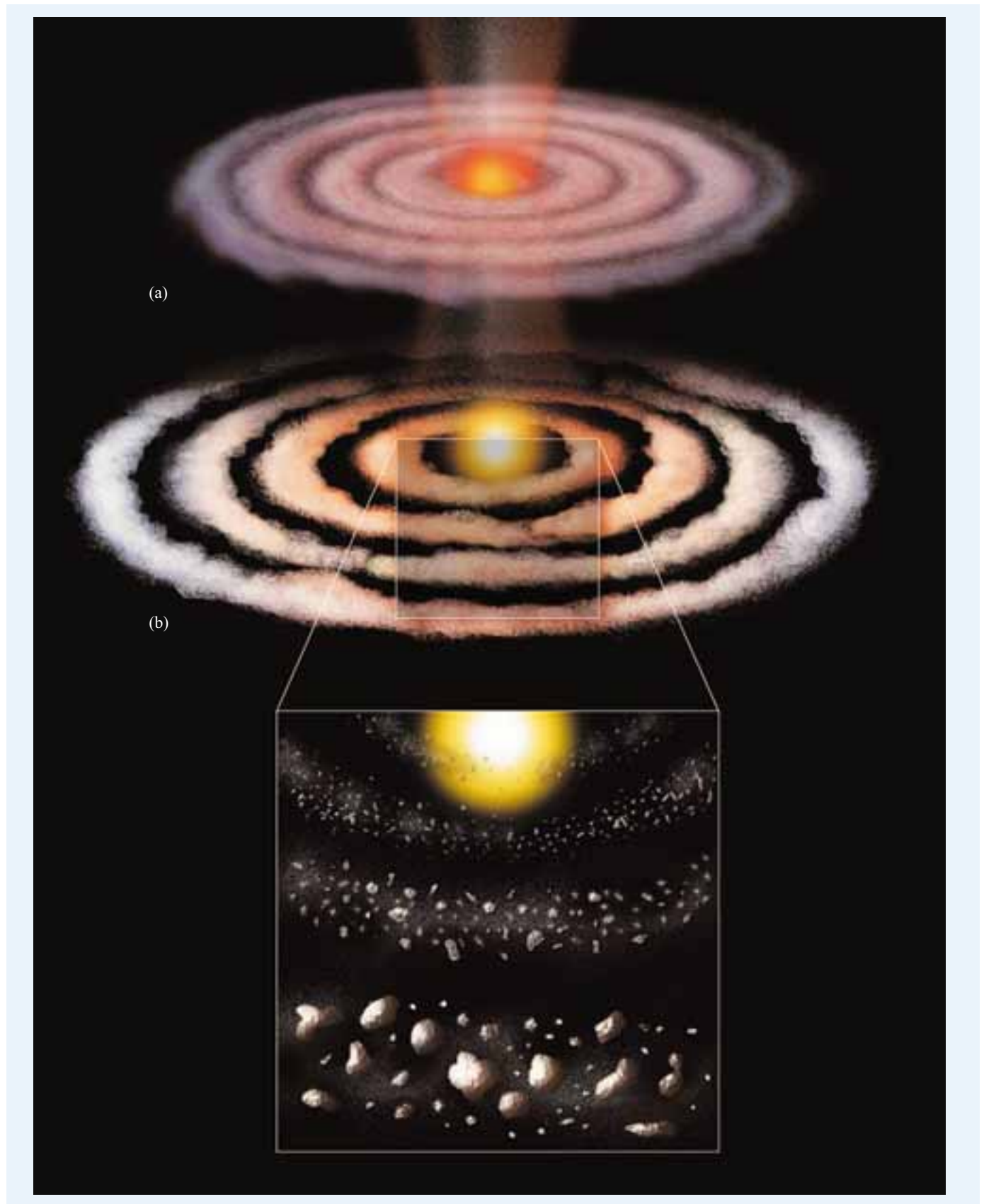
Differentiation of planetary embryos and the assembly of a planet

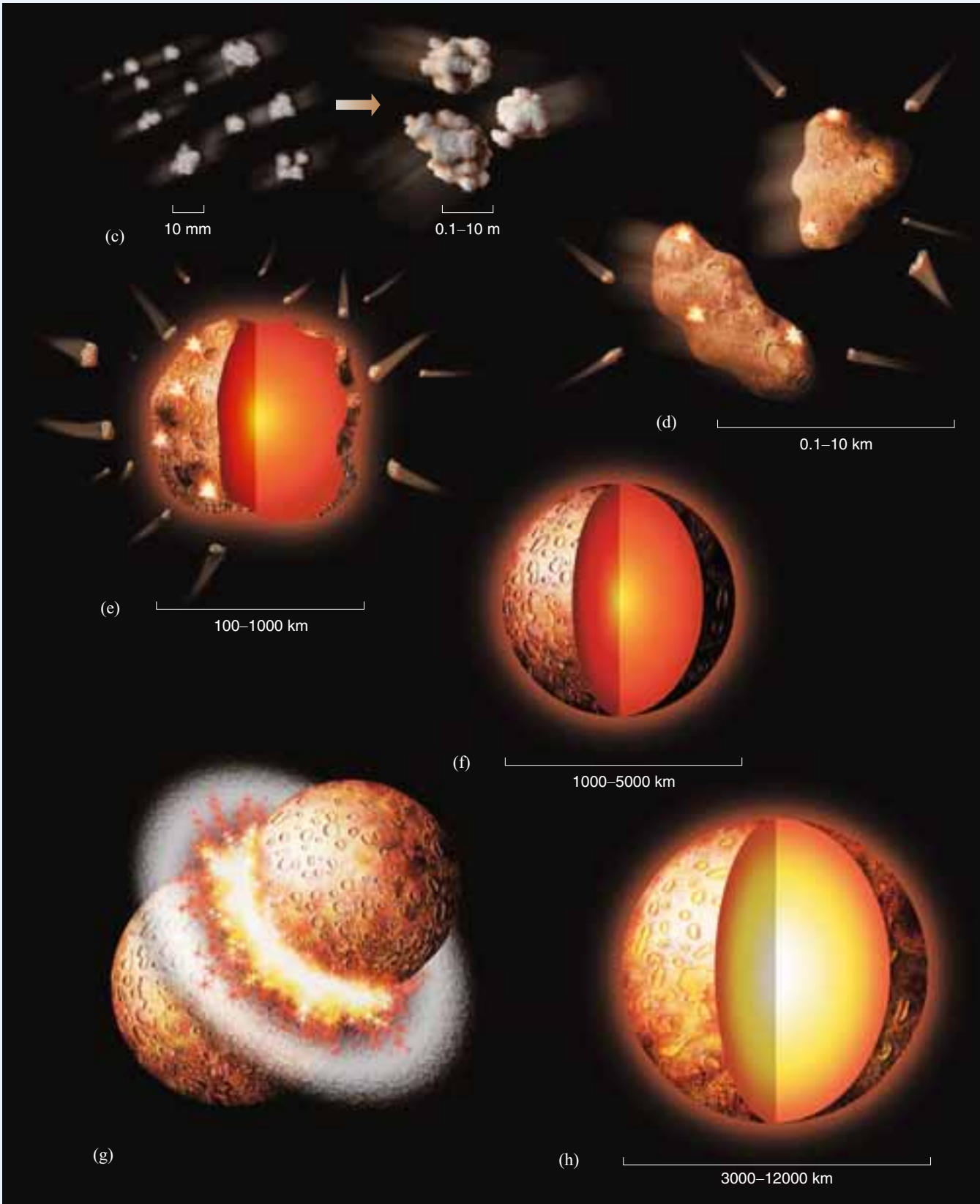
Giant impacts would have continued to occur on an ever increasing scale between these partially differentiated planetary bodies and it is estimated that it would have taken about 10 Ma for the terrestrial planets to reach half their current mass and about 100 Ma to complete their growth and build an Earth-sized planet (Figure 1.11h).

Completion of terrestrial planet formation

Once the last giant impact had occurred (for the Earth this resulted in the formation of the Moon (Section 2.3)), there were just five surviving planetary bodies (Mercury, Venus, Earth, Mars and the Moon). Of course, accretion continues to the present day in the guise of meteorite impacts, but fortunately most of these are too small as to be noticeable; but even so, the estimated accretion rate to the Earth today, including dust particles, is about 10^7 kg y^{-1} .

► **Figure 1.14** Artist's impression of the stages of planetary growth. See text for explanation.





Images like those shown in Figure 1.13 offer a visual link between astronomy and Earth sciences, providing clues as to our own astronomical origins. Those origins involved:

- the condensation of gas to form dust particles
- the accretion of dust to form planetesimals, followed by decreasing numbers of increasingly larger collisions between planetesimals to form planetary embryos
- collisions between planetary embryos during a period of giant impacts to form the planets.

The evidence for the violence of planet formation is nowhere more graphically apparent than on the surface of the Moon (Section 1.1). The composition of the Earth is therefore related to the composition of the solar nebula and, as most of the material from the solar nebula condensed to form the Sun, the composition of the Earth is also related to that of the Sun.

1.3.2 The Sun

The composition of the Sun is known from studying the electromagnetic radiation it emits, which is dominantly white light – a continuous spectrum at visible, near-ultra-violet and infra-red wavelengths. As this light passes through the Sun's atmosphere, the different chemical elements absorb radiation at specific wavelengths, producing dark bands in the spectrum as recorded by earthbound instruments. This is the same method by which astronomers gain knowledge of the compositions of distant stars and galaxies. The amount of light absorbed is proportional to the amount of an element present in the Sun's atmosphere and, while such measurements do not readily give the absolute abundances of elements in the Sun, they do provide valuable information on the relative abundances. Some of this information is summarised in Figure 1.15, in which element abundances are shown relative to a nominal abundance of silicon (Si), in this case 10^6 atoms.

Study Figure 1.15 and compare the relative solar abundances of two everyday elements, such as iron (Fe)

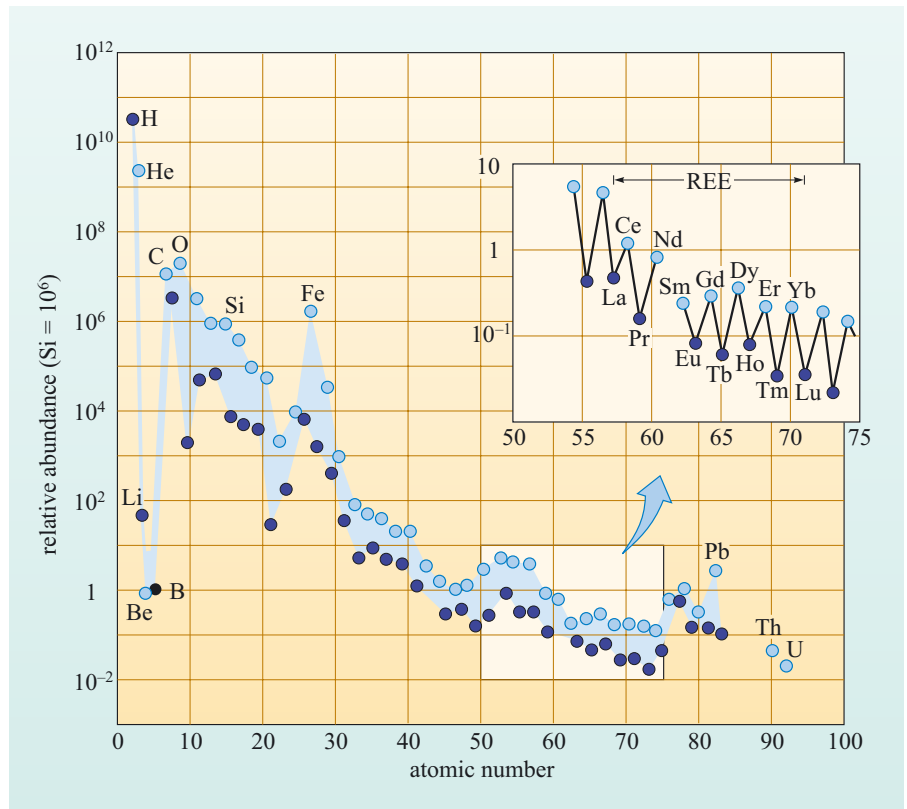


Figure 1.15 The relative abundances of the elements in the Sun. The abundances are expressed relative to silicon (Si), the abundance of which is arbitrarily set at 10^6 atoms. Thus, in the Sun, for every million atoms of Si there are $\sim 5 \times 10^{10}$ atoms of hydrogen (H_2) and about 1 atom of beryllium (Be) and boron (B). The elements are ordered by atomic number from left to right and selected elements are labelled with their chemical symbols. Note that the vertical scale is logarithmic, such that each increment on the scale is 10 times that of the previous one. REE = rare earth elements. (Gill, 1982)

and lead (Pb). Both of these elements are used in many modern-day applications, both have to be won from ore bodies by industrial processes, yet iron is almost 1 000 000 times more abundant in the Sun than is Pb. Indeed Pb is only slightly more abundant in the Sun than elements such as gold (Au; atomic number 79) and platinum (Pt; atomic number 78) which, on Earth, are regarded as being much rarer and of greater value.

- Why do you think that Pb seems to be more abundant than the precious metals?
- The Earth's crust is the source of our mineral wealth and its composition is the product of many processes that have occurred during and since planetary accretion.

Different elements have different geochemical properties and so respond in contrasting ways during geological and planetary processes. It so happens that Pb has become concentrated in the Earth's continental crust, whereas Fe has remained largely in the mantle and core, as you will see below. Our everyday view of the Earth is dominated by what we observe and find in the crust, but the crust, as you already know, is but a thin skin on the surface of our planet. If we could see right through to the core, our view of what the Earth is made of and which elements are abundant and which are rare would be very different.

1.3.3 Element condensation from the solar nebula

The solar nebula is considered to have been hot to start with and as it cooled the different chemical elements gradually condensed according to their individual volatilities. Many of the chemical elements are metals and most have boiling points much too high to be considered significant in our everyday experience. But under the low-pressure conditions of the solar nebula, differences in their boiling points become important. Because the pressure in the nebula was low relative to that experienced at the Earth's surface today, the temperatures at which different elements vaporised (or condensed on cooling) were lower than they are on the present-day surface of the Earth. For example, alumina (Al_2O_3) has a vaporisation temperature (boiling point) of about 3500 K under atmospheric conditions (10^5 N m^{-2}), but in the lower pressures of the solar nebula (10 N m^{-2} or 10 000 times lower than atmospheric pressure) this is reduced to 1700 K.

Figure 1.16 shows the temperature at which various compounds and elements condense out of a nebula with a solar composition, plotted against the fraction of the nebula condensed. As you can see, some elements condense as specific compounds, whereas others condense as the elements themselves. This sequence allows the subdivision of the different elements and compounds into those that belong to an early condensate, a metal–silicate fraction, and volatiles. At even lower temperatures, compounds such as water, methane, ammonia and nitrogen ices condense.

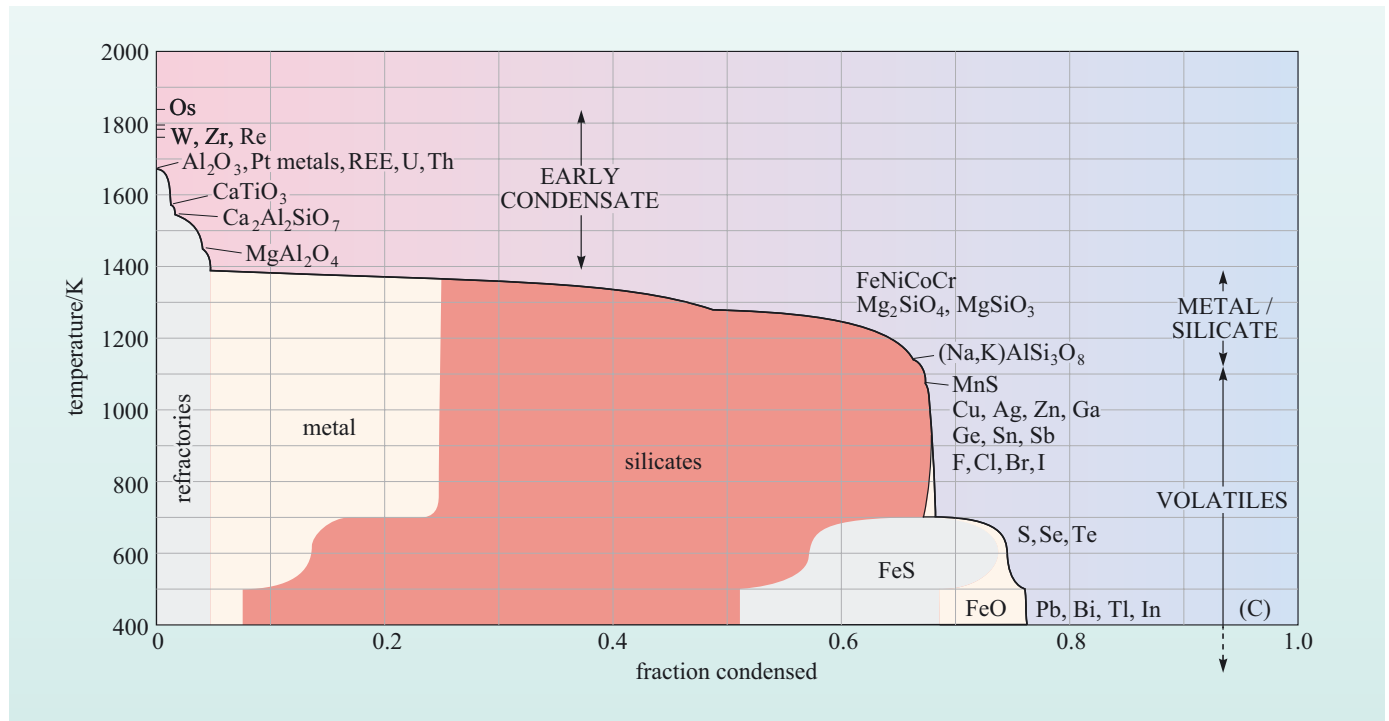


Figure 1.16 The condensation sequence of the solar nebula at a pressure of about 10 N m^{-2} . The x -axis represents the fraction of the nebula that has condensed at the temperature given on the y -axis. The curve is annotated to show which elements and compounds condense at which temperatures. At this low pressure, all materials condense from gases to solids directly without an intervening liquid phase. ($K = \text{kelvin} = T ^\circ\text{C} + 273$) (Morgan and Anders, 1980)

- What are the proportions of the dominant components of the condensate?
- The early condensates, or refractories, comprise 5% of the total. This is followed by metals, which comprise 20% of the total, and silicates, which comprise 40% of the total. The remainder includes volatiles and ices. The condensate is dominated by silicates and metals.
- If the temperature in the solar nebula decreased outwards from the proto-Sun, can you suggest a mechanism that might have produced the terrestrial planets (Mercury, Venus, Earth and Mars) and gas giant planets (Jupiter, Saturn, Uranus and Neptune)?
- Given the sequence of condensation in Figure 1.16, then volatile elements and ices would only condense when the temperature of the solar nebula dropped below 300 K. Perhaps such cold conditions only occurred beyond the orbit of Mars.

Element (and compound) volatility is one of the important properties that controls the behaviour of the elements in planetary formation and, consequently, the composition of all planetary bodies.

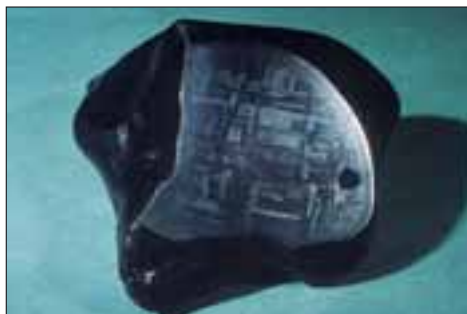
1.3.4 Meteorites

A second source of information about the composition of the Solar System and the primary material of the planets comes from meteorites. These are astronomical objects of varying sizes, made of metal or rock, that fall to Earth. They have been studied for many years, not only because they are of intrinsic interest but also because they contain material thought to be representative of the early solar nebula. Meteorites are derived from the asteroid belt, which comprises many small planetary bodies (possibly planetary embryos) that did not accrete into a single larger body. This probably resulted from gravitational disturbances caused by Jupiter, perturbing the orbits of individual asteroids and causing repeated violent collisions that resulted in further fragmentation rather than accretion.

Meteorites come in a variety of compositions, but can be broadly classified into:

- **stony meteorites**, consisting dominantly of silicate minerals
- **iron meteorites**, primarily composed of metallic iron
- **stony-iron meteorites**, a hybrid of the other two, containing variable amounts of silicates and metal.

Figure 1.17 Different types of meteorites: (a) a cut and polished surface of an iron meteorite that has been etched with acid to reveal details of its internal texture; (b) a carbonaceous chondrite; (c) a stony meteorite with a basaltic composition; (d) a stony-iron meteorite, or pallasite, with olivine crystals set in a metal matrix; (e) a relatively large mm-sized chondrule from a carbonaceous chondrite (Bokkeveld meteorite), showing individual crystals within it; (f) a collection of individual chondrules, each less than 1 mm in diameter, from an ordinary chondrite (Sharps meteorite). (Natural History Museum)



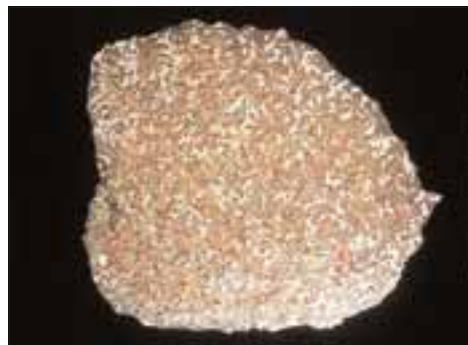
(a)



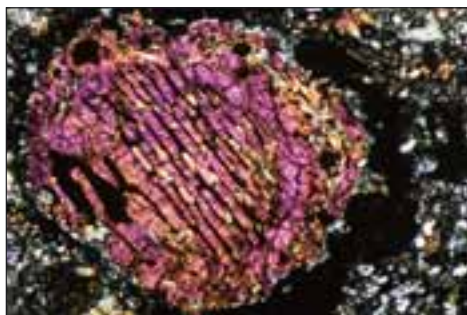
(b)



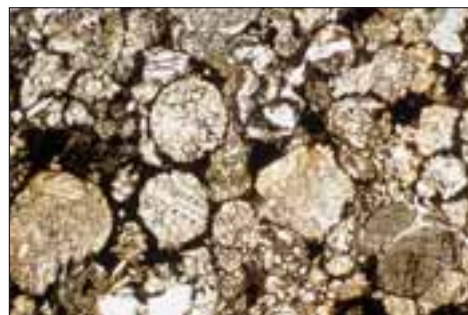
(c)



(d)



(e)



(f)

The iron and stony-iron meteorites have undergone an amount of chemical processing and are described as **differentiated**; their compositions have been affected by removal of some or all of the silicate minerals to leave a bulk composition enriched in iron (Fe) and nickel (Ni), the major metallic elements.

Stony meteorites account for up to 95% of all known meteorite falls and are subdivided into **chondrites** and **achondrites**, depending on whether they contain **chondrules** or not. Chondrules are small, roughly spherical globules of silicate minerals that are between 0.1 mm and 2 mm in diameter (Figure 1.17e and f). Their shape and crystallinity suggest that they were once molten in a low gravitational field, indicating that they formed away from major planetary bodies, either on the surface of planetesimals or even within the solar nebula.

Achondrites constitute only about 10% of all stony meteorite falls. Their textures are reminiscent of terrestrial igneous rocks and they are classified as differentiated meteorites, along with the irons and stony-irons.

Chondrites are subject to further subdivision and classification based on their mineralogy. The three main classes are:

- **ordinary chondrites**, so-called because they are the most abundant type
- **enstatite (or E-) chondrites**, which are rich in the magnesium silicate mineral enstatite (MgSiO_3)
- **carbonaceous (or C-) chondrites**, which contain non-biogenic carbon-rich organic compounds in addition to silicate minerals.

The ordinary chondrites are further subdivided according to their iron contents and their oxidation state, reflected in the amount of iron in silicate minerals relative to that in chemically reduced phases such as metallic iron and sulfides. These groups are referred to as **H-, L- and LL-chondrites**. This rather complex classification is summarised in Figure 1.18.

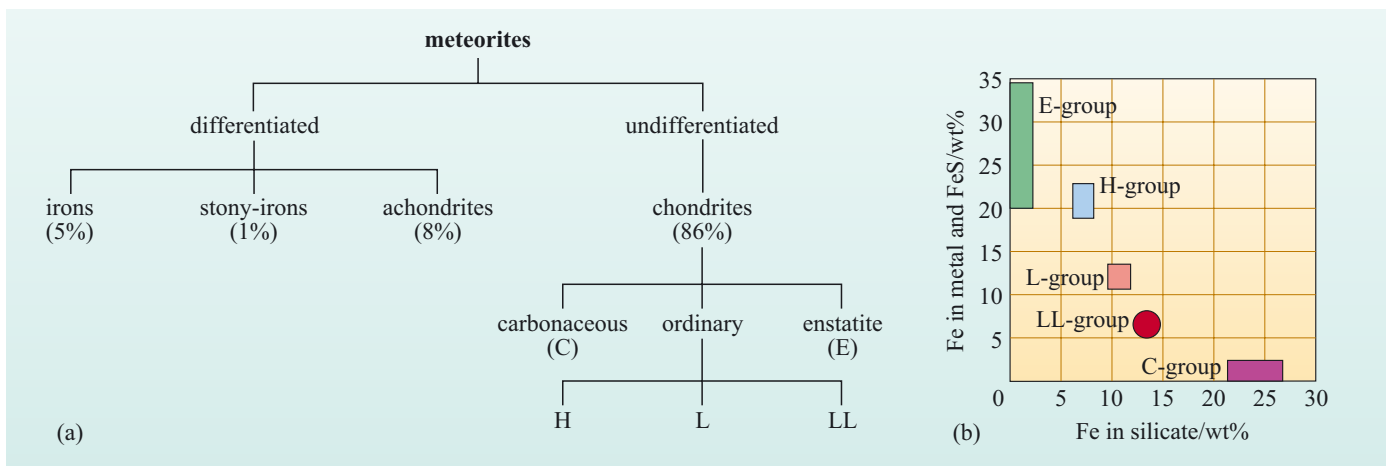


Figure 1.18 (a) Meteorite classification based on observable mineralogy and textures. (b) Classification of chondritic meteorites based on the oxidation state of iron.

- Considering this overall classification, what do you consider to be the dominant process in meteorite differentiation?
- The separation of a metallic phase from the silicate minerals.

The compositions of iron meteorites reflect processes that have taken place within a planet-sized body (or planetary embryo) where temperatures were at one time high enough to melt the silicates and the metallic phases and allow them to separate as a result of gravity. Thus, in our search for the most primitive composition, the differentiated meteorites can be rejected, leaving the chondrites. But which of these represents the most primitive material?

- If you return to a consideration of the processes that occur within the solar nebula, which are the elements that would be most easily lost from a meteorite?
- The volatile elements and ices, because they have the lowest condensation temperatures.

Assuming that condensation from the solar nebula continued without disturbance to low temperatures, then the most primitive meteorites should contain abundances of the volatile elements in the same proportion as in the Sun. A comparison of the compositions of the different meteorite groups reveals an overall similarity in the abundances of elements found in high-temperature early condensates and increasingly diverse concentrations of the more volatile elements. This feature is illustrated in Figure 1.19, where the abundances of selected elements in enstatite chondrites and ordinary chondrites are compared with those in average carbonaceous chondrites.

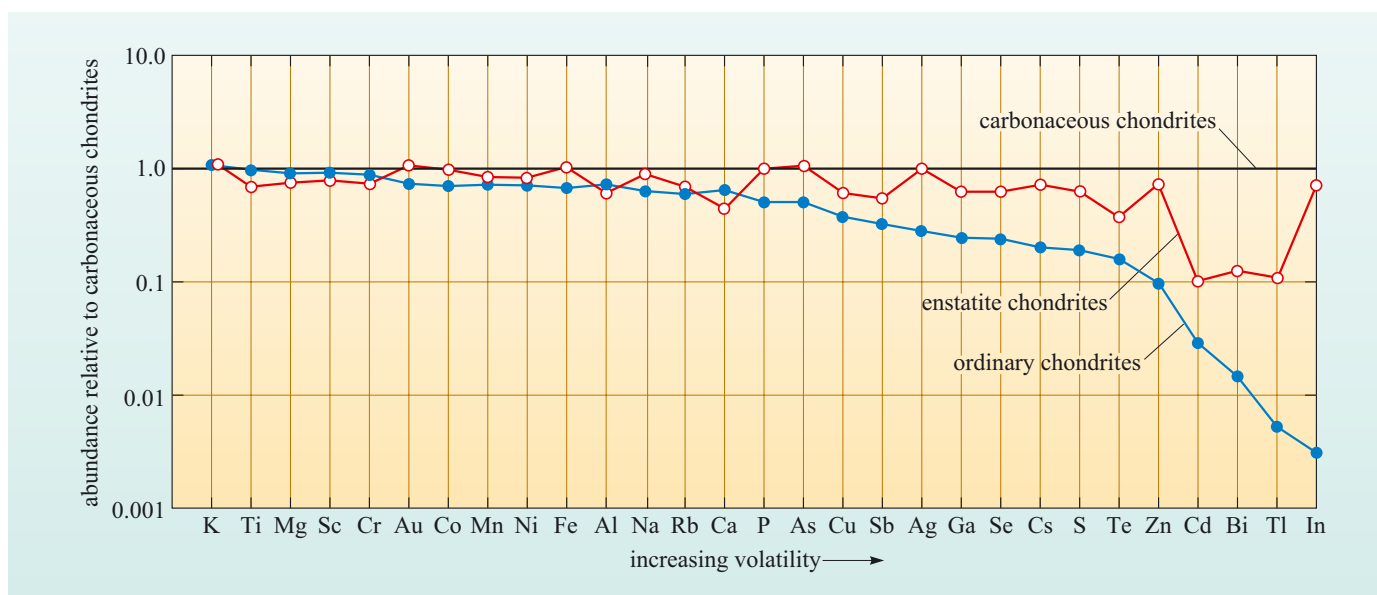


Figure 1.19 Comparison of the abundances of selected volatile elements in enstatite, ordinary and carbonaceous chondrites. Element abundances are normalised to carbonaceous chondrites. (McSween, 1987)

- Which types of chondrite do you consider to be the most differentiated and which type the least differentiated?
- The ordinary chondrites have lost a greater proportion of their volatile elements and so are more differentiated than the E-chondrites. The C-chondrites are therefore the least differentiated because they have not lost their volatile elements.

The C-chondrites thus appear to have been least affected by processes after their formation. This conclusion is borne out when their elemental compositions are compared with those of the **solar photosphere**, which is the outer layer of the Sun. This is shown in Figure 1.20 for a wide range of elements with different volatilities and different chemical properties.

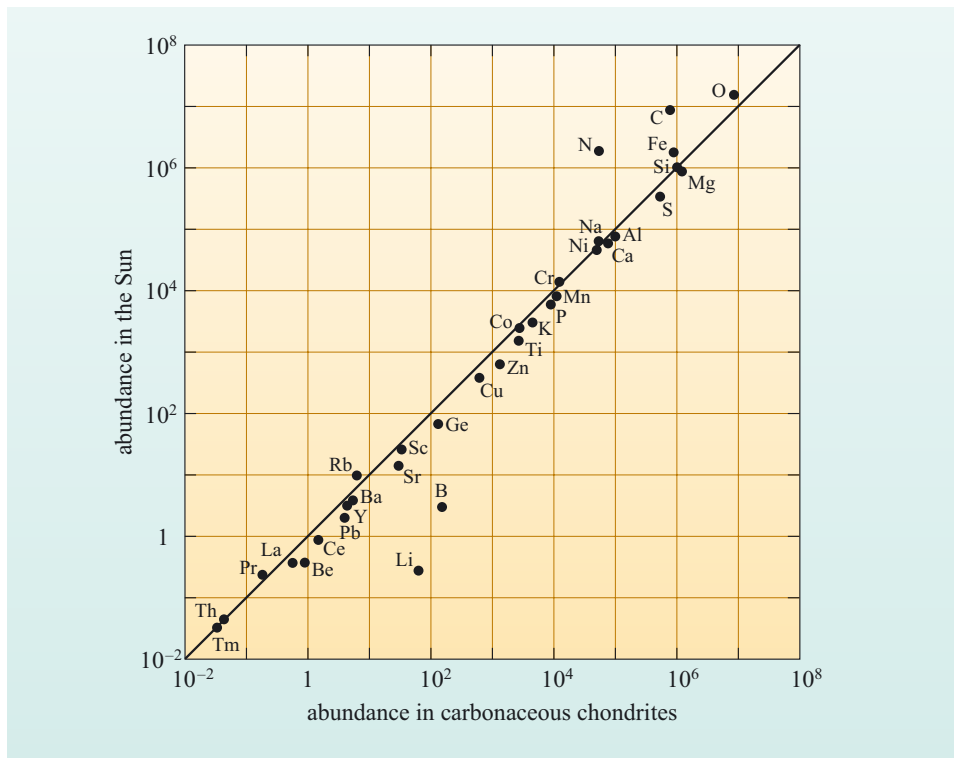


Figure 1.20 The relative abundances of the elements in the solar photosphere plotted against those in C-chondrites, based on Si = 10^6 atoms. (McSween, 1987)

The correlation for most elements is very good, ranging from the most abundant, such as oxygen, magnesium and iron, to the least abundant such as thorium (Th) and thulium (Tm). There are some exceptions – for example lithium (Li) and boron (B) are depleted in the solar composition relative to C-chondrites because they are used up in nuclear fusion reactions in the Sun. Conversely, carbon (C) and nitrogen (N) are relatively depleted in carbonaceous chondrites, possibly because they are amongst the most volatile elements. Despite these differences, the match between carbonaceous chondrites and the Sun is remarkable and gives us some confidence that this really is the composition of the primordial material of the Solar System.

I.4 The composition of Earth layers

The results of seismic investigations into the Earth reveal a twofold division into a dense core and an overlying mantle, with the mantle comprising about 68% by mass of the whole Earth. By contrast, the crust comprises about 0.4% of the Earth's mass and is insignificant in this context, even though it is the part of the Earth with which we are most familiar.

- Recalling the variety of meteorite compositions, which types of meteorite do you think will give the best indications of the compositions of planetary layers?
- The Earth is a differentiated planetary body, so the differentiated meteorites will give the best indications of the compositions of planetary layers.

The likely compositions of the core and mantle should now be fairly obvious – the core is made of a dense alloy of iron and nickel while the mantle is composed of silicates which, by analogy with achondrites, will be rich in magnesium (Mg). How can we test these ideas? Fortunately, there are places where mantle rocks are brought to the surface by volcanic and tectonic processes and so at least these samples can be compared with meteorites. However, there are no such samples of the core so comparison of the core and meteorites is less direct, relying instead on properties such as seismic velocity and density.

I.4.1 The Earth's mantle

The Earth's mantle is almost everywhere obscured by the crust, but occasionally geological processes carry mantle rocks to the surface. Mantle samples occur in two main geological situations. The first is as accidental fragments of rock known as **xenoliths** brought to the surface in volcanic rocks that erupt from great depth. They are distinctive because they have a green colour that contrasts with the dark grey or black of the host volcanic rock and, for many years, their significance was unclear. However, as ideas of the composition of the Earth's interior developed they became the subject of more intensive study. These rocks are termed **peridotites** and are rich in the minerals olivine and pyroxene, which are both silicates rich in magnesium and iron.

Similar rock types are also found in the deeper layers of rock sequences called **ophiolites**. Within an ophiolite, a thin layer of sedimentary rocks overlies volcanic lavas, dykes and gabbros, all of basaltic composition, which in turn overlie peridotite. Ophiolites are found in young mountain belts and are interpreted as being segments of old oceanic crust that have been trapped between colliding continents.

- How could you confirm that a sample of peridotite from a xenolith or ophiolite is from the upper mantle?
- The density or seismic velocities could be measured and compared with the Earth reference model (Section 1.2).

Laboratory measurements of the density of peridotite (about 3200 kg m^{-3}), and their P-wave speeds (approximately 8.0 km s^{-1}), are a close match for the measured properties of the upper mantle.

Direct comparisons between the composition of peridotites and carbonaceous chondrites are less easy because the mantle represents the silicate residue after core formation. However, some compositional parameters, particularly those involving element ratios, can be exploited to make a useful comparison. An example is shown in Figure 1.21, which plots the Mg/Si ratio against Al/Si in both meteorites and mantle samples.

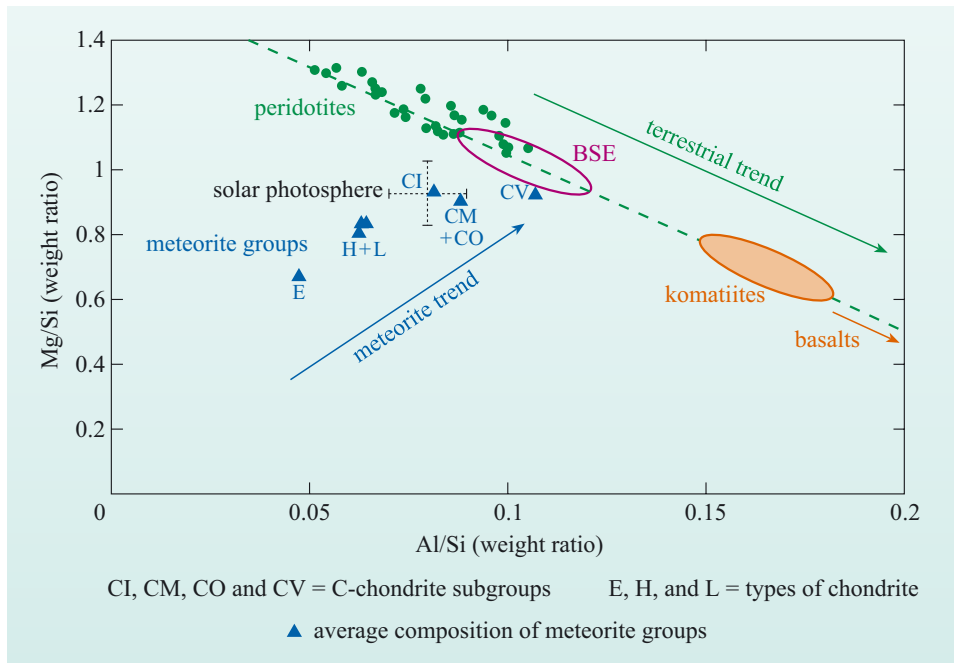


Figure 1.21 The variation of the Mg/Si ratio with the Al/Si ratio in terrestrial mantle samples and different classes of meteorites. The composition of the solar photosphere is shown for reference. The negative trend defined by the terrestrial mantle (peridotite) samples is related to the generation and extraction of basaltic magma. Komatiites are Mg-rich, ultramafic magmas that were produced by extensive melting of the mantle during the Archaean. BSE represents the range of estimates of the composition of the bulk silicate Earth (sometimes referred to as the primitive upper mantle). (Adapted from Drake and Righter, 2002)

In Figure 1.21, peridotites define a negative trend along which the Mg/Si ratio decreases as the Al/Si ratio increases. Firstly, this shows that the composition of the mantle as revealed by peridotites is not constant – the mantle is heterogeneous. The causes of this variation are indicated by the position of basalts and **komatiites** on this diagram, which lie on an extension of the peridotite trend at higher values of Al/Si and lower values of Mg/Si. Basaltic magma is derived from the mantle by a process known as partial melting, and this will be explored in greater detail in Chapter 4. For now, it is sufficient to appreciate that the variation in peridotite composition is a direct consequence of the extraction of basaltic magma.

- What does this relationship between basalts and the mantle indicate about the most probable location of primitive mantle within the array of peridotite data?
- Mantle samples with the most primitive compositions will be those that have not experienced the extraction of basaltic magma and so they will plot at the low Mg/Si end of the peridotite array, closest to the composition of basalt.

Average meteorite compositions, in contrast to peridotites, define a positive trend on Figure 1.21: both Mg/Si and Al/Si increase together from E-chondrites that have the lowest values of both ratios, increasing through H- and L-chondrites and then into the C-chondrites, which have the highest values of both ratios. Notice that the C-chondrites are further subdivided into a number of different groups, designated CI, CM, CO and CV. Although, the origin of this variation remains unclear, the intersection of the positive meteorite trend with the terrestrial array has been used to define the composition of the primitive mantle.

- Do the variations illustrated in Figure 1.21 suggest that the Earth has a composition directly comparable with C-chondrites?
- No. C-chondrites have lower values of both Mg/Si and Al/Si than the most primitive mantle composition.

Although the chondritic Earth model has been used as a standard for many decades, it is now well established that for certain elements, of which Si is one, the **bulk silicate Earth** does not have a composition that is represented by any meteorite group and so it is not directly comparable with the solar photosphere and hence the Sun. The only class of meteorites that do fall on the terrestrial array are an unusual group of meteorites designated CV.

Other elements can be investigated in a similar way, and the abundances of selected elements in the mantle relative to CI meteorites are illustrated in Figure 1.22. The abundances of the different elements are all normalised to that of Mg, and it shows that elements such as aluminium (Al), calcium (Ca) and uranium (U) retain the same relative abundances in the Earth's mantle as they have in CI meteorites and hence the Sun. Other elements, such as Si, the alkali elements (sodium (Na), potassium (K), rubidium (Rb) and caesium (Cs)) and transition elements such as copper (Cu), silver (Ag) and zinc (Zn) are all depleted by up to two orders of magnitude relative to CI meteorites.

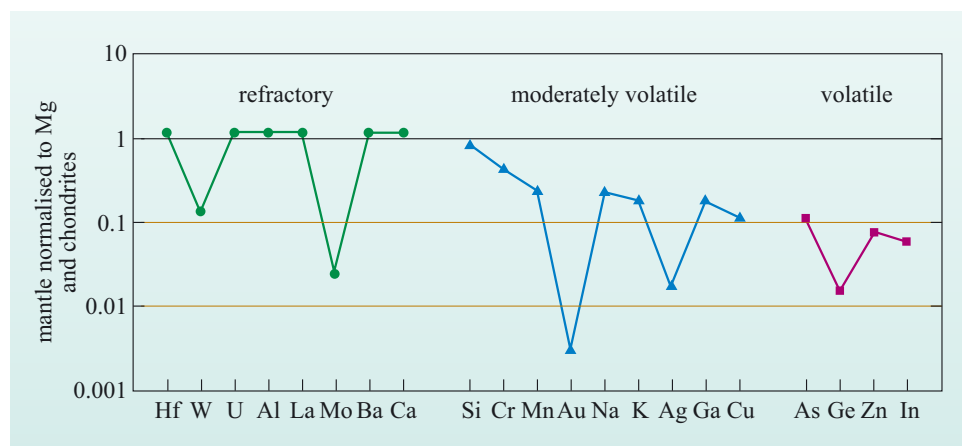


Figure 1.22 The abundances of selected trace elements in the bulk silicate Earth normalised to the composition of CI meteorites, in order of increasing volatility from left to right. (Abundances taken from McDonough and Sun, 1995)

- Referring back to the dominant processes in the solar nebula, suggest a property that might cause these depletions.
- All of the elements listed as depleted are more volatile than Al, Mg, Ca and U so their depletions may relate to conditions in the solar nebula or to a subsequent loss of a volatile component after planetary formation.

Relative to CI meteorites, the Earth's mantle is depleted in volatile elements and this may in part explain the difference between the estimated bulk composition of the Earth and the different meteorite classes – the Earth has lost a greater proportion of its volatile components compared with meteorites. But that is not all.

- Consider the element tungsten (W). Using the information in Figures 1.16 and 1.23, does its relative abundance relate to its volatility?
- No. W is a refractory element with a condensation temperature similar to Al, Ca and U, yet it is depleted in the mantle with an abundance of only ~13% of that expected, assuming a chondritic starting composition.

Clearly, the abundances of elements in the bulk silicate Earth are not controlled solely by their volatility, and other properties govern their distribution within the Earth. Box 1.3 describes these properties in relation to the position an element occupies in the Periodic Table.

Box 1.3 Element properties and distribution coefficients

The distribution of the elements between the different phases (metal, silicate and sulfide) in meteorites and other natural materials has led to a broad classification of their geochemical (and cosmochemical) properties. Elements that are found in the metallic phase of a natural system are referred to as **siderophiles** (from the Greek, *sideros*, meaning iron and *philos*, meaning like or love – literally iron-loving). They contrast with those elements that preferentially bond with oxygen, especially in silicate or oxide structures, which are known as **lithophiles** (from the Greek *lithos*, meaning stone). There is also a subgroup of lithophile elements that tend to be gaseous at the Earth's surface, notably H, C, N, O and the noble gases, that are referred to as **atmophiles**. A third major grouping refers to elements that frequently occur bound with sulfur and these are known as **chalcophiles** (from the Greek *khailos*, meaning copper, an element commonly found as a sulfide).

Figure 1.23 shows how these properties are distributed across the Periodic Table and that some elements can behave in different ways. For example, in the absence of a separate metal phase, cobalt (Co) will behave as a lithophile element, but it partitions preferentially into the metal liquid as soon as a molten metal is introduced into the system. Similarly, iron (Fe) can be siderophile, lithophile or chalcophile, depending on the prevailing conditions. Despite this overlap of properties, the broad distinction between siderophile, lithophile and chalcophile has proven useful in the development of geochemistry.

group period	I	II	transition elements (+ lanthanides + actinides)										III	IV	V	VI	VII	noble gases
1	H 2.1																	He —
2	Li 1.0	Be 1.5											B 2.0	C 2.5	N 3.0	O 3.5	F 4.0	Ne —
3	Na 0.9	Mg 1.2											Al 1.5	Si 1.8	P 2.1	S 2.5	Cl 3.0	Ar —
4	K 0.8	Ca 1.0	Sc 1.3	Ti 1.5	V 1.6	Cr 1.6	Mn 1.5	Fe 1.8	Co 1.8	Ni 1.8	Cu 1.9	Zn 1.6	Ga 1.6	Ge 1.8	As 2.0	Se 2.4	Br 2.8	Kr —
5	Rb 0.8	Sr 1.0	Y 1.2	Zr 1.4	Nb 1.6	Mo 1.8	Tc 1.9	Ru 2.2	Rh 2.2	Pd 2.2	Ag 1.9	Cd 1.7	In 1.7	Sn 1.8	Sb 1.9	Te 2.1	I 2.5	Xe —
6	Cs 0.7	Ba 0.9	La-Lu 1.1-1.2	Hf 1.3	Ta 1.5	W 1.7	Re 1.9	Os 2.2	Ir 2.2	Pt 2.2	Au 2.4	Hg 1.9	Tl 1.8	Pb 1.8	Bi 1.9	Po 2.0	At 2.2	Rn —
7	Fr 0.7	Ra 0.9	Ac 1.1	Th 1.3	Pa 1.5	U 1.7	Np-Lr 1.3											





 atmophile
  lithophile
  chalcophile
  siderophile

Figure 1.23 The Periodic Table subdivided according to the geochemical properties of the elements. Colour shows the dominant property and the shading shows subsidiary properties. The number beneath each element symbol refers to electronegativity.

The differences in behaviour are related to a property known as **electronegativity (E)**, which is based on a relative scale from 0 to 4.0 and related to the ability of an atom to attract electrons and so form a negatively charged ion. Broadly speaking, those elements with low electronegativities ($E < 1.6$) are lithophiles, forming positive ions that bond with negative oxygen ions. Those elements with $1.6 < E < 2.0$ tend to be chalcophile, because they more readily form covalent bonds with elements such as S. Finally, siderophile elements have $2.0 < E < 2.4$ and most readily form metallic bonds. Elements with $E > 2.4$, such as O, N and the halogen elements, F, Cl, Br and I, readily attract electrons to form negative ions and form ionic bonds with the lithophile elements with low electronegativity.

The way in which elements distribute themselves between solid and molten states is known as **element partitioning** and is quantitatively expressed in the form of a number known as a **partition** or **distribution coefficient**, variously defined as K , K_d , K_D or D . It is defined by the following simple equation:

$$D = \frac{\text{concentration of element } i \text{ in phase } a}{\text{concentration of element } i \text{ in phase } b}$$

$$D = \frac{C_a^i}{C_b^i} \quad (1.7)$$

In the case we are considering in this chapter, the distribution coefficient is between molten metallic iron and solid silicate minerals. The convention is for the numerator (C_a^i) to be allocated to the liquid and the denominator (C_b^i) to be allocated to the solid. Thus, a lithophile element will have a D value of < 1 and a siderophile element will have a D value of > 1 .

Most commonly, the distribution coefficient is used to describe the way in which an element partitions between crystals and liquids in silicate systems. In contrast with this chapter, in these systems elements that partition into a solid phase ($D \geq 1$) are said to be **compatible**, whereas those that are excluded and concentrate in the silicate liquid ($D < 1$) are said to be **incompatible**.

- Having studied Box 1.3, suggest why elements such as W, Mo, Au and Ge have lower abundances in the bulk silicate Earth than expected for their volatilities.
- These elements can behave as siderophile elements in the presence of a molten metal. Therefore they have probably been concentrated into the Earth's core.

Core formation involved the gravitational separation of molten metallic iron from silicate melt or solid crystals. By analogy with iron meteorites, which are considered to be representative of the cores of small planetary bodies, this iron selectively removes a number of other elements that, because of their chemical properties, bond more easily with iron than they do with silicates; W is one of these so-called siderophile elements (Box 1.3).

Figure 1.24 shows how the abundances of the siderophile elements in the mantle, once again normalised to CI meteorites. In this case, the most volatile elements have been excluded and the order of the elements is according to their increasing siderophile properties as determined by laboratory experiment, so that elements on the right are increasingly siderophile. Clearly, the more siderophile elements are more depleted in the Earth's mantle.

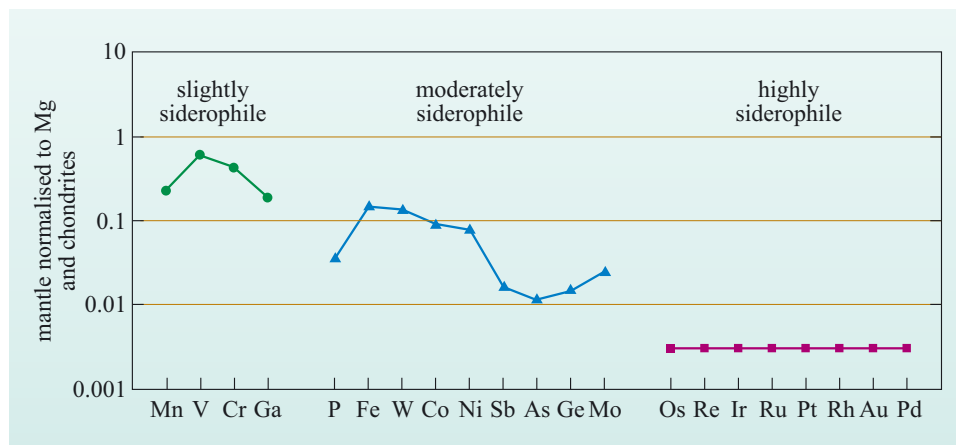


Figure 1.24 Abundances of the siderophile elements in the Earth's mantle relative to CI meteorites. (Data from McDonough and Sun, 1995)

To summarise all of this information, the present-day composition of the Earth's mantle may be similar to that of some C-chondrites. The ratios of refractory lithophile elements (e.g. Ca, Al, Mg, U, Th and rare earth elements) in the mantle are similar to those in chondritic meteorites generally, indicating a close link with primordial (solar) abundances. But the mantle is depleted in both volatile and siderophile elements. The latter are now presumably concentrated in the core, whereas the volatile elements were probably lost during accretion or even earlier during condensation from the solar nebula. The important result, however, is that the abundances of elements within the mantle are controlled systematically by their geochemical and cosmochemical properties.

1.4.2 The Earth's core

From the various lines of evidence, it is clear that the Earth's core is primarily made of iron, according to both the model developed above and the requirements of the seismic Earth reference model (Section 1.2). Samples of the core are not available at the Earth's surface, so any scientific investigation of the composition of the core has to be done either remotely (e.g. with seismology) or by using laboratory analogues and simulations. The latter involves subjecting materials to the temperatures and pressures of the core in the laboratory, measuring their properties and comparing them with Earth reference models.

One of the most useful properties is density; Figure 1.25 shows the variation of density with depth, and hence pressure, for the Earth's core and for pure iron.

- Do the curves match?
- No. The density of the core is less than that experimentally determined for pure iron.

The mismatch between the curves for the core and pure iron requires the core to include other elements that reduce its overall density.

- Which other elements might the core contain?
- Siderophile elements.

The core must contain significant amounts of siderophile elements, but can these elements account for the lower density of the core? In short, the answer is no.

Most of the siderophile elements have atomic masses, and hence densities, either similar to or even greater than that of iron – if anything, siderophile elements make the problem of the core's low density even worse! The unavoidable conclusion (assuming experiments have given the correct results) is that the core includes a significant fraction of an element (or elements) with an atomic number and density less than liquid Fe–Ni alloy.

- Recalling the comparison between chondrites and the mantle, which of the abundant elements is depleted in the mantle?
- Silicon.

Figures 1.21 and 1.22 show that Si is depleted in the mantle by ~10% and although that depletion may be related to its volatility, it is also possible that Si could have partitioned into the core. Certainly some Si has been identified in the metal phase of iron and stony-iron meteorites, but the concentrations are far too low – and are measured in ppm rather than the 10–15% required to reduce the density sufficiently. However, some experiments reveal that Si has increasing siderophile tendencies at very high pressures. Other elements that have been suggested include O and S, in the forms of FeO and FeS respectively. Both FeO and

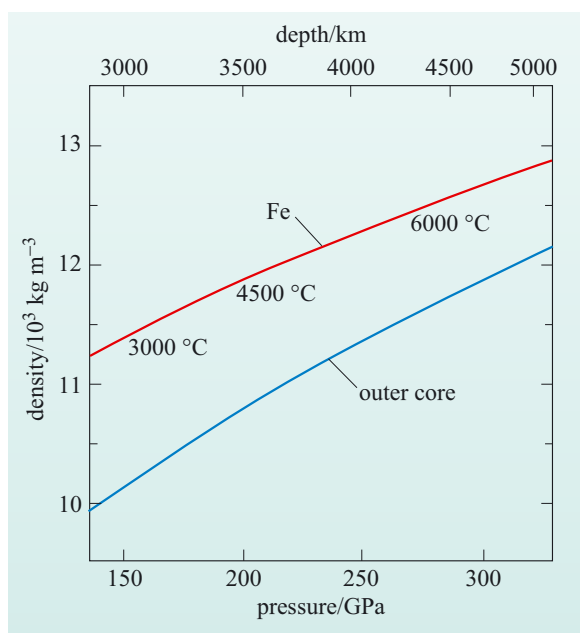


Figure 1.25 Variation of density with pressure in the outer core and the density of iron as determined experimentally for conditions in the core. Temperatures are indicated along the curve for pure iron.

FeS can dissolve in molten Fe in sufficient quantities to reduce its density enough to match that of the core.

- From what you have understood so far about the behaviour of the different elements, can you see why either of these suggestions might *not* be reasonable?
- Sulfur is a volatile element and so may have been lost from the Earth before the core formed, whereas the very high electronegativity of oxygen (Box 1.3) means that it is dominantly lithophile.

A major problem that must be overcome in determining whether or not any particular element could be incorporated in the Earth's core relates once again to its volatility during accretion. Sulfur is a particularly volatile element, condensing out of the solar nebula in the form of FeS at a temperature of ~ 600 K (Figure 1.16). Its relative abundance in the Earth's mantle is broadly similar to that of other similarly volatile elements, such as the halogens, and more volatile elements such as Pb, Bi and thallium (Tl). These observations indicate that S is behaving as a typical volatile element, that it is neither more nor less depleted in the mantle than it should be and that, by analogy with these similarly volatile elements, it is unlikely to have been present in the early Earth in a high enough concentration to reduce the density of the core sufficiently.

This leaves oxygen. Experimental studies have shown that, at the conditions appropriate for the core–mantle boundary, oxygen can dissolve in molten Fe as FeO. The mantle contains abundant oxygen bound in silicates that dissociate into dense oxide phases in the deep mantle, and a variety of reactions between these oxides and oxygen-free molten iron have been suggested, all of which lead to the production of FeO (wüstite). (You will encounter these high-pressure transformations in Chapter 6, but for the moment it is sufficient to know that they exist.) Thus, mechanisms for transferring oxygen from the mantle to the core can be demonstrated on the basis of high-pressure experiments, and oxygen remains the main candidate for the light element in the Earth's core.

The Earth's core and the magnetic field

Although the core is a very remote component of the Earth, it has a profound effect on the surface environment as it is the source of the Earth's magnetic field. The Earth has one of the strongest magnetic fields in the Solar System and certainly the strongest magnetic field associated with any of the terrestrial planets. Continuing on the theme of what makes the Earth unique in the Solar System, it is an important aspect of maintaining and preserving a benign environment on the Earth's surface by protecting the surface from cosmic rays and high-energy particles.

The Earth's magnetic field cannot result from the presence of a solid magnet in the interior because permanent magnetism cannot remain above the **Curie point** of any magnetic material. For example if a solid iron magnet is heated to about 600°C it loses its magnetism. Given that the Earth's core is considerably hotter than this ($<4000^\circ\text{C}$) the field must be generated by some other mechanism.

Some clues to where and how the field is generated can be found in the record of observations of the magnetic field over time. For example, the position of

magnetic North moves at a perceptible rate over periods of decades to centuries, and the geological record reveals rapid flips in the Earth's magnetic polarity many times in the recent and more distant geological past (Chapter 3). These observations suggest that the magnetic field must originate in a region of the Earth that can move and respond rapidly – in other words it must originate in a very fluid part of the Earth, the outer core. However, the mechanism whereby the field is generated is still not fully understood. It is thought to be related to an interaction between the Earth's rotation and convection within the outer core that, in turn, is driven by a combination of solidification of the inner core and secular cooling.

Traces of ancient magnetism, known as **remnant palaeomagnetism**, have been frozen into some of the Earth's oldest volcanic rocks and demonstrate that the Earth's magnetic field has been in existence for at least ~3.8 Ga. If **compositional convection** is the principal means of generating the Earth's magnetic field, then the inner core must be at least as old as the magnetic field. Compositional convection caused by crystallisation of the inner core also involves cooling of the core. However, the estimated cooling rate of the core is too great for the core to maintain heat for 3.8 Ga, and suggests that the inner core cannot be older than about 1 Ga. In order to account for the age of the magnetic field and this cooling rate, it has been suggested that there may be significant radioactive heating of the core (for example, from the decay of the light element potassium (K), which may be incorporated in the core in trace amounts). Alternatively, it has been suggested that cooling rates, and compositional convection, may have been slower earlier in Earth's history.

1.5 The Earth's crust: continents and oceans

To bring this introductory chapter to an end, we return to the differences between the continents and the oceans, not just that one is covered with water and the other not, but the differences in the composition and physical structure of oceanic and continental crust and how they relate to the bimodal hypsometric curve described earlier in this chapter.

The geology of the Earth's surface is known so well in places that it is difficult to treat it as a single compositional entity as is routinely done for the mantle and core. Geological observations show that the continental crust is highly complex and variable on a variety of length scales – from a single outcrop a few metres across to the width of a continent. The variation in rock types has kept field geologists busy for over 200 years, and still large tracts remain unmapped. However, where the deeper layers of the continents have been exposed by tectonic processes and erosion, the predominant rock type is a variety of granite. Similarly, seismic velocities in the uppermost layer of the crust are highly variable, reflecting its geological complexity, but below this surface layer velocities and densities become more uniform, implying that a single composition might be justified.

But that is not the whole story, because within many crustal sections a marked increase in seismic velocities occurs at ~15 km depth. Known as the **Conrad discontinuity**, it separates the upper crust from the lower crust and, while not apparent in all sections, it is best developed in regions of older rocks of Archaean or Proterozoic age. Examples of seismic crustal profiles are shown in Figure 1.26.

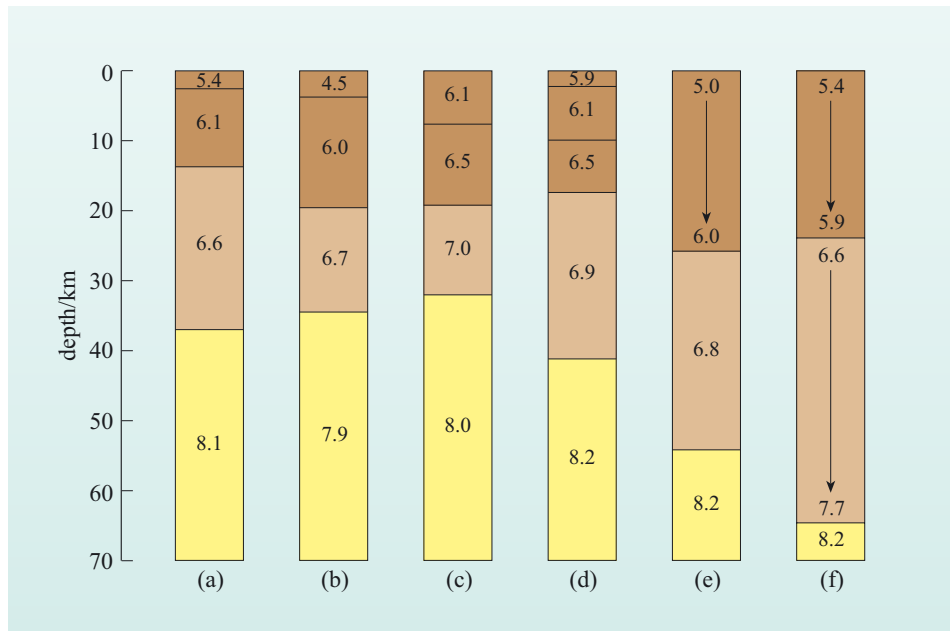


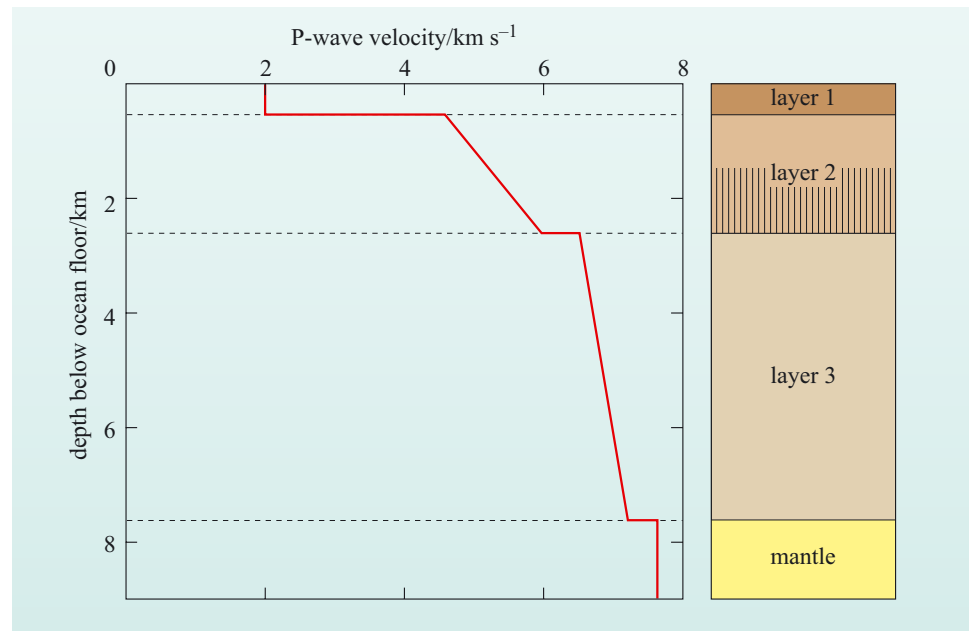
Figure 1.26 Sections through six representative segments of the continents: (a) a stable continental shield, Wisconsin, USA; (b) a modern continental rift, Basin and Range Province, USA; (c) a 400 Ma continental collision zone in northern Scotland; (d) a 100 Ma old ocean–continent plate boundary in southern California; (e) a modern continent–ocean plate boundary, Central Andes, South America; (f) a modern continent collision zone in the central Alps. The arrows in (e) and (f) indicate smooth velocity increases between the values indicated at either end. Seismic velocities of 6.0–6.5 km s⁻¹ are typical of the upper crust and are comparable with those of granodiorite, a variety of granite. Higher velocities of 6.6–7.0 km s⁻¹ are typical of the lower crust, while a sharp increase to velocities of about ≥8 km s⁻¹ marks the Moho at the base of the crust.

The general consensus is that the upper crust has a composition approximating **granodiorite**, which is a type of granite containing plagioclase and alkali feldspar in roughly equal proportions, 20% quartz and small amounts of biotite and hornblende. The lower crust, by contrast, is thought to be made up of granulites with a slightly more mafic composition, i.e. more Fe and Mg and less Si than in granodiorite. **Granulites** are metamorphic rocks that have been subjected to high pressures and temperatures such that they have lost most of their volatile components (largely water) and their mineralogy is dominated by plagioclase feldspar, pyroxenes and garnets, which give the lower crust its higher density. Overall, the bulk composition of the continental crust is thought to be close to that of the intermediate rock type, andesite, or to its plutonic equivalent, diorite.

The oceanic crust contrasts with the continental crust in almost all physical and compositional characteristics. Whereas the continental crust is of variable thickness, the oceanic crust is much thinner and of a more uniform thickness. Where the continents show a diversity of seismic structure, the oceans reveal a structure that can be consistently applied across most of the ocean basins.

The most important difference between continental and oceanic crust is its thickness. Beneath the oceans, the Moho is located on average at <10 km depth, compared with an average of 35 km beneath the continents. Secondly, the oceanic crust has a pronounced layered structure, as illustrated in Figure 1.27. You should recall from Section 1.4 that one of the sources of mantle samples was ophiolites, which are slices of oceanic crust and mantle emplaced by tectonic processes in mountain belts.

Figure 1.27 Variation of seismic velocities with depth in a typical section through the oceanic crust.



- Given the brief description of an ophiolite in Section 1.4, what are the likely compositions of the different seismic layers in Figure 1.27?
- Layer 1 is made up of sediments, layer 2 of basaltic lavas and small intrusions (dykes) and layer 3 of gabbros. These three overlie mantle peridotite.
- What is the average composition of the oceanic crust?
- Disregarding the thin layer of sediments, both layers 2 and 3 have a basaltic composition.

Overall, oceanic crust is basaltic compared to the andesitic composition of the continental crust. Importantly, the continental crust contains much less Fe, Mg and Ca and more Na, K and Si than the oceanic crust.

The combination of low density and greater thickness has important implications for the average elevation of regions underlain by continental and oceanic crust. Superficially, the Earth is composed of solid rocks that behave in a rigid way, and that is certainly correct on length scales of metres and kilometres, but that style of behaviour breaks down over longer length scales and in the mantle where temperatures are much higher. At depth, rocks begin to behave more as fluids than as rigid solids, and when loaded with ice sheets, volcanoes or mountains they begin to flow. This leads to a new mechanical, as opposed to compositional, division of the outer layers of the Earth, the importance of which will be developed in later chapters. The rigid, cold surface layer is known as the **lithosphere** and it includes the crust and the uppermost mantle. Below this is the more mobile mantle known as the **asthenosphere** (from the Greek *asthenos*, meaning weak), which can deform like a viscous fluid when under stress. When a load is applied to the Earth's surface, the lithosphere subsides into the underlying asthenosphere; by contrast, when the load is removed the lithosphere rises again. In essence this is simply Archimedes principle applied on a grand scale, but when considering the Earth it is known as **isostasy** and it controls the surface topography of the Earth.

1.5.1 Isostasy

The principles behind isostasy can be illustrated by considering blocks floating in a liquid, such as wood floating in water or, as in Figure 1.28, metal blocks floating in a container of mercury. Quite clearly, the situation illustrated in Figure 1.28 is probably one that few people will ever witness for real, but it is a representation of an experiment carried out by an eminent Earth scientist, W. Bowie (1872–1940), to illustrate the effects of isostasy for other Earth scientists who found the concept problematic!

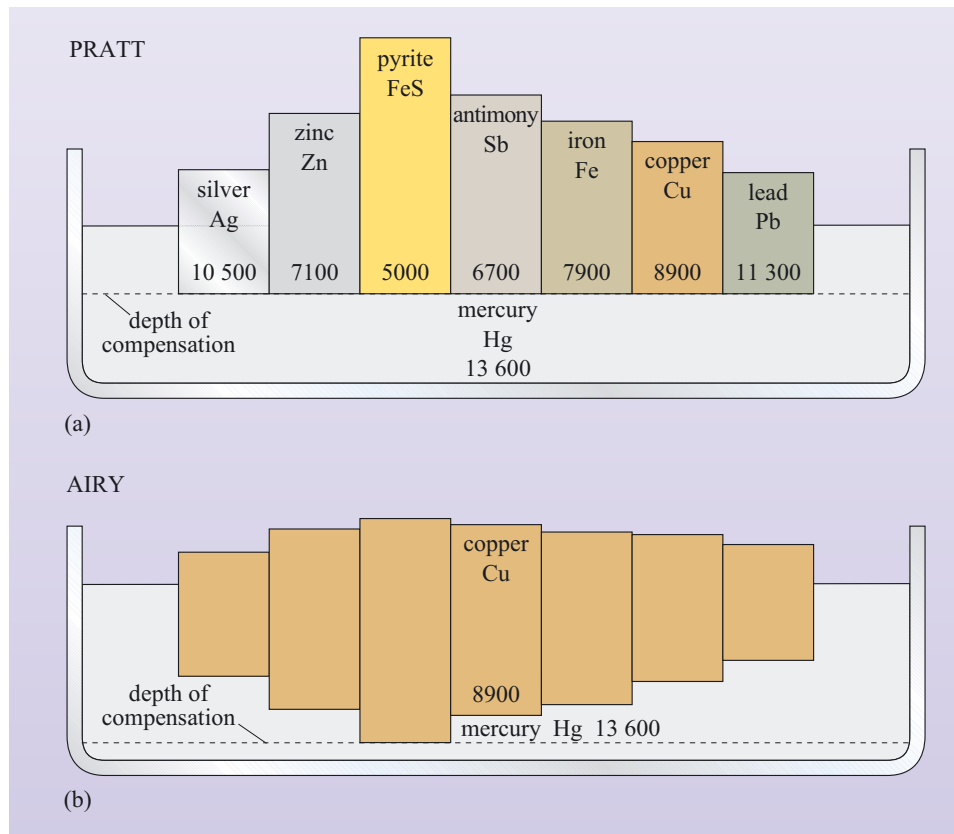


Figure 1.28 The principles of isostasy: (a) a series of metal blocks with different densities (numbers indicate densities in kg m^{-3}) floating in a container of mercury; (b) copper blocks of varying lengths floating in mercury. Note the differences in the depth of compensation. (Watts, 2000)

Figure 1.28 shows how differences in elevation, as reflected by the height of the upper surfaces of the blocks above the surface of the mercury, can be achieved in two ways. In Figure 1.28a, a series of metal blocks with the same cross-sectional area have been cast from identical masses of different metals with different densities; as a result, the blocks have different lengths (heights). When placed in the mercury, according to Archimedes principle each block displaces the same mass and hence the same volume of mercury. Consequently, the blocks exert the same pressure (mass per unit area) on the mercury on which they rest and the base of each block is at the same level in the mercury. This depth is known as the **depth of compensation**. In Figure 1.28b, only one metal is used, in this case copper, but the blocks have different heights. As a result the longer blocks not only stand higher than the shorter ones, they also project deeper into the mercury. Defining the depth of compensation in this case is more difficult, but one way is to relate it to the depth of the lower surface of the longest block. The main difference between this example and that in Fig 1.28a is that the pressure at the compensation level beneath all of the blocks except the longest is made up of the mass of the blocks *and* any overlying mercury.

Fig 1.28 illustrates the principles behind the two models of isostasy as applied to the Earth, proposed in the 19th century by J.H. Pratt (1809–1871) and the then Astronomer Royal, G.B. Airy (1801–1892). At the time, the level of compensation was thought to be the base of the Earth's crust and the Airy model was invoked to explain the presence of a thick, low-density crustal root beneath the Himalaya and Tibet. Seismic profiles beneath this region and other mountainous areas have shown that this is the case and that high mountains are indeed underlain by thicker crust, as shown in Figure 1.26. However, it is now known that the level of compensation in the Earth is probably considerably deeper than the Moho and corresponds at least to the base of the lithosphere or even deeper in the underlying asthenosphere. Thus, within the continents, different elevations could be explained by greater than average lithosphere thicknesses beneath mountains, with thinner lithosphere underlying sedimentary basins, such as shelf seas.

In the oceans, by contrast, the structure is different and the Pratt model of isostatic compensation becomes significant.

- Can you see why this is the case?
- The oceanic crust has a higher density than the continental crust so the comparison is more like Figure 1.29a

The differences in density between oceanic and continental crust and the average thickness of the oceanic crust produce a lithospheric column with a higher average density than the lithospheric column beneath the continents. Assuming a similar depth of compensation beneath both continents and oceans this results in the overall lower elevation for the surface of the oceanic crust.

Thus oceanic crust reaches isostatic equilibrium at a lower elevation than continental crust because it is both thinner and denser. And this is the underlying cause of the terrestrial hypsometric plot discussed earlier in Section 1.1. The lack of a similar bimodal plot for Mars and Venus, suggests that continental and oceanic crust as we understand them in the geological sense are absent from both of these planets. This almost certainly reflects a difference in the behaviour of their interiors from that of the Earth. It may be, for example, that the mantles of both Mars and Venus are much more viscous than the Earth's mantle and do not allow isostatic adjustments in surface elevation in response to different rock densities. Alternatively, as you will discover later in this book, since both continental and oceanic crust are the products of plate tectonic processes, the lack of crustal differences suggests that the internal processes of those two planets are different. Either way, this primary observation suggests that the internal processes of planet Earth are unique in the Solar System and have contributed significantly to the way the Earth and its surface environments have evolved over geological time.

Summary of Chapter I

Having reached the end of this introduction to the Earth, you should now appreciate that the differences between the Earth and the other terrestrial planets are more than just skin deep. The uniqueness of the Earth goes beyond its atmosphere and hydrosphere, being expressed in the contrasts between continental and oceanic crust, the dominance of plate tectonics and the presence of a strong magnetic field. Other characteristics not covered in detail here include the presence of an unusually large satellite, the Moon, and the Earth's relatively rapid rotation rate. These all influence the surface environment of the Earth and the evolution of the biosphere.

After this introduction, you are now in a position to explore aspects of the Earth's structure and evolution in more detail. The following chapters develop some of the themes established here in addition to introducing new ideas concerning the movement of mass and energy within the Earth and how that leads to the dominance of plate tectonics and the cycling of material between the interior of the Earth and its surface and back again.

Learning outcomes for Chapter I

You should now be able to demonstrate a knowledge and understanding of:

- 1.1 The physical and chemical features of the Earth that make it unique compared with other terrestrial bodies in the Solar System.
- 1.2 The internal structure of the Earth and how it is divided into the core, mantle and crust.
- 1.3 How seismic refraction can be used to identify and define different layers within the Earth.
- 1.4 The principle of isostasy and why the Earth's crust is divided into continental highlands and oceanic basins.
- 1.5 The classification scheme used for the different types of meteorites.
- 1.6 How the chemical composition of the bulk Earth and each of its constituent layers have been obtained from various sources including meteorites and samples obtained directly from the mantle.

The Solar System is thought to have begun after one or more local supernova explosions about 4.6 Ga ago. In one widely accepted scenario of the later stages of formation of the Solar System it is thought that there were hundreds of planetesimals in the region occupied by the inner planets. Once the planetesimals began to attain the proportions approaching those of planetary embryos it is likely that the heat generated by collisions would have been sufficient to allow both melting and, as denser materials began to sink inwards, separation of the original constituents. However, the development of a more evolved layering, such as that seen on Earth, would require this separation to have been more or less complete once giant impacts had ceased. This is because from that point onwards further evolution of the Earth would be mainly driven by processes from within the planet.

The first 660 million years of the Earth's existence, known as the Hadean, was the stage during which the metallic core separated from the silicate mantle, the atmosphere and hydrosphere were formed, and melting of the silicate mantle produced the earliest crust. The early Earth was violent and hot, and giant impacts would have been devastating. The heat released would have been capable of melting the outer part of the Earth to form a global **magma ocean**. Convective currents and degassing would have also destroyed any solidified surface areas. Consequently, there is virtually no direct rock record of the Hadean. This means that we have to turn to theoretical modelling and geochemistry to reveal the mechanisms of formation of the different layers in the Earth and the timescales involved.

In this chapter, you will be introduced to the evidence for the development of the early Earth. The first part reviews the various heat sources necessary to drive planetary differentiation. The second part investigates the mechanisms of core formation, the evidence for the presence of a magma ocean, and the timing of accretion and core formation. The third part explores the evidence for the origin and age of the Moon. In the fourth part, you will learn about the formation and evolution of the atmosphere and hydrosphere, and finally, the evidence for the nature and formation of the earliest continental crust will be reviewed. You will discover that the processes that formed the layering in the early Earth, atmosphere and hydrosphere effectively shaped the planet as we know it today, and ultimately provided conditions suitable for life to develop. You will also see that water played a key role, not only in the formation of the atmosphere and hydrosphere, but also in the early development of the solid Earth.

2.1 Heating and differentiation of the Earth

Differentiation is the process by which planets develop concentric layering, with zones that differ in their chemical and mineralogical compositions. The generation of such zones results from a differential mobility of elements due to differences in their physical and chemical properties.

One obvious way of mobilising constituents in a planetary body is by allowing them to melt. When a rock is heated, different minerals within the rock will melt

at different temperatures. This phenomenon is known as **partial melting** and is a key process in the formation of liquid rock or magma, as you will see in Chapter 4. Once elements have been mobilised in this manner, they will begin to migrate under the influence of pressure or gravity.

- Imagine a body the size of a planetary embryo that had accreted from nickel–iron and silicate minerals. Nickel–iron has a density of about $7.9 \times 10^3 \text{ kg m}^{-3}$ (compared to $\sim 3.0 \times 10^3 \text{ kg m}^{-3}$ for silicate minerals) and a melting point some hundreds of degrees higher than silicates. What would happen if temperatures within this planetary embryo were increased to a point at which silicates began to melt?
- Since nickel–iron has a higher melting point, it would remain solid after the silicates had begun to melt and, because it is much denser than any silicate minerals, it would begin to sink towards the centre of the body.

In addition to the separation of a metallic core from a silicate mantle, partial melting of planetary bodies also produces residual solid silicate minerals in contact with a silicate melt that have different compositions, depending on how extensive the melting is. In such a system, incompatible elements are partitioned into the melt more readily than compatible elements (Box 1.3). This process, known as element partitioning, is the principal mechanism by which incompatible elements first become concentrated into the melt. Then, if the magma is buoyant, they migrate upwards to form the overlying crust. If element partitioning continues over time, this concentration and migration will gradually create a crust and mantle with different compositions.

2.1.1 Heat sources

If partial melting is the principal cause of differentiation, then the Earth needs to be heated before layering can begin to develop. There are several sources of heat that can arise during Earth's evolution. The most important are:

- the so-called primordial heat sources, which develop in the early stages of planetary evolution (i.e. those associated with accretion, collision and core formation)
- tidal and radiogenic heating processes, which can operate long after the planet has been formed.

Accretional heating

During accretion, any planetesimal falling towards the Earth will acquire a velocity because of gravitational attraction from the Earth. The body will thus have a kinetic energy due to its motion. If the velocity of the body immediately before impact is v , the kinetic energy (E) at that point will be:

$$E = \frac{1}{2}mv^2 \quad (2.1)$$

where m is the mass of the body. Upon hitting the Earth, if all the kinetic energy of motion is converted into heat, then the increase in temperature, ΔT , can be calculated:

$$\Delta T = \frac{mv^2}{2(m+M)C} \quad (2.2)$$

where the body (of mass m) impacts the Earth (of mass M) and C is the specific heat capacity of Earth material (i.e. the amount of heat required to raise the temperature of 1 kg of material through 1 K).

Question 2.1

- (a) A planetesimal of mass 10^{15} kg impacts the Earth with a velocity of $10\,000\text{ m s}^{-1}$. Calculate the rise in temperature in the Earth assuming that the heat generated by the impact spreads rapidly and uniformly throughout the whole Earth. Because m is much smaller than M , the effect of m is negligible and can be ignored, so Equation 2.2 can be simplified to:

$$\Delta T = \frac{mv^2}{2MC}$$

Take the total mass of the Earth to be 6×10^{24} kg, and the average specific heat capacity of the Earth to be $750\text{ J kg}^{-1}\text{ K}^{-1}$. (Note: $1\text{ J} = 1\text{ kg m}^2\text{ s}^{-2}$)

- (b) Suppose that the Earth was constructed entirely of 10^{15} kg planetesimals, each of which generated the temperature rise obtained in part (a). What would be the total temperature rise?

In practice, the various uncertainties involved make it difficult to determine how much heating occurred during accretion. For one thing, as you will see later in this chapter, not all of the impacting material arrived at the same time: accretion took place over $\sim 10^7$ years. More importantly, not all of the kinetic energy would be converted to heat. For example, some of the impact energy would be used in the excavation of large craters. Finally, much of the heat would have been radiated into space. Nevertheless, most estimates predict temperatures to have risen above the melting point of silicate minerals and iron–nickel, which means that the Earth is likely to have gone through an early molten stage.

Core formation

If the Earth went through an early molten phase, allowing the metals and silicates to separate, then the ‘falling inwards’ of the nickel–iron-rich fraction to form the core would have released **potential energy**.

- What do you think would have happened to this gravitational potential energy?
- The gravitational energy lost by the inward movement of nickel–iron would have been converted first to kinetic energy and then into thermal energy. It is estimated that the core-forming process would have contributed significantly to the Earth’s primordial heating (though it would still have been an order of magnitude less than that generated by collision and accretion).

As you will see later in this chapter, geochemical evidence suggests that accretion and core formation were completed very early in the Earth's history but, rather than being a single 'catastrophic event', it is likely to have been a gradual process, with progressive segregation of the core as more material was accreted.

Since both accretion and core formation relate to events that occurred early in the history of planetary evolution, they are primordial processes and the heat generated by these processes was primordial heat. However, if these primordial heat sources had remained the only way of heating the Earth, their intensity would have waned through time due to continual radiative heat loss to space. Since heat drives fundamental processes such as volcanism, the fact that the Earth has remained volcanically active to the present day requires additional processes of internal heat generation.

Tidal heating

One heat source known to be generated within planetary bodies is tidal heating, which is created by the distortion of shape resulting from mutual gravitational attraction. These effects are readily observed upon the Earth's oceans where the attraction created by the Sun and Moon produces 'bulges' in the ocean water masses that are then dragged around the planet as the Earth rotates. This process produces the ebb and flow of tides seen around the coast. In much the same way, the solid Earth is also distorted by these forces and produces tides that reach a maximum amplitude of about 1 m on the rocky surface. This deformation causes heating within the planet, though precisely where this heating is concentrated depends upon the planet's internal properties. In Earth's case, it is thought to occur largely within the crust and mantle.

Question 2.2

The current rate of heating generated within the Earth by tidal distortion is estimated at $3.0 \times 10^{19} \text{ J y}^{-1}$. Given the mass of the Earth is approximately $6.0 \times 10^{24} \text{ kg}$, determine the rate of tidal heating. Express your answer in W kg^{-1} ($1 \text{ W} = 1 \text{ J s}^{-1}$) and to an appropriate number of significant figures.

Radiogenic heating

During the latter half of the 19th century, the eminent physicist Lord Kelvin (1824–1907) attempted to determine the age of the Earth. He believed the Earth had cooled slowly after its formation from a molten body and assumed the main sources of energy were from primordial heat and tidal friction. Taking many factors into account, including the mass of the Earth, the current rate of surface heat loss, and the melting points of various rock types, he concluded that the planet could not be much older than about 20–40 Ma.

This conclusion was contrary to the ideas of eminent geologists such as James Hutton (1726–1797) and Charles Lyell (1797–1875), who had already argued that immense spans of time were required to complete the changes produced by the action of tectonic, erosional and depositional processes. It was also greatly at odds with the then emerging theories of the evolution of life because scholars such as Darwin (1809–1882) also argued for much longer periods based upon the

evidence of biological speciation. As a consequence, a ‘heated’ controversy continued for many years. Even though Kelvin’s calculations did not gain wide acceptance with geologists he was a powerful scientific influence of the time, and it was not until much later in the 1950s that accurate radiometric dating experiments eventually proved him wrong. These experiments were conducted on primordial lead in meteorites and demonstrated that the formation of the Earth occurred about 4.6 Ga ago (See Section 2.2).

So why did Kelvin get the age of the Earth so wrong? The answer lies partly in the decay of unstable isotopes of certain radioactive elements, the discovery of which was not made until some decades after Kelvin’s initial calculations. John Joly (1857–1933), an Irish physicist, was one of the first to suggest that radioactive decay, leading to radiogenic heating, was an important heat source within the Earth. It is now known that radioactive decay creates an important independent source of heat within the Earth that supplements that remaining from primordial sources (Box 2.1). This radiogenic heating is something that Kelvin could not possibly have known about when making his calculations.

Box 2.1 Radioactive decay and heating

Most elements have different isotopes (i.e. atoms having the same number of protons but different numbers of neutrons). Some of these isotopes are unstable and decay to stable forms. For example, isotopes deficient in protons decay by the transformation of a neutron into a proton and an electron, which is expelled from the nucleus. During this process, known as beta-decay, the mass of the nuclide does not change significantly. In contrast, during alpha-decay, heavy atoms decay through the emission of an α -particle, which consists of two protons and two neutrons (He^{2+}). This process reduces the overall mass of the nuclide. α - and β -particle collision with adjacent nuclei during decay causes heating through the loss of kinetic energy.

The rate of decay of a radioactive parent nuclide to form a stable daughter product is proportional to the number of atoms, n , present at any time, t :

$$\frac{dn}{dt} = -\lambda n \quad (2.3)$$

where λ is the **decay constant** characteristic of the radionuclide in question (expressed in terms of reciprocal time). The decay constant states the probability that a given atom of the radionuclide will decay within a stated time.

The term $\frac{dn}{dt}$ is the rate of change of the number of parent atoms, and is negative because this rate decreases with time. Equation 2.3 can be integrated from $t = 0$ to t , where the number of atoms present at time $t = 0$ is n_0 :

$$n = n_0 e^{-\lambda t} \quad (2.4)$$

A useful way of referring to the rate of decay of a radionuclide is by its **half-life** ($t_{1/2}$), which is the time required for half of the parent atoms to decay.

On substituting $n = \frac{n_0}{2}$ and $t = t_{1/2}$ into Equation 2.4:

$$t_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda}, \text{ where } \ln 2 \text{ is the natural log of } 2. \quad (2.5)$$

The number of radiogenic daughter atoms formed (D^*) is equal to the number of parent atoms consumed (Figure 2.1). So:

$$D^* = n_0 - n \quad (2.6)$$

The total number of daughter atoms (D) consists of those produced by radioactive decay after time t (i.e. D^*) plus those already present at time $t = 0$ (i.e. D_0), that is:

$$D = D_0 + D^*. \quad (2.7)$$

As $D^* = n_0 - n$ (Equation 2.6) we can substitute D^* in Equation 2.7:

$$D = D_0 + n_0 - n$$

Since $n_0 = ne^{\lambda t}$ (from Equation 2.4):

$$D = D_0 + ne^{\lambda t} - n$$

$$D = D_0 + n(e^{\lambda t} - 1) \quad (2.8)$$

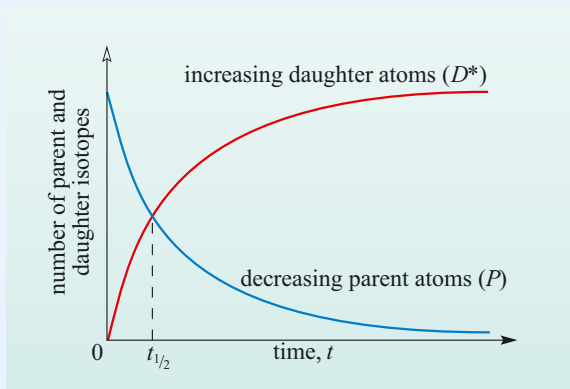


Figure 2.1 The changing number of parent and daughter atoms during radioactive decay. This illustrates that the growth of radiogenic daughter atoms (D^*) is the mirror image of the decay curve of the number of parent atoms (n_0). The half-life ($t_{1/2}$) is the time taken for half of the parent atoms to decay (at which point $n_0 = D^*$).

Table 2.1 lists a number of isotopes that either occur in the Earth or for which there is evidence of their having been active at some time during Earth history.

Table 2.1 Naturally occurring radioactive decay systems of geochemical and cosmochemical interest.

Parent	Decay mode*	Decay constant (λ)/ y^{-1}	Half-life/y	Present heat production [‡]	Daughter isotopes	Measured ratio
^{40}K	β^+ , e.c., β^-	5.54×10^{-10}	1.28×10^9	2.8	^{40}Ar , ^{40}Ca	$^{40}\text{Ar}/^{36}\text{Ar}$
^{87}Rb	β^-	1.42×10^{-11}	4.88×10^{10}		^{87}Sr	$^{87}\text{Sr}/^{86}\text{Sr}$
^{147}Sm	α^\dagger	6.54×10^{-12}	1.06×10^{11}		^{143}Nd	$^{143}\text{Nd}/^{144}\text{Nd}$
^{187}Re	β	1.59×10^{-11}	4.35×10^{10}		^{187}Os	$^{187}\text{Os}/^{188}\text{Os}$
^{232}Th	α	4.95×10^{-11}	1.39×10^{10}	1.04	^{208}Pb , ^4He	$^{208}\text{Pb}/^{204}\text{Pb}$, $^3\text{He}/^4\text{He}$
^{235}U	α	9.85×10^{-10}	7.07×10^8	0.04	^{207}Pb , ^4He	$^{207}\text{Pb}/^{204}\text{Pb}$, $^3\text{He}/^4\text{He}$
^{238}U	α	1.55×10^{-10}	4.47×10^9	0.96	^{206}Pb , ^4He	$^{206}\text{Pb}/^{204}\text{Pb}$, $^3\text{He}/^4\text{He}$
^{26}Al	β^-	9.5×10^{-7}	0.73×10^6		^{26}Mg	$^{26}\text{Mg}/^{24}\text{Mg}$
^{129}I	β^-	4.41×10^{-8}	1.57×10^7		^{129}Xe	$^{129}\text{Xe}/^{130}\text{Xe}$
^{146}Sm	α	6.73×10^{-9}	1.03×10^8		^{142}Nd	$^{142}\text{Nd}/^{144}\text{Nd}$
^{182}Hf	β^-	7.78×10^{-8}	8.9×10^6		^{182}W	$^{182}\text{W}/^{184}\text{W}$
^{244}Pu	α , SF	8.45×10^{-9}	82×10^6		^nXe	$^n\text{Xe}/^{130}\text{Xe}^{**}$

* α = alpha decay (^4He); β^- = beta decay (electron or positron); e.c. is electron capture; SF is spontaneous fission.

† The production of ^4He from ^{147}Sm decay is insignificant compared with that produced by decay of U and Th.

‡ Heat production averaged for the whole Earth in units of $10^{-12} \text{ W kg}^{-1}$ of Earth material (not of the isotope).

** n can be 124, 126, 128 or 129, all of which are produced by ^{244}Pu fission. Element symbols are listed in the Appendix.

■ Which of the isotopes listed in Table 2.1 remain active today and which are extinct?

■ All those with half-lives significantly less than the age of the Earth, i.e. 4.6 Ga, are extinct, namely: ^{26}Al , ^{129}I , ^{146}Sm , ^{182}Hf and ^{244}Pu . The others, principally isotopes of ^{40}K , ^{87}Rb , ^{147}Sm , ^{232}Th , ^{235}U and ^{238}U , are still active today.

If radionuclides have short half-lives and are not replenished by the decay of other isotopes, then they may be lost altogether. One such **short-lived extinct nuclide** is ^{26}Al , which has a half-life of 0.73 Ma.

■ What evidence would you look for to support the presence of ^{26}Al in the early Solar System?

■ ^{26}Al decays to ^{26}Mg (Table 2.1), so you would expect to see anomalously high abundances of ^{26}Mg relative to other isotopes of Mg in materials from the early Solar System.

Studies of primitive meteorites, notably carbonaceous chondrites, do indeed show slightly high $^{26}\text{Mg}/^{24}\text{Mg}$ ratios, which suggests that a significant proportion of the aluminium present at the time of condensation of the solar nebula was the unstable isotope ^{26}Al . This was originally created during a supernova explosion predating

the birth of our Solar System, and cannot be replenished by the spontaneous decay of other radiogenic elements.

- What does the observation that ^{26}Al was present in chondritic meteorites tell you about the timescale of the formation of the Solar System?
- The half-life of ^{26}Al is only 0.73 Ma, so the time between the supernova explosion that generated the ^{26}Al and the accretion of the meteorite parent body must have happened on a similar timescale of a few million years.

Given that after 10 half-lives only 2^{-10} (or $\frac{1}{1024}$ th) of the original number of ^{26}Al atoms remain, then for any measurable amount of radiogenic ^{26}Mg to be found, chondritic meteorites must have formed within, at most, 7.3 Ma of the supernova. Similar arguments are used later in this chapter for the timescales of formation of other planetary bodies and layers within the Earth. The decay of ^{26}Al may have contributed significantly to the heating of planetary embryos but, because of its short half-life, any remaining ^{26}Al has long since vanished within the terrestrial planets.

Whilst such short-lived isotopes may have been important heat sources during the early stages of terrestrial planet evolution, study of Earth material indicates that it is the isotopes of the elements U, Th and K that are responsible for most of the radiogenic heating that has occurred throughout the history of the planet. These isotopes, which all have particularly long half-lives (see Table 2.1), are termed **long-lived radiogenic nuclides** and were present in sufficient quantities after condensation and accretion to ensure that they have remained abundant within present-day Earth.

Most common minerals contain small amounts of the elements with these unstable isotopes, the most important of which are summarised in the upper half of Table 2.1. Their decay to form more stable isotopes releases tiny increments of heat, as described in Box 2.1. This decay has produced a continuous source of heat within our planet since the Earth's formation. Of course, the total amount of heat produced will depend upon the concentration and types of radiogenic isotope present in the parts of the Earth, and the mass of suitable material present in its different layers.

- U, Th and K are incompatible elements (i.e. they are preferentially partitioned into silicate liquids), so where will they be concentrated in the Earth?
- The elements U, Th and K (and their radiogenic isotopes) are particularly concentrated in the silicate-dominated outer layers of the Earth, and in particular within the continental crust. They are thought to be virtually absent from the core.

As a result of the incompatibility of the heat producing elements, the radiogenic heat produced per unit mass of the continental crust is, on average, over one hundred times greater than that of the underlying mantle. But because the mantle is so much more massive than the crust, in effect this means that the overall radiogenic heating budget is roughly split equally between the mantle and crust

despite the much greater mass of mantle material. It is the decay of these long-lived isotopes that provides sufficient heat energy to keep the Earth geologically active. Therefore, the surface heat flux is not simply the slow cooling of a once molten body, as originally envisaged by Kelvin.

Finally, whilst the rate of radiogenic decay is constant for each isotope system (Table 2.1), the total amount of radioactive decay, and hence heat generation, will decline over time as the reserves of the original radioactive materials are gradually used up. This gradual depletion of radioactive materials is expressed in half-lives, and each isotopic decay system has its own unique half-life. For example ^{235}U decays through a series of α -particle emissions to the daughter isotope ^{207}Pb . The data in Table 2.1 indicate that, after 7.07×10^8 years, half of the ^{235}U originally present will have decayed to ^{207}Pb and the remainder will continue to halve every 7.07×10^8 years. Over the age of the Earth (4.6 Ga), approximately 6.5 half lives of ^{235}U have elapsed, so the heat production from ^{235}U is now $(\frac{1}{2})^{6.5} \approx 0.01$ (i.e. 1%) of what it was originally.

Question 2.3

- What proportions of the original ^{40}K and ^{232}Th currently remain since the formation of the Earth?
- Based upon the data in Table 2.1, what was the amount of radiogenic heating in the Earth at 4.6 Ga, and how does it compare with that of today?
- What proportion of Earth's surface heat flux loss is due to radioactive decay, compared with the $1.5 \times 10^{-13} \text{ W kg}^{-1}$ (the value determined in Question 2.2) created by tidal heating effects?

2.1.2 Heat transfer within the Earth

In the previous section you have seen how accretion, core formation and radioactive decay heated the Earth. But how is this internal heat transferred to the surface? Three main mechanisms of heat transfer operate within the Earth; these are **conduction**, **convection** and **advection**.

Conduction This is perhaps the most familiar mechanism, since it is the process of heat transfer experienced when the handle of a pan becomes hot. Heat is conducted from the stove to the pan and then to its handle. Different materials, such as rocks of various compositions, conduct heat at different rates, and the efficiency of heat transfer in this manner is known as conductivity. This method of heat transfer is the most important in the outermost layer of the Earth (i.e. the lithosphere).

Convection This involves the movement of hot material from regions that are hotter to those that are cooler and the return of cool material to warmer regions. During this transfer the material gives up its heat. It is a particularly efficient method of heat transfer, but the medium through which transfer takes place must be fluid. It should be noted, however, that the term 'fluid' describes any substance capable of flowing and is not just restricted to liquids and gases. Under the right conditions, even solid rocks can flow, albeit at a very slow rate. Over long periods of geological time, the effects of such solid-state flow become a highly significant way of transferring heat towards a planet's surface.

Advection The final process of transferring heat is when molten material (magma) moves up through fractures in the lithosphere and remains there. This is termed advection and operates when magma spreads out at the surface as a lava flow or, if it is injected, cools and crystallises within the lithosphere itself. The effect is the same in both cases, since heat is transferred by the molten rock from deeper levels where melting is taking place to shallower levels where it solidifies, losing its heat by conduction into the overlying crust. Any planetary body that exhibits, or has exhibited, volcanic activity must have lost some of its internal heat in this manner.

In Chapter 1 you learned that under the conditions prevailing deep within the Earth the solid rocks of the mantle can flow when subject to surface loads, leading to isostatic readjustment of surface elevations. The mantle can also flow when subject to temperature differences in a process known as **solid-state convection** and, whilst rates may be no more than a few centimetres per year, it is the most efficient form of heat transfer within all but the outermost part of the mantle.

Near the Earth's surface the rocks are too cold and rigid to permit convection, so conduction is the most significant process. You should recall from Chapter 1 that this zone of the uppermost mantle and all of the overlying crust is called the lithosphere. In the underlying asthenosphere the principal process of heat transfer is convection and this process of heat transfer applies to much of the mantle thickness down to the core (Figure 2.2a). The marked difference in strength between the lithosphere and asthenosphere is caused by increasing temperature with depth below the Earth's surface (the **geotherm**) (Figure 2.2b). So the thickness of the lithosphere depends on the rate at which temperature increases with depth.

Finally in this section, it is of interest to return to Kelvin and the debate concerning the age of the Earth. Kelvin assumed that the Earth cools by conduction alone. You can see from Figure 2.2b that the geotherm for the convecting mantle is much shallower than that in the lithosphere. In this way, convection in the asthenosphere maintains a *higher* geothermal gradient within the lithosphere than would occur by conduction alone, thus mimicking the geotherm of a planet that is much younger

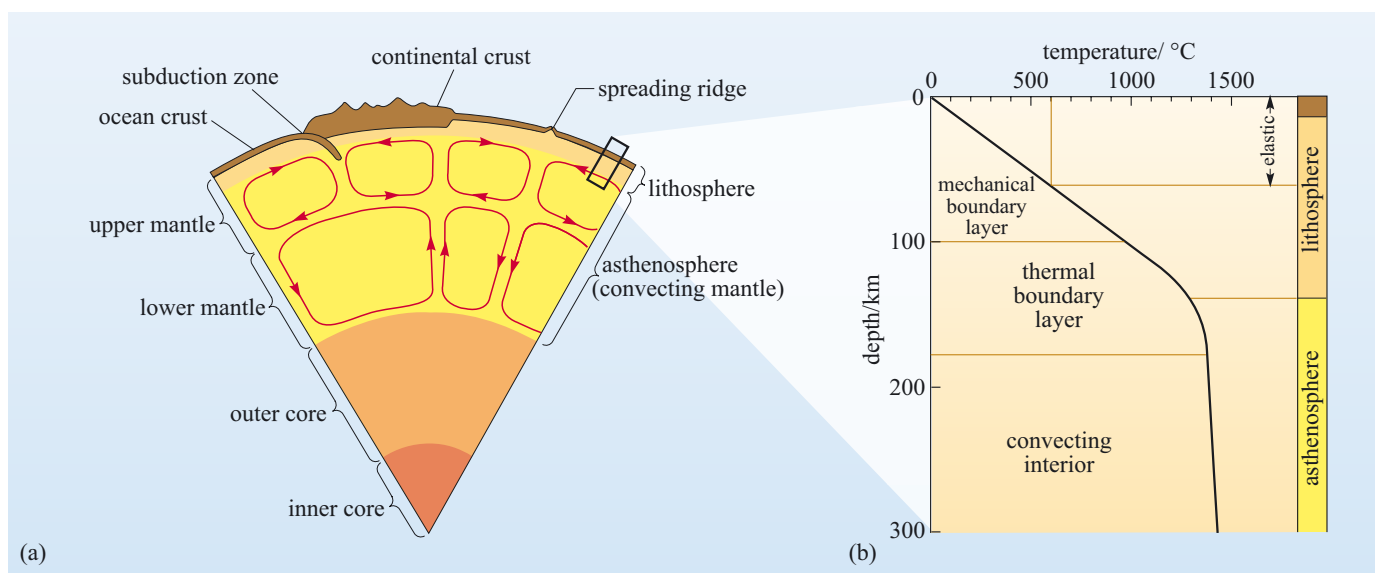


Figure 2.2 (a) Section through the Earth showing the division of the mantle into the uppermost rigid lithosphere and the mobile, convecting asthenosphere. (b) Geotherm through the lithosphere and uppermost asthenosphere.

than the Earth. By assuming that the surface, conductive geotherm applied to the whole Earth, Kelvin arrived at an erroneously young age for the Earth – a result he would have obtained even if he had included the correct estimates for radiogenic heat production!

2.2 The age of the Earth and its layers

Throughout this and the preceding chapter the age of the Earth has been given as being around 4.6 Ga. But where is the evidence for this? To find out just how old the Earth is we once again have to return to meteorites and radioactivity, for, in addition to being sources of heat in planetary systems, radioactivity also allows absolute ages to be determined from measurements of long-lived radioactive isotopes and their daughters.

Several isotope systems are used to date events and processes from throughout Earth history, but the three most commonly used are the K–Ar, U–Th–Pb and Rb–Sr systems. The principles of radiometric dating are most clearly illustrated using the Rb–Sr system, as outlined in Box 2.2, and isotope data from this and the U–Th–Pb system are most frequently illustrated on an **isochron plot** (or **isochron diagram** or **isotope evolution diagram**), examples of which are shown in Figure 2.3.

Figure 2.3a shows Rb–Sr isotope data from a series of ordinary chondrites that define an isochron age of ~4.50 Ga. This age relates to the last time the Rb and Sr were fractionated from each other by a particular process. In the case of Rb and Sr, both elements are lithophile (Box 1.3), so it is unlikely that they were fractionated by the separation of a metallic phase from a silicate fraction. However, Rb, being a Group 1 alkali metal, is significantly more volatile than Sr, a Group 2 element similar to Ca, which is one of the early condensing elements (Figure 1.16). Hence the Rb/Sr fractionation may relate to the loss of a volatile phase; the age indicates when the Rb/Sr ratio in ordinary chondrites was last disturbed.

Figure 2.3b shows a slightly more complex plot of data relating to the U–Pb system. You should notice from Table 2.1 that the U–Pb system has two parent

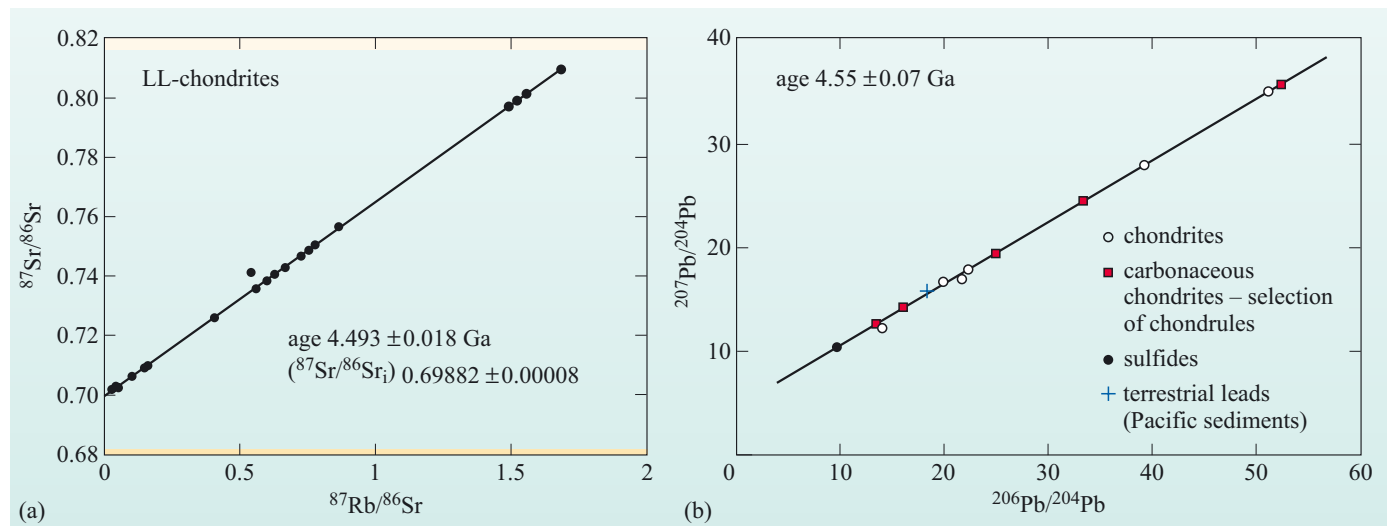


Figure 2.3 (a) Rb–Sr isochron plot for a suite of ordinary chondrites, giving an age of $4.49 \pm 0.02 \text{ Ga}$. (b) Isochron plot of $^{206}\text{Pb}/^{204}\text{Pb}$ against $^{207}\text{Pb}/^{204}\text{Pb}$ for chondritic meteorites giving an age of $4.55 \pm 0.07 \text{ Ga}$. (Minster and Allègre, 1980; Murty and Patterson, 1962)

isotopes, ^{235}U and ^{238}U , decaying to ^{207}Pb and ^{206}Pb respectively. By combining these two it is possible to eliminate the U/Pb ratio and determine an age from the plot of $^{207}\text{Pb}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$. In this case, the ages represent the time at which U was fractionated from Pb and, as Pb is a moderately volatile element and can be lithophile, siderophile or chalcophile in different environments, it is less easy to define the process that led to U/Pb fractionation. However, iron meteorites are rich in Pb and poor in lithophile U, so the age probably represents the timing of the separation of a metallic phase. Given that the chondrite isochron passes through the Pb isotope ratio of most iron meteorites, it adds further support to this idea.

The data illustrated in Figure 2.3 are for whole meteorite samples but, as we have seen in Chapter 1, meteorites are far from homogeneous, comprising a number of different components. Primitive carbonaceous chondrites are thought to be amongst the least differentiated material in the Solar System. Among other things, they contain chondrules and calcium- and aluminium-rich inclusions (CAIs). Chondrules are millimetre-sized spherical droplets believed to have been produced when mineral grain assemblages were flash heated and cooled quickly. CAIs are typically centimetre-sized and consist of the first minerals to condense at equilibrium from a gas of solar composition. A detailed study of CAIs and chondrules yielded a $^{206}\text{Pb}/^{207}\text{Pb}$ isotope age for CAIs is 4567.2 ± 0.6 Ma, whereas that of chondrules is 4564.0 ± 1.2 Ma.

- What is the difference between the ages of CAIs and chondrules, and how old then are carbonaceous chondrites?
- The data give an interval of 3.2 ± 1.8 Ma between formation of the CAIs and chondrules – carbonaceous chondrites must have formed at or after the time of formation of the chondrules i.e. 4564 Ma.

Even though the difference between these two ages is small, it is greater than the combined uncertainty associated with the two ages – they are significantly different. The difference represents a real difference in the timing of the formation of the CAIs and chondrules.

These data show that the oldest components of meteorites, and hence the Solar System, must be close to 4.57 Ga old, but how do we know that this age also applies to the Earth?

Part of the answer to this question lies in Figure 2.3b, where the average Pb isotope ratios of Pacific sediments are compared with the data from chondrules. The sediment data fall on or close to the meteorite isochron, implying ultimate derivation from a similar source or common parent. Other evidence is found in lunar samples returned by the Apollo missions, which have ages that extend back through the Hadean. The oldest igneous rocks from the Moon are samples of anorthosite. One clast has been dated at 4.56 ± 0.07 Ga using the ^{147}Sm – ^{143}Nd system, placing the formation of the Moon to within 70 Ma of the start of the Solar System. But this in itself raises the question of what we are trying to date – what do we mean by the age of the Earth? In particular, given that radiometric dating systems date the time of element fractionation and therefore reflect the effects of different processes during Earth accretion and differentiation, it is probably more precise to consider the timing of major processes in the formation of the Earth.

Hence the following sections focus on determining the timing of important events in the history of the early Earth in relation to the age of the Solar System, which is given as 4.57 Ga.

Box 2.2 Radioactivity applied to dating

The rubidium–strontium isotope system provides a good illustration of the principles of isotope dating, and will be used here to demonstrate those principles. The number of ^{87}Sr daughter atoms produced by the decay of ^{87}Rb in a rock or mineral since its formation t years ago is given by substitution into the radioactive decay equation (Equation 2.8).

$$^{87}\text{Sr} = ^{87}\text{Sr}_i + ^{87}\text{Rb}(e^{\lambda t} - 1) \quad (2.9)$$

where $^{87}\text{Sr}_i$ is the number of ^{87}Sr atoms initially present. It is, however, difficult to measure precisely the absolute abundance of a given nuclide. Mass spectrometers can, however, measure isotope ratios to very high accuracy and precision, and so it is more convenient to work with isotope ratios by dividing by the number of atoms of ^{86}Sr (which is a stable isotope and therefore remains constant with time).

$$\left(\frac{^{87}\text{Sr}}{^{86}\text{Sr}}\right)_p = \left(\frac{^{87}\text{Sr}}{^{86}\text{Sr}}\right)_i + (e^{\lambda t} - 1) \left(\frac{^{87}\text{Rb}}{^{86}\text{Sr}}\right) \quad (2.10)$$

y c m x

The present-day Sr isotope ratio $(^{87}\text{Sr}/^{86}\text{Sr})_p$ is measured by mass spectrometry, and the $^{87}\text{Rb}/^{86}\text{Sr}$ ratio can be calculated from the measured concentrations of Rb and Sr. If the initial ratio $(^{87}\text{Sr}/^{86}\text{Sr})_i$ is known or can be estimated then t can be determined, if it is assumed that the system has been closed to Rb and Sr mobility from the time t to the present.

Most rocks are many millions of years old, in which case it is difficult to estimate the initial Sr isotope ratio. However, examination of Equation 2.10 shows that it is equivalent to the equation for a straight line.

$$y = c + mx \quad (2.11)$$

By plotting $(^{87}\text{Sr}/^{86}\text{Sr})_p$ on the y -axis against $^{87}\text{Rb}/^{86}\text{Sr}$ on the x -axis, the intercept c is then the **initial ratio** of the system (Figure 2.4). On such a diagram, a suite of co-genetic rocks or minerals having the same age define a line termed an **isochron**, and the diagram is called an isochron plot. The slope of the isochron, $m = e^{\lambda t} - 1$, yields the age of the rocks or minerals. If one of the rocks or minerals is Rb-poor then this may yield the initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio directly. Otherwise, the initial ratio is determined by extrapolating back to the y -axis using a best-fit line through the available data points.

The isotope evolution of a suite of hypothetical minerals is shown in Figure 2.4. At the time of crystallisation of the minerals all four have the same $^{87}\text{Sr}/^{86}\text{Sr}$ ratio, and plot as points on a horizontal line (AB). After each of these minerals becomes closed to exchange of Rb and Sr isotopes, evolution begins. The sloping line (AC) represents the $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{87}\text{Rb}/^{86}\text{Sr}$ ratios measured today.

- Study Figure 2.4. What has happened after time t ?
- ^{87}Rb has decayed to ^{87}Sr over the period of time t to the present day. This has had two effects: (i) it has reduced the amount of ^{87}Rb present, so the $^{87}\text{Rb}/^{86}\text{Sr}$ ratios have *decreased* slightly; (ii) it has increased the amount of ^{87}Sr present and so the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios have *increased*.
- But why have the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios increased by differing amounts along the line AC?
- The amount of ^{87}Sr produced by radioactive decay depends on the amount of ^{87}Rb present. However, Figure 2.4 uses isotope ratios; hence the *relative* increase in ^{87}Sr (the increase in $^{87}\text{Sr}/^{86}\text{Sr}$) depends on the *relative* amount of ^{87}Rb (or the $^{87}\text{Rb}/^{86}\text{Sr}$ ratio). Thus, after a given period of time the samples with the highest $^{87}\text{Rb}/^{86}\text{Sr}$ ratio will also be those with the highest $^{87}\text{Sr}/^{86}\text{Sr}$ ratio.

The four samples shown on Figure 2.4 represent an idealised case. For an isochron to yield a slope that reflects a true age then the following assumptions must be valid.

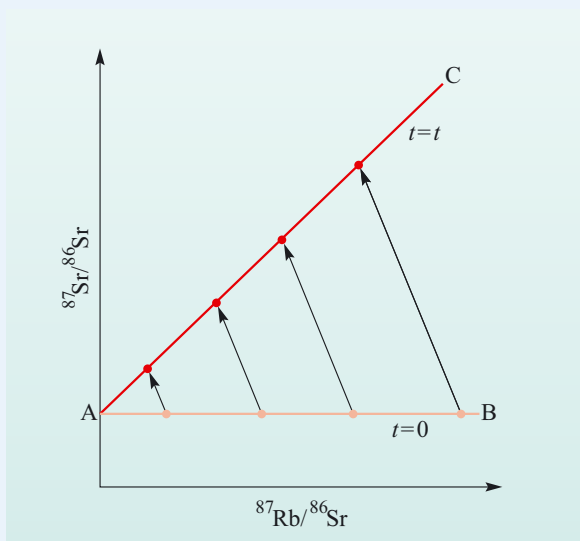


Figure 2.4 Isochron plot of $^{87}\text{Sr}/^{86}\text{Sr}$ against $^{87}\text{Rb}/^{86}\text{Sr}$ illustrating how four samples of the same age but different Rb/Sr ratios evolve from a horizontal line (AB) at the time of their formation, to plot on a straight line (AC) with a slope $m = e^{\lambda t} - 1$ at the present time.

- 1 All samples must be of the same age. If the samples are from the same igneous intrusion or are minerals from the same rock then this is a reasonable assumption. However, if the samples are from a large area where the geology is poorly understood, then they may be of a different origin and may not be the same age and will not plot on a straight line.
- 2 All samples must have the same initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio when they formed, otherwise they will not plot on the same line on the isochron diagram (AB on Figure 2.4) even if they are of the same age. This assumption is likely to be valid for a suite of igneous rocks that have crystallised from the same parental melt, but it is less likely for sediments that are composed of differing amounts of recycled material.
- 3 The $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{87}\text{Rb}/^{86}\text{Sr}$ ratios of the samples have only changed by the process of radioactive decay – no Rb or Sr has been added to or lost from them between the time of formation and the present day. Such **open-system behaviour** might occur due to the infiltration or loss of fluids, or diffusional exchange between minerals after their crystallisation.

2.2.1 Core formation and magma oceans

The formation of the metallic core is the biggest differentiation event that has affected the Earth, resulting in a large-scale change in the distribution of composition, density and heat production. One would think that such a fundamental feature would be well understood, but the physical mechanism by which metal separates from a silicate mantle and accumulates at the centre of the Earth remains poorly understood.

One potential mechanism for Fe–Ni metal separation or **segregation** is that the metal melts and forms an interconnected network.

Whether or not this happens depends on a property known as the **dihedral angle**, θ

(Figure 2.5). The dihedral angle is that formed by the liquid in contact with two solid grains,

which in the case of the mantle will be silicate or oxide grains.

If the dihedral angle θ is $<60^\circ$, the melt will fill channels between the solid grains and form an interconnecting network, even in small melt fractions. On the other hand, if θ is $>60^\circ$, the melt is confined to pockets at grain corners and cannot easily move, unless the melt fraction is greater than 10%.

If melt is able to connect, its rate of migration is quite rapid, and can be calculated using Darcy's law:

$$v = \frac{k}{\eta} \Delta \rho g \quad (2.12)$$

where v is the velocity of the melt relative to the solid matrix, k is the permeability, η is the viscosity of the melt measured in Pa s, $\Delta \rho$ is the density difference between the melt and the solid, and g is the acceleration due to gravity.

Permeability can be defined as:

$$k = \frac{a^2 \Phi}{24\pi} \quad (2.13)$$

where a is the mean grain radius and Φ is the melt fraction.

Question 2.4

Taking a grain radius, a , of 10^{-3} m (1 mm), Φ of 0.1 (10% volume melt), $\Delta \rho$ of 3500 kg m^{-3} , g of 9.8 m s^{-2} (the acceleration due to gravity on Earth) and a viscosity, η , of 0.005 Pa s , calculate the migration velocity of Fe–Ni metal (give your answer in kilometres per year). (Note: $1 \text{ Pa s} = 1 \text{ kg m}^{-1} \text{ s}^{-1}$)

Such calculations show that any metallic melt that can form an interconnected network ought to sink rapidly to form a core. The key question then is the extent to which the metal connects.

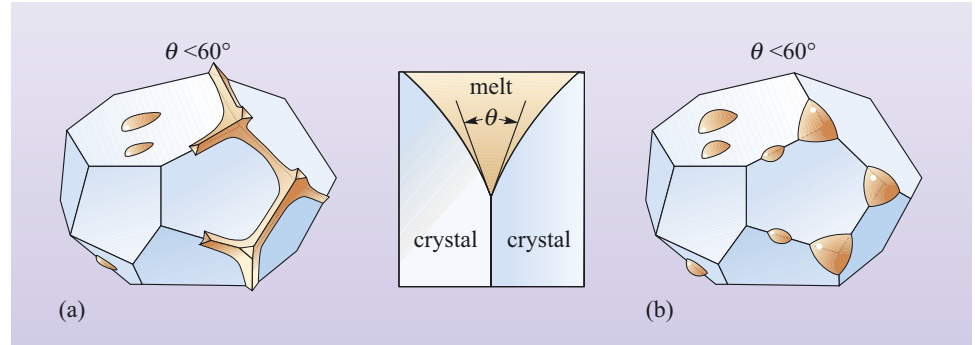


Figure 2.5 The definition of the dihedral angle θ , and the depiction of two different microstructures for static systems: (a) wetting, where melt forms an interconnected network along grain edges, and (b) non-wetting, where melt forms isolated pockets at grain corners. (Rushmer et al., 2000)

- Experiments indicate that the dihedral angle θ is $>60^\circ$ for metallic melts. To what extent will such melts form an interconnected network?
- If θ is $>60^\circ$ then melts will be isolated at grain corners, creating an impermeable silicate framework through which metallic melts cannot segregate.

For this reason core formation is thought by many to occur only after the silicate framework has broken down after extensive silicate melting ($>40\%$). At these high degrees of melting the grain boundary framework will no longer be interlocked, but rather crystals will be floating in a silicate liquid – a crystal mush. In such a mush, dense molten metal droplets would sink, but to achieve such high degrees of melting requires enormous amounts of heat.

- How might such heating have occurred on the early Earth?
- As you have previously seen, accretional heating during planetesimal–embryo Earth collisions would result in extensive melting, which is likely to have resulted in a global magma ocean.

It is important to note that there is no independent evidence that a magma ocean ever existed on Earth. Any early formed crust has long since been destroyed by impacts, erosion and plate recycling. Nevertheless, a deep magma ocean is thought to be a likely consequence if accretion included giant impacts as predicted in modern theory. The extent and depth of any magma ocean depends on many factors, including the impactor/Earth mass ratio, impact velocity and initial temperatures of the impactors.

The evidence also suggests that the Earth had a huge proto-atmosphere, formed by degassing of the Earth's interior. This would have provided a thermal blanket that retained the heat generated during accretion and sustained the magma ocean. Finally, as you will see later in Section 2.3, a giant collision is the most popular theory for the origin of the Moon, and such an impact would have delivered sufficient energy to melt the entire planet. These accretional and impact sources of heat would also have been supplemented by radioactive decay, which would have been much greater than at the present day.

If the entire silicate mantle were molten, the metal blobs would sink directly to the centre of the Earth, but if the base of the magma ocean were solid silicate then liquid metal might pool at this boundary until gravitational instability permits movement of the metal to the core in large **diapirs**. Alternatively, metal from the impactor and perhaps even the Earth's proto-core may have become highly fragmented and emulsified in the magma ocean, in which case small metallic droplets could equilibrate rapidly with silicate melt.

Cooling of a peridotite magma ocean would eventually lead to crystallisation from the bottom up. In the absence of an atmosphere, heat radiation to space would have been very efficient and the deeper part of the magma ocean would have cooled very quickly (i.e. in less than 1000 years). However, the upper part of the magma ocean could have remained hot and molten much longer, maybe for more than 10^7 – 10^8 years, especially if a cooled crust formed at the surface. This situation may have been complicated by the continuous supply of new material from meteorites that remelted on impact, and tidal heating.

2.2.2 Core–mantle equilibration

You should recall from Chapter 1 that one consequence of core–mantle separation is that the metal-loving siderophile elements would be strongly partitioned into the metallic core. However, trace amounts of siderophile elements are retained in the mantle and if metal segregation were an equilibrium process then these elements would provide important clues for deducing the conditions of core formation.

Figure 2.6 shows the abundances of siderophile elements subdivided into slightly, moderately and strongly siderophile, in the Earth's mantle normalised to chondritic meteorites, revealing the stepped abundance profile as discussed in Section 1.4. An assumption of many early core formation models was that metal segregation was contemporaneous with accretion, and that metal and silicate equilibrated at near-surface, low-temperature (T) and low-pressure (P) conditions. Low T – P metal silicate distribution coefficients (Box 1.3) for highly siderophile elements have been determined experimentally and found to lie between 10^{-7} and 10^{-15} . These values lead to the expected abundances in the mantle as shown in Figure 2.6.

- How do the abundances of the highly siderophile elements observed in the silicate mantle compare with what would be expected from the experimentally determined low-pressure partition coefficients?
- The observed data (triangles in Figure 2.6) indicate a depletion of about 2×10^{-3} relative to chondrites (i.e. present at $\sim 0.2\%$ of the chondritic abundance), whereas the partition coefficients predict depletions of 10^{-5} and 10^{-6} (filled circles in Figure 2.6).

As you have seen, metal segregation in a magma ocean would probably occur over a range of temperatures and pressures, and so equilibrium metal segregation at high temperature and high pressure in a deep magma ocean becomes a realistic possibility. Applying experimentally determined partition coefficients to

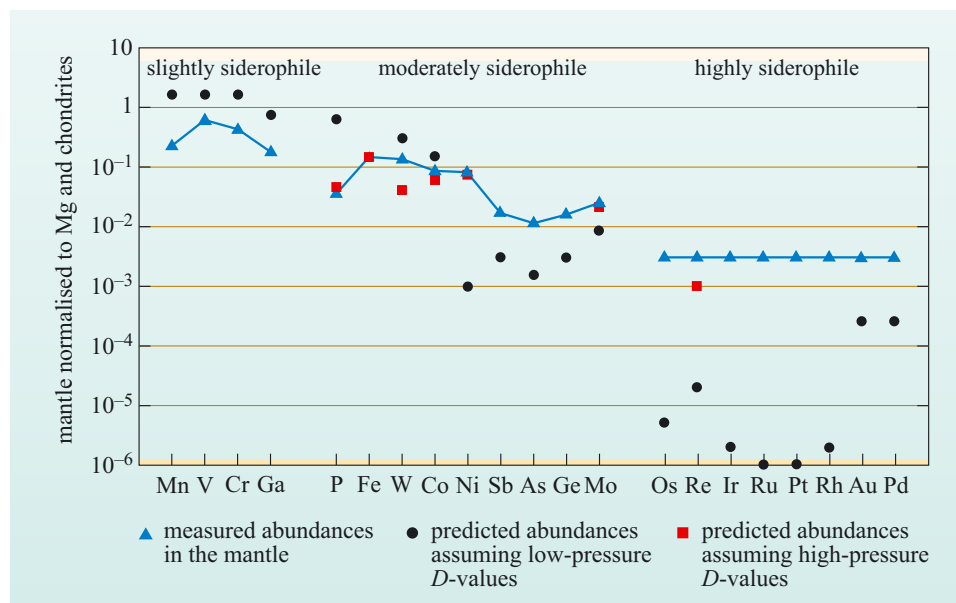


Figure 2.6 The abundance of siderophile elements in the present-day Earth's upper mantle. Also shown are the mantle abundances expected for selected elements in the mantle, assuming separation of the core as metal droplets sinking through a magma ocean and equilibrating according to low- and high-pressure partition coefficients respectively. (Modified after Drake and Righter, 2002)

metal segregation at high pressures can reproduce the abundance of Re, but as yet experimental data on other elements are lacking. However, given their dramatically different partitioning behaviour at low temperature and pressure it is unlikely that there exists any set of conditions at which all siderophile partition coefficients converged to a single value, which is required by the uniform depletion of the most highly siderophile elements.

The failure of low-temperature, low-pressure metal/silicate equilibration models to explain the siderophile excess inspired a number of alternative models, including:

- partitioning between a sulfur-rich liquid metal and silicate
 - inefficient core formation, where small amounts of metal or sulfide remain behind in the mantle
 - the heterogeneous accretion or **‘late veneer’ model** in which core formation effectively strips out all the siderophile elements from the mantle, which are subsequently raised to the observed values by another process. This is the most popular model.
- What process could have raised siderophile element abundances in the silicate mantle following core formation?
 - Continued accretion of meteoritic materials from new impacts of meteorites with chondritic proportions of the siderophile elements.

In contrast to the highly siderophile elements, the abundances of moderately siderophile elements would not have been significantly altered by the later addition of meteoritic material because they would not have been so efficiently stripped from the mantle by core formation. Potentially, moderately siderophile elements can provide much more information on the conditions of core formation. Of these elements, Ni and Co provide some key constraints because their abundances in the mantle are accurately and precisely known, and their partitioning behaviour has been studied over a wide range of conditions.

- With reference to Figure 2.6, estimate the Ni/Co ratio of Earth’s mantle at the present day.
- Figure 2.6 shows that within uncertainty, Ni and Co are present in proportions that are close to chondritic, i.e. both at $\sim 0.1 \times$ chondrite.

The chondritic ratio of Ni to Co in the mantle requires the ratio of the two partition coefficients $D_{\text{Ni}}/D_{\text{Co}}$ to be about 1.1. Experiments show that an increase in pressure and/or temperature causes both Ni and Co to become less siderophile, but at different rates (Figure 2.7).

- Assuming a $D_{\text{Ni}}/D_{\text{Co}}$ ratio of about 1.1 is required to explain the Ni/Co ratio of the Earth’s mantle, use the experimental data in Figure 2.7 to estimate the pressure of metal–silicate equilibration.
- From Figure 2.7, a $D_{\text{Ni}}/D_{\text{Co}}$ ratio of 1.1 occurs at a pressure of about 28 GPa equivalent to a depth of 900–1000 km, implying high temperature and pressure metal–silicate equilibration and core segregation.

Figure 2.8 summarises the conceptual model of metal segregation and metal/silicate equilibrium in a deep magma ocean. For the model geotherm shown, the entire upper mantle would be molten (a magma ocean), whereas the lower mantle would be solid.

In the upper part of the mantle, equilibrium metal segregation from the upper mantle would occur by the ‘rain-out’ of small, liquid metal globules over a wide range of temperature and pressure conditions, and the metal would accumulate at the magma ocean floor. At the boundary between the lower and upper region, the metal would equilibrate a final time, giving the upper mantle its present siderophile element signature. Finally, gravitational instability would cause the formation of large metal diapirs that sink through the lower region, with or without re-equilibration. (Note: these divisions do not necessarily correspond to the present-day upper and lower mantle.)

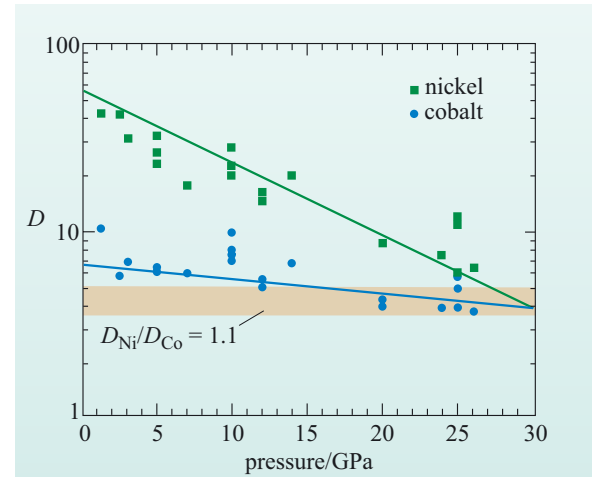


Figure 2.7 Metal/silicate partition coefficients (D) as a function of pressure for nickel (Ni) and cobalt (Co). See text for explanation. (Walter et al., 2000)

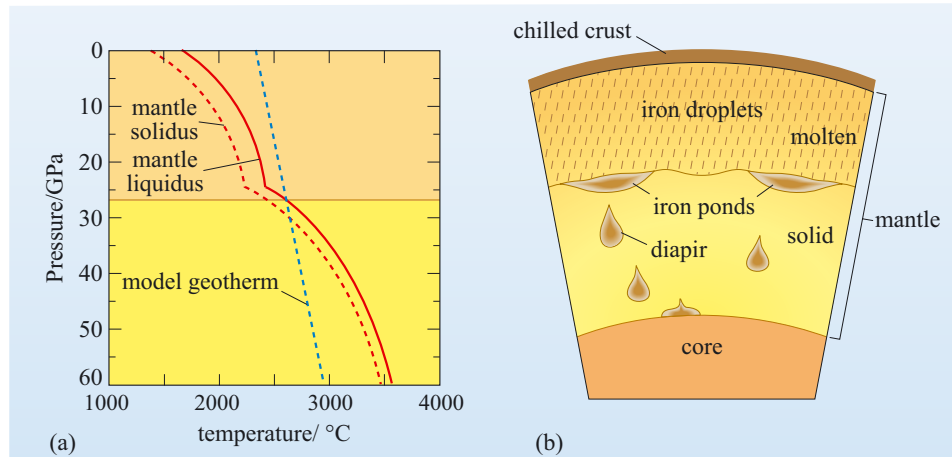


Figure 2.8 Pressure–temperature diagram showing metal/silicate equilibrium and metal segregation in a deep magma ocean.

(a) Pressure–temperature plot of the mantle solidus and liquidus (Chapter 4) compared with the mantle geotherm. Where the geotherm lies above the liquidus, the mantle is completely molten (upper region); where it lies below the solidus it is completely solid (lower region). (b) Cartoon section through the Earth’s mantle early in Earth history showing metal droplets falling to the base of the mantle magma. The droplets accumulate into larger bodies of dense, molten metal that eventually sink through the solid lower mantle as diapirs and accumulate in the core. (Modified from Walter et al., 2000)

2.2.3 The timing of core formation

In order for any radioactive decay system to be of use in dating a process, that process must be able to fractionate the parent element from the daughter element. Thus, in order to investigate the timing of core formation, radioactive systems are needed in which one of either the parent or the daughter elements is siderophile and the other is lithophile.

- Which of the radioactive decay schemes in Table 2.1 satisfy this criterion? (*Hint*: use Figure 1.24 to determine the dominant geochemical properties of the elements involved.)
- The two decay schemes with elements of contrasting properties are hafnium–tungsten (Hf–W) and uranium–lead (U–Pb).

Of the two systems, U–Pb involves long-lived isotopes, whereas the Hf–W scheme has a much shorter half-life – it is much easier to illustrate the principles of the approach to this problem using the Hf–W system.

The hafnium–tungsten system

In addition to the contrasting geochemical properties of Hf and W during core formation, both elements are refractory and so were accreted to the Earth in chondritic proportions, therefore the Hf/W ratio of the bulk silicate Earth is known relatively well.

- Assuming the bulk silicate Earth has a chondritic Hf/W ratio, how will the Hf/W ratio of the core and mantle differ from that in chondrites?
- The core will have an Hf/W ratio lower than chondrites because W is siderophile and will be enriched in the core. Hf, being lithophile, will remain in the mantle, which will have an Hf/W ratio greater than in chondrites.

^{182}Hf decays by β decay to ^{182}W with a half-life of 8.9 Ma (Table 2.1).

- If the core separated from the mantle magma ocean while ^{182}Hf was sufficiently abundant, what would happen to the $^{182}\text{W}/^{184}\text{W}$ ratio of (a) the mantle and (b) the core?
- At the time of element fractionation, the $^{182}\text{W}/^{184}\text{W}$ ratio of both core and mantle would be similar. However, with time, that of the mantle would increase rapidly because of its high Hf/W ratio, while that of the core would increase less rapidly because of its low Hf/W ratio.

The effect of Hf/W fractionation on W isotopes is well illustrated by measurements of W isotopes in iron meteorites. The data are illustrated in Figure 2.9 and reveal

that the metal from iron meteorites has low $^{182}\text{W}/^{184}\text{W}$ ratios (low $\epsilon^{182}\text{W}$) (see Box 2.3), lower than both chondritic meteorites and the silicate Earth (the mantle). The simplest explanation for this is that these metals sampled early Solar System tungsten before live ^{182}Hf had decayed.

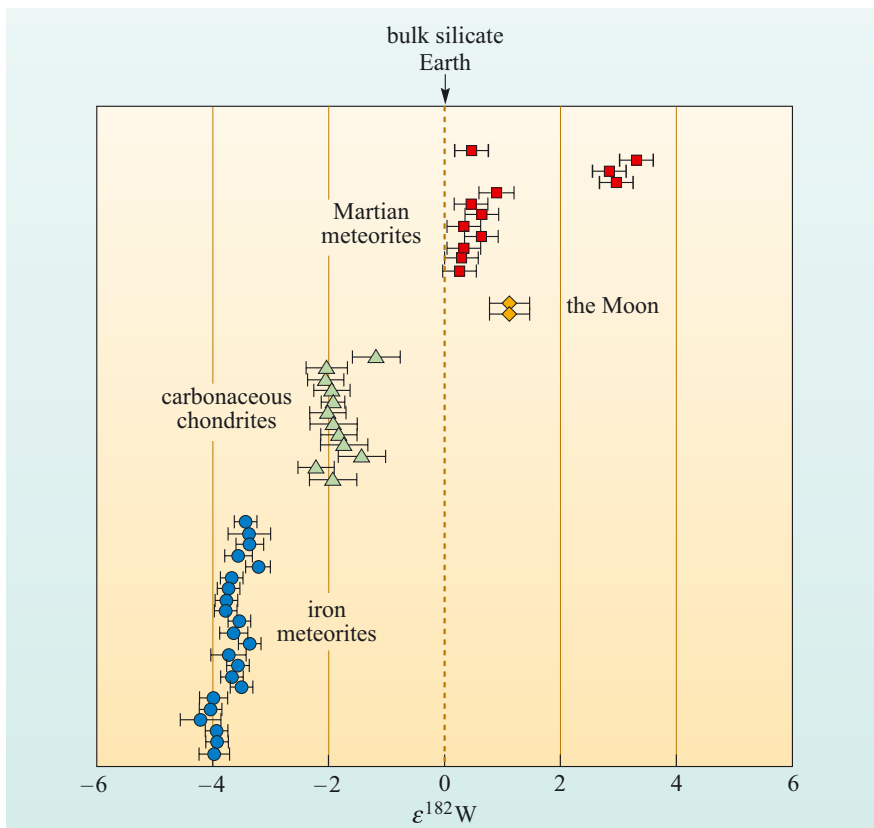


Figure 2.9 $\epsilon^{182}\text{W}$ for iron meteorites, carbonaceous chondrites, the bulk silicate Earth and the Moon. These show a well-defined deficiency in ^{182}W in early metals and carbonaceous chondrites relative to the bulk silicate Earth (BSE). Error bars denote analytical uncertainty on individual measurements. (Data from Foley et al., 2005; Klein et al., 2002; Yin et al., 2002; Scherstein et al., 2006; Lee et al., 2002)

Box 2.3 Tungsten isotope ratios and their notation

The beta decay of ^{182}Hf to ^{182}W ($t = 9 \text{ Ma}$) has proven to be of enormous value in determining the relative timing of events during planetary accretion and core formation. Because ^{182}Hf is now extinct, evidence for its original presence is recorded in the isotope ratios of its daughter element, tungsten (W). Tungsten consists of five stable isotopes but only one of these, ^{182}W , has been partly produced by the radioactive decay of ^{182}Hf . Therefore, to use this geochronometer, geochemists need to measure the abundance ratio of ^{182}W to another of the isotopes of W, conventionally ^{184}W .

The magnitude of the variations in the $^{182}\text{W}/^{184}\text{W}$ ratio relate to the parent/daughter ratio, $^{182}\text{Hf}/^{184}\text{W}$, in the original material. This was of the order of 10^{-4} . Consequently, variations in the $^{182}\text{W}/^{184}\text{W}$ ratio will be of a similar magnitude.

Modern mass spectrometers are capable of measuring isotope ratios to very high precision, in some cases to better than 10 parts in a million (0.001%). This means that isotope ratios can be and often are quoted to six decimal places. Dealing with such numbers is inconvenient, especially when significant differences occur in the fourth and fifth decimal places. To cope with this, geochemists have adopted a convention in which differences in isotope ratios are compared to a standard value, usually one that relates to the bulk silicate Earth or meteorites. In the case of tungsten isotopes, differences in isotope ratios are measured in the number of parts in 10^4 (10 000) and are designated by the Greek letter ε (epsilon), defined according to the following equation:

$$\varepsilon^{182}\text{W} = \left[\frac{\left(\frac{^{182}\text{W}}{^{184}\text{W}} \right)_{\text{sample}}}{\left(\frac{^{182}\text{W}}{^{184}\text{W}} \right)_{\text{BSE}}} - 1 \right] \times 10^4 \quad (2.14)$$

in which $\left(\frac{^{182}\text{W}}{^{184}\text{W}} \right)_{\text{sample}}$ is the measured $^{182}\text{W}/^{184}\text{W}$ ratio of the sample and $\left(\frac{^{182}\text{W}}{^{184}\text{W}} \right)_{\text{BSE}}$ is the measured $^{182}\text{W}/^{184}\text{W}$ ratio of the bulk silicate Earth.

Consider the following example. A chondritic meteorite has a measured $^{182}\text{W}/^{184}\text{W}$ ratio of 0.864640. In the same experiment, a standard representative of the bulk silicate Earth has a measured $^{182}\text{W}/^{184}\text{W}$ ratio of 0.864810.

What is the $\varepsilon^{182}\text{W}$ value of the meteorite?

$$\begin{aligned} \varepsilon^{182}\text{W} &= \left[\left(\frac{0.864640}{0.864810} \right) - 1 \right] \times 10^4 \\ &= (0.999803 - 1) \times 10^4 \\ &= -0.000196 \times 10^4 \\ &= -1.96 = -2.0 \text{ (to 2 sig. figs)} \end{aligned}$$

This result shows that the chondrite measured has a slightly lower amount of radiogenic ^{182}W than the bulk silicate Earth by 0.02%, or 2 parts in 10 000. Quite clearly, saying it has a $\varepsilon^{182}\text{W}$ value of -2 is much easier than dealing with a six figure decimal, while the sign conveys that it is less radiogenic than BSE. A sample with a positive value has more radiogenic W than the BSE.

The ε notation also has another advantage in that it allows results from different laboratories to be compared more easily.

Question 2.5

Calculate the $\varepsilon^{182}\text{W}$ value for the following data from another laboratory in which the same meteorite has a measured $^{182}\text{W}/^{184}\text{W}$ ratio of 0.864523 and a standard representative of the BSE has a measured $^{182}\text{W}/^{184}\text{W}$ ratio of 0.864696.

The answers to the above calculations are given to 2 significant figures because the precision with which individual measurements are made is of the order of 0.3ε units, so any result is subject to this degree of uncertainty. Both sets of original data give the same result within the limits of experimental uncertainty, despite giving different absolute values for the measured ratios. This illustrates that while different laboratories may get different absolute values of $^{182}\text{W}/^{184}\text{W}$ ratios, even though they are measuring the same samples, the calculation of $\varepsilon^{182}\text{W}$ values overcomes these differences, as long as the samples and standards are measured in the same laboratory.

But what about the Earth's core? Unfortunately we do not have access to the core but we can compare W isotope ratios in the silicate part of the Earth (the mantle and crust) with chondritic meteorites – which are representative of the bulk silicate Earth at least as far as Hf and W are concerned. The results in Figure 2.9 show that the bulk silicate Earth has a different $\varepsilon^{182}\text{W}$ from chondritic meteorites.

- Is the difference in W isotopes between chondrites and the bulk silicate Earth consistent with core formation?
- Yes, the bulk silicate Earth has a higher $\varepsilon^{182}\text{W}$ and hence a higher $^{182}\text{W}/^{184}\text{W}$ ratio than chondrites, signifying that it has a higher Hf/W ratio.

The tungsten isotope difference between early metals, carbonaceous chondrites and the bulk silicate Earth reflects the Hf/W ratio of the material that formed the Earth and its fractionation during the lifetime of ^{182}Hf . As a result of core formation, the bulk silicate Earth has an elevated Hf/W ratio (~ 15) relative to chondrite meteorites (Hf/W ratio ~ 1). Therefore, provided that the Earth's core formed early, i.e. during the lifetime of ^{182}Hf , an excess in the ^{182}W atomic abundance in the bulk silicate Earth relative to chondrites or iron meteorites will be generated. The conclusion from the extinct Hf–W system is that planetary differentiation occurred while ^{182}Hf was still present. As its half-life is 8.9 Ma, this implies that the core and mantle must have separated within the first few tens of millions of years of Earth history.

Needless to say, the detailed interpretation of W isotopes in planetary bodies is more complex than described here, but the principles remain the same. The data for chondrite meteorites, some of which are independently dated, yield an $\varepsilon^{182}\text{W}$ value of around -2 , relative to the bulk silicate Earth (and the Moon), which has a much higher Hf/W ratio. These differences have been modelled with core separation between 30 Ma and 50 Ma after the start of the Solar System.

As a footnote to this section, studies of Pb isotopes in the Earth and meteorites have also been exploited to investigate the timescales of core segregation, again based on the contrasting properties of U and Pb. The U–Pb system has a long half-life and so any measurements of Pb isotopes in modern Earth materials reflect a long history of U–Pb fractionation. However, it can be argued that the major U–Pb fractionation occurred early in Earth history, that it was related to core formation, and that it occurred over a period of 70–150 Ma after the start of the Solar System.

2.3 The Moon

The Moon is a significant part of the terrestrial system and no consideration of the history of the early Earth would be complete without at least a brief consideration of its formation. One of the fundamental discoveries of the Apollo missions was that the Moon consists of a variety of igneous rock types that differ widely in their mineralogy, composition and age. The most visible evidence of these differences is the existence of two distinct terrains on the Moon:

- the light-coloured feldspathic rocks of the highlands
- the dark basalts of the maria (see Figure 1.2b).

Both rock types are igneous and indicate that at some time during its history the Moon was extensively molten. Indeed the anorthosites, which are also very old, are thought to have formed by a process of floatation on a magma ocean, their dominant mineral, plagioclase, being lighter than the iron-rich basalts from which they crystallised.

As discussed previously, the oldest igneous rocks are generally considered to be the iron-rich anorthosites, which may be as old as 4.56 ± 0.07 Ga. Since anorthosites are igneous rocks, they provide a lower limit for the age of the Moon, but again imply that the Moon must have formed within the first 70 Ma of the start of the Solar System.

It is also possible to use Hf–W model ages to date the age of the Moon. One lunar basalt yields a positive $\varepsilon^{182}\text{W}$, i.e. it has more radiogenic W than the bulk silicate Earth (Figure 2.9). However, the Moon is also thought to possess a higher Hf/W ratio than Earth (see below), so the derived age is similar to that of the bulk silicate Earth, and ages for Hf/W fractionation in the Moon range from between 30 Ma and 45 Ma after the start of the Solar System.

Structurally, the Moon contrasts with the Earth in that it has a very small core and is totally solid – there is no convecting outer core and hence no magnetic field. The Moon, being much smaller than the Earth and having a much larger surface to volume ratio, cooled very rapidly after formation. As a result, the lithosphere is very thick, comprising in excess of 75% of the thickness of the Moon's mantle.

The compositions of lunar rocks have been used to develop models of the bulk composition of the Moon, which can be compared with that of the bulk silicate Earth and meteorites. These are summarised for a selection of elements with contrasting properties in Table 2.2. (Element names and symbols are listed in the Appendix.)

Table 2.2 Comparison of elemental abundances in primitive meteorites, Earth and Moon.

	CI chondrite (primitive meteorite)	Earth (crust + mantle)	Moon (crust + mantle)	Ratio of trace element abundance Moon/Earth
<i>Volatile elements</i>				
K/ppm	545	180	83	0.46
Rb/ppm	2.32	0.55	0.28	0.51
Cs/ppb	279	18	12	0.67
<i>Moderately volatile</i>				
Mn/ppm	1500	1000	1200	1.20
<i>Refractory elements</i>				
Cr/ppm	3975	3000	4200	1.40
Th/ppb	30	80	112	1.40
Eu/ppb	87	131	210	1.60
La/ppb	367	551	900	1.63
Sr/ppm	7.26	17.8	30	1.69
U/ppb	12	18	33	1.83
<i>Siderophile elements</i>				
Ni/ppm	16500	2000	400	0.200
Mo/ppb	1380	59	1.4	0.024
Ir/ppb	710	3	0.01	0.003
Ge/ppm	48	1.2	0.0035	0.003

Note: some elements are parts per million (ppm) and some are in parts per billion (ppb).

- Relative to the Earth, what do you notice about the abundances of the volatile elements in the Moon?

- They are all lower than those of the Earth.

One of the primary observations of the Moon is that it is depleted in the most volatile elements and enriched in refractory elements. This has been interpreted as relating to a very high temperature origin for the material that makes up the Moon.

- How do the siderophile element abundances in the Moon compare with those of the bulk silicate Earth?

- They are much lower.

Explaining this difference is more complex, especially as the Moon may not possess a metallic core. The extreme depletion of the siderophile elements in the silicate portion of the Moon strongly suggests that the material of which the Moon is made was already differentiated – it had already lost a metallic fraction and hence its inventory of siderophile elements.

In addition to these characteristics, any model of lunar formation must also take into account the following:

- the **angular momentum** of the Earth–Moon system. Angular momentum is a property of rotating systems that depends upon mass and its distribution, angular velocity, and radius. The angular momentum of the Earth–Moon system is contained in the Earth’s rotation and the Moon’s orbital motion, and is unusually high compared with the other terrestrial planets
- the Earth and Moon have indistinguishable oxygen isotope compositions, whereas most planetary bodies have different and distinct oxygen isotope compositions.

2.3.1 The formation of the Moon

The origin of the Moon has been debated for over a century, but particularly since the Apollo mission provided samples to study. Several theories of formation have been suggested.

Co-accretion

This theory proposes that the Earth and Moon simply accreted side by side. The difficulty with this model is that it does not explain the angular momentum of the Earth–Moon system. This model explains neither the difference in density nor the difference in the depletion of volatile elements.

Capture

This theory proposes that the Moon was originally in a heliocentric orbit and was captured following a close approach to the Earth. However, it is difficult to do this without the Moon spiralling into the Earth and colliding. It is also difficult to explain the indistinguishable oxygen isotope compositions of the Earth and Moon with this model.

Fission

This theory proposes that the Moon split off as a blob during the rapid rotation of a molten Earth. George Howard Darwin (1845–1912), the son of Charles Darwin, originally championed this idea. At one time (before the young age of the ocean floor was known) it was thought by some that the Pacific Ocean might have been the residual space vacated by the loss of material. This model does have some attractive features: it explains why the Earth and Moon have identical oxygen isotope compositions. It also explains, for example, why the Moon has a lower density, because the outer part of the Earth would be deficient in iron due to core formation, and it explains why so much of the angular momentum of the Earth–Moon system is in the Moon’s motion. However, it is not clear why the Moon should spontaneously split away from the Earth without some large input of energy.

Impact models

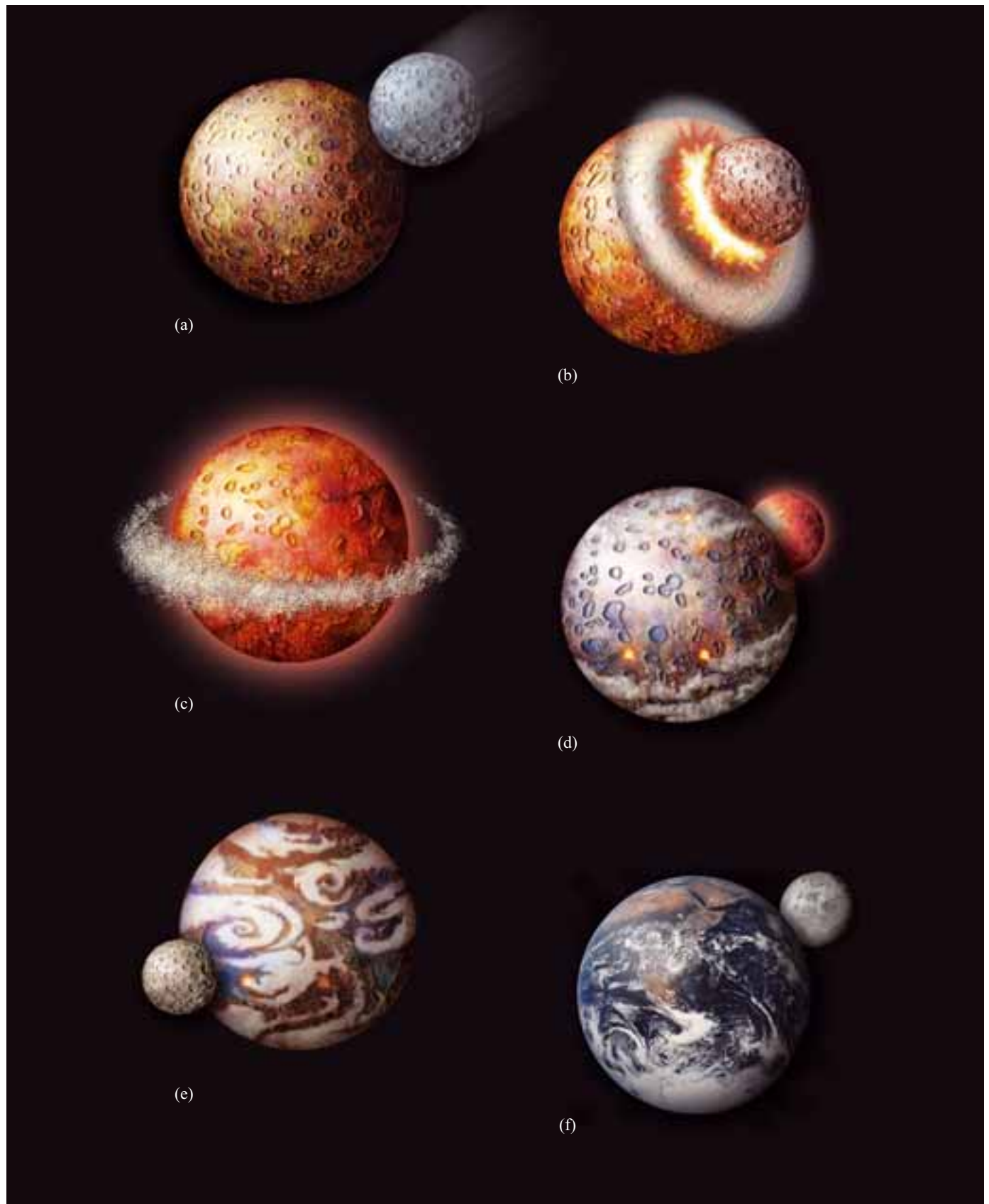
Following the Apollo missions it was proposed that the Moon formed as a result of a major impact on the Earth that propelled sufficient debris into orbit to produce the Moon. Such models are now the most widely accepted (Figure 2.10).

Important information that came from sample-return was that more than 80% of the lunar crust was composed of anorthosite, indicating the presence of a magma ocean early in lunar history. The presence of a magma ocean requires significant degrees of melting, which occurred in response to an impact – provided the accreting body was large and that subsequent lunar accretion was rapid (1–100 Ma). Also, it is necessary to link the dynamics of the Moon with that of the Earth's spin. This led to the proposal of a series of single giant impact models in which the Moon was the product of a glancing blow collision with another differentiated planet (Figure 2.10a–b). A ring of debris would have been produced from the outer silicate portions of the Earth and the impactor (named Theia, the mother of Selene, the goddess of the Moon), which was roughly the size of Mars (Figure 2.10c). This model explains the angular momentum, the extensive early melting, the isotopic similarities and the density difference. The most recent model simulations indicate that the giant impact that formed the Moon probably occurred at the very end of Earth's accretion, when the Earth was 90% of its present size.

Certain features of the Moon may be a consequence of prior differentiation of Earth and Theia, and of the giant impact itself. For example, the depleted volatile elements require that both Earth and Theia had already become at least partially differentiated due to earlier heating. In addition, the low abundances of siderophile elements can be attributed to prior core separation in both the Earth and Theia.

- How is the sequence of events described above consistent with the observation that the Moon's core (if any exists) is very small relative to the size of the whole Moon?
- The Moon formed from mantle material derived from the colliding, partially differentiated planetary embryos. This mantle material was already depleted in Ni, Fe, and S due to the development of cores within the embryos. Therefore, there would have been relatively little Ni, Fe and S left to differentiate inwards once the Moon had formed.

► **Figure 2.10** The formation of the Moon. (a)–(c) illustrate the collision and aftermath between the proto-Earth and Theia. Both bodies were large enough to have differentiated into core, mantle and primary crust as a result of accretionary heating. Following collision (a and b), the cores of the two bodies are thought to have combined and the mantles became mixed while some material was fragmented and vaporised and scattered into orbit around the Earth. (c) Some of the debris fell back to Earth while the remainder accreted under its own gravity to form the Moon. (d) The heat of accretion would again have resulted in wholesale lunar melting. The Moon then cooled and differentiated into mantle, primary crust and possibly a small core, depending on how much of the core of Theia was dispersed around the Earth and how much merged with the Earth's core. (e) The Moon and, presumably, the Earth were subject to further meteorite bombardment and the formation of large craters. Some of these impacts were intense enough to initiate melting of the lunar mantle, flooding the larger impact structures with basalt and forming the lunar maria. (f) Over time the Moon's orbit has slowly decayed, its rate of rotation becoming synchronised with its orbit as the distance between the two bodies progressively increases.



For the Moon's crust and mantle, this depletion was probably further augmented by further differentiation immediately after its formation. Moreover, the Moon's depletion in volatile elements, relative to the Earth, can be explained if the Moon accreted from the partially vaporised debris coalescing after the impact. In these circumstances, the more volatile elements would have had the opportunity to escape into space prior to accretion.

Question 2.6

Given the arguments regarding planetary accretion, volatile elements, and their behaviour following giant impacts, suggest a reason why the concentrations of Rb, K, and Na differ between chondrites and the Earth's mantle.

2.3.2 The late heavy bombardment

The final major event that affected the early Moon is known as the late heavy bombardment. The chronology of impact events on the Moon has been based on telescopic observations, crater counting and disturbed radiogenic isotope systematics. The majority of impact-melt rocks returned from the Moon yield ages around 3.85 Ga (Figure 2.11). These ages have been interpreted as either representing a short and intense heavy bombardment or as the tail end of a prolonged post-accretionary bombardment. In either case the bulk of this bombardment, which produced craters many hundreds of kilometres across, preceded 3.8 Ga. Within this time span, the Earth must have been subjected to a significantly greater bombardment than the Moon, as it has a larger diameter and a much greater gravitational cross-section, making it an easier target to hit. It has been estimated that the impact rate for the Earth would have been more than 20

times greater than the Moon, with both more and larger impact events. The consequences for the hydrosphere, atmosphere and even the lithosphere must have been devastating. It has been suggested that the absence of any rocks on Earth older than ~4.0 Ga is the result of this late heavy bombardment, during which impact-induced mixing recycled early crustal fragments back into the upper mantle.

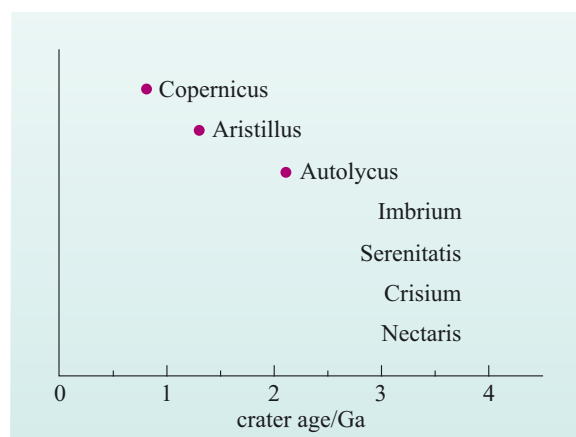


Figure 2.11 Crater ages on the Moon.

2.4 The origins of the atmosphere and hydrosphere

The atmosphere and hydrosphere represent the products of the outgassing of the Earth over geological time, primarily from volcanic activity. Yet the present-day composition of the atmosphere does not reflect that of volcanic volatiles. These are dominated by water vapour, carbon dioxide (CO₂) and sulfur dioxide (SO₂) with smaller amounts of nitrogen, halogens, hydrides and other more exotic volatile compounds. Most of these volatile compounds are soluble in water or, in the case of methane, are easily oxidised to water-soluble compounds. They therefore do not accumulate in the atmosphere but dissolve in the oceans and react with the oceanic crust. The remaining less-reactive and less-soluble elements and compounds, notably nitrogen and the inert gases (He, Ne, Ar, Kr and Xe), accumulate in the gaseous atmosphere.

- Which important atmospheric gas is *not* abundant in volcanic volatiles?
- Oxygen.

For about the first billion years of Earth's history, oxygen was only present in the atmosphere in trace amounts as a result of the breakdown of water vapour by UV radiation. However, this inorganic mechanism of releasing oxygen into the atmosphere produces only tiny amounts of free oxygen. Since the Earth's present atmosphere is oxygen rich and because all higher forms of life require free oxygen, there is an obvious need for some other source of oxygen. The most plausible source is oxygen-producing photosynthesis, a process that first evolved in cyanobacteria during the Archaean era (from 3.8 Ga to 2.5 Ga ago). So, in part, our currently breathable atmosphere is a by-product of life and not a primary feature of the geosphere alone.

The evolution of the hydrosphere is also intimately linked with that of the atmosphere – water is a volatile compound and is only present on the Earth's surface because the surface temperature is below 100 °C. So to understand the origins of the atmosphere and hydrosphere our attention needs to be focused on those components of the present-day atmosphere that have a limited or negligible interaction with the modern biosphere, namely nitrogen and the inert gases.

An important factor dictating whether or not an object in the Solar System can retain an atmosphere is the strength of the gravitational field at its surface – the stronger the field, the stronger the gravitational forces acting on the molecules in the atmosphere. Without the gravitational field, those molecules moving away from the planet would be lost. Even with the gravitational field, those molecules with particularly high velocities can still escape. This leads to the notion of **escape velocity**, which is the minimum velocity needed before a body has enough *kinetic energy* to escape from the surface of a planet (i.e. overcome its gravitational field). It can be shown that the escape velocity (V_{esc}) for a body of mass M and radius r is given by:

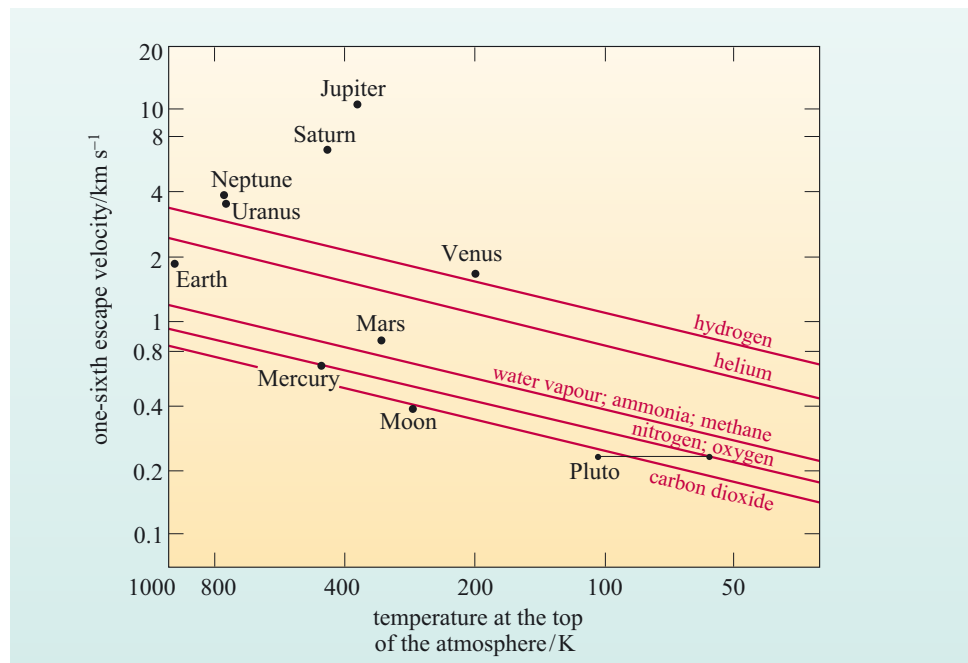
$$V_{\text{esc}} = \sqrt{\frac{2GM}{r}} \quad (2.15)$$

where G is the universal gravitational constant. Whether atmospheric molecules have sufficient velocity depends on the temperature. As the temperature of a gas increases, its molecules move around more quickly, and the average velocity of its molecules increases. Some fraction of the molecules will always be travelling fast enough to overcome gravitational forces, allowing them to escape to space. At low temperatures, this proportion is negligible, but at higher temperatures it becomes progressively more significant, until most molecules exceed the escape velocity for the planetary body. Note that the relevant temperature is that at a level in the upper atmosphere above which the atmosphere is so thin that a molecule moving outwards has little chance of colliding with another, and so *will* escape if it has sufficient velocity.

Different gases have different molecular masses, so their average velocities are different at a given temperature. In order for a planetary body to retain a particular gas in its atmosphere for a period of time of the same order as the age of the Solar System, the average velocity of the molecules in the gas should be less than about one-sixth of the escape velocity. (If the average velocity exceeds one-sixth of the escape velocity, a significant proportion of molecules will be moving faster, and will be lost.) This condition is achieved on only a few planets and satellites.

Figure 2.12 explores these relationships further. For each of the planets (and Pluto and the Moon), the temperature is plotted along the horizontal axis and one-sixth of the corresponding escape velocity on the vertical axis. Note that in order to cover the range of values needed, the scales are not linear, a particular distance along an axis corresponds to a doubling of the quantity. For the named gases, the sloping lines show at each temperature the average molecular velocities of each gas. Figure 2.12 thus defines the conditions under which a planet would lose or retain that gas over geological timescales, i.e. thousands of millions of years. The giant planets plot well above all the lines; they can therefore retain any of the gases. The Moon plots below all the lines – it can retain no gases.

Figure 2.12 Graph summarising conditions of absolute temperature (K) and escape velocity for which planetary bodies can retain common gases in their atmospheres for long periods. For planetary bodies with substantial atmospheres, the temperature is that at the top of the atmosphere. For other bodies it is the mean surface temperature. Pluto's temperature varies greatly according to its position around its orbit. *Note:* both axes are logarithmic scales.



- Which gases should Earth be able to retain?
- Earth plots below the hydrogen and helium lines so it cannot retain these gases. It plots above the lines for water vapour, nitrogen, ammonia, methane, oxygen and carbon dioxide, so it can retain these gases.

2.4.1 Nitrogen and the inert gases

Nitrogen is the most abundant gas (78.1%) in the atmosphere; the third most abundant is argon (0.93%), one of the inert gases. The remaining inert gases, Ne, Kr, and Xe are also found in small traces in the atmosphere. All are relatively unreactive elements in inorganic systems and, although nitrogen can be removed from the atmosphere (fixed) by bacterial activity, its presence in the atmosphere in relatively large quantities is a reflection of its inorganic chemical inertness.

Much attention has been paid to the isotopic composition of the inert gases both in the atmosphere and in the mantle because they can tell us much about the sources of the Earth's volatile elements. In particular, the isotopes of Ar and Xe are significant because they are the daughter products of radioactive decay schemes (Table 2.1), which can be exploited to tell us about the timescales of planetary degassing. The reason why Ar is one of the more abundant atmospheric gases is that it is dominated by one isotope, ^{40}Ar . ^{40}Ar is the daughter of radioactive ^{40}K , which has a half-life of 1.28 Ga (Table 2.1). Xe is much less abundant than Ar, but one of its isotopes, ^{129}Xe , is the daughter of ^{129}I , but in this case the radioactive parent has a half-life of only 15.7 Ma.

- What information can these two radioactive systems provide?
- By analogy with short-lived isotopes in the solid Earth (e.g. Hf–W) and in meteorites (Mg–Al), the I–Xe couple provides information on the timescales of outgassing in the early phases of Earth evolution. By contrast, the long half-life of ^{40}K means that the isotope ratio of Ar reflects the history of planetary outgassing over the whole age of the Earth.

How does this work? The principles behind these isotopic systems are described in Box 2.4. This explains how the radiogenic isotope ratios of the atmosphere and the mantle evolve with time and why measurements of the isotope ratios of Xe and Ar in both are necessary to understand global outgassing.

Box 2.4 Ar and Xe isotopes and the evolution of the atmosphere

The effects of planetary outgassing on the K–Ar and I–Xe radioactive systems are shown graphically in Figure 2.13, in which global outgassing is regarded to be a single event that occurred early in Earth history. Prior to this event, the mantle had low I/Xe and K/Ar ratios and so the ratio of the daughter isotope to a non-radiogenic isotope increased steadily. After degassing, the two parent–daughter ratios in the mantle increased markedly, because Xe and Ar have been lost to the atmosphere and the isotope ratios of Xe and Ar in the mantle both increased at a greater rate than before.

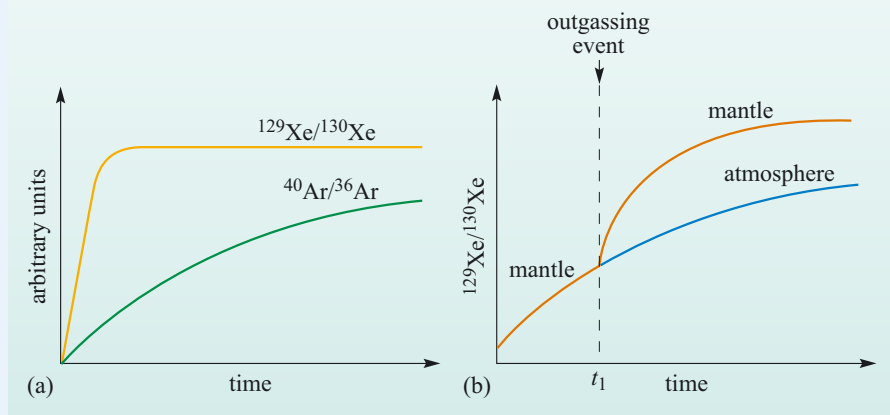


Figure 2.13 Schematic diagrams showing the effect of planetary degassing on the isotopic evolution of Xe and Ar in the Earth's atmosphere and mantle. (a) The simple situation in which the atmosphere gradually forms from the mantle by degassing. Early in Earth history there is a period when the $^{129}\text{Xe}/^{130}\text{Xe}$ ratio evolves rapidly, because of the decay of ^{129}I , until it reaches a maximum value when all the ^{129}I has decayed. Because ^{129}I has a half life of 15.7 Ma this occurs between 100 and 150 Ma after the formation of the Solar System. (b) If the atmosphere formed as the result of a major outgassing event, t_1 , early in Earth history while ^{129}I was active, then the fractionation of Xe into the atmosphere and the preferential retention of I in the mantle would result in greater ^{129}Xe production in the mantle. See text for further discussion.

The critical issue concerns the timing of the degassing event relative to the half-life of ^{129}I . The example illustrated in Figure 2.13b shows how the isotope ratio of $^{129}\text{Xe}/^{130}\text{Xe}$ varies if degassing occurs before all the ^{129}I has decayed. Subsequent to degassing, the mantle and the atmosphere evolve along different paths.

- What will be the consequences for the $^{129}\text{Xe}/^{130}\text{Xe}$ ratios of the mantle and atmosphere if degassing occurs after all the ^{129}I has decayed?
- They will be identical.

Note that the important measurement to make is not the Ar and Xe isotope ratios of the atmosphere, but those of the mantle. The reason for this is that any outgassing event or process extracts virtually all of the Xe or Ar from the mantle and so the radioactive clock is reset. Because all the volatile elements eventually end up in the atmosphere, its isotope composition reflects the sum total of all outgassing events and processes throughout Earth history. Hence the measurement of Xe isotopes in the Earth's mantle provides a means of distinguishing between early, catastrophic planetary outgassing as opposed to gradual outgassing over the whole of Earth history.

As you may imagine, making isotope measurements of Ar and Xe in the mantle is extremely difficult and enormous care needs to be taken to avoid atmospheric contamination, both of the samples and in the measuring instruments. However, the most successful measurements show that both the $^{129}\text{Xe}/^{130}\text{Xe}$ and the $^{40}\text{Ar}/^{36}\text{Ar}$ ratios of the mantle are high relative to the atmosphere.

This observation requires early and extensive degassing of the mantle to generate high parent/daughter ratios in the mantle early enough to have an effect on the I–Xe system. Taken together, the Xe and Ar isotope ratios of the present-day mantle show that between 80% and 85% of the atmosphere was outgassed extremely early in Earth history, and given that the half-life of ^{129}I is only 15.7 Ma this suggests that outgassing occurred within the first few tens of millions of years after accretion.

- Which major events occurred within the first few tens of millions of years of Earth evolution?
- Core formation and the giant impact that formed the Moon both occurred within 30–50 Ma of Earth accretion.

The short timescale of these two major events and the development of the earliest atmosphere were all part of the primary differentiation of the Earth. However, Earth history is never that simple – some outgassing continues to the present day, largely via mid-ocean ridge and within-plate volcanic activity, but the Xe and Ar isotope record of the atmosphere and the mantle show that little if any of the noble gases and, by analogy, nitrogen, are recycled via plate tectonics or any other process into the mantle.

2.4.2 Loss of the earliest atmosphere

Xenon isotope data also provide evidence that much (>99%) of the Earth's early atmosphere was lost within the first 100 Ma. The present I/Xe ratio of the Earth is an order of magnitude higher than chondritic values at the start of the Solar System. We know the abundance of ^{129}I at that time, and as all of this ^{129}I formed ^{129}Xe , it should have produced xenon that was highly enriched in ^{129}Xe , given the Earth's high I/Xe ratio. However, instead, the Earth has xenon that is only slightly more radiogenic than is found in meteorites that are rich in primordial gases. This provides evidence that the Earth had a low I/Xe ratio that kept its xenon isotope compositions close to the chondritic value. At some point xenon was lost, but by this time ^{129}I was close to being extinct so that the xenon did not become radiogenic despite a very high I/Xe ratio.

We have already seen that gravitational forces are likely to prevent many gases from escaping Earth (Figure 2.12), so it is difficult to see how a large early atmosphere would have been lost. One possibility is that the atmosphere was blown off by a major impact like the Moon-forming giant impact. Alternatively, hydrogen and helium may have been present in substantial amounts in the early atmosphere but, as Figure 2.12 shows, were easily lost through thermal escape. If large amounts of these gases escaped rapidly, then other heavier gases could be lost at the same time by being carried along with them.

Thus, the noble gas data indicate that >80% of the atmosphere was formed extremely early in Earth history, and that much of this early atmosphere was lost in the first 100 million years of Earth history. This suggests that initially the atmosphere was far denser and more massive than today's atmosphere. The presence of such an atmosphere would have an insulating effect, leading to the build up of temperatures in the outer parts of the Earth, facilitating the formation of magma oceans and core formation.

2.4.3 Water

The evolution of the hydrosphere is intimately linked with that of the atmosphere because water is also a volatile compound. Until recently, there were competing views as to the origin of the Earth's water. One was that the Earth accreted as a dry body and its water was subsequently added through cometary impact. The alternative view was that the Earth inherited its water from water-bearing minerals in the undegassed interiors of planetary embryos, and that this was outgassed, along with Xe and Ar, early in Earth history. As noted earlier, the evidence suggests that the early Earth experienced intense meteoritic bombardment and must have been hot. It might be assumed, therefore, that surface conditions were too extreme for the young Earth to have a liquid hydrosphere and that much of the initial water was lost to space. However, current models suggest that with the presence of a dense early atmosphere the pressures at the surface of the Earth beneath this atmosphere would have been high enough to ensure that a significant proportion of water and other volatiles were retained in solution. This debate was partially resolved with the measurement of the deuterium/hydrogen (D/H) ratio in three comets, using both space probe measurements (the Giotto probe to comet Halley) and two ground-based measurements of radio and infra-red emissions. All three measurements agree within experimental uncertainty and show that deuterium (heavy hydrogen) is twice as abundant relative to hydrogen in comets as it is in the terrestrial hydrosphere. Such a major distinction effectively rules out comets as a major source of the Earth's water.

Thus the preferred model for the evolution of the hydrosphere is that it degassed from the mantle, and that this material was ultimately derived from water-bearing grains that became incorporated into planetesimals and eventually into planetary embryos.

2.5 The earliest continents

2.5.1 Isotopic evidence

As stated earlier, evidence suggests that, in its early stages, Earth may have had a magma ocean, sustained by heat from accretion and the blanketing effects of a dense early atmosphere. With the loss of the early atmosphere during planetary collisions, the Earth would have cooled quickly, the outer portions would have solidified, and it would have developed its first crust.

We have little idea of what such a crust might have looked like. This is because, unlike the Moon, the Earth appears to have no rock preserved that is more than 4.0 Ga old. There was intense bombardment of the Moon until about 3.85 Ga

(Section 2.3) and Earth's earliest crust may therefore have been destroyed by such impacts. It may also be that because Earth was hotter this crust was highly unstable. Some have argued that the earliest crust was like the lunar highlands – made from a welded mush of crystals that had previously floated on a magma ocean. Others have suggested that it was made of denser rocks more like the basalts that presently form the Earth's ocean floor.

Some of the oldest rocks on Earth come from exposures in western Greenland near Isua. These include ancient sediments and volcanic lavas that have since been subjected to folding and intrusion by younger igneous rocks.

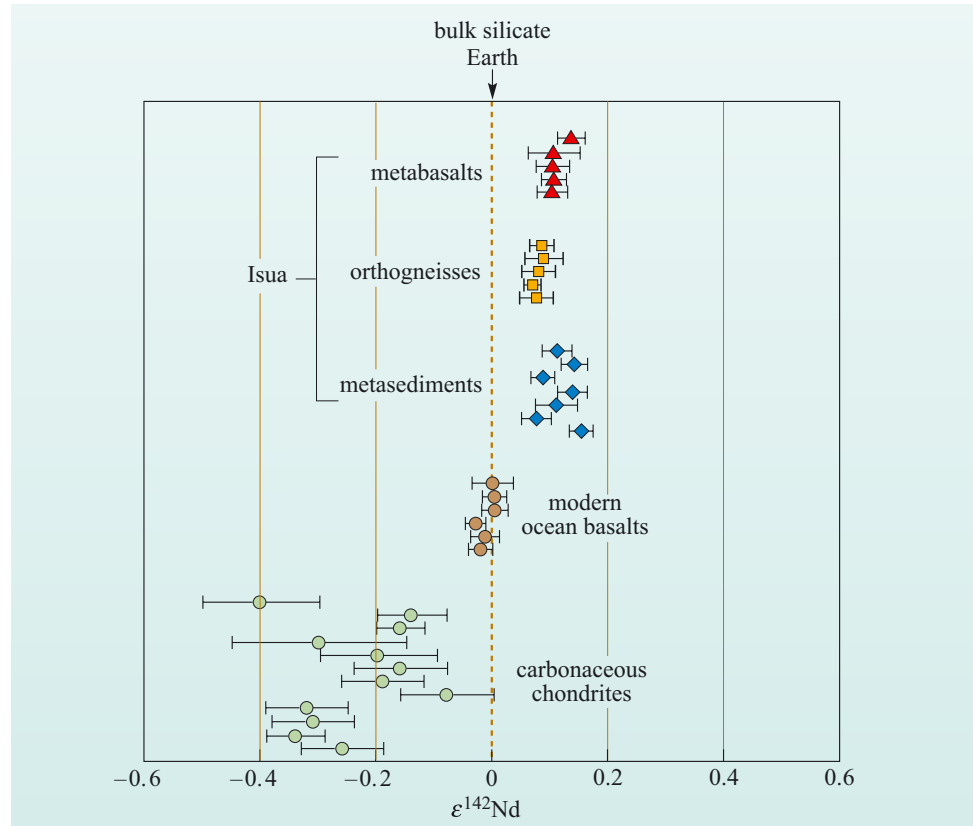
Metamorphosed sediments from Isua give an age of about 3.9 Ga, which is regarded as a minimum age of sediment deposition. However, these rocks were formed by weathering and erosion of pre-existing crustal material, and may preserve clues as to the age and origin of this older crust.

Just as U–Pb and Hf–W are ideal for studying the rates of accretion and core formation, and I–Xe is useful for studying the rate of formation of the atmosphere, the ^{146}Sm – ^{142}Nd (samarium–neodymium) system is useful for studying the early history of melting in the silicate Earth. ^{146}Sm decays to ^{142}Nd with a half-life of 103 million years (Table 2.1), but unlike Hf–W or U–Pb systems, both Sm and Nd are lithophile elements and remain in the mantle during core formation. However, during melting of the silicate mantle to produce the crust, Nd is more incompatible than Sm (that is, Nd is preferentially partitioned into the silicate melt) producing a crust with a low Sm/Nd ratio and leaving the mantle with a complementary high Sm/Nd ratio. If melting of the silicate Earth occurred within the lifetime of the now extinct ^{146}Sm (about 500 Ma), then this would fractionate Sm/Nd and could potentially produce differences in the abundance of ^{142}Nd in early crust. If subsequent recycling of such crust by weathering and erosion or meteorite impact failed to eradicate the difference in ^{142}Nd , then this might be detectable in early Archaean rocks.

However, such ^{142}Nd differences (sometimes termed anomalies), if they ever existed, were always likely to be small because of the low abundance of ^{146}Sm in the early Solar System. As a consequence their measurement is extremely difficult and for many years only one sample showed the hint of such an effect, and these data were questioned by many. Study of ^{142}Nd anomalies has, needless to say, focused on the Isua rocks and improved measurement techniques have recently confirmed the existence of differences in ^{142}Nd (from the present-day mantle) in many rock types seen at Isua, including metamorphosed sediments and volcanic rocks (Figure 2.14). These results indicate that the precursor materials of the Isua rocks were derived from material that differentiated from the silicate mantle within 50 Ma of the start of the Solar System.

Surprisingly, it was also discovered that an excess of ^{142}Nd is ubiquitous in all modern terrestrial rocks – at least all those measured so far – relative to chondritic meteorites (Figure 2.14). If it is assumed that chondritic meteorites are representative of the material from which the Earth was made, then this excess ^{142}Nd indicates that the silicate Earth must have experienced a global chemical differentiation within 50 Ma of the start of the Solar System.

Figure 2.14 $^{142}\text{Nd}/^{144}\text{Nd}$ ratios measured for chondrites, present-day terrestrial rocks and Archaean rocks at Isua.



2.5.2 Geological evidence

The rocks themselves at Isua allow us to make some inferences about conditions on and in Earth 3.8 Ga ago. They include lavas known as komatiites (Box 2.5) that have what is called a *pillow structure*, which indicates cooling under water, and sediments rich in quartz pebbles and volcanic debris (Figure 2.15). The observation that sediments and lavas were deposited and erupted under water indicates that there must have been bodies of liquid water at the Earth's surface, and perhaps even ocean basins by 3.9 Ga. Land areas would have to have been exposed to weathering and erosion to form such sediments. The presence of quartz pebbles implies that at least some of the land being weathered was similar to the upper parts of present-day continental areas. For example, granite is a common rock that occurs in continental areas and is rich in quartz. It seems, therefore, that the geological processes and cycles that are recognised today were operating on the Earth at least 3.9 Ga ago, albeit with some differences in detail.

- What does the occurrence of sedimentary rocks at Isua tell us about the surface temperature of the Earth 3.9 Ga ago?
- For weathering, erosion and deposition of sediments to occur there must have been both rain and liquid water present, implying surface temperatures of between 273 K and 373 K (i.e. 0–100 °C).



(a)



(b)

Figure 2.15 (a) Metamorphosed pillow lava at Isua showing pillows with dark rims in a matrix of chert (fine-grained silica). (b) Metamorphosed conglomerate at Isua, showing pebbles of quartz.

Box 2.5 Komatiites

Komatiites are rare ultramafic lavas that were produced most commonly during the Archean and are very rare in the Phanerozoic. These magmas are thought to provide a record of the thermal and chemical characteristics of the mantle through time. Komatiites are distinguished by their high MgO content (>18 wt%) and olivines showing a *spinifex* texture (elongated, skeletal, branching crystals) (Figure 2.16), which were initially thought to be indicative of the rapid quenching of magma. When komatiites were subjected to melting experiments in the laboratory they were found to possess very high liquidus temperatures – a result that was initially interpreted as indicating deep melting at high temperatures. The progressive decline of komatiites was then used as evidence for progressive cooling of the Earth's mantle.



Figure 2.16 Olivine needles in a spinifex-textured komatiite. Each crystal is 2–3 cm long.

However, over recent years an alternative interpretation of komatiite formation has gained favour. This is because komatiites are rich in water, and experiments indicate that the liquidus temperature is dramatically reduced under hydrous conditions. Similarly, in the presence of water spinifex textures can develop even at a low cooling rate. Therefore it has been proposed that the high water contents in komatiite lavas are due either to the Archaean mantle containing a higher volatile content, left over from accretion, or that water was introduced into the mantle as hydrated crust in a subduction zone environment. This alternative interpretation of komatiites indicates that water played a key role in their formation, and as a consequence the Archaean mantle was only slightly hotter than the present day. It also indicates that subduction may have operated in the early Archaean.

2.5.3 Continents and the hydrosphere during the Hadean

The Archaean sediments and volcanic rocks at Isua and at other locations across the world show that by about 3.9 Ga water was present on the surface of the Earth, and that a cycle of erosion and deposition of continental rocks had been established. Is there any evidence for water further back in time?

The possibly surprising answer to this question is ‘yes’, and the source of this evidence is from a lonely outcrop of Archaean sandstones and conglomerates in the Murchison district of Western Australia. These ancient rocks, which are about 3 Ga old, are known as the Mount Narryer and Jack Hills quartzite units. They contain detrital grains of the mineral zircon. Zircons are of enormous value to Earth scientists because, once formed, they are almost indestructible and contain high concentrations of U and Th and other important trace elements. Hence they are one of the primary sources of high-precision ages from Pb isotopes of crustal rocks and their trace element contents provide further information on their formation conditions and provenance. Zircons form as accessory minerals in igneous rocks, usually granites, but, because of their indestructibility, they can survive the rock cycle of erosion, transport and deposition and are frequently preserved in coarse-grained sedimentary rocks such as sandstones, grits and conglomerates, and their metamorphic equivalents, e.g. quartzite. Thus, the inescapable conclusion is that the Jack Hills zircons must have been inherited from an older granite and so from older continental crust. So far this is not very exciting – rocks are known that are almost 1 Ga older than the Jack Hills quartzites, so why the excitement?

Detailed isotopic analysis of the Pb isotope composition of these zircons (Figure 2.17) has revealed that they range in ages from 3.05 Ga, which is close to the age of deposition, up to much older ages – a significant number of crystals are in excess of 4 Ga. Indeed, the oldest has an age of 4.4 Ga and formed only 170 Ma after the start of the Solar System. These zircons are the oldest solid material yet found on Earth and, because zircon can only form in high temperature igneous processes in a high-silica magma, they are thought to represent a tiny sample inherited from some of the Earth’s original continental crust.

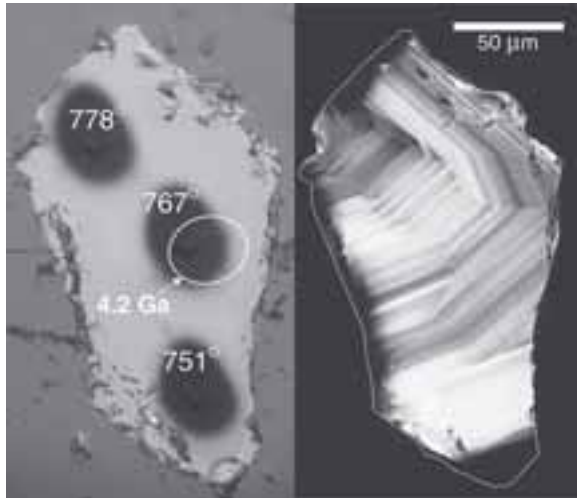


Figure 2.17 Images of a zircon from the Jack Hills quartzites in Western Australia: (a) shows the damage produced during ion-microprobe analyses. Temperatures are in degrees Celsius and refer to results from an assessment of their Ti contents. The white ellipse shows the region of the crystal dated at 4.2 Ga; (b) shows the same crystal from cathode-luminescence, revealing compositional banding. (Watson and Harrison, 2005)

That conclusion is in itself exciting – solid material from the Hadean and evidence for continental crust – but apart from their age, the Jack Hills zircons show compositional variations that allow conditions of crystallisation to be inferred. It turns out that the zircons crystallised at a temperature of around 700 °C, a temperature that is no greater than the temperature that granites crystallise today. As you will see in Chapter 6, granites can only form at this temperature *if water is present*. They form by *dehydration* reactions in which mica breaks down, releasing substantial amounts of water that dissolve in the silicate melt so generated. The presence of granite and its crystallisation at low temperatures is powerful evidence that water was present in the earliest crust of the Earth.

This argument may appear a little convoluted, and the leap from the tiny Jack Hills zircons to a planet with continents and an active low-temperature hydrosphere is large, but the logic is sound. The information gleaned from these zircons is that the surface environment during the Hadean may have been remarkably similar to the Earth today!

2.5.4 The Archaean atmosphere, hydrosphere and biosphere

The nature of the Earth's atmosphere and hydrosphere after the Earth had cooled and the first rocks were preserved at the surface is the subject of considerable debate. It has been suggested that the warm conditions and presence of water may have provided conditions suitable for life to develop. Several geological settings might have hosted such early life. It may have been supported in ocean basins during transient heating after a major meteorite impact,

or possibly life could have existed in a hydrothermal system. Submarine systems would have also offered a protective setting against UV radiation.

Although the rocks at Isua show clear evidence that water was present at the Earth's surface, there are no fossils. Again, evidence for life comes from geochemical clues. Carbonate in some of the volcanic rocks contains carbon with a heavy isotope signature, i.e. it shows an enrichment in ^{13}C over ^{12}C . This is complementary to the ubiquitous depletion in ^{13}C relative to ^{12}C , which is characteristic of 'light' carbon that has been involved in biological processes.

However, the carbonate is all non-biological in origin and secondary (i.e. precipitated after the crystallisation of the volcanic rock) and therefore may be much younger than the host volcanic rock. Isotopically light carbon also occurs in graphite flakes in some rocks and this chemical signature is thought by some to be of biological origin. However, this interpretation remains controversial, as the graphite is found in highly metamorphosed and deformed rocks, and is usually associated with secondary carbonate deposition.

The sequence at Isua also contains rocks that are thought by some to provide possible evidence for the changing nature of the Earth's atmosphere as a result of the appearance of life. **Banded iron formations (BIFs)** are characterised by finely banded dark-brown, iron-rich layers alternating with lighter-coloured iron-poor layers (Figure 2.18). The layers range in thickness from less than a millimetre to about a centimetre. The iron-rich bands contain the highly insoluble iron oxides: hematite (Fe_2O_3), limonite ($\text{Fe}_2\text{O}_3 \cdot 3\text{H}_2\text{O}$) and magnetite (Fe_3O_4). Chert, a rock composed of precipitated silica, occupies the iron-poor bands. Individual bands, often only a few millimetres thick, can extend for several kilometres. How BIFs formed is not entirely clear, as we have no modern analogues to guide us. However, the involvement of iron oxides in their formation suggests that the process that led to the formation of BIFs must have affected the oxidation state of iron and hence such rocks may contain information about the oxidation state of the Earth's ocean and atmosphere at that time.



Figure 2.18 Close-up view of a banded iron formation. The light bands are dominated by silica-rich chert and the darkest bands by red-brown oxides of iron. Each band is ~1 cm thick.

Whatever the chemistry occurring in the formation of BIFs, large amounts of oxygen were incorporated into BIFs very early in the Earth's history. One interpretation is that this suggests oxygen was available in the shallow seas where most BIFs were formed. Most theories for the origin of BIFs involve a significant role for hydrothermal activity on the early Earth. The seawater

flowing through these hydrothermal systems would have dissolved iron-containing minerals, so that iron in a reduced form was subsequently injected into the deep ocean through hydrothermal vents. It is generally accepted that the deep ocean on the early Earth was extremely oxygen deficient or anoxic, so that iron escaped oxidation and precipitation at the vents themselves but was deposited in much shallower, more oxygenated water.

How did the iron get from the deep ocean to shallow water, crossing large expanses of oceans in so doing? One idea is that the iron was actually consumed by bacteria that flourished near the vents and that these bacteria then drifted away in vast colonies into shallow water where they died, depositing a thin film of organic-rich material. After a while the organic material would have been recycled, leaving the iron behind in its highly insoluble oxide form.

- Hydrothermal vents provide a potential source for the large amounts of iron involved in the formation of BIFs, but where might the large amounts of oxygen come from?
- Various mechanisms have been proposed to account for the oxidation of BIFs and the precipitation of Fe(III):
 - abiotic breakdown of water by UV radiation
 - direct oxidation by anaerobic bacteria
 - photosynthesis, which generates free oxygen.

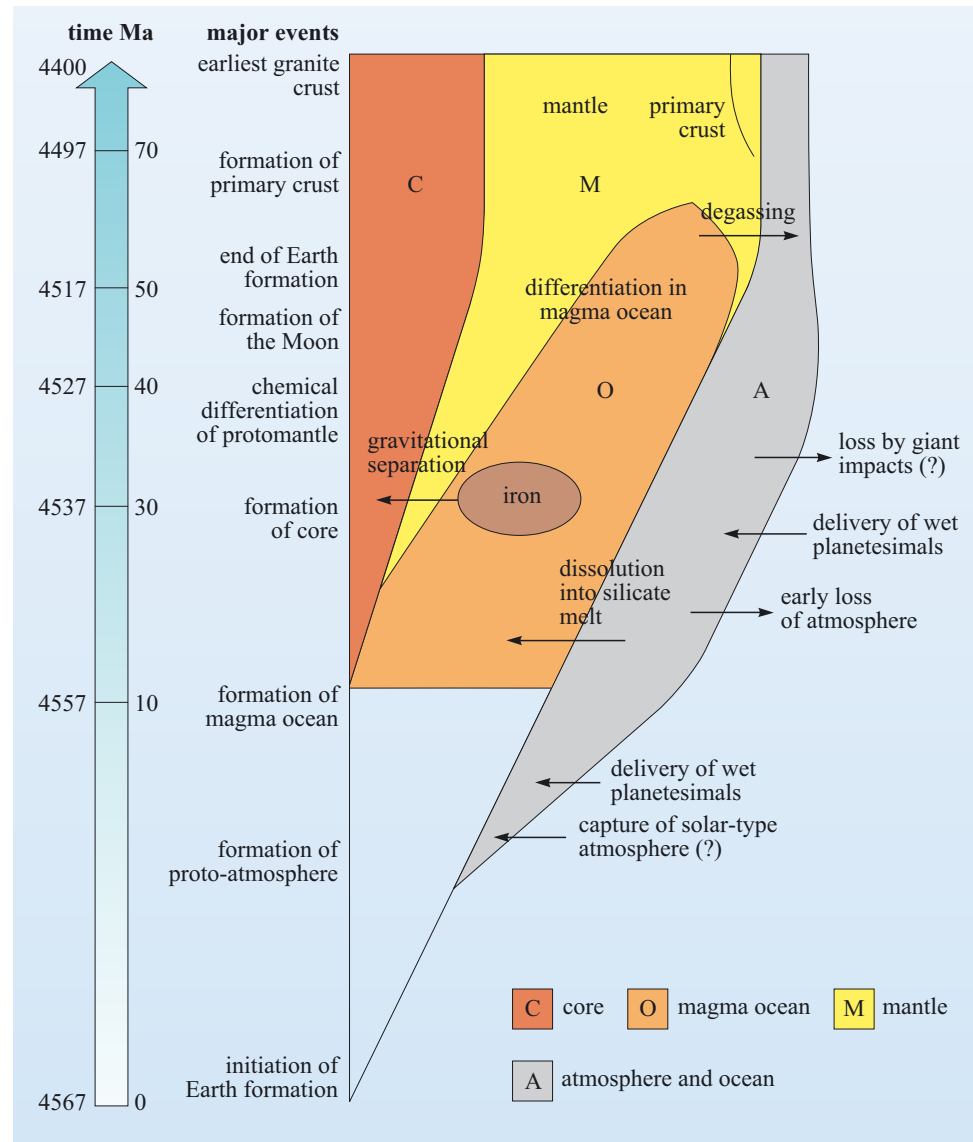
Summary of Chapter 2

The sequence of events accompanying the evolution of the early Earth is summarised in Figure 2.19. Taken together, the evidence suggests that the development of the core and magma ocean, and outgassing to form the atmosphere, were contemporaneous and interdependent processes.

The issue of whether accretion of the Earth was homogeneous or heterogeneous has long been debated. To some extent the Earth must have accreted heterogeneously because early planetesimals are likely to have experienced different degrees of differentiation. Furthermore, meteorites with different chemical compositions from Earth continue to accrete right up to the present day. However, if the early Earth were covered by a deep magma ocean, any effect of accreting material with diverse compositions would be erased in the magma ocean. A possible exception to this might be the highly siderophile elements. The abundances of these elements in the silicate mantle might be explained by metal–silicate equilibration in a magma ocean or may have been added as a ‘late veneer’ after core formation.

Finally, it would appear inescapable that most of the water on Earth arrived as part of the accretion process and the presence of that water influenced primary differentiation. Accretion of hydrous materials and subsequent outgassing led to the formation of a dense ‘steam’ atmosphere, which in turn served as a thermal blanket allowing the long-term persistence of a terrestrial magma ocean. The higher pressure of water in the atmosphere would have allowed the presence of

Figure 2.19 Schematic diagram showing the timescale of major events in the evolution of the early Earth. The timescale is indicative rather than accurate. (Adapted from Abe et al., 2000)



liquid water at the Earth's surface, and a significant amount of water would have been able to dissolve in the silicate liquid of the magma ocean. Cooling and crystallisation of the magma ocean would have provided a further source of water for the atmosphere and hydrosphere, and a mantle reservoir of stable hydrous phases.

The formation of the hydrosphere probably occurred soon after the atmosphere because water is also a volatile molecule. The oldest zircons indicate that water must have been present at the Earth's surface 170 Ma after accretion. This is because these ancient zircons crystallised from granite that formed at a temperature only possible if water was present. It is also possible that hydrous komatiites may have been produced either by an Archaean mantle with a high volatile content, or by melting of hydrous crust in an early subduction zone environment. Following the end of heavy bombardment of the Earth around 3.85 Ga, it is possible that early life may have been supported in ocean basins heated by meteorite impact or hydrothermal systems.

Learning outcomes for Chapter 2

You should now be able to demonstrate a knowledge and understanding of:

- 2.1 The different heat sources and transfer mechanisms involved in driving planetary differentiation and the early chemical and physical evolution of the Earth.
- 2.2 The current theories associated with the mechanisms and timing of accretion, core formation, and the formation and existence of early magma ocean.
- 2.3 The evidence for the origin and age of the Moon.
- 2.4 The evidence for the formation and chemical evolution of the Earth's atmosphere and hydrosphere.
- 2.5 The nature, composition and age of the earliest continental crust.
- 2.6 The role of partial melting and the presence of water in generating silicic crustal material from the mantle.

Plate tectonics

The Earth's face is changing all the time, but at barely perceivable rates. It is now known that the Earth is a highly dynamic planet – far more so than the other terrestrial planets (Mercury, Venus and Mars) and the Moon – and one that has altered its outward surface many times over geological time. On Earth, this dynamism is manifest in the opening and closure of ocean basins and the associated movements of continents called **continental drift**. At times, continental drift has resulted in the continents fragmenting into many smaller land masses, whilst at other times collisions have assembled vast supercontinents with immense mountain chains along their joins (Figure 3.1). Such constant rearrangement has had a profound effect upon the surface geology of our planet. It has also affected the hydrosphere through the changes in the shape and size of the oceans, the ocean–atmosphere circulation, the configuration and extremities of the Earth's climate zones and, perhaps, even the nature of the biosphere and the course of the evolution of life itself.

In Chapters 1 and 2 you discovered how the Earth formed from a primordial nebula and then developed its layered structure (i.e. core, mantle and crust). From your studies of those chapters you will have begun to discover the physical and chemical differences between those three layers, the differences that distinguish the mantle and the crust, the heat that is generated within them, and how this heat then makes its way to the surface to be lost into space. It is the movement of this internal heat that drives the forces that result in the formation and destruction of ocean basins and, ultimately, the movement of the continents.

In this chapter, you will examine how the evidence for the movement of continents was gathered and how this movement relates to, and generates, geological features and phenomena such as ocean basins, mountain ranges, volcanoes and earthquakes. You will learn how and why the continents have moved, and continue to move, and the forces that drive them around our globe.



(a)



(b)

Figures 3.1 An example of (a) a continental reconstruction for the late Carboniferous showing the supercontinent of Pangaea compared with (b) today's more fragmented arrangement.

3.1 From continental drift to plate tectonics

The remarkable notion that the continents have been constantly broken apart and reassembled throughout Earth's history is now widely accepted. The greatest revolution in 20th century understanding of how our planet works, known as **plate tectonics**, happened in the 1960s, and has been so profound that it can be likened to the huge advances in physics that followed Einstein's theory of relativity. According to the theory of plate tectonics, the Earth's surface is divided into rigid plates of continental and oceanic lithosphere that, through time, move relative to each other, and which increase or decrease in area. The growth, destruction and movement of these lithospheric plates are the major topics of this chapter, but it is first worth considering how the theory actually developed from its beginnings as an earlier idea of 'continental drift'.

3.1.1 Continental drift

The German meteorologist Alfred Wegener (1880–1930) is largely credited with establishing the fundamentals of the theory that we now call plate tectonics. The idea that continents may have originally occupied different positions was not a new one (Box 3.1), but Wegener was the first to present the evidence in a diligent and scientific manner.

Box 3.1 Continental drift to plate tectonics: the evolution of a theory

1620 Francis Bacon commented upon the 'conformable instances' along the mapped Atlantic coastlines.

1858 Antonio Snider-Pellegrini suggested that continents were linked during the Carboniferous Period, because plant fossils in coal-bearing strata of that age were so similar in both Europe and North America.

1885 Austrian geologist Edward Seuss identified similarities between plant fossils from South America, India, Australia, Africa and Antarctica. He suggested the name 'Gondwana' (after the indigenous homeland of the Gond people of north-central India), for the ancient **supercontinent** that comprised these land masses.

1910 American physicist and glaciologist Frank Bursley Taylor proposed the concept of 'continental drift' to explain the apparent geological continuity of the American Appalachian mountain belt (extending from Alabama to Newfoundland) with the Caledonian Mountains of NW Europe (Scotland and Scandinavia), which now occur on opposite sides of the Atlantic Ocean.

1912 Alfred Wegener repropoed the theory of continental drift. He had initially become fascinated by the near-perfect fit between the coastlines of Africa and South America, and by the commonality among their

geological features, fossils, and evidence of a glaciation having affected these two separate continents. He compiled a considerable amount of data in a concerted exposition of his theory, and suggested that during the late Permian all the continents were once assembled into a supercontinent that he named **Pangaea**, meaning 'all Earth'. He drew maps showing how the continents have since moved to today's positions. He proposed that Pangaea began to break apart just after the beginning of the Mesozoic Era, about 200 Ma ago, and that the continents then slowly drifted into their current positions.

1920–1960 A range of geophysical arguments was used to contest Wegener's theory. Most importantly, the lack of a mechanism strong enough to 'drive continents across the ocean basins' seriously undermined the credibility of his ideas. The theory of continental drift remained a highly controversial idea.

1937 South African geologist Alexander du Toit provided support through the years of controversy by drawing maps illustrating a northern supercontinent called Laurasia (i.e. the assembled land mass of what was to become North America, Greenland, Europe and Asia). The idea of the Laurasian continent provided an explanation for the distribution of the

remains of equatorial, coal-forming plants, and thus the widely scattered coal deposits in the Northern Hemisphere.

1944 Wegener's theory was consistently championed throughout the 1930s and 1940s by Arthur Holmes, an eminent British geologist and geomorphologist. Holmes had performed the first uranium–lead radiometric dating to measure the age of a rock during his graduate studies, and furthered the newly created discipline of geochronology through his renowned book *The Age of the Earth*. Importantly, his second famous book *Principles of Physical Geology* did not follow the traditional viewpoints and concluded with a chapter describing continental drift.

1940–1960 The complexity of ocean floor topography was realised through improvements to sonar equipment during World War II. Accordingly, there was a resurgence of interest in Wegener's theory by a new generation of geophysicists, such as Harry Hess (captain in the US Navy, later professor at Princeton), through their investigations of the magnetic properties of the sea floor. In addition, an increasing body of data concerning the magnetism recorded in ancient continental rocks indicated that the magnetic poles appeared to have moved or 'wandered' over geological time. This **apparent polar wander** was explained by the movement of the continents, and not the magnetic poles.

1961 The American geologists Robert Dietz, Bruce Heezen and Harry Hess proposed that linear volcanic chains (mid-ocean ridges) identified in the ocean basins are sites where new sea floor is erupted. Once formed,

this new sea floor moves toward the sides of the ridges and is replaced at the ridge axis by the eruption of even younger material.

1963 Two British geoscientists, Fred Vine and Drummond Matthews, propose a hypothesis that elegantly explained magnetic reversal stripes on the ocean floor (Section 3.1.2). They suggested that the new oceanic crust, formed by the solidification of basalt magma extruded at mid-ocean ridges, acquired its magnetisation in the same orientation as the prevailing global magnetic field. These palaeomagnetic stripes provide a chronological record of the opening of ocean basins. By linking these observations to Hess's sea-floor spreading model, they lay the foundation for modern plate tectonics.

1965 The Canadian J. Tuzo Wilson offered a fundamental reinterpretation of Wegener's continental drift theory and became the first person to use the term 'plates' to describe the division and pattern of relative movement between different regions of the Earth's surface (i.e. plate tectonics). He also proposed a tectonic cycle (the Wilson cycle) to describe the lifespan of an ocean basin: from its initial opening, through its widening, shrinking and final closure through a continent–continent collision (Section 3.5.1).

1960s–present day. There was an increasingly wide acceptance of the theory of plate tectonics. A concerted research effort was made into gaining a better understanding of the boundaries and structure of Earth's major lithospheric plates, and the identification of numerous minor plates.

Evidence for continental drift

Wegener's ideas on continental drift based on the following evidence remain the root of modern continental reconstructions.

1 Geometric continental reconstructions

Ever since the first global maps were drawn following the great voyages of discovery of the 15th and 16th centuries, it has been realised that the coastline geography of the continents on either side of the Atlantic Ocean form a pattern that can be fitted back together; in particular, the coastlines of western Africa and eastern South America have a jigsaw-like fit (Box 3.1).

Although some coastline fits are striking, it is important to note that the current coastlines are a result of relative sea level rather than the actual line along which land masses have broken apart. Indeed, coastline-fit is a common misconception – Wegner himself pointed out that it is the edge of the submerged **continental**

shelf, i.e. the boundary between continental and oceanic crust that actually marks the line along which continents have originally been joined.

It was not until 1965 that the first computer-drawn reassembly of the continents around the Atlantic Ocean was produced by the British geophysicist Edward Bullard and his colleagues at Cambridge University. They used spherical geometry to generate a reconstruction of Africa with South America, and Western Europe with North America, which were all fitted together at the 500 fathom (about 1000 m) contour, which corresponded to the edge of the continental shelf. This method revealed the fit to be excellent (Figure 3.2), with few gaps or overlaps.

- Figure 3.2 shows some overlaps in the way in which the continents fit together around the Atlantic. Why might these exist?
- Most of the overlaps are caused by features that have formed since the continents broke up or rifted apart, such as coral banks (Florida), recent river deltas (Niger) and volcanoes (Iceland).



Figure 3.2 A computer-generated spherical fit at the 500 fathom contour (i.e. edge of continental shelf) showing Bullard's fit of the continents surrounding the Atlantic.

2 Geological match and continuity of structure

Previous configurations of continents can also be recognised by the degree of geological continuity between them. These include similar rock types found on either side of an ocean or, more commonly, successions of strata or igneous bodies that have otherwise unique characteristics. Taylor (Box 3.1) was first prompted to consider continental drift by noting the similarity of the rock strata and geological structures of the Appalachian and Caledonian mountain belts of eastern USA and NW Europe respectively. Similarly, Wegener investigated the continuity of Precambrian rocks and geological structures between South America and Africa (Figure 3.3).

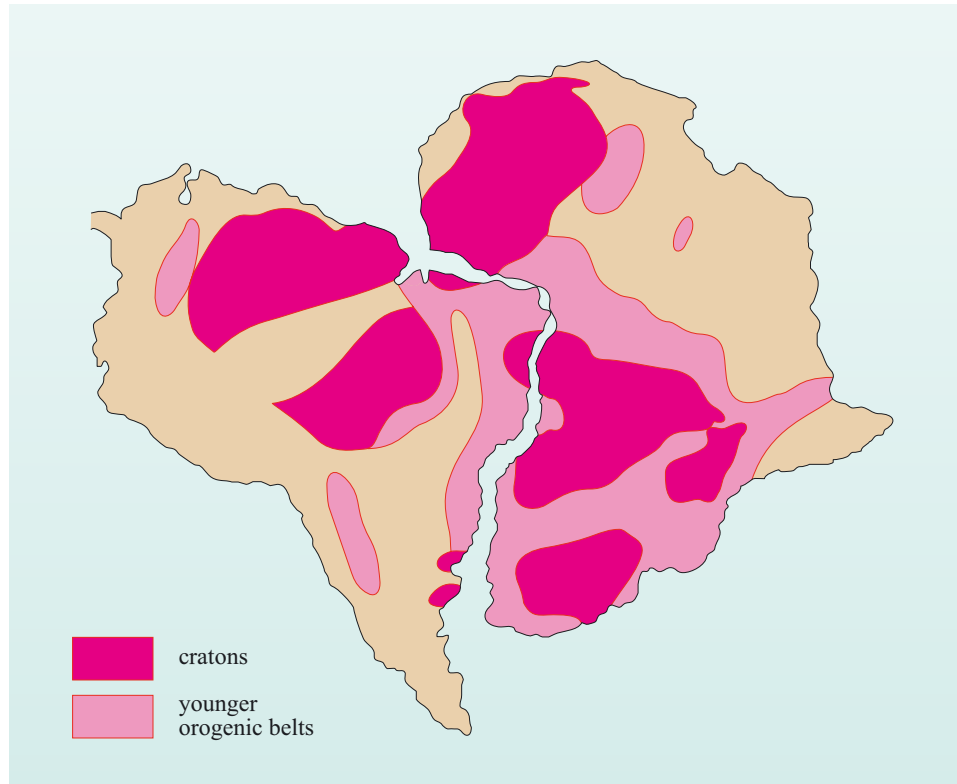


Figure 3.3 Continuity of Precambrian rocks. There is good correlation between these geological units when the continents are fitted along their opposing margins. The immense periods of time over which these Archaean and Precambrian units were formed (>2 Ga) indicate that South America and Africa had together formed a single land mass for a considerable part of the Earth's history. (Adapted from Hallam, 1975)

3 Climate, sediment and the mismatch of sedimentary deposits with latitude

The climate of modern Earth may be divided into different belts that have cold arctic conditions at high latitudes and hot tropical conditions at equatorial and low latitudes. The nature and style of rock weathering and erosion varies according to these climate belts, such that glaciation and freeze–thaw action predominate at present-day high latitudes, whilst chemical alteration, aeolian and/or fluvial processes are more typical of present-day low latitudes. Once a rock is weathered and eroded, each climatically controlled suite of processes gives rise to its own type of sedimentary succession and landforms:

- sand dunes form in hot, dry deserts
- coal and sandstone successions form in tropical swamps and river deltas
- boulder clay deposits and 'U-shaped' valleys form where there are ice sheets and glaciers.

It has long been recognised that geologically ancient glacial-type features are not just restricted to the present-day, high-latitude locations, but also occur in many warm-climate continents such as Africa, India and South America. Similarly, warm-climate deposits may be found in northern Europe, Canada and even Antarctica. For instance, coal is one of our most familiar geological materials, yet the European and North American coal deposits are derived from plants that grew and decayed in hot, steamy tropical swamps 320–270 Ma ago during the late Carboniferous and early Permian Periods. Reasons for these unusual distributions are often provided by reconstructing the ancient continental areas and determining their original positions when the deposits or landforms were created.

Question 3.1

Late Carboniferous coalfields are found in northern Britain around latitude 55° N. If these coals formed from plants that grew in the tropics between 23° N and 23° S, what is the *minimum* distance Britain has travelled in 300 Ma? At what rate has it travelled (in mm y^{-1})? (Assume the radius of the Earth is 6370 km.)

4 Palaeontological evidence

Palaeontological remains of fossil plants and animals are amongst the most compelling evidence for continental drift. In many instances, similar fossil assemblages are preserved in rocks of the same age in different continents; the most famous of these assemblages is the so-called *Glossopteris* flora. This flora marks a change in environmental conditions. In the southern continents, the Permian glacial deposits were succeeded by beds containing flora that was distinct from that which had developed in the climatically warm, northern land masses of Laurasia. The new southern flora grew under cold, wet conditions, and was characterised by the ferns *Glossopteris* and *Gangamopteris*, the former giving its name to the general floral assemblage. Today, this readily identifiable flora is preserved only in the Permian deposits of the now widely separated fragments of Gondwana.

5 Palaeomagnetic evidence and ‘polar wander’

The Earth has the strongest magnetic field of all the terrestrial planets, with similar properties to a magnetic dipole or bar magnet. As newly erupted volcanic rocks cool, or sediments slowly settle in lakes or deep ocean basins, the magnetic minerals within them become aligned according to the Earth’s ambient magnetic field. This magnetic orientation becomes preserved in the rock. The ancient inclination and declination of these rocks can then be measured using sensitive analytical equipment.

As a continent moves over the Earth’s surface, successively younger rocks forming on and within that continent will record different palaeomagnetic positions, which will vary according to the location of the continent when the rock was formed. As a result, the position of the poles preserved in rocks of different ages will apparently deviate from the current magnetic pole position (Figure 3.4a). By joining up the apparent positions of these earlier poles, an **apparent polar wander** (APW) path is generated. It is now known that the Earth’s magnetic poles do not really deviate in this manner, and the changes depicted in APW paths are simply a result of the continent moving over time (Figure 3.4b).

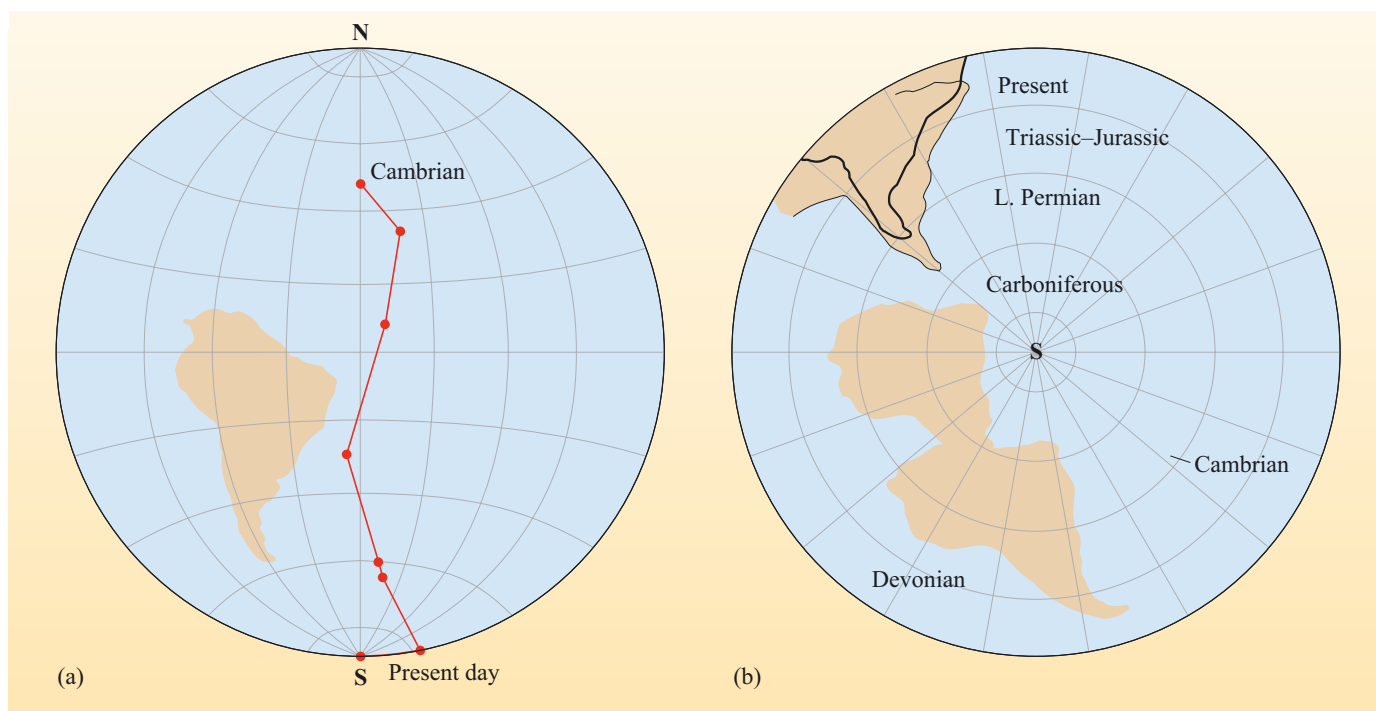
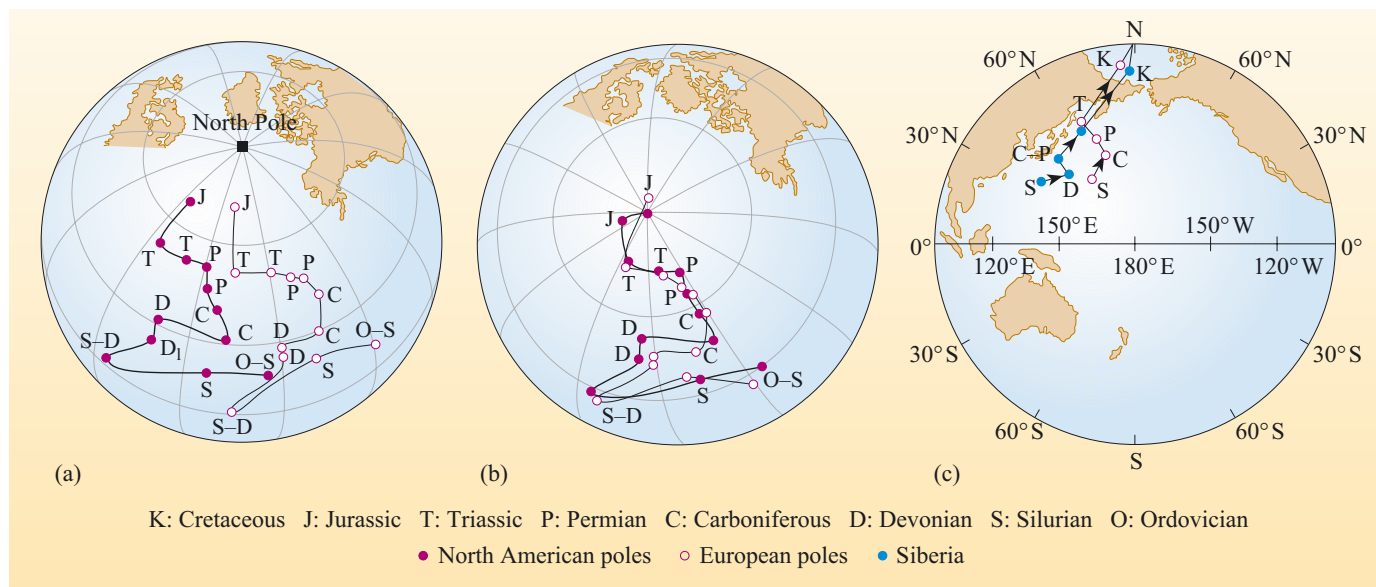


Figure 3.4 Two methods of displaying palaeomagnetic data: (a) assumes that the continent has remained fixed over time, and records the apparent polar wandering path of the South Pole; (b) assumes the magnetic poles are fixed over time, and records the latitude drift of a continent. (Adapted from Creer, 1965)

Nevertheless, APW paths remain a commonly used tool because they provide a useful method of comparing palaeomagnetic data from different locations. They are especially useful in charting the rifting and suturing of continents.

Figure 3.5a shows North America and Europe have individual apparent polar wander paths. However, they are broadly alike in that they have similar changes in direction at the same time. Figure 3.5b shows the APW paths if the Atlantic Ocean is closed by matching the continental shelves.

▼ **Figure 3.5** (a) Apparent polar wander paths for North America and Europe, as measured. (b) Apparent polar wander paths for North America and Europe with the Atlantic closed. Poles for successive geological periods are shown. (c) The apparent polar wander paths for Europe and Siberia. (Adapted from Mussett and Khan, 2000)



- What does this tell you about the North American and European continental masses during the periods spanned by these palaeomagnetic records?
- The two continents were moving together as one mass from the Ordovician right through to the opening of the Atlantic Ocean during the Jurassic Period.

Conversely, if the APW paths of two regions were different to begin with, but became similar later on, one explanation would be that the two regions were originally on independent land masses that then collided and subsequently began to move together as a single continental unit.

Question 3.2

What do the APW paths in Figure 3.5c tell you about the way in which Europe and Siberia have drifted from the Silurian Period to the present day?

Despite Wegener's amassed evidence and the increasing body of geological, palaeontological and palaeomagnetic information, there remained strong opposition to his theory of continental drift, leaving just a few forward-thinking individuals to continue seeking evidence to support this theory (Box 3.1).

The scientific opposition reasoned that if continents move apart, then surely they must either leave a gap at the site they once occupied or, alternatively, must push through the surrounding sea floor during their movement. The geophysicists of the day quickly presented calculations demonstrating that the continents could not behave in this way and, more importantly, no one could conceive of a physical mechanism for driving the continents in the manner Wegener had proposed. Consequently, the theory of continental drift did not gain scientific popularity at the time and became increasingly neglected for several decades. To gain a wider scientific acceptance, Wegener's ideas had to await a greater understanding in the internal structure of the Earth and the processes controlling the loss of its internal heat.

3.1.2 Sea-floor spreading

During and just after World War II, the technological improvement to submarines led to an improvement in underwater navigation and surveying that revealed many intriguing underwater features. The most important of these were immense, continuous chains of volcanic mountains running along the ocean basins. These features are now termed **mid-ocean ridges** or more accurately, **oceanic ridge systems**.

Using this new information, three American scientists Hess, Dietz and Heezen (Box 3.1) proposed that the sea floor was actually spreading apart along the ocean ridges where hot magma was oozing up from volcanic vents. They further suggested that the oceanic ridges were the sites of generation of new ocean lithosphere, formed by partial melting of the underlying mantle followed by magmatic upwelling. They named the process **sea-floor spreading**. Moreover, they proposed that the topographic contrast between the ridges and the oceanic **abyssal plains** was as a consequence of the thermal contraction of the crust as it cooled and spread away from either side of the ridge axis. Most importantly,

because new oceanic crust is generated at the ridge, the ocean must grow wider over time and, as a consequence, the continents at its margin move further apart. The evidence to support this model was found, once again, in the magnetic record of the rocks, but this time using rocks from the ocean floor.

Linear magnetic anomalies – a record of tectonic movement

At the time that sea-floor spreading was proposed, it was also known from palaeomagnetic studies of volcanic rocks erupted on land that the Earth's magnetic polarity has reversed numerous times in the geological past. During such **magnetic reversals**, the positions of the north and south magnetic poles exchange places. In the late 1950s, a series of oceanographic expeditions was commissioned to map the magnetic character of the ocean floor, with the expectation that the ocean floors would display largely uniform magnetic properties. Surprisingly, results showed that the basaltic sea floor has a striped magnetic pattern, and that the stripes run essentially parallel to the mid-ocean ridges (Figure 3.6). Moreover, the stripes on one side of a mid-ocean ridge are symmetrically matched to others of similar width and polarity on the opposite side.

In 1963, two British geoscientists, Vine and Matthews (Box 3.1), proposed a hypothesis that elegantly explained how these magnetic reversal stripes formed by linking them to the new idea of sea-floor spreading. They suggested that as new oceanic crust forms by the solidification of basalt magma, it acquires a magnetisation in the same orientation as the prevailing global magnetic field. As sea-floor spreading continues, new oceanic crust is generated along the ridge axis. If the polarity of the magnetic field then reverses, any newly erupted basalt becomes magnetised in the opposite direction to that of the earlier crust and so records the opposite polarity. Since sea-floor spreading is a continuous process on a geological timescale, the process preserves rocks of alternating polarity across the ocean floor (Figure 3.7a). Reading outwards in one direction from the mid-ocean ridge gives a record of reversals over time, and this can be matched with the record read in the opposite direction.

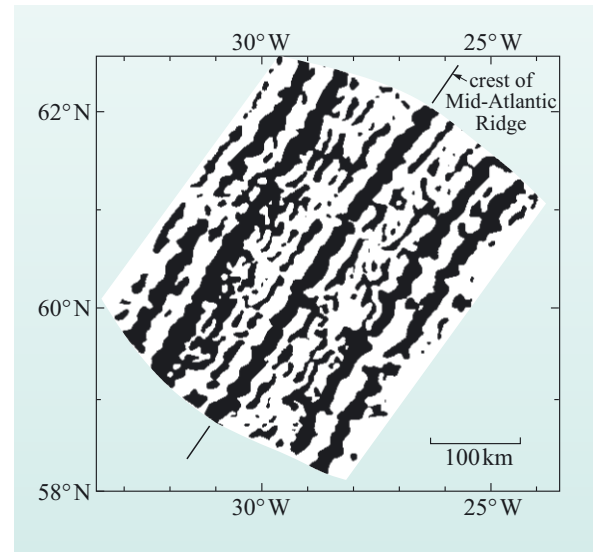


Figure 3.6 A modern map of symmetrical magnetic anomalies about the Atlantic Ridge (the Reykjanes Ridge), south of Iceland. (Adapted from Hirtzler et al., 1966)

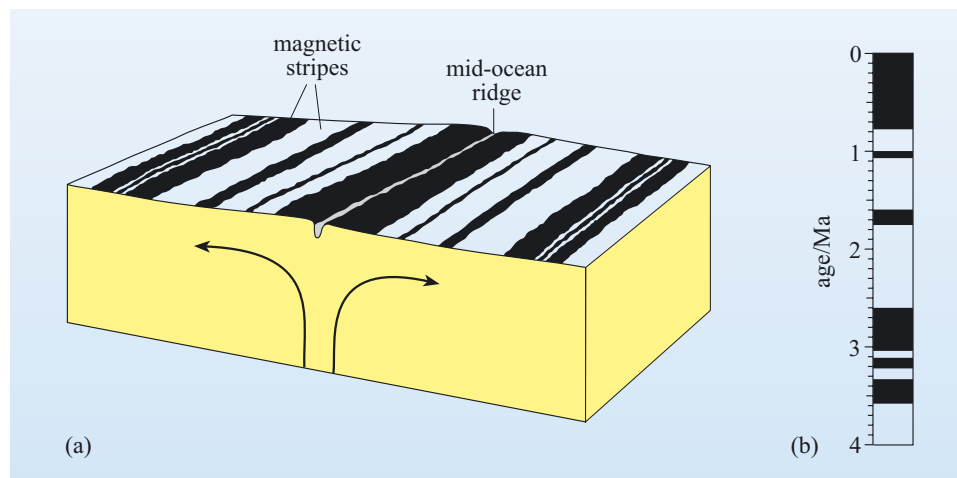


Figure 3.7 (a) Pattern of magnetic reversals on either side of a mid-ocean ridge. Black = normal magnetic field; white = reversed field. When these reversal data are combined with age data (derived by radiometric dating of rocks dredged from the sea floor), a geomagnetic timescale (b) can be produced. Detailed geomagnetic timescales have now been produced for all of the geological time since the Jurassic Period.

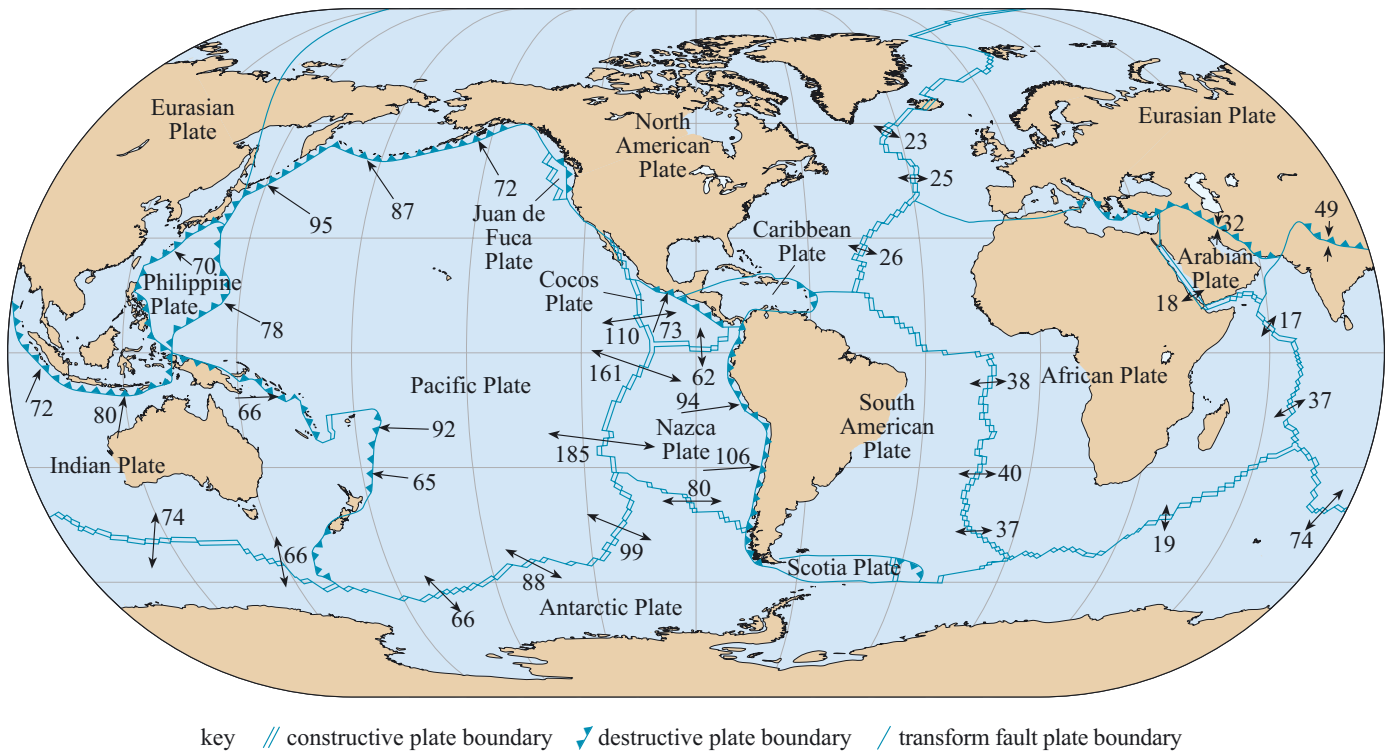


Figure 3.8 Map showing the global distribution of tectonic plates and plate boundaries. The black arrows and numbers give the direction and speed of relative motion between plates. Speed of motion is given in mm y^{-1} . (Adapted from Bott, 1982)

Magnetic and oceanographic surveys of the ocean floor have collected information on both its palaeomagnetic polarity and its absolute age (by radiometric dating of retrieved sea-floor samples). Combining these two records has helped establish a **geomagnetic timescale** (Figure 3.7b) and, by using samples from the oldest sea floor, this timescale has now been extended back into the Jurassic Period, allowing the ages and rates of sea-floor spreading to be established for all the world's oceans, as shown in Figure 3.8. (The different types of plate boundary shown in Figure 3.8 are discussed later in the text.)

Two measures of spreading rate are commonly cited:

- where the rate of spreading is determined on one side (i.e. the rate of movement away from the ridge axis), this is termed the **half spreading rate**
- where the rate is determined on both sides (i.e. the combined rate of divergence), the combined value is termed the **full spreading rate**.

■ Assuming symmetrical spreading rates, use the data given on Figure 3.8 to discover the maximum and minimum spreading rates, and half spreading rates for the ocean ridge system of (i) the Atlantic Ocean (ii) the Pacific Ocean.

■ The maximum and minimum spreading rates for (i) the Atlantic Ocean are 40 mm y^{-1} and 23 mm y^{-1} (as shown by the double-ended arrows along the central Atlantic ridge); these represent half spreading rates of 20 mm y^{-1} and 11.5 mm y^{-1} respectively. The maximum and minimum spreading rates for the Pacific are 185 mm y^{-1} and 66 mm y^{-1} (as shown by the double-ended arrows along the ridge). These represent half spreading rates of 92.5 mm y^{-1} and 33 mm y^{-1} respectively.

Question 3.3

The width of ocean floor between the spreading ridge in the South Atlantic Ocean at 30° S and the edge of the continental shelves along the east coast of South America and the west coast of southern Africa at 3 °S is approximately 3100 and 2700 km respectively. Assuming that the spreading rate on this segment of the ridge is 38 mm y⁻¹, estimate the maximum age of the sea floor on either side of the South Atlantic.

Magnetic stripes not only tell us about the age of the oceans, they can also reveal the timing and location of initial continental break-up. The oldest oceanic crust that borders a continent must have formed after the continent broke apart initially, and just as sea-floor spreading began. In effect, it records the age when that continent separated from its neighbour. In the northern Atlantic, for example, oceanic crust older than 140 Ma is restricted to the eastern USA and western Saharan Africa, therefore separation of North America from this part of Africa must have commenced at this time. The oldest oceanic crust that borders South America and sub-equatorial Africa is only about 120 Ma old. Accordingly, it follows that the North Atlantic Ocean started to form before the South Atlantic Ocean.

If new sea floor is being created at spreading centres, then old sea floor must be being destroyed somewhere else. The oldest sea floor lies adjacent to deep ocean trenches, which are major topographic features that partially surround the Pacific Ocean and are found in the peripheral regions of other major ocean basins. The best known example is the Marianas Trench where the sea floor plunges to more than 11 km depth. Importantly, ocean trenches cut across existing magnetic anomalies, showing that they mark the boundary between lithosphere of differing ages. Once this association had been recognised, the fate of old oceanic crust became clear – it is cycled back into the mantle, thus preserving the constant surface area of the Earth.

...and on to plate tectonics

The combination of evidence for continental drift with the increasing evidence in favour of sea-floor spreading finally led to the development of plate tectonics (Box 3.1). Ideas developed in the 1960s and 70s have survived largely unaltered to the present day, albeit modified by more sophisticated data and modelling methods.

3.2 The theory of plate tectonics

The surface of the Earth is divided into a number of rigid plates that extend from the surface to the base of the lithosphere. A plate can comprise both oceanic and continental lithosphere. As you already know, continental drift is a consequence of the movement of these plates across the surface of the Earth. Thus the need for the continents to plough through the surrounding oceans is removed, as is the problem of the gap left in the wake of a continent as it drifts – both issues that led to scientific opposition to Wegener's ideas (Section 3.1.1).

The theory of plate tectonics is based on several assumptions, the most important of which are:

- 1 New plate material is generated at ocean ridges, or **constructive plate boundaries**, by sea-floor spreading.
- 2 The Earth's surface area is constant, therefore the generation of new plate material must be balanced by the destruction of plate material elsewhere at **destructive plate boundaries**. Such boundaries are marked by the presence of deep ocean trenches and volcanic island arcs in the oceans and, when continental lithosphere is involved, mountain chains.
- 3 Plates are rigid and can transmit stress over long distances without internal deformation – relative motion between plates is accommodated only at plate boundaries.

As a consequence of these three assumptions, and particularly the third assumption, much of the Earth's geological activity, especially seismic and volcanic, is concentrated at plate boundaries (Figure 3.9). For example, the position of the Earth's constructive, destructive and conservative plate boundaries can be mapped largely on the basis of seismic activity. However, it is not enough just to know where boundaries are. In order to understand the implications that plate tectonics has for Earth evolution and structure, you first need to explore the structure of lithospheric plates, their motion – both relative and real – and the forces that propel the plates across the Earth's surface.

3.2.1 What is a plate?

In order to understand how and why the lithospheric plates move it is first necessary to understand their physical and thermal structure. In Chapter 1 you learned that:

- the Earth can be divided into the core, mantle and crust based on its physical and chemical properties
- the lithosphere comprises the Earth's crust and the upper, brittle part of the mantle.

The thickness of the lithosphere is variable, being up to 120 km thick beneath the oceans; it is considerably thicker beneath ancient continental (cratonic) crust. However, the thermal structure of a plate is best illustrated with reference to the ocean basins and how their thermal characteristics change with time.

3.2.2 Heat flow within plates

As newly formed lithosphere moves away from an oceanic ridge, it gradually cools and heat flow (Box 3.2) decreases away from constructive plate boundaries.

- If a body cools, what happens to its density?
- It contracts, so its volume decreases, resulting in an increase in density.

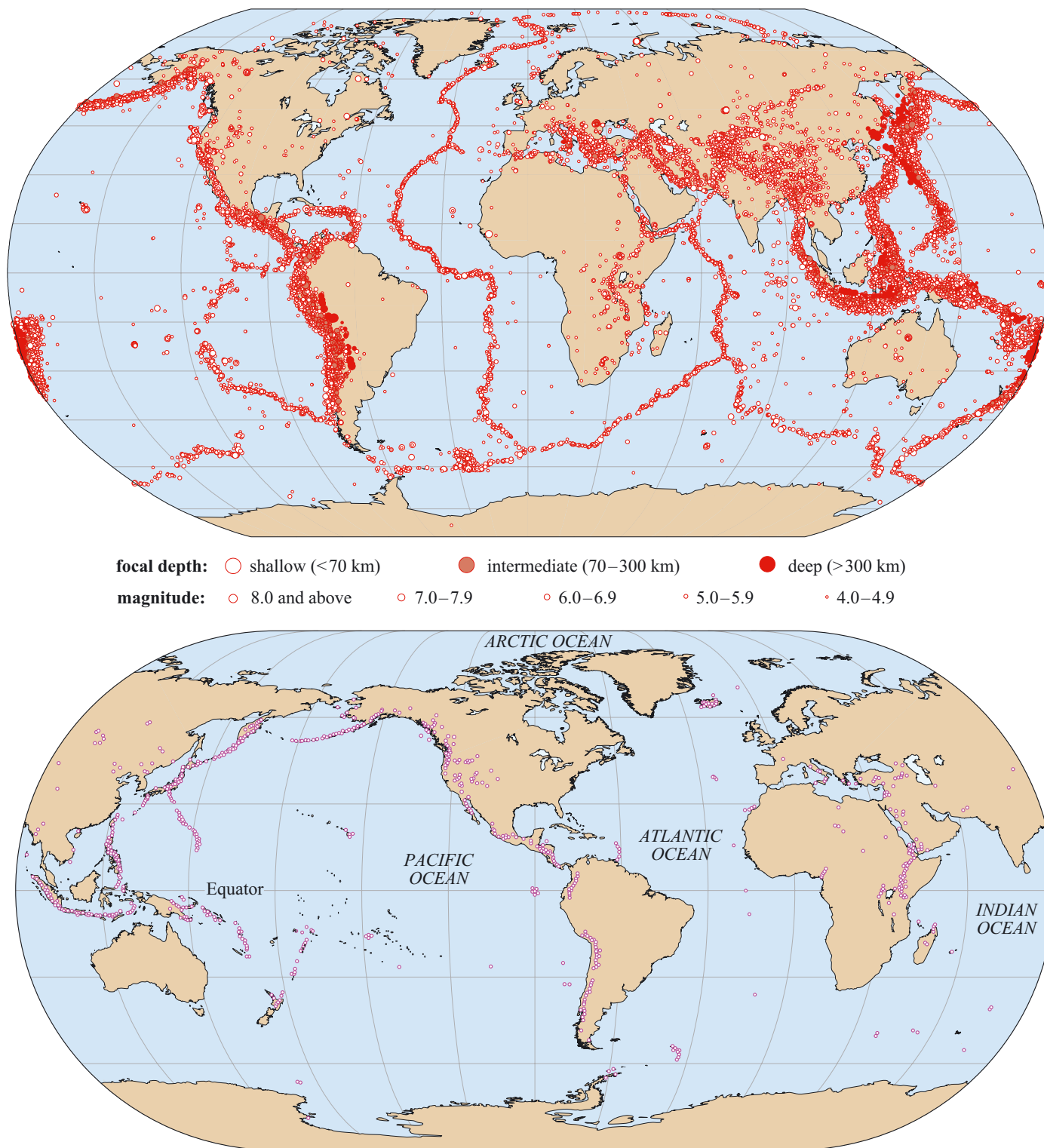


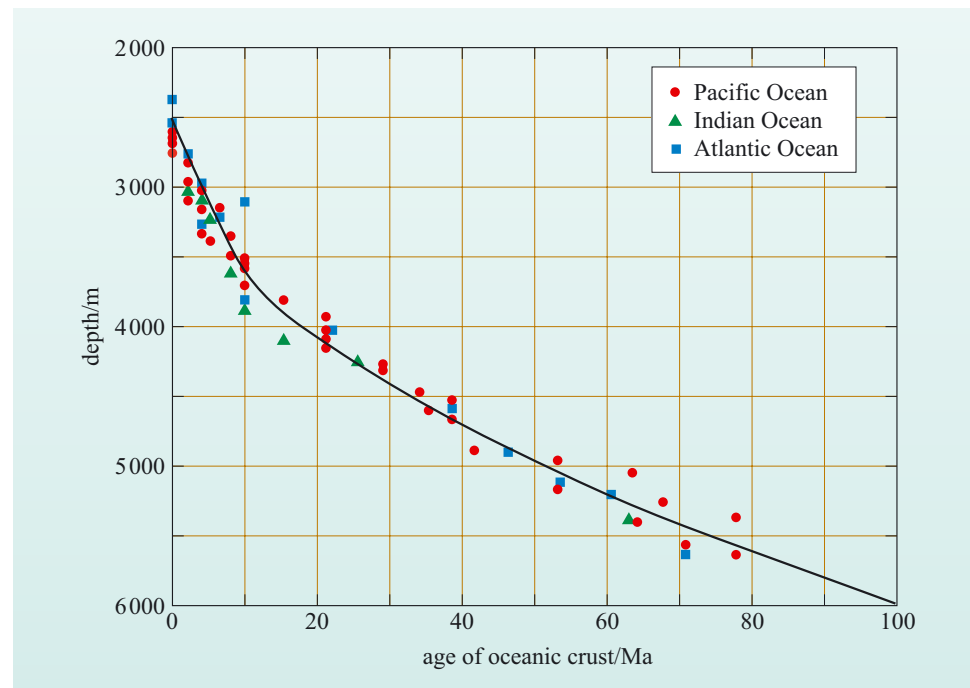
Figure 3.9 (a) Global earthquake epicentres between 1980 and 1996. Only earthquakes of magnitude 4 and above are included. (b) Map showing locations of active, sub-aerial volcanoes. Enlarged versions of Figures 3.9a and b are in the Appendix. ((a) BGS; (b) adapted from Johnson, 1993)

The cooling and shrinking of the lithosphere result in an increase in its density and so, as a result of isostasy (Section 1.5.1), it subsides into the asthenosphere and ocean depth increases away from the ridge, from about 2–3 km at oceanic ridges to about 5–6 km for abyssal plains. Indeed, one of the more remarkable observations of ocean-floor bathymetry is that ocean floor of similar age always occurs at similar depths beneath sea level (Figure 3.10). The relationship between mean oceanic depth (d in metres) and lithosphere age (t in Ma) can be expressed as:

$$d = 2500 + 350t^{1/2} \quad (3.1)$$

If the depth of the ocean floor can be determined, then the approximate age of the volcanic rocks from which it formed may also be estimated, and vice versa.

Figure 3.10 Observed relationship between depth of the ocean floor (both for ocean ridges and abyssal plains) and age of formation of the oceanic crust for the Pacific, Indian and Atlantic Oceans. The solid curve shows the relationship between age and ocean depth according to Equation 3.1.



Question 3.4

The ocean depth at a distance of 1600 km from the Mid-Atlantic Ridge is 4700 m.

(a) Calculate:

- (i) the age of the crust at this location
- (ii) the mean spreading rate represented by this age.

(b) Is this a half or a full spreading rate?

Box 3.2 Earth's heat flow

As you know from Section 2.1, the sources of Earth's internal heat are:

- heat remaining from the initial accretion of the Earth
- gravitational energy released from the formation of the core
- tidal heating
- radiogenic heating within the mantle and crust.

Although the proportion of each heat source cannot be determined accurately, radiogenic heat is considered to have been the major component for much of the Earth's history. There are three main processes by which this internal heat gets to the Earth's surface, these being conduction, convection and advection (Section 2.1.2).

Heat flow (or heat flux), q , is a measure of the heat energy being transferred through a material (measured in units of watts per square metre; W m^{-2}). It may be determined by taking the difference between two or more temperature readings (ΔT) at different depths down a borehole (d), and then determining the thermal conductivities (k) of the rocks in between. q can then be calculated according to the relationship:

$$q = \frac{k\Delta T}{d} \quad (3.2)$$

Earth scientists are interested in the heat flow measured at the Earth's surface because it reveals important information concerning the nature of the rocks and the processes that affect the lithosphere.

The total annual global heat loss from the Earth's surface is estimated as $4.1\text{--}4.3 \times 10^{13} \text{ W}$. This yields an average of $q \geq 100 \text{ mW m}^{-2}$ (milliwatts per square metre), though individual measurements may be much higher than this. However, values of q decrease to less than 50 mW m^{-2} for oceanic crust older than 100 Ma (Figure 3.11). In continental areas, the younger crust (i.e. mountain belts that are less than 100 Ma) have relatively high values of q , which are $60\text{--}75 \text{ mW m}^{-2}$, whilst old continental crust and cratons have much lower heat flow values, averaging $q = 38 \text{ mW m}^{-2}$. Thus, variations in heat flow are closely related to different types of crustal materials and, importantly, different types of tectonic plate boundary.

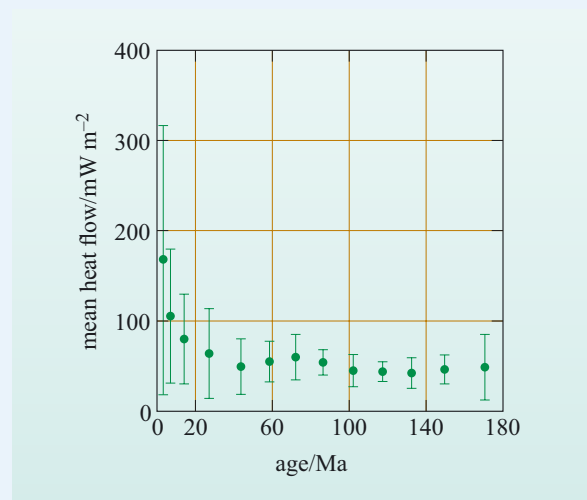


Figure 3.11 Mean heat flow (q) and associated standard deviation (vertical lines) plotted against the age of the oceanic lithosphere for the North Pacific Ocean.

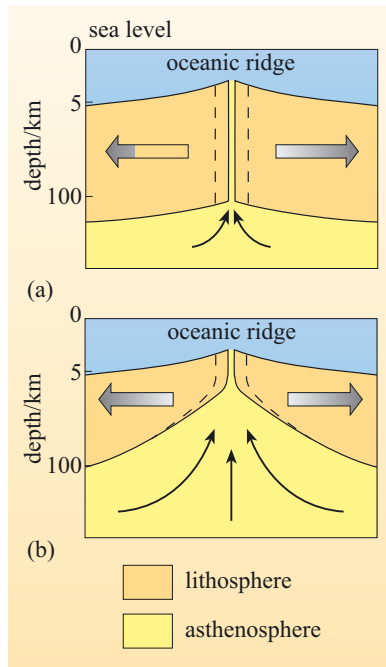


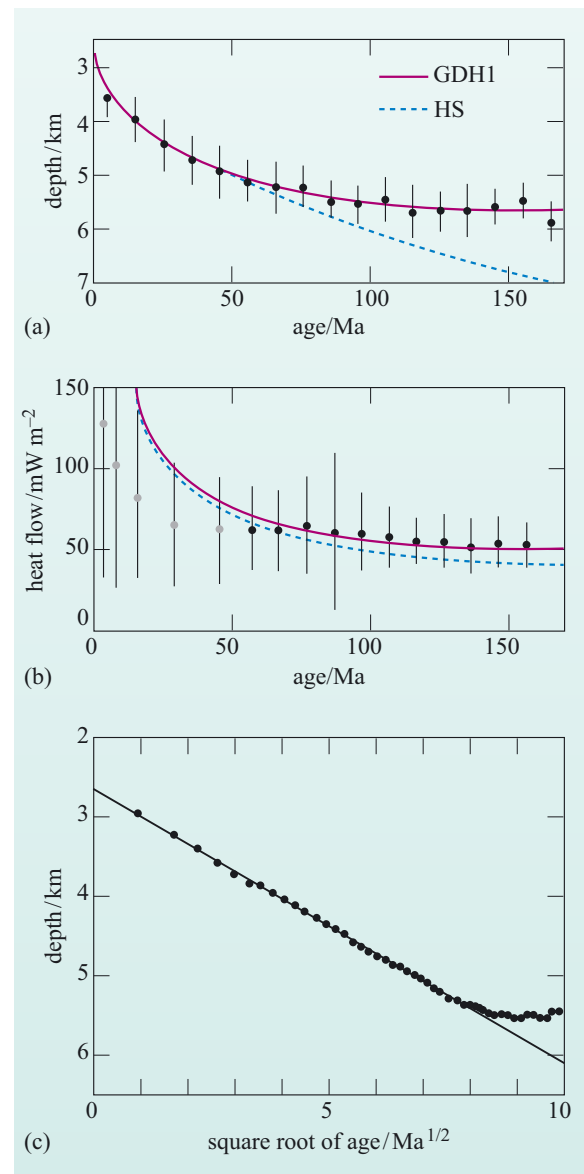
Figure 3.12 Schematic sections through oceanic lithosphere formed at an oceanic ridge: (a) plate model, in which oceanic lithosphere thickness remains constant as the lithosphere moves away from the ridge; (b) boundary-layer model, in which the lithosphere thickens as it ages and cools. The dashed lines show where new lithosphere forms in both models.

There are two general models for the thermal evolution of the oceanic lithosphere: the plate model and the boundary-layer (or half-space) model.

- The **plate model** (Figure 3.12a) assumes that the lithosphere is produced at a mid-ocean ridge with constant thickness and that the temperature at the base of the plate corresponds to its temperature of formation.
- The **boundary-layer model** (Figure 3.12b) assumes that the lithosphere does not have a constant thickness, but thickens and subsides as it cools and moves away from the ridge. This is achieved by loss of heat from the underlying asthenosphere, which progressively cools below the temperature at which it can undergo solid-state creep and is transformed from asthenosphere to lithosphere.

The thermal consequences of these two models can be calculated from a knowledge of the temperature of the mantle at depth (the geotherm) and the thermal conductivity of the rocks in the lithosphere. As it turns out, both models predict similar results for both heat flow and ocean depth, as shown in Figure 3.13.

Figure 3.13 Graphical plots of (a) depth and (b) heat flow against age for the Pacific Ocean. The line labelled GDH1 refers to the plate model, whereas the curve labelled HS refers to the boundary layer model. (c) Graphical plot of ocean bathymetry (depth) against the square root of the age of the lithosphere for the Pacific Ocean. Note the good linear relationship for young lithosphere and the deviation from the simple linear trend for older lithosphere. (Fowler, 2005)



Question 3.5

Study Figure 3.13(c) and then answer the following questions:

- Is the correlation between depth and the square root of the age of the lithosphere positive or negative?
- At what age does the relationship depart from a linear correlation?
- Does this departure imply that older oceanic lithosphere is warmer or cooler than predicted from the simple, linear boundary-layer model?

Both models predict the observed linear variation between ocean depth and the square root of age of the lithosphere, showing that the oceanic lithosphere cools and subsides as it ages away from a spreading centre. However, the plate model fits the data better for older lithosphere (>60 Ma), suggesting that once lithosphere has cooled to a certain thickness, the thickness remains more-or-less constant until the plate is subducted.

Both models also predict greater heat flow from young oceanic crust than that observed in the ocean basins, as shown in the example in Figure 3.13b.

- What do you think the cause of this difference might be?
- The thermal models are based on the assumption that heat is lost by conduction only, whereas in reality other mechanisms of heat transfer might be in operation.

The formation of oceanic lithosphere involves contact between hot rocks and cold seawater. As the rocks cool and fracture, they allow seawater to penetrate the young, hot crust to depths of at least a few kilometres. During its passage through the crust the seawater is heated before being cycled back to the oceans. This process is known as **hydrothermal circulation** and the development of submarine hydrothermal vents ('black smokers') close to mid-ocean ridges is the most dramatic expression of this heat transfer mechanism. However, less dramatic but probably equally significant lower temperature circulation continues well beyond the limits of oceanic ridges and contributes to heat loss from the crust up to 60 Ma after formation.

The geophysical evidence from seismology and isostasy suggests that the oceanic lithosphere increases in thickness as it ages until it reaches a maximum of about 100 km. By contrast, bathymetry and heat flow indicate a more constant thickness for older plates. To explain these observations, a plate structure such as that shown in Figure 3.14 has achieved wide acceptance. The plate is divided into two layers: an upper, rigid **mechanical boundary layer** and a lower, viscous **thermal boundary layer**. Both layers thicken progressively with time until the lithosphere is about 80 Ma old, after which the thermal boundary layer becomes unstable and starts to convect. Convection within this layer provides a constant heat flow to the base of the mechanical boundary layer. Thus old lithosphere maintains a constant thickness because the heat flow into its base, and hence the basal temperature, is maintained by convection.

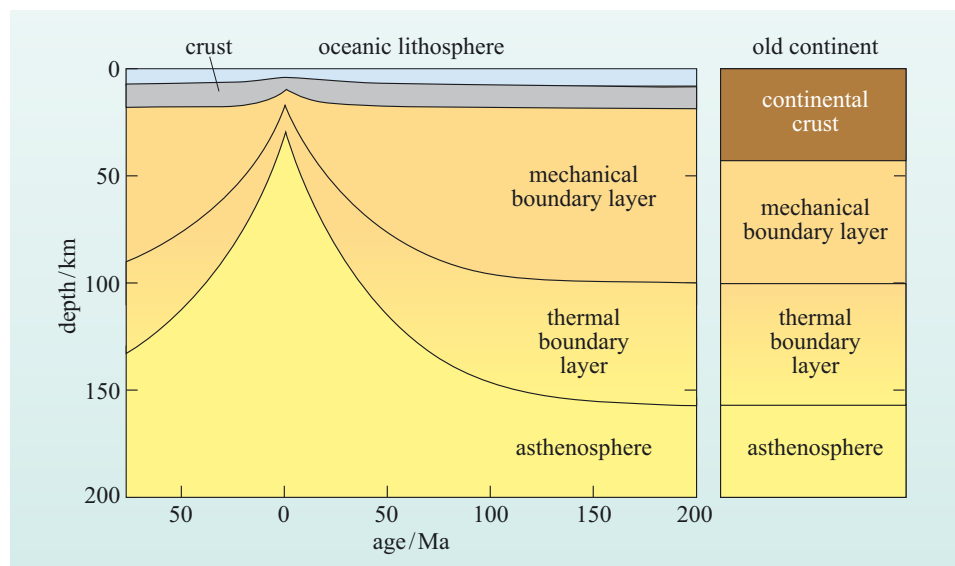


Figure 3.14 Thermal model of lithospheric plates beneath oceans and continents. Note the thicker crust and thinner mantle lithosphere in the continental section. (Fowler, 2005)

3.2.3 Constructive plate boundaries

Constructive plate boundaries or margins are regions where new oceanic crust is being generated. However, in order for the magma to ascend to the surface and build new lithosphere, the earlier formed crust must be pulled apart and fractured to create a new magma pathway. Hence constructive plate boundaries are regions of extensional stresses and **extensional tectonics**. The process of fracturing, injection and eruption is repeated frequently, so that tensional stresses do not have time to accumulate significantly and, as a result, constructive plate boundaries are characterised by frequent, low-magnitude seismicity (typically less than magnitude 5), occurring at shallow crustal depths (<60 km) along the ocean ridge systems.

- Why do you think earthquakes are restricted to shallow depths beneath ocean ridges?
- Earthquakes can only occur in brittle rocks that fail by fracture rather than solid-state creep. As a result of high geothermal gradients beneath oceanic ridges, the brittle lithosphere is thin and so earthquakes are restricted to shallow depths.

Sonar surveying, and direct investigation by sea-floor drilling or deep sea submersibles, has revealed that volcanism along the ridge systems typically consists of a series of individual, active eruption centres. Each eruptive centre is no more than about 2–3 km long, and along the ridge axis they are often separated from each other by an inactive gap of about 1 km. Beneath the spreading ridge the feeder magma chambers that supply the volcanic centres are more continuous, often linking between and across the active segments. This means that magma generation occurs along much of the length of the ridge even though it is erupted via a chain of individual volcanic centres.

Plates move away from constructive boundaries at speeds that can be as low as $<10 \text{ mm y}^{-1}$ to so-called ultra-fast spreading ridges where half spreading rates can exceed 100 mm y^{-1} . Examples of slow spreading ridges include parts of the mid-Atlantic Ridge and the southwest Indian Ridge. Fast and ultra-fast ridges occur in the East Pacific, along the East Pacific Rise and the Galapagos spreading centres (Figure 3.8).

The depth structure of a constructive plate boundary can be further defined from the variation in the Earth's gravity (Box 3.3). Figure 3.15 shows gravity anomalies across the Mid-Atlantic Ridge. Despite the topographic rise associated with the ridge, the **free-air gravity anomaly** is relatively flat and close to zero across the whole structure. This indicates that there is no mass deficit or excess down to the level of isostatic compensation, i.e. the ridge is in isostatic equilibrium with the lithosphere of the abyssal plains. By contrast, when the free-air anomaly is corrected for the effects of the ridge topography and overall altitude, the resulting **Bouguer gravity anomaly** is strongly positive but with a local dip across the ridge axis. The positive anomaly occurs because of the raised topography of the ridge, but the ridge zone itself is underlain by lower-density material. A possible density model is shown in Figure 3.15b.

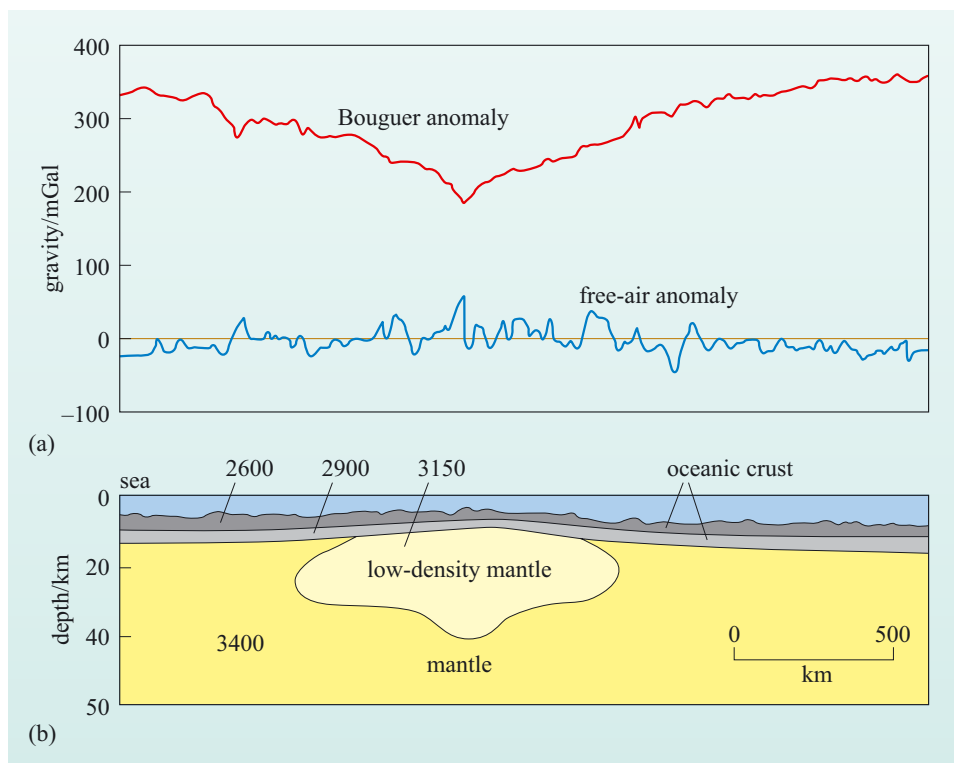


Figure 3.15 (a) Bouguer and free-air gravity anomalies across part of the Mid-Atlantic Ridge. (b) One possible density model that could produce the observed anomalies and satisfies other constraints (e.g. seismic structure). Figures give the densities of different layers in kg m^{-3} .

- What do you think the low-density material beneath the ridge might represent?
- Hot, possibly partially molten mantle that feeds the basaltic volcanism of the ridge axis.

Box 3.3 Gravity and gravity anomalies

Gravity is the attractive force experienced by all objects simply as a consequence of their mass. The magnitude of the attraction is determined from Equation 3.3:

$$F = \frac{Gm_1m_2}{r^2} \quad (3.3)$$

where m_1 and m_2 are the masses (measured in kg) of two objects, r is the distance (measured in m) between them and G is the universal gravitational constant ($6.672 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$).

If one object is the Earth, with mass M , and the other is a much smaller object with mass m , Equation 3.3 can be rewritten:

$$F = \frac{GMm}{d^2} \quad (3.4)$$

where d is the radius of the Earth, i.e. the distance to the Earth's centre of gravity.

However, the force experienced by an object at the surface of the Earth can be measured:

$$F = mg \quad (3.5)$$

where m is its mass and g is the acceleration due to gravity. Hence:

$$F = mg = \frac{GMm}{d^2} \quad (3.6)$$

Cancelling m gives

$$g = \frac{GM}{d^2} \quad (3.7)$$

Hence g is proportional to the mass of the Earth and inversely proportional to the square of the distance from the centre of the Earth. Note that Equation 3.7 gives a way of measuring the mass of the Earth (M) if g can be measured and G and d are known.

- Given that the Earth is not truly spherical and that the poles are closer to the centre than the Equator is, will g at the poles be greater than, less than or similar to g at the Equator?
- Because the poles are closer to the centre of the Earth than the Equator, d^2 in Equation 3.7 will be lower, therefore g will be slightly greater.

Variations in gravity are measured in milligals (mGal) and 1 mGal is equivalent to 10^{-5} m s^{-2} . Since $g = 9.81 \text{ m s}^{-2}$, this means that $1 \text{ mGal} \sim 10^{-6} g$.

Measured variations in gravity across the surface of the Earth relate to the mass in the vicinity of the point of measurement. Thus a region underlain by dense rocks, such as basalt, will exhibit a slightly stronger gravitational pull than those underlain by less dense rocks, such as granite or sediments. In addition, gravity is affected by the underlying topography and the altitude at which the measurement was made. Thus measurements of the variation in the Earth's gravitational field require numerous corrections for latitude and topography, the details of which are beyond the scope of this book (but see the texts recommended in the Further Reading section). The resultant gravitational anomaly is known as a **Bouguer anomaly** and this reflects the variations in the Earth's gravity due to the underlying geology. Bouguer gravity is usually calculated over continental regions where the surface topography and the underlying geology are both highly variable. In marine surveys, fewer corrections are applied to the measured value of g , and gravitational anomalies are conventionally referred to as **free-air anomalies**. Gravity measurements, particularly over the oceans, are now routinely recorded from satellites and have resulted in accurate and detailed maps of the free-air anomaly over most of the Earth.

Hot material is generally less dense than cold material, so the low density of the mantle beneath the ridge is related to the locally high geothermal gradient, as also indicated by the presence of basaltic magma and the restriction of earthquakes to the upper levels of the lithosphere.

Finally, it should be noted that constructive plate boundaries by definition cannot occur within continental lithosphere as they must be bounded by new oceanic lithosphere. There are regions of the Earth's crust where a constructive boundary (or boundaries) can be traced into a continental region, for example at the southern end of the Red Sea and the Gulf of Aden the marine basins join and extend into the Ethiopian segment of the African Rift Valley by way of the Afar Depression. While these three features are all part of an extensional tectonic regime, the African Rift Valley cannot be considered to be a true constructive plate boundary, although, in future, if plate configurations are suitable, it may provide the site for the opening of a new ocean.

3.2.4 Destructive plate boundaries

Destructive plate boundaries are regions where two lithospheric plates converge. This situation provides a more varied range of tectonic settings than do constructive plate boundaries. Firstly, and in contrast to constructive plate boundaries, destructive plate boundaries are asymmetrical with regard to plate speeds, age and large-scale structures. Secondly, whereas true constructive boundaries occur almost invariably in oceanic lithosphere, destructive boundaries also affect continental lithosphere – they can occur entirely within continental lithosphere. Consequently, there are three possible types of destructive plate boundary:

- those involving the convergence of two oceanic plates (ocean–ocean subduction)
- those where an oceanic plate converges with a continental plate (ocean–continent subduction)
- collisions between two continental plates (continent–continent destructive boundaries).

These can be thought of as representing three stages in the evolution of destructive boundaries.

In addition to the disappearance of old lithosphere, destructive boundaries associated with ocean–ocean subduction and ocean–continent subduction are also characterised by:

- ocean trenches, generally 5–8 km deep, but sometimes up to 11 km deep. The sea floor slopes into the trenches from both the landward and oceanward sides. They are continuous for many hundreds of kilometres, occurring both adjacent to continents and wholly within oceans.
- a belt of earthquakes that are shallow-centred closest to the trench and deeper further away. Earthquakes can occur as deep as 600–700 km.
- most destructive boundaries are associated with a belt of active volcanoes that, in the case of intra-oceanic boundaries, form chains of islands known as **island arcs**.

1 Ocean–ocean (island-arc) subduction

The convergence of two oceanic plates represents the simplest type of destructive plate boundary and exemplifies most of the features associated with the destruction of oceanic lithosphere. Around the northern and western edges of the Pacific Ocean, many islands are arranged in gently curved archipelagos: anticlockwise these include the Aleutian Islands, the Kuril Islands, Japan, the Mariana Islands, the Solomon–New Hebrides archipelagos and the Tonga–Kermadec Islands north of New Zealand. These all occur some distance off the edge of the continental areas, but lie adjacent to a deep ocean trench. To the oceanward side of the deep trenches the ocean lithosphere is amongst the oldest on Earth. For example, the oceanic crust adjacent to the Marianas Trench, the deepest trench on Earth, is Jurassic in age and up to 180 Ma old. The trenches are sites where old oceanic lithosphere is being destroyed, or **subducted**, beneath younger lithosphere. For this reason, destructive boundaries are often referred to by their alternative name of **subduction zones**.

The typical pattern of earthquakes associated with ocean–ocean subduction is well illustrated in Figure 3.16, which shows the distribution of earthquakes associated with the Tonga Trench in the southwest Pacific. The earthquake data summarised in this diagram clearly define a zone of earthquakes deepening to the west away from the Tonga Trench, beneath the Tonga volcanic islands, and reaching a final depth in excess of 600 km. This inclined plane of earthquakes associated with the Tonga Trench (and every other deep ocean trench) is known as a **Wadati–Benioff zone**, after the first seismologists to recognise its existence.

- Using the information in Figure 3.16, estimate the angle of the Wadati–Benioff zone beneath the Tonga island arc at $\sim 20^\circ$ S. (Assume the first occurrence of >400 km earthquakes along this line is representative of earthquakes with a minimum depth of 400 km.)
- Earthquakes at about 400 km depth are located ~ 400 km west of the Tonga Trench. Hence the tangent of the angle of subduction is $400/400 = 1$, so the angle of subduction is $\sim 45^\circ$ ($\tan 45^\circ = 1$).
- Is the angle constant along the whole length of the Tonga Trench?
- No. To the north at $\sim 16^\circ$ S the horizontal distance between trench and the deepest earthquakes is much greater than at 20° S, so the angle of subduction is shallower.

The presence of earthquakes at great depth in Wadati–Benioff zones shows that rocks are undergoing brittle failure throughout this depth range.

- What does this observation tell you about the thermal state of subducted lithosphere?
- Subducted lithosphere must remain cool to fail seismically.

Subduction involves the recycling of old, and therefore cold, oceanic lithosphere back into the mantle. A common misconception is that earthquakes represent failure between the subducted lithosphere and the overlying mantle. While this

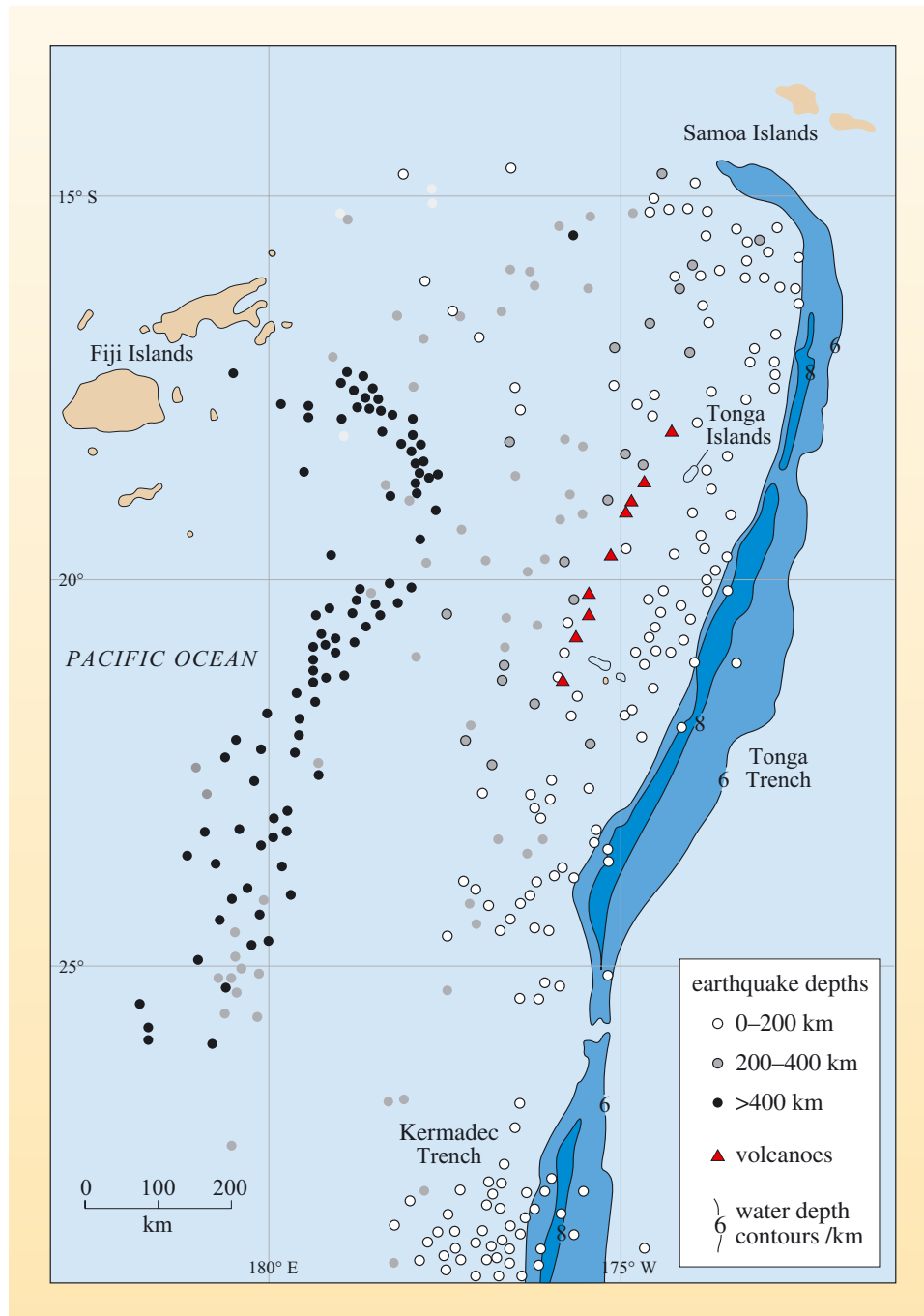


Figure 3.16 Map showing the distribution of earthquake foci associated with the Tonga Trench in the SW Pacific Ocean. (Sykes, 1966)

may be the case for the shallowest levels of subduction, analysis of earthquake waves from deeper seismic events indicate that earthquake foci lie *within* the subducted lithosphere and reflect differential movement between rigid blocks in the subducted lithosphere as it heats up and expands.

The fact that the subducted slab remains colder than ambient mantle at a comparable depth raises a problem for the generation of volcanic activity – a diagnostic feature of subduction. If subduction zones are regions where cold material is being cycled back into the mantle, why are they the site of volcanic activity? Constructive plate boundaries are underlain by regions of hot mantle

that rises in response to the separation of the overlying plates and so it is easy to see how magma can be generated from the mantle. Beneath island arcs there is less evidence for hot mantle and so melting must be triggered by some alternative mechanism. Seismic profiles of subduction zones show that melt is generated immediately above the subducted plate, and most volcanic arcs are located approximately 100 km above its surface, so there is clearly a relationship between subducted oceanic lithosphere and the presence of island-arc volcanism.

The link between subduction and volcanism lies in the composition of the subducted lithosphere.

- In addition to basaltic crust and mantle rocks, what other rocks would you expect in the subducted plate?
- As the plate ages it accumulates a veneer of sediments on its surface. Also hydrothermal processes close to the ridge add water to the upper layers of the basaltic crust. So the subducting plate will include sediments and altered basalt crust.

Subduction provides a mechanism for introducing water-bearing sediments into the mantle, and as the subducted lithosphere heats up the water is gradually released. Water has the effect of reducing the melting temperature of the mantle. It is this process that allows the generation of magma at depth that feeds surface volcanism. As a result, subduction-related magmas are also richer in volatiles than similar rocks from other tectonic environments, such as constructive plate boundaries. This subject will be covered in more detail in Chapter 5.

All of the above characteristics are more or less diagnostic of an oceanic destructive plate boundary. There are, however, a number of other structural features that may or may not be present, but reflect different processes associated with subduction. Figure 3.17b shows the mean ocean depth across the Kuril Trench in the NW Pacific Ocean. On the oceanward side of the Kuril Trench, as with all deep ocean trenches, the ocean depth is between 4 km and 6 km, whereas the trench is 2–4 km deeper still. Note that the vertical scale has been exaggerated fifty times in Figure 3.17b and the actual angles of the sides of the trench are quite shallow, being between 20° and 5°.

- What happens to ocean depth immediately to the ocean side of the trench?
- It decreases slightly.

The decrease in ocean depth towards the trench is characteristic of all island-arc systems and can elevate the ocean floor by as much as 0.5 km. It is caused by the flexure of the lithosphere in response to its entry into the subduction zone and is known as a **flexural bulge**. It is analogous to flexing a ruler over the edge of a table. If you place an ordinary plastic ruler on the edge of a table so that about one-third of it protrudes over the edge and then apply pressure to the extreme tip of the ruler while holding the other end firmly on the table, the ruler will flex and the part that was lying flat on the table will rise slightly. When the pressure is released the ruler will return to its original position because it is a rigid but elastic material. (Imagine trying this experiment with plasticine.)

The flexural bulge is a common feature of ocean-trench systems and is marked by a small increase in free-air gravity, while the trench itself is marked by a large

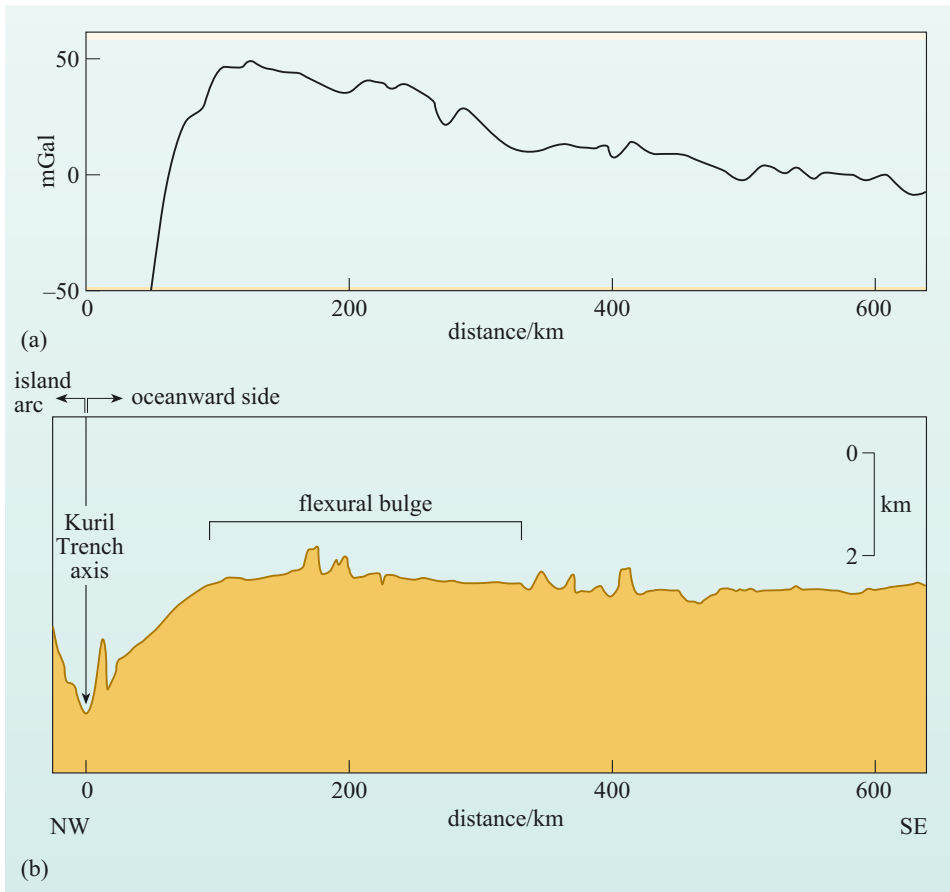


Figure 3.17 (a) Free-air gravity variations and (b) topography across the Kuril Trench. (Adapted from Watts, 2001)

decrease in free-air gravity (Figure 3.17a). Such gravity variations imply that arc systems are out of isostatic equilibrium – the negative anomaly over the trench reflects a mass deficit, meaning that the crust in the trench must be being held down, while the increase over the flexural bulge implies that it is underlain by dense material beneath the plate. The interpretation is that as the plate flexes upwards it ‘pulls in’ the asthenospheric mantle beneath. The isostatic imbalance, with the trench held down and the bulge supported, is due largely to the rigidity of the subducting plate.

While many ocean trenches are particularly deep, others are not. However, they are still characterised by a strongly negative free-air gravity anomaly, implying that they are filled with low-density material.

■ What do think this low-density material might be?

■ Sediments.

As a plate ages, it accumulates a veneer of deep-sea sediments made up of clays and the remains of micro-organisms in the oceans. At the subduction zone, this sedimentary cover is partly scraped off against the overriding plate to form huge wedges of deformed sediment that can eventually fill the trench system. This material is often known as an **accretionary prism**. Not all the sediment is removed, however, and some remains attached to the descending oceanic plate and may become attached to the base of the overlying plates or even be carried into the upper mantle.

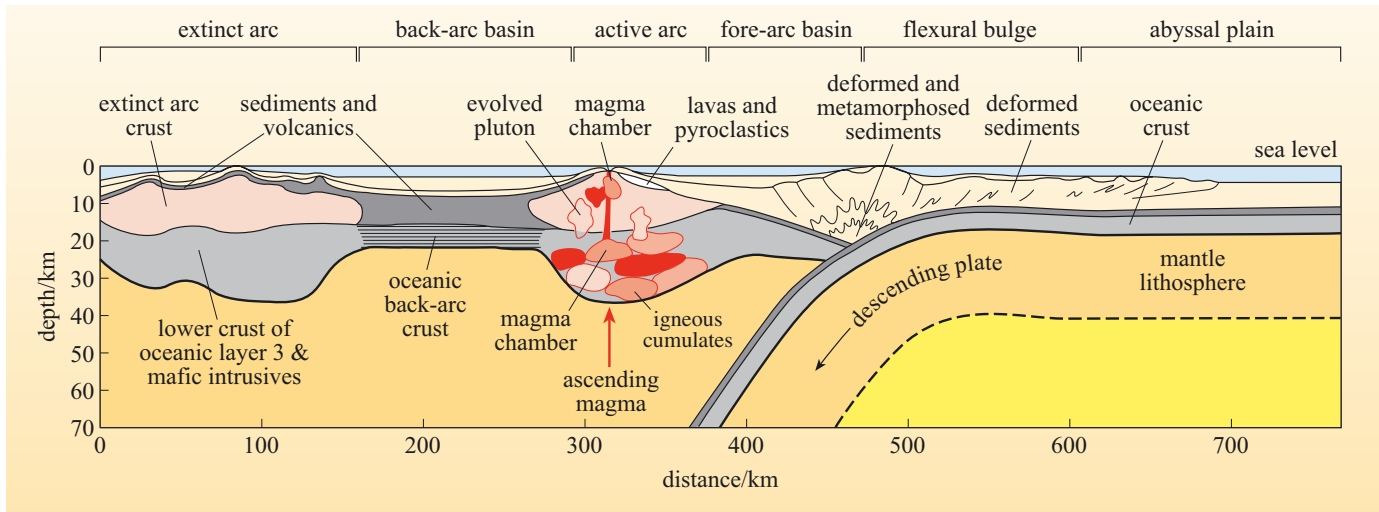


Figure 3.18 Schematic cross-section through an ideal island arc. Note that not all of the features shown here will be present in any one arc system. (Note: 2× vertical exaggeration.)

Behind many island-arc systems, especially those in the western Pacific, small ocean basins open up between the arc and the adjacent continent. Typical examples include the ocean basins immediately to the west of the Tonga and Marianas island arcs. Various lines of evidence show that these regions of ocean crust are very young and characterised by active spreading centres. Such features are known as **back-arc basins**.

- Does the existence of young oceanic crust suggest an extensional or a compressional tectonic regime in back-arc basins?
- An extensional tectonic regime. New ocean crust is only produced when lithospheric plates move apart.

The presence of an extensional regime in the back-arc region basin may appear counter-intuitive because where two plates are converging the dominant tectonic regime should be compressional. The mechanisms that give rise to back-arc tension may relate to convection in the asthenosphere underlying the back-arc region. Alternatively, it has been suggested that old, dense slabs may subside into the mantle at a faster rate than the plate is moving, causing the trench to migrate towards the spreading centre (euphemistically called ‘slab roll-back’). This gives rise to an extensional regime not only in the back-arc basin but also across the whole arc, even to the extent of suggesting that back-arc basins may originate as arcs that have been split by extension as a consequence of slab roll-back.

All of the major features of an oceanic destructive boundary are included in Figure 3.18, which is an idealised cross-section through an oceanic island-arc system.

2 Ocean–continent (Andean type) subduction

When an oceanic plate converges with a continental plate, it is always the ocean plate that subducts beneath the continental plate. You should recall from Chapter 1 that continental lithosphere lies at a higher surface elevation than oceanic lithosphere because of its lower overall density. The resistance shown by continental lithosphere to subduction is simply a further reflection of its lower density.

This type of destructive plate boundary is characterised by the west coast margin of South America. Here, the oceanic lithosphere of the Nazca Plate is being subducted beneath the overriding continental lithosphere that forms the western part of the South American Plate. The overriding continental edge is uplifted to form mountains (the Andes) and the collision zone itself is marked by a deep ocean trench that runs parallel to the continental margin. A chain of active volcanoes runs along much of the length of the South American Andes from Colombia to southern Chile. The ocean trench is similarly characterised by a dipping Wadati–Benioff zone, and is marked by earthquakes reaching depths of several hundred kilometres. The shallowest earthquakes (<60 km) occur at, or near, the trench, with the deepest ones recorded as being up to 700 km deep – the depth before which the plate becomes sufficiently heated to prevent further brittle behaviour. The overall structure of such **Andean margins** is very similar to those described in the previous section, with the added influence of the greater thickness of the overriding continental lithosphere and the probable increased flux of sedimentary material into the system as a result of continental erosion.

3 Continent–continent destructive boundaries

When two continental plates meet at a destructive boundary, the continents themselves collide. These types of continental collision are typically the result of an earlier phase of subduction of intervening oceanic lithosphere that has resulted in the closure of an ocean. Perhaps the best known and most spectacular example is the collision of peninsular India with Asia, which began 50 Ma ago, following the closure of an intervening ocean and produced the Himalayas and Tibetan Plateau (Chapter 6). Even today, India continues to move northwards, indenting the southern edge of Asia at a rate of 40–50 mm y⁻¹. Such collisions result in intense deformation at the edges of the colliding plates, and those sea-floor sediments that were not subducted become folded and compressed into immense mountain chains or **orogenic belts**. Active mountain belts, such as the Alps and Himalayas in Eurasia, and the Rocky Mountains in the USA and Canada, are generally much wider than mountain belts associated with Andean-style arc systems, with deformation belts occurring many hundreds of kilometres into continental interiors.

- What does this observation suggest about the strength of continental lithosphere relative to oceanic lithosphere?
- It suggests that the continents are less strong and less rigid than oceanic lithosphere.

The continents are made up of less-dense rock than oceanic lithosphere and are dominated by quartz and feldspars. At elevated temperatures, these minerals are much weaker than the olivine and pyroxene characteristic of the oceanic crust and mantle. Moreover, continental crust contains a higher concentration of the heat-producing elements K, U and Th. The overall higher heat production conspires with the dominance of weaker minerals to make the continental crust much less rigid than the crust beneath the oceans and, therefore, easier to deform. During deformation the continental crust is thickened and this gives rise to the dramatic topography of active mountain ranges. However, once the forces that drive collision are removed, erosion takes over and the high topography is

reduced to more modest elevations. Older mountain belts, such as the Appalachian and Caledonian orogenic belts, which are the products of continental collisions that occurred hundreds of millions of years ago, are now supported by a balance between isostatic support of their thickened crustal roots and erosion that is controlled by climate. This topic will be dealt with in more detail in Chapter 6.

3.2.5 Conservative plate boundaries and transform faults

Conservative plate boundaries and transform faults occur when plates slide past each other in opposite directions, but without creating or destroying lithosphere. Transform faults connect the end of one plate boundary to the end of another plate boundary, so there are potentially three types of transform fault:

- those that link two segments of a constructive boundary
- those that link two destructive boundaries
- those that link a destructive boundary with a constructive boundary.

Transform faults linking two constructive boundaries are the most common, and account for the displacements between adjacent segments of mid-ocean ridges. Accordingly, this type of ocean transform fault forms an integral part of constructive plate boundaries, and their position is made obvious by the jagged shape of parts of the ocean-ridge system that are split into several segments by series of so-called fracture zones. Examples can be easily seen on the Cocos–Nazca Ridge (also known as the Galapagos Spreading Centre), and the Pacific Ocean spreading ridge (i.e. East Pacific Rise) between 10° N and 10° S, and 40° S and 55° S respectively, or manifest as shorter segments along the Atlantic Ocean spreading ridge between 0° and 40° S. Generally, oceanic transform faults occur at right angles to spreading ridges and, therefore, their orientation is indicative of the direction of plate motion.

Transform faults are seismically active – but only where two different plates are adjacent to one another. In Figure 3.19, the fault trace marks the boundary between plates A and B. Plate A is moving towards the east while plate B is moving towards the west.

- Describe the sense of relative movement along the length of the fault between W and X, X and Y, and Y and Z.
- Between W and X, the fault separates different parts of plate B and so there is no differential movement. Between X and Y it separates plate A from plate B, which are moving in opposite directions. Between Y and Z it separates different parts of plate A and there is, again, no differential movement.
- Which part of the fault between W and Z will be seismically active?
- Only that segment between X and Y where two different plates are adjacent to one another.

Only those sections of transform faults between two segments of constructive boundaries (e.g. the segment between X and Y in Figure 3.19) are seismically

active and therefore real plate boundaries. Transform faults continue to exhibit a topographic expression beyond the constructive plate boundaries, even though only a short length of a transform fault is active. This topographic expression is simply a result of the different ages of adjacent oceanic lithosphere: younger lithosphere rests at a higher elevation than older lithosphere – this situation is illustrated schematically in Figure 3.20.

Transform faults associated with subduction zones are much less common, and destructive plate boundaries do not, in general, show the segmented structure so common in constructive boundaries. An example occurs at the eastern end of the Cocos–Nazca Ridge, where a heavily faulted seismic zone delineates a transform fault (the Panama Fracture Zone), connecting a constructive boundary (the eastern end of the Costa Rica Rift, which is the easternmost part of the Cocos–Nazca Ridge) with the eastern end of a destructive boundary (the Middle America Trench). Similarly, the Scotia arc in the southern Atlantic is terminated in the north by a long transform fault along the North Scotia Ridge that marks the boundary between the South American Plate and the Scotia Plate.

Occasionally, conservative plate boundaries occur in continental plates. The most famous example is the San Andreas Fault of California, which marks a segment of the boundary between the North American and Pacific Plates. Here, Baja and southern California (including Los Angeles) are moving slowly northwards relative to the rest of California. This type of transform boundary produces shallow earthquakes and accompanying ground faulting. The friction between the two plates is often so great that the two sliding margins become ‘stuck’ together, allowing stresses to build up, which are then relieved by large earthquakes.

3.2.6 Triple junctions

All of the plate boundaries discussed so far have involved junctions between two plates. However, there are some localities where three plates are in contact, and these are termed **triple junctions**. Triple junctions between three ocean ridges, such as that in the South Atlantic between the African, South American and Antarctic Plates, are known as ridge–ridge–ridge, or RRR triple junctions. A similar notation can be used to identify triple junctions involving ocean trenches (T) or transform faults (F). For instance, a ridge–ridge–trench junction would be termed an RRT triple junction. The ordering of the letters is not significant. Considering all the geometric possibilities of fitting together three plate boundaries and their relative motions, there are actually only ten possible triple junctions. Some of these, such as RRR junctions, are termed **stable triple junctions**, which means they maintain their form over time. However, some can only exist briefly before they evolve into another plate configuration and these are termed **unstable triple junctions**. Figure 3.21 shows the evolution of three triple junctions over time.

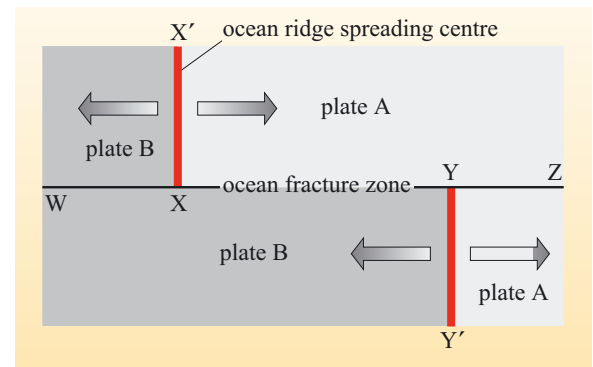


Figure 3.19 Diagram showing relative movements across an oceanic transform fault W, X, Y and Z off-setting a constructive plate boundary. The large arrows indicate the sense of plate motion away from the ridges XX' and YY'.

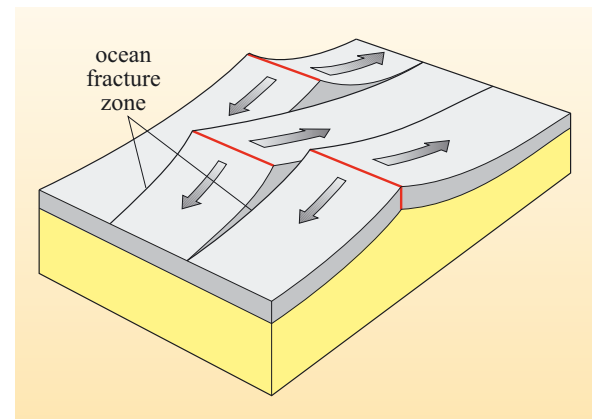
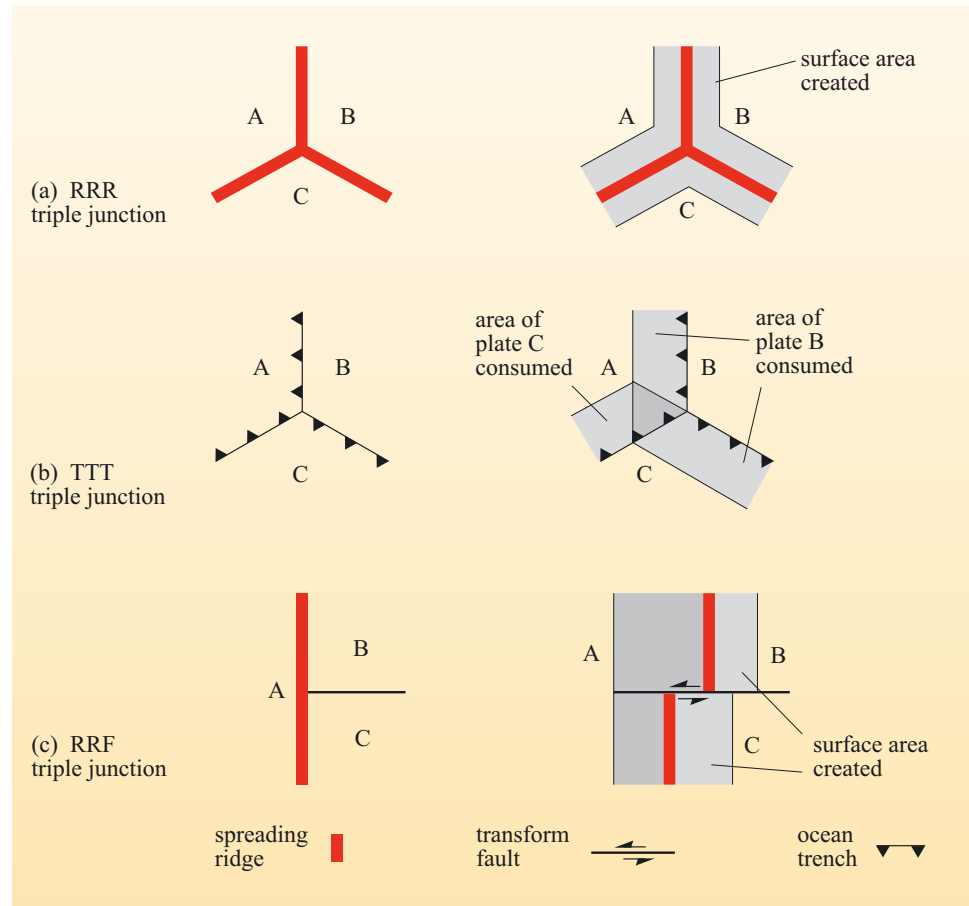


Figure 3.20 Block diagram showing how the topographic contrast across an ocean fracture zone (transform fault) develops as a consequence of lithospheric age as opposed to differential vertical movement.

Figure 3.21 The evolution of triple junctions with time. (a) A triple junction involving three ridges (RRR triple junction) is always stable, and the magnetic anomalies within the surface area created have Y-shaped patterns around the spreading ridges. (b) A triple junction between three trenches (TTT) is almost always *unstable* except in the special circumstance shown here when the relative motion of plates A and C is parallel to the plate boundary between B and C. (c) A triple junction between two ridges and a transform fault (RRF) can only exist for a short instant in geological time, and decays immediately to two FFR stable plate junctions.



The RRR junction shown in Figure 3.21a is always stable, regardless of the relative rates of spreading at each of the three ridges. The TTT junction in Figure 3.21b is basically unstable except if, by coincidence, the movement rates are the same and if the direction of subduction of plate C below plate A is exactly parallel to the boundary between plates B and C. The triple junction in Figure 3.21c is an RRF junction, and is unstable because there is relative motion between plate B and plate C. The RRF triple junction evolves immediately to form two FFR junctions. FFR and RRF junctions are always unstable.

Seven types of triple junction exist in the present plate tectonic configuration. These are:

- RRR (e.g. in the South Atlantic, the Indian Ocean and west of the Galapagos Islands in the Pacific)
- TTT (e.g. central Japan)
- TTF (e.g. off the coast of Chile)
- TTR (e.g. off Moresby Island, western North America)
- FFR, FFT (e.g. junction of the San Andreas Fault and the Mendocino Transform Fault off western USA)
- RTF (e.g. southern end of the Gulf of California).

3.3 Plate tectonic motion

Plates move relative to one another and relative to a fixed reference frame, such as the rotational axis of the Earth. Plates also move across the curved surface of the Earth and so should not be considered as flat sheets on a flat surface but as caps moving across the roughly spherical surface of the Earth. Consequently, plate motion is not as simple as it might at first appear. This section begins with a consideration of relative plate motions and how they can be measured before moving on to the more complex assessment of plate motions across a sphere and how true plate motions may be measured.

3.3.1 Relative plate motions

In previous sections you have already tackled the problem of assessing the full and half spreading rates of ocean ridges.

- Recall two different methods of determining spreading rates.
- The use of dated magnetic anomalies, and the known relationship between ocean depth and age (Sections 3.1 and 3.2).

A simple calculation dividing the distance from the ridge by the age gives the plate speed and, combined with the direction of travel, its velocity.

Question 3.6

Figure 3.22 shows a section through the Earth from the Atlantic to the Indian Ocean, cutting across three different plates and two constructive plate boundaries. The half spreading rates are shown for each plate at each plate boundary. For the situations in Table 3.1, estimate the relative rates and relative directions of motion of the African Plate.

Figure 3.22 Section through the Earth from the Atlantic Ocean to the Indian Ocean showing the relative spreading rates across the Mid-Atlantic and Carlsberg Ridges.

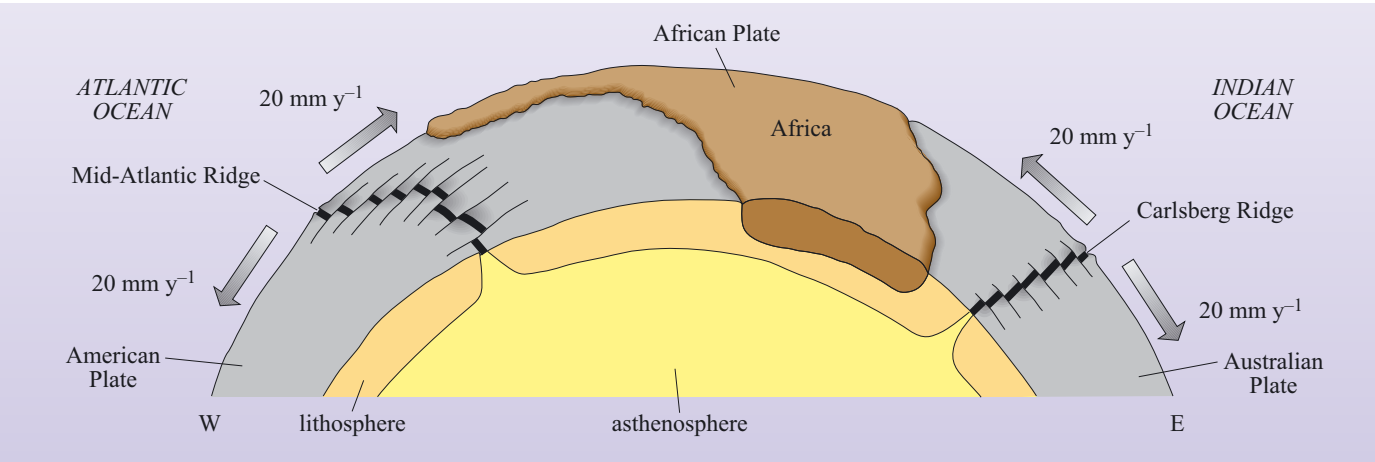


Table 3.1 For use with Question 3.6.

	Relative plate motion/mm y ⁻¹	Direction of plate motion
American Plate fixed		
Australian Plate fixed		

The answer to this question reveals two important points about plate motion:

- measured plate velocities (speed and direction) must be stated relative to one another
- plates and their boundaries cannot be fixed in relation to the mantle.

The second point derives from a consideration of the African Plate – as it grows in size, at least one of its constructive boundaries must have moved. So how can we determine the true movement of a plate against a truly fixed frame of reference? One possibility would be to assume that one plate is stationary and to determine plate movement relative to that ‘fixed’ reference (as in Question 3.6). However, polar wander curves for all of the continents show that all plates bearing continents have moved relative to the Earth’s magnetic pole over periods of tens of millions of years. Another way would be to determine plate motion relative to surface features that might be more firmly rooted in the deeper mantle and for this we turn to volcanism that is not dependent on plate boundary interactions – the so-called within-plate or **hot-spot volcanism** on ocean islands.

3.3.2 Hot-spot trails and true plate motions

In addition to volcanism associated with constructive and destructive plate boundaries there is a third important component to global volcanism. This occurs in the interior of plates and is associated with broad surface up-doming, which is often 1000 km across and hundreds of metres in elevation. Gravity anomalies across these domes show that they are not in isostatic equilibrium, but are supported from sub-lithospheric depths, presumably by upwelling mantle. Perhaps the best-known example is located beneath the active volcanoes of Hawaii (Box 3.4), whose long history of volcanism has been related to a structure in the deep mantle known as a **mantle plume**. Mantle plumes are an important feature of mantle convection and will be discussed in more detail in Chapter 7, but for now it is sufficient to know that they produce surface volcanism that is not necessarily associated with plate boundaries. There are numerous plumes of different sizes recognised around the globe. Some are associated with chains of islands and seamounts, whereas others have produced long ridges in the ocean floor. A good example is the Ninetyeast Ridge in the Indian Ocean. Termed **aseismic ridges** because of a lack of seismicity along their lengths, these ridges are very different structures from the ocean ridges associated with constructive plate boundaries.

Hawaii is part of an extensive chain of islands and submarine volcanic peaks (called seamounts) stretching almost 6000 km across the floor of the Pacific Ocean. The chain forms an ‘L’-shaped chain of volcanic islands and seamounts across the sea floor that increase in age northwards from Hawaii (Figure 3.23).

Hawaii rises from the sea floor some 6 km below the Pacific Ocean to a summit elevation of about 4 km above sea level, making it taller than Mt Everest. Although all of the Hawaiian Islands are volcanic in origin, only Hawaii is currently active and still growing

in size. The islands are situated within the Pacific Plate some 4000 km from the nearest plate boundary. The magma that has caused the volcanism is the result of a plume of anomalously hot material rising through the mantle. Where this plume impinges upon the base of the lithosphere, magma finds its way to the surface to produce a so-called **hot spot**. Other island and seamount chains located on the Pacific Plate (Figure 3.23) show a similar pattern of age progression, and are related to different hot spots. In fact, hot-spot volcanoes may be found dotted around the world and most of them are remote from plate boundaries.



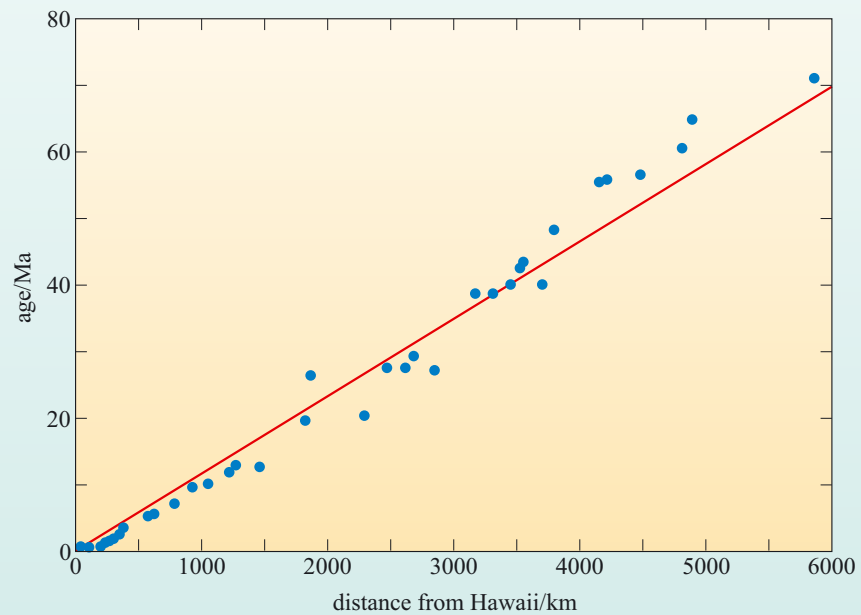
Figure 3.23 Map showing three of the major seamount and island chains in the Pacific Ocean. These are the Kodiak Island–Cobb Seamount Chain in the northeastern Pacific, the Marshall–Ellice Islands–Austral Seamount Chain of the southern Pacific, and the well-known Hawaiian Ridge–Emperor Seamount Chain in the central Pacific. The Hawaiian Islands are located at the southeastern end of the Hawaiian Ridge.

The ages of the islands and seamounts are proportional to their distance away from the currently active site of Hawaii, as shown in the graph in Figure 3.24. The best-fit line through the data points indicates that the site of volcanic activity has apparently migrated at a constant speed along the chain. Each island or seamount has been constructed as the Pacific Plate has moved over the stationary hot spot. If it is assumed that the Hawaiian hot spot has remained stationary with respect to the Earth's axis, then the rate of migration of volcanism along the chain gives the rate of plate movement across the Hawaiian hot spot.

- The Hawaii–Emperor chain of islands and seamounts is not straight, but kinked. What do you think the kink represents?
- If the hot spot is stationary then this must represent a change in direction of the Pacific Plate.

Careful interpretation of the age progression along the Hawaii–Emperor chain suggests that this change in direction occurred 43–50 Ma ago, which ties in with a series of tectonic adjustments around the world, including the start of continental collision between India and Asia.

Figure 3.24 Graph of age versus distance from Hawaii measured along the Hawaii–Emperor chain of islands and seamounts. A best-fit line has been drawn through the data.



Question 3.7

Use Figure 3.24 to estimate the average rate at which volcanic activity has appeared to move along the Hawaiian–Emperor Seamount Chain. Express your answer in mm y^{-1} to two significant figures. With reference to Figure 3.23, determine in which direction the active volcanism has moved, and from this, the direction the plate has moved. *Note:* the speed at which volcanic activity moves is the inverse of the gradient of the best-fit straight line.

By assuming mantle plumes and hot spots are stationary, the motions derived from age progressions such as those along the Hawaii–Emperor Seamount Chain represent **true plate motions**. However, if the mantle plumes underlying hot spots have also moved with respect to the Earth’s axis, but at a rate different from that of plate movement, then the motions defined by a hot-spot trace (see Figure 3.23) may be misleading.

- How do you think plate motions could be verified?
- By measuring true plate motion from another hot-spot trace.

For the Pacific Plate, the Hawaiian hot spot may be the largest hot spot, but it is not the only example of a hot-spot trace on the ocean floor. Two other examples related to the mantle plumes currently beneath the Cobb and the Macdonald Seamounts are shown in Figure 3.23. You should be able to appreciate from this figure that both of these chains of seamounts and islands are broadly parallel with the Hawaiian–Emperor Seamount Chain, and the Austral–Marshall Islands Seamount Chain shows a similar bend. Detailed geochronology of these chains reveals that the age progression along them is also consistent with the rates of

northwesterly plate motion derived from the Hawaiian–Emperor Seamount Chain. Since the chances of three (and more) hot-spot traces moving independently of the plates and one another giving such similar results is remote, these results strongly indicate that hot spots do provide a reference frame within which true plate motions can be measured.

This methodology has been extended to studies of numerous hot-spot trails on other plates. Combined with a knowledge of relative plate motions between adjacent plates, such measurements have allowed the development of a framework of true plate motions across the globe, which are summarised in Figure 3.25.

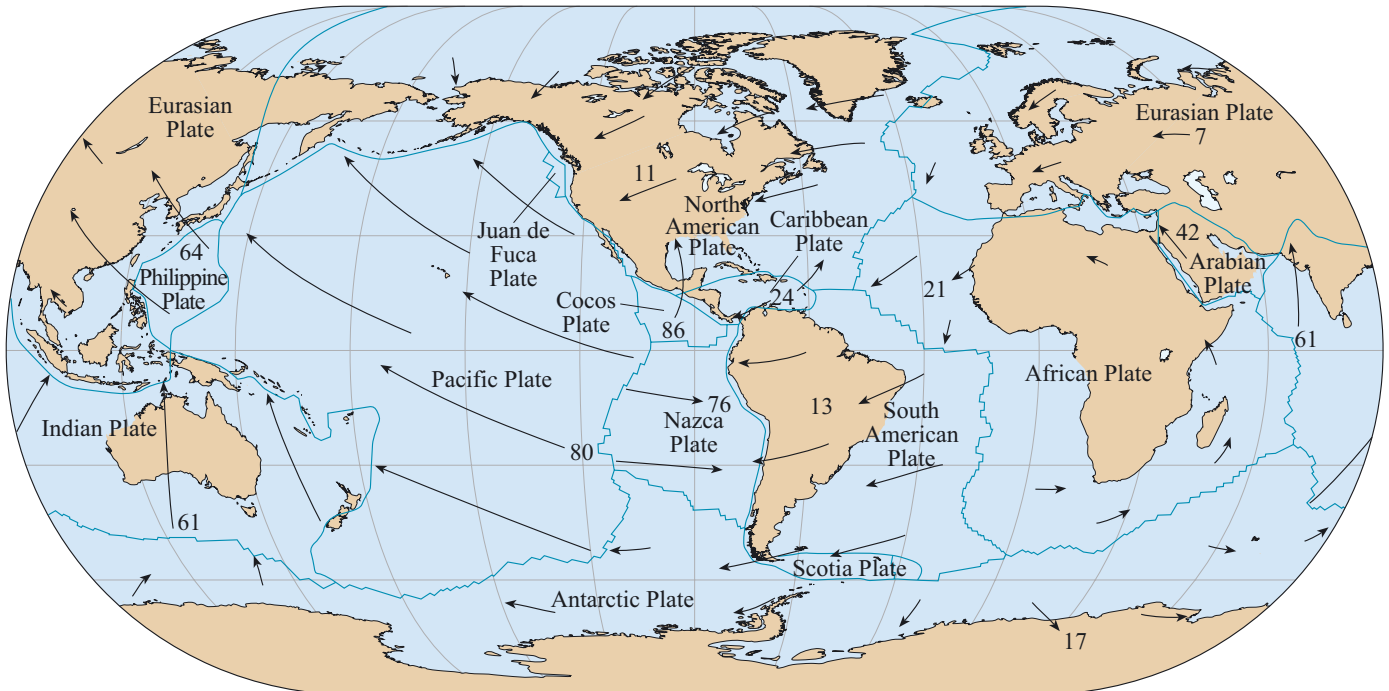


Figure 3.25 Map showing present-day motion of lithospheric plates (indicated by arrows) relative to hot spots (i.e. true plate motion). The lengths of the arrows indicate the amount of movement that would occur over a period of 50 Ma and the figures represent the current mean true plate speed.

3.3.3 Plate motion on a spherical Earth

Earth's tectonic plates are continuously in motion with respect to each other, and together they form the closed surface of a sphere (i.e. the Earth's surface). Understanding the movement of plates, therefore, requires a geometrical analysis of motions over a spherical surface. This is described in the Euler (pronounced 'oiler') geometrical theorem, which shows that every displacement of a plate from one position to another on the Earth's surface can be regarded as a simple rotation of that plate about a suitably chosen axis, known as an Euler pole or **pole of rotation**, which passes through the centre of the Earth.

Plate movement on a spherical Earth is illustrated in Figure 3.26. Two plates, A and B, are separating at a constructive plate boundary and rotating about a pole of rotation (which in this case is the North Pole). The constructive plate boundaries lie along lines of longitude, whereas the transform faults are parallel

to lines of latitude. Lines of longitude are great circles with centres that coincide with the centre of the Earth; lines of latitude, which are at right angles to lines of longitude, are small circles with centres displaced from the Earth's centre. When the plates separate, the gap between them increases but, quite obviously, they separate a smaller amount close to the North Pole than at the Equator.

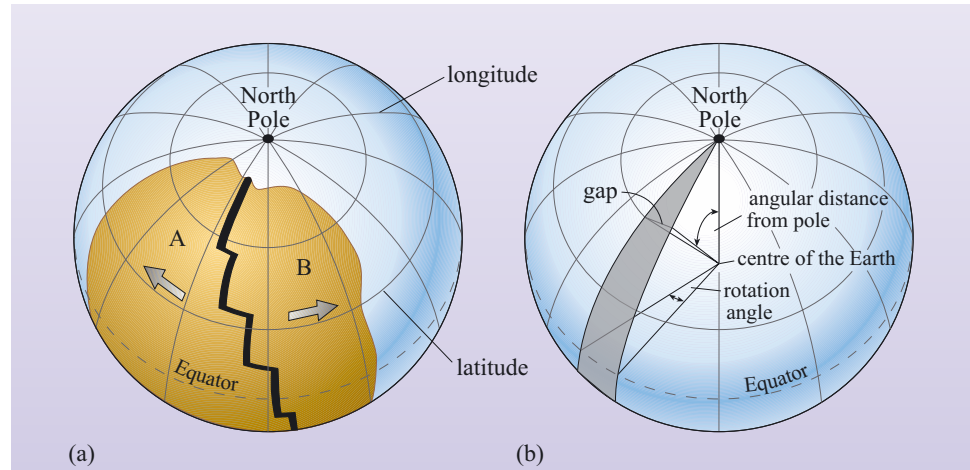


Figure 3.26 (a) The geometry of a constructive boundary in which two plates are separating. (b) The geometrical relationships that may be used to describe plate movement. Note that whilst the rotation angle remains the same along the length of the split, the width of the gap increases with angular distance from the pole of rotation (in this case the North Pole).

- What effect does this have on local full and half spreading rates along the length of the constructive plate boundary?
- Spreading rates measured close to the pole of rotation will be small while those further away will be much greater.

This aspect of spherical geometry explains why relative plate motions expressed simply in terms of mm y^{-1} vary along the length of a plate margin. Needless to say, poles of rotation are not all located at the geographical poles but can fall anywhere on the Earth's surface. For example, the Mid-Atlantic Ridge has variable spreading rates along its length, with values of $37\text{--}40 \text{ mm y}^{-1}$ in the South Atlantic, to lower values of $23\text{--}26 \text{ mm y}^{-1}$ in the North Atlantic (Figure 3.8).

- What do these variations in spreading rate tell you about the location of the pole of rotation of the South American Plate?
- Spreading rates are lower in the North Atlantic and so the pole of rotation must be located somewhere in the northernmost Atlantic Ocean.

The location of the pole of rotation can be determined more accurately from the orientation of the transform faults that cut the Mid-Atlantic Ridge. Transform faults lie along small circles and, by definition, lines at right angles to them will pass through the pole of rotation (Figure 3.27a). Using this method, the pole of rotation of the South Atlantic Plate can be shown to be located in the North Atlantic, south of Greenland (Figure 3.27b).

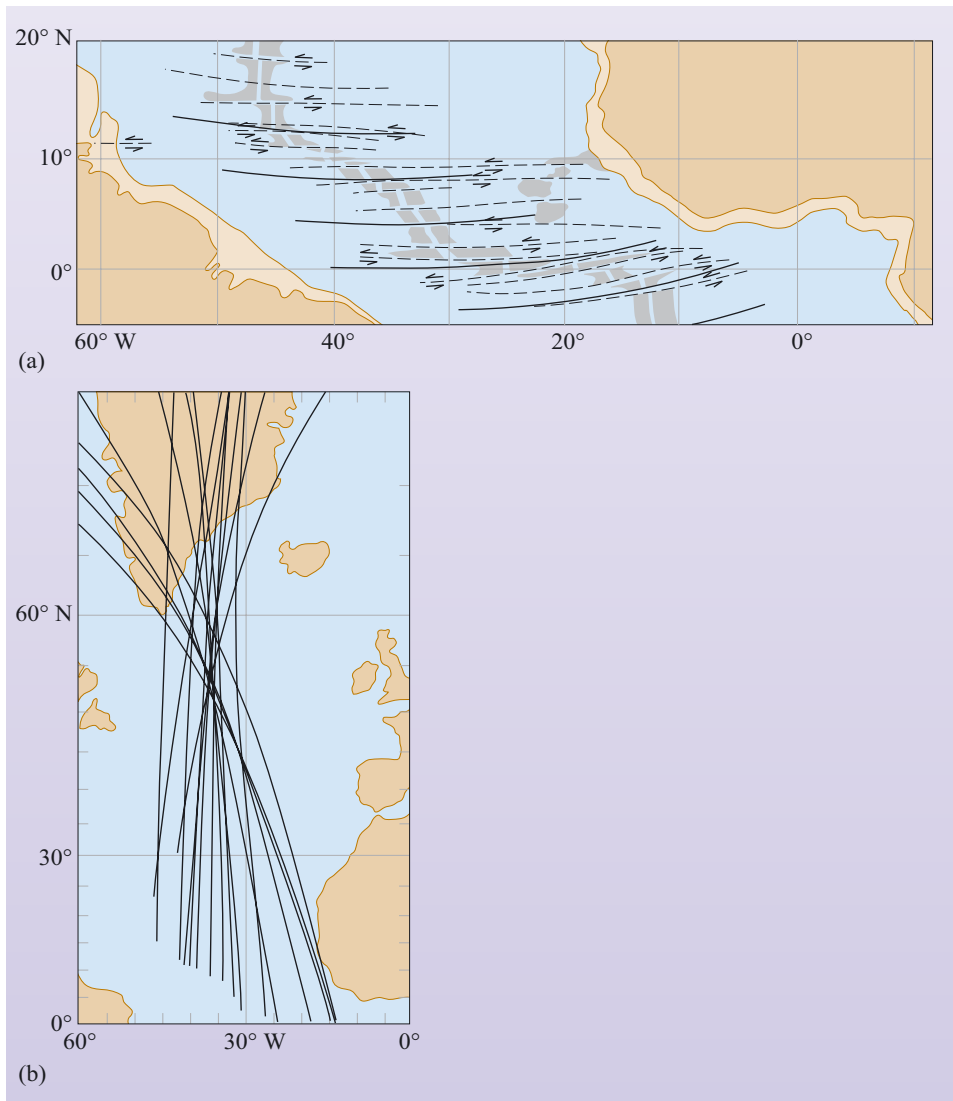


Figure 3.27 (a) Transform faults (dashed lines) in the equatorial region of the Atlantic Ocean compared with small circles (solid lines) concentric about a rotation pole at 58° N , 36° W . (b) Great circles drawn perpendicular to the transform faults that offset the Atlantic spreading ridge. The great circles converge on an area in the North Atlantic, defining the general position of the pole of rotation of the plate.

Finally, some plates have an internal pole of rotation that results in some complex consequences for relative plate motions. The first, and possibly most obvious, is that the real motion of the plate becomes one of rotation. The African Plate (Figure 3.25) is a good example. The southern part of the African Plate is moving to the east while the northern part is moving to the west – the plate is rotating anticlockwise about a pole located somewhere near the Canary Islands, off the northwest coast of Africa. The second consequence is that an internal pole can lead to a given plate boundary transforming from a constructive plate boundary through a conservative plate boundary to a destructive plate boundary along its length, although this is not apparent around the African Plate.

3.4 Plate driving forces

One of the key questions associated with plate tectonics is why plates move and what drives them. Plate tectonics is an expression of the thermal state of the Earth's interior and is the way that the Earth is currently losing a large proportion of its internal heat. Hot lithosphere generated at constructive plate boundaries loses its heat to the oceans, atmosphere and, ultimately, space by conductive cooling as it ages and spreads away from the ridge before being recycled into the mantle by subduction. It is perhaps, tempting to relate this cycling of material to a convective cycle, with hot material upwelling beneath ocean ridges and cold material sinking in subduction zones.

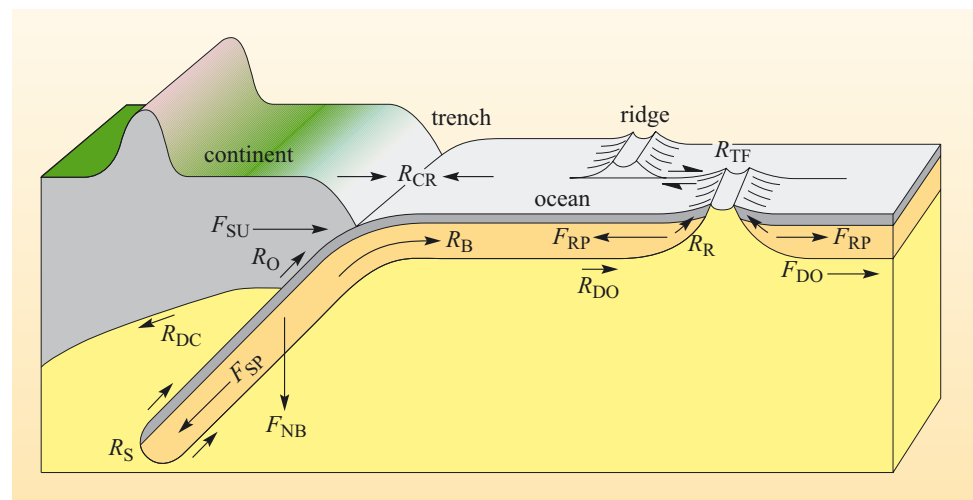
- Can you think of a reason why this might not be the case?
- Plate boundaries migrate over time and it is unlikely that convection cells within the mantle would migrate with them.

The migration of plate boundaries across the surface of the Earth means that they are not firmly fixed into the underlying convective motions of the mantle – from our analysis of real plate motions mantle convection is more likely to be related to the location of hot spots and mantle plumes, a subject to which you will return in Chapter 7. So what, then, drives the plates? An answer to this question may lie in an analysis of the forces acting on plates, both on their undersides and at their boundaries.

3.4.1 Forces acting upon lithospheric plates

Figure 3.28 provides a very simplified overview of the forces that are thought to affect the movement of lithospheric plates. The relative contribution of these forces to plate motion needs to be established before their roles can be explored.

Figure 3.28 Forces acting upon a plate. F denotes a driving force, whereas R denotes a retarding force. See text for explanation. (Bott, 1982)



Forces acting on the underside of lithospheric plates

Lithospheric plates are decoupled from the rest of the mantle because the underlying asthenosphere is weak. However, plates may be driven, at least in part, by forces imparted by the convecting mantle. For instance, if a lithospheric plate is being carried along by a faster moving asthenosphere, then the force acting along

the bottom surface of the oceanic plate can be considered as an **ocean driving force**, F_{DO} , which helps the plate to move. By contrast, if the asthenosphere is moving slower than the plate in the direction of plate movement or even in the opposite direction, then the force acting along the bottom surface of the oceanic plate can be considered as an **ocean drag force**, R_{DO} , retarding the movement of the plate.

Continental lithosphere is thicker than oceanic lithosphere, so continents almost always have a 'keel' of lithospheric material projecting downward. As a result, the resistance to movement might be greater beneath continental plates than oceanic plates. Accordingly, continental plates might be associated with an additional **continental drag force**, R_{DC} , and so the resistive force acting on the base of a continental plate would be the sum of both oceanic and continental drag forces, $R_{DO} + R_{DC}$.

Forces acting at plate margins

Other forces that act on plates must be generated at their boundaries. These forces push from the ridge, drag the plates down at the trenches, or act along the sides of plates at conservative boundaries (Figure 3.28).

At constructive boundaries, the upwelling of hot material at ocean ridges generates a buoyant effect that produces the ocean ridge, which stand some 2–3 km above the surrounding ocean floor. Here, oceanic plates experience a force that acts away from the ridge, the so called **ridge-push force** (F_{RP}), which is a result of gravity acting down the slope of the ridge. The occurrence of shallow earthquakes, resulting from the repeated tearing apart of newly formed oceanic crust, indicates there is also some frictional resistance to this force at ridges. This can be called **ridge resistance**, R_R . Bounding the ridge segments, the oceanic transform faults, where the plate segments slide past each other, encounter resistance to movement, and produce a series of earthquakes: this retarding force is the **transform fault resistance**, R_{TF} .

The situation at destructive plate boundaries is more complex. A major component is the downward gravitational force acting on the cold and dense descending slab as it sinks into the mantle. This gravity-generated force pulls the whole oceanic plate down as a result of the negative buoyancy of the slab: this is the **negative buoyancy force**, F_{NB} . The component of this downward-acting force that is transmitted to the plate is the **slab-pull force**, F_{SP} . The magnitude of the slab pull is related to the angle at which the plate descends, and is greater for steeply dipping plates. However, the sinking slab encounters resistance as it descends, both from the frictional drag on its upper and lower surfaces and from the viscosity of mantle material that is being displaced: this combined resistive force is termed **slab resistance**, R_S . A further complication is that the downwards moving plate must flex at the trench before it begins to slide beneath the opposing plate. This provides a further resistance to the plate motion; this is labelled **bending resistance**, R_B . Due to the pushing of the subducting slab against the overriding plate, there is in addition, frictional resistance that gives rise to shallow and deep earthquakes at subduction zones. These frictional forces can be labelled collectively as **overriding plate resistance**, R_O .

For the overriding plate, another theoretical force analogous to the ocean driving force has been proposed, which is derived from convection induced in the mantle

above the subducted plate. Cooling of the mantle wedge against the upper surface of the subducting plate induces convection that sucks more mantle into the wedge. This is the **trench suction force**, F_{SU} , which serves to pull the plate towards the trench. The collision of plates and the associated deformation processes generate a **collisional resistance force**, R_{CR} . This acts in the opposite direction within each converging plate, but it is equal in magnitude in both.

The velocities of present-day plate motions appear to be constant, indicating a state of dynamic equilibrium where a balance exists between the driving and resistive forces. However, each plate moves at its own rate – which suggests that the relative importance of the driving and retarding forces must vary from plate to plate. It seems unlikely that any single force is the sole driving mechanism of plate motions. For example, if the ridge-push force is the only driving force, why does the Philippine Plate, with no ridge on its boundary, move at a similar rate (70 mm y^{-1}) to the Indian Plate, which is bounded by both the Carlsberg and South Indian Ocean spreading ridges? Plate motions must, therefore, be controlled by a combination of forces.

Question 3.8

For each of the forces listed in Table 3.2, indicate whether they are likely to act as a driving force or a resistive force, and place a tick in the appropriate space. Note that some forces can act as either a driving force or resistive force.

Table 3.2 For use with Question 3.8.

Force	Acts as a driving force	Acts as a resistive force	Might act as <i>either</i> a driving force <i>or</i> a resistive force
oceanic drag			
continental drag			
ridge-push			
transform fault			
slab-pull			
slab resistance			
trench suction			

To investigate which are the most important forces that act on plates, the reasoning process applied in the previous paragraph can be adopted and the plate speed compared quantitatively with factors that relate to the different driving forces. For example, if the dominant driving force is ridge-push (F_{RP}), then the fastest moving plates should be the oceanic plates with the highest ratio of ridge length to surface area. The forces acting at plate boundaries should have magnitudes that are proportional to the length of ocean ridge (in the case of ridge-push), the length of ocean trench (for slab-pull (F_{SP}), trench suction (F_{SU}) and inter-plate resistances) and the length of transform fault (for transform fault resistance (R_{TF})). Oceanic and continental drag act over the lower surface of the plate, and so

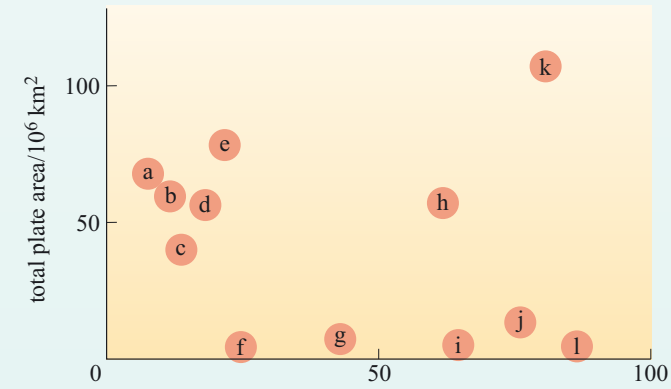
they should be proportional to the area of the plate. The rates of true plate motion in relation to the relative lengths of ocean ridge, ocean trench, transform fault, and plate area must, therefore, be examined to estimate the effects of each of the plate driving forces. The motions of the twelve plates listed in Table 3.3 are shown graphically in Figure 3.29a–e. Some of the important physical properties of each plate are listed in the table, together with the average true velocity of each plate calculated relative to a hot-spot reference frame. To explore whether or not each of these physical attributes of plate configuration are significant in producing plate motion, the degree and character of data correlation need to be examined.

Table 3.3 Dimensions and true velocities of lithospheric plates.

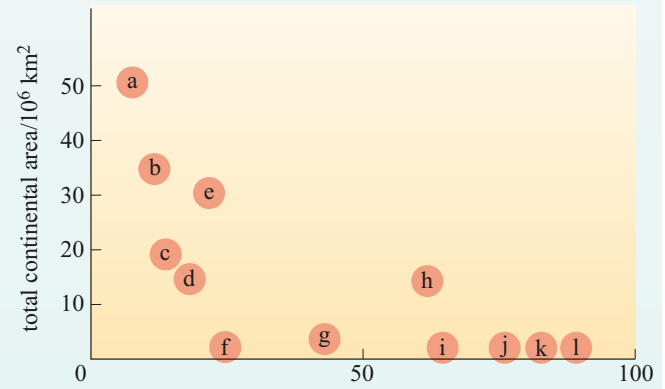
Plate	Total plate area $\times 10^6$ /km ²	Total continental area $\times 10^6$ /km ²	Average true velocity /mm y ⁻¹	Circumference $\times 10^2$ /km	Length (effective length)* $\times 10^2$ /km		
					Ocean ridge	Ocean trench	Transform fault
(a) Eurasian	69	51	7	421	90 (35)	0	56
(b) N. American	60	36	11	388	146 (86)	12 (10)	122
(c) S. American	41	20	13	305	87 (71)	5 (3)	107
(d) Antarctic	59	15	17	356	208 (17)	0	131
(e) African	79	31	21	418	230 (58)	10 (9)	119
(f) Caribbean	4	0	24	88	0	0	44
(g) Arabian	5	4	42	98	30 (27)	0	36
(h) Indian	60	15	61	420	124 (108)	91 (83)	125
(i) Philippine	5	0	64	103	0	41 (30)	32
(j) Nazca	15	0	76	187	76 (54)	53 (52)	48
(k) Pacific	108	0	80	499	152 (119)	124 (113)	180
(l) Cocos	3	0	86	88	40 (29)	25 (25)	16

*Effective lengths (in brackets) are the lengths of plate boundary that are capable of exerting a net driving or resistive force.

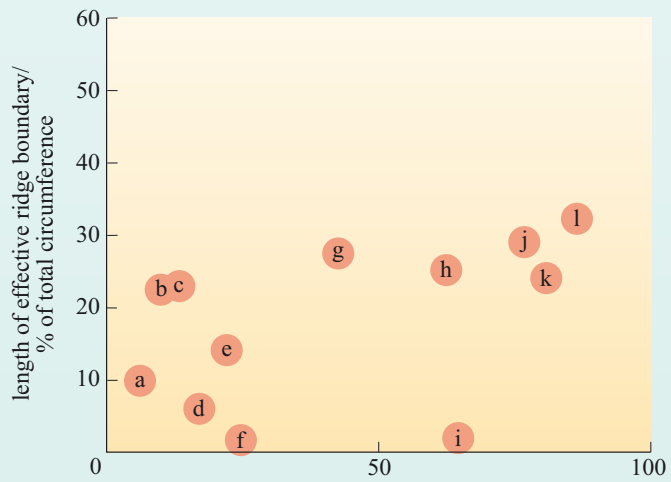
Figure 3.29a is a plot of total plate area against average true plate velocity. The graph reveals that whilst it is true that the plate with the largest area (k, the Pacific Plate) has one of the highest velocities, it does not always follow that large plates have high velocities – the plate with the second largest area (e, the African Plate) has a relatively low velocity. Moreover, the plate with the highest velocity (l, the Cocos Plate) is one of the smallest. Whilst it would be possible to draw a best-fit line through some of the data points (for instance, a, b, d, e and k) that showed reasonable positive correlation (i.e. that larger plates have faster velocities and smaller plates have slower velocities), there is little obvious pattern if the data are taken as a whole. Accordingly, it is reasonable to conclude that there is no real correlation between these two variables of plate area and true plate velocity. This logic can be extended to examine the role of oceanic drag: if this were a significant motive force then these same data would show a positive correlation; if it were a significant plate-retarding force then the data would show a negative correlation. Since neither is true, the role of oceanic drag appears to have an insignificant effect on plate motion.



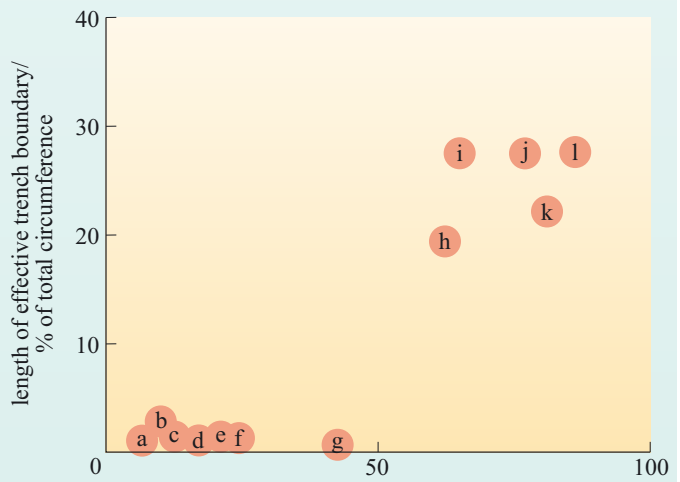
(a)



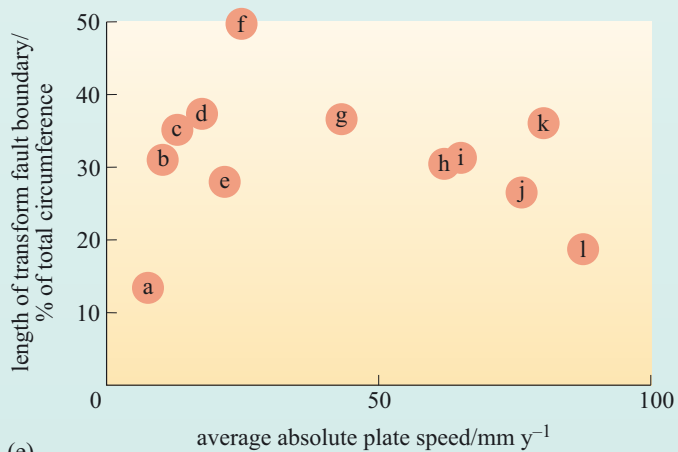
(b)



(c)



(d)



(e)

- (a) Eurasian
- (b) N. American
- (c) S. American
- (d) Antarctic
- (e) African
- (f) Caribbean
- (g) Arabian
- (h) Indian
- (i) Philippine
- (j) Nazca
- (k) Pacific
- (l) Cocos

◀ **Figure 3.29** (a) Plot of total plate area against average true plate speed. The letters for the plates are as in Table 3.3. (b) Plot of total continental area of each plate against average true plate speed. (c) Plot of length of effective ocean ridge boundary (expressed as a percentage of the total circumference of the plate) against true plate speed. (d) Plot of length of effective ocean trench boundary (expressed as a percentage of the total circumference of the plate) against true plate speed. *Note:* plates a–g have relatively little or no boundary along which the plate is being subducted. (e) Plot of total length of transform fault boundary for each plate (expressed as a percentage of the total circumference of the plate) against true plate speed.

Figure 3.29b is a graph showing the continental area of each plate compared with the average true velocity of that plate. If the continental ‘keel’ were acting to retard plate motion, those plates with the largest proportion of continental lithosphere might be expected to display the slowest velocities. Interestingly, the plate with the largest *continental* area (a, the Eurasian Plate) does have the lowest velocity, while the plate with the lowest continental area (l, the Cocos Plate) has the highest velocity; this is the negative correlation predicted above. Clearly, the data points in Figure 3.29b do not fall on a line, but the general trend is for low speed to be associated with a large continental area. Accordingly, these data appear to indicate that continental drag is an effective resistance to plate movement.

Question 3.9

Examine Figure 3.29 and, by ticking the appropriate columns in Table 3.4, indicate the type of correlation shown by the *y*-axis variables listed in the table with average true plate velocity (*x*-axis).

From the best positive and negative correlations that you identified, what do you conclude about the driving mechanism of plate motions?

Table 3.4 For use with Question 3.13.

	Good positive correlation	Poor correlation	Good negative correlation
Total plate area			
Total continental area			
Effective length of ridge boundary			
Percentage circumference of plate attached to descending slab			
Percentage circumference of plate that is a transform fault			

Figure 3.29 shows that the most striking positive data correlation is between true plate velocity and the length of ocean trench boundary, which includes a subducted slab (Figure 3.29d). In fact, the plates form two distinct data groups:

- the Indian, Philippine, Nazca, Pacific, and Cocos Plates (h–k), which are all connected to descending slabs and are all moving at an absolute rate of 60–90 mm y^{-1} relative to the asthenosphere

- the remaining plates (i.e. Eurasian, North American, South American, Antarctic, Caribbean and Indian Plates (a–g)) are moving much more slowly at 0–40 mm y^{-1} and do not have significant connections with descending slabs.

It is reasonable, therefore, to conclude that slab-pull is the dominant plate-driving force. By contrast, there is a poorer correlation between ridge length and plate velocity (Figure 3.29c), from which it can be concluded that the so-called ‘ridge-push’ may also contribute to plate speed.

The best correlations from the data shown in Figure 3.29, and summarised in Table 3.7 (in the answer to Question 3.9), suggest that true plate velocity depends largely on the slab-pull force associated with the descent of oceanic lithosphere, and to a lesser extent on the ridge-push force; continental plate drag is a major retarding force.

Finally, it should be noted that slab-pull and ridge-push forces are largely a consequence of density differences and gravity—slab pull relates to the gravitational sinking of a cold slab of lithospheric material into the mantle whereas ridge-push relates to the gravitational potential energy of the ocean ridge standing 2–3 km above the surrounding ocean floor. So the answer to the question, ‘What drives plate tectonics?’ is not directly related to mantle convection, but to gravity acting on density differences in the lithosphere that have resulted from its own thermal history.

3.5 Implications of plate tectonics

3.5.1 The Wilson cycle

High-quality, palaeomagnetic data are now sufficiently abundant that it is possible to reconstruct the movement of the continents throughout the past 500–600 million years (i.e. the Phanerozoic) and, with increasing uncertainty, back to 750 Ma and possibly earlier. From these reconstructions it became apparent that the continental masses have been assembled previously into supercontinents that have broken apart, dispersed, and then later reassembled in a different configuration to form another supercontinent. This observation was noted by Wilson (Box 3.1) who proposed that an ocean basin has a lifespan with several stages: it begins with the initial opening, and then goes through a widening phase before starting to close and on to its ultimate destruction. This theory accounts for the cycle of continental break up and reassembly, and became known as the **Wilson cycle** in his honour. From the palaeomagnetic reconstructions, it appears that the cycle of supercontinent assembly – break-up and subsequent reassembly – takes about 500 million years to complete. This time period can be further explained by a simple calculation.

Question 3.10

Imagine that a roughly circular supercontinent, 5000 km in radius, and located about the Equator, rifts in two along a north–south line. A new spreading centre between the rifted halves spreads at an average rate of about 3 cm y^{-1} . How long would it take for the two halves to first meet again on the opposite side of the globe? (Assume that the circumference of the Earth is 40 000 km.)

Clearly, this is an average estimate because spreading rates vary, and continental configurations are far more complex than the simple two-continent rifting model outlined in Question 3.10. But if it is correct, then, given that Pangaea formed about 300 Ma ago, the next supercontinent is due to begin to assemble in about 200 million years, perhaps once the Pacific Ocean has been closed by the subduction zones that surround it.

Various stages have been identified for the Wilson cycle, and all of these stages can be recognised in different parts of the Earth today.

- 1 The earliest stage, called the *embryonic stage*, involves uplift and crustal extension of continental areas with the formation of rift valleys (e.g. the East African Rift System).
- 2 The *young stage* involves the evolution of rift valleys into spreading centres with thin strips of ocean crust between the rifted continental segments. The result is a narrow, parallel-sided sea, for example the Red Sea that is opening between NE Africa and Arabia.
- 3 The *mature stage* is exemplified by widening of the growing basin and its continued development into a major ocean flanked by continental shelves and with the continual production of new, hot, oceanic crust along the ridge system (e.g. Atlantic Ocean).
- 4 Eventually, this expanding system becomes unstable and, away from the ridge, the oldest oceanic lithosphere sinks back into the asthenosphere, forming an oceanic trench subduction system with a Wadati–Benioff zone demarking the descending plate and associated island arcs, such as the situation in the western Pacific Ocean, or Andean-type volcanism. The onset of subduction at the ocean boundary marks the *subduction stage* of the cycle (e.g. the Pacific Ocean).
- 5 Once subduction outpaces the formation of new crust at the constructive boundary, the ocean begins to contract. Island arc complexes, complete with their inventory of sedimentary and volcanic rocks, collide and create young mountain ranges around the periphery of the contracting ocean. These features mark the *terminal stage* of the cycle (e.g. the Mediterranean).
- 6 The *end stage* occurs once all the oceanic crust between the continental masses has subducted, and the continents converge along a collision zone characterised by an active fold mountain belt, such as the Himalayas. Finally the plate boundary becomes inactive, but the site of the join, or suture, between the two continental masses is a zone of weakness in the lithosphere that has the potential to become the site of a new rift and so the cycle continues.

3.5.2 Plate tectonics and climate change

This chapter began by considering the evidence in the Earth's past for the existence of supercontinents and how evidence of past climates recorded in continental rocks can be used to reassemble ancient continental configurations. The evidence was interpreted in such a way that the continents were considered as passive recorders of the surface conditions that they have experienced on

their inexorable passage across the Earth's surface. While such an assumption is broadly correct, it does not take more than a momentary glance at a map of the world today to realise that the disposition of the continents has a marked effect on both local and global climate. Not the least of these effects results from the difference in the thermal properties of land versus ocean – a continental region will be colder in winter and warmer in summer than an oceanic region at any given latitude. Moreover mountain belts formed as a consequence of plate tectonic activity dramatically modify rainfall through the effects of **orography** – the development of a rain shadow on the leeward side of mountain belts.

Global climate is also strongly controlled by ocean currents. For example, northwestern Europe is significantly warmer than other regions at similar latitudes because of the warming effects of the Gulf Stream and North Atlantic Drift. The reversal of oceanic currents in the equatorial Pacific – a phenomenon known as El Niño – has a far-reaching effect on climate around the Pacific. Ocean currents depend on the geometry of the oceans and this is controlled by plate tectonics. Hence, over geological timescales the movement of plates and continents has a profound effect on the distribution of land masses, mountain ranges and the connectivity of the oceans. As a consequence, plate tectonics has a very direct and fundamental influence on global climate.

To illustrate this effect, Box 3.5 briefly describes the opening of a seaway between the southern tip of South America and Antarctica, and how that affected global climate.

The climate of modern Antarctica is extreme. Located over the South Pole and in total darkness for six months of the year, the continent is covered by glacial ice to depths in excess of 3 km in places. Yet this has not always been the case. 50 Ma ago, even though Antarctica was in more or less the same position over the pole, the climate was much more temperate – there were no glaciers and the continent was covered with lush vegetation and forests. So how did this extreme change come about?

The modern climate of Antarctica depends upon its complete isolation from the rest of the planet as a consequence of the Antarctic Circumpolar Current that completely encircles Antarctica and gives rise to the stormy region of the Southern Ocean known as the roaring forties. The onset of this current is related to the opening of seaways between obstructing continents. Antarctica and South America were once joined together as part of Gondwana and were the last parts of this original supercontinent to separate. By reconstructing continental positions

from magnetic and other features of the sea floor in this region, geologists have shown that the Drake Passage opened in three phases between 50 Ma and 20 Ma, as illustrated in Figure 3.30. At 50 Ma there was possibly a shallow seaway between Antarctica and South America, but both continents were moving together. At 34 Ma the seaway was still narrow, but differential movement between the Antarctic and South American Plates created a deeper channel between the two continents that began to allow deep ocean water to circulate around the continent. Finally, at 20 Ma there was a major shift in local plate boundaries that allowed the rapid development of a deep-water channel between the two continental masses.

- What other major change in global plate motions occurred between 43 Ma and 50 Ma?
- The change of orientation of the Hawaiian hot-spot trace shows that at this time the Pacific Plate changed from a northward velocity direction to a northwestward direction.

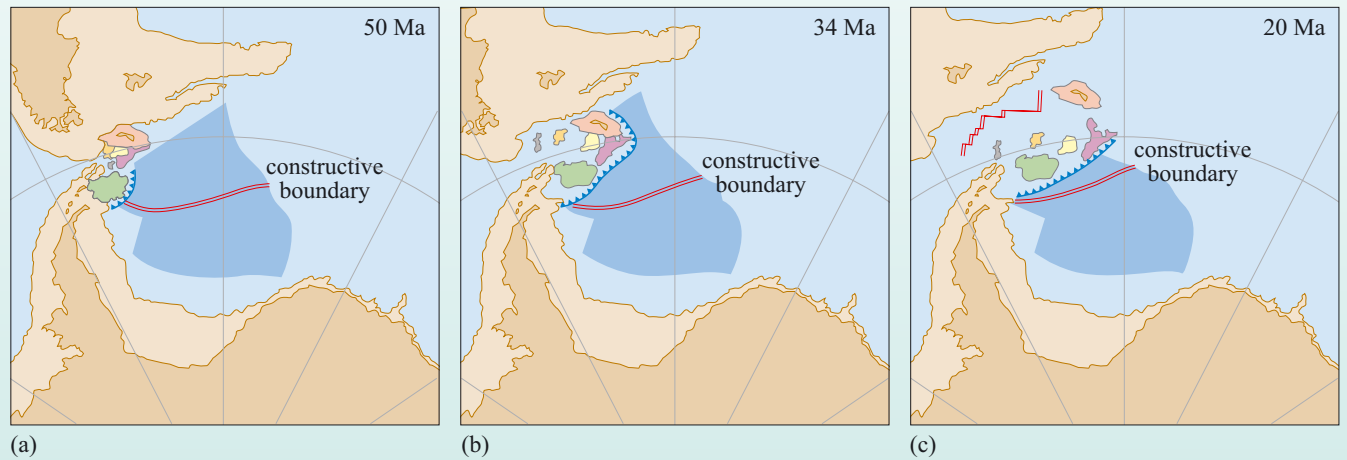


Figure 3.30 Plate reconstruction of the southern tip of South America and the Antarctic Peninsula at (a) 50 Ma, (b) 34 Ma and (c) 20 Ma. The development of a destructive plate boundary at 50 Ma results in the destruction of the ocean lithosphere previously generated at the spreading centre as shown. Continued subduction leads to the development of a back-arc basin and the southward migration of continental fragments shown in different colours, so that by 20 Ma there is a deep seaway floored by oceanic crust beneath the two continental masses. The darker area represents oceanic lithosphere. (Adapted from Livermore et al., 2005)

The coincidence of the change in motion of the Pacific Plate with changes in plate motions between S. America and Antarctica shows how the motions of all the plates are interconnected – a change in the true motion of one plate leads to changes in the true motions of many others.

While these plate motions were taking place the effect on Antarctica was profound. By 34 Ma the climate cooled from the temperate conditions that previously existed. This was sufficient for glaciers to begin their advance, and was followed by a period of continued cooling until at about 20 Ma, glaciation was complete. Even though the Drake Passage first opened at 50 Ma it was not until it opened to deep water at 34 Ma that glaciation really took hold

Today, the Antarctic Circumpolar Current is the strongest deep ocean current and its strength is responsible for the ‘icehouse’ climate that grips the planet. The opening of the Drake Passage had both a local and a global effect, initially cooling the climate of Antarctica from temperate to cold and ultimately playing an important role in the change from global ‘greenhouse’ conditions 50 Ma ago to the global ‘icehouse’ of today.

This example shows how plate tectonics, continental drift and the opening and closing of seaways can have a profound influence on both local and global climate. Throughout the Phanerozoic there were long periods when the Earth was much warmer than today – often called a ‘greenhouse’ climate – and other times when it was cold – called an ‘icehouse’ climate. These cycles, like the Wilson cycle, occur over periods of 100 Ma, reflecting the timescale of plate movements and the growth and destruction of oceans. Given the clear link between ocean circulation and climate, and the similar timescales of global climate change and plate motions, it is inescapable that one of the chief controls on long-term changes in the global climate must be plate tectonics.

Summary of Chapter 3

Plate tectonics is the grand, unifying theory of Earth sciences, combining the concepts of continental drift and sea-floor spreading into one holistic theory that explains many of the major structural features of the Earth's surface. It explains why the oceanic lithosphere is never older than about 180 Ma and why only the continents have preserved the Earth's geological record for the past 4000 Ma. It provides the framework to explain the distribution of earthquakes and volcanoes and a mechanism for the slow drift of the continents across the Earth's surface. The theory has now reached such a level of scientific acceptance that the movement of plates, both relative to one another and to the hot-spot reference frame, are being used to infer movement of the hot-spot reference frame with respect to the Earth's rotational axis.

Plate tectonics is an expression of the convective regime in the underlying mantle, but the link between individual convection cells and plate boundaries is not direct because plate boundaries are not fixed and, like the plates, move relative to one another. Plate movements are driven by gravity, largely by cold, dense lithospheric slabs pulling younger lithosphere towards a destructive boundary. A less-powerful driving force is generated by the potential energy of spreading centres, elevated some 2–3 km above the general level of the abyssal plains. You will learn more about mantle convection and how it relates to plate tectonics in Chapter 7.

As ideas concerning plate tectonics have evolved since the 1970s, it has become apparent that while the theory can be applied rigorously to the oceans, the same cannot be said of the continents. Because of the strength and rigidity of oceanic plates, deformation is focused into narrow linear zones along plate margins. By contrast, when continental lithosphere approaches a plate boundary, deformation can extend hundreds of kilometres into the continental interior because continental plates are less strong. Such deformation gives rise to the major mountain belts of the Earth, as exemplified by the Alpine Himalayan Chain, part of which will be described in Chapter 6. Before that, however, you will focus on the processes that occur at constructive (Chapter 4) and destructive plate boundaries (Chapter 5), the major sites of interaction and exchange between the interior of the Earth and the hydrosphere, atmosphere and biosphere.

Learning outcomes for Chapter 3

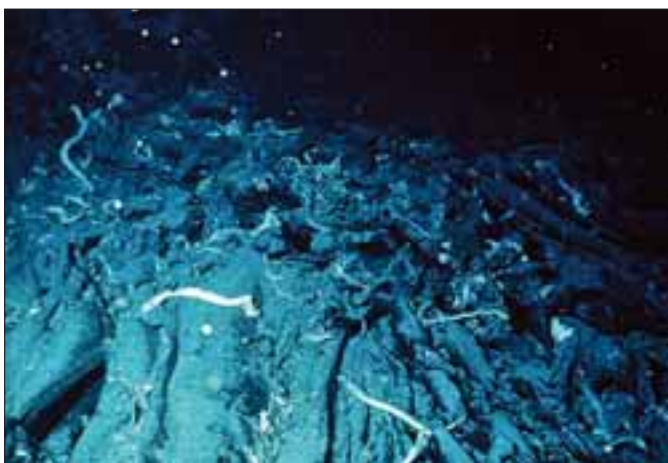
You should now be able to demonstrate a knowledge and understanding of:

- 3.1 The theory of tectonic plates and the different forms of evidence (e.g. palaeontology, palaeomagnetism, continuity of structures etc.) that can be used to understand the movement of the lithospheric plates over geological time.
- 3.2 The mechanisms of crustal growth and transfer of heat at spreading ocean ridges.
- 3.3 The three main types of plate boundary (constructive, destructive and conservative) and how they interact at triple junctions.
- 3.4 The difference between relative and true plate motion.
- 3.5 The driving and retarding forces that influence plate motion at constructive, destructive and conservative plate boundaries.

Processes at constructive plate boundaries

Chapter 4

Two-thirds of the Earth's surface (about $300 \times 10^6 \text{ km}^2$) is made of oceanic lithosphere, and all of that lithosphere has formed during the last 4% of the Earth's history. On average, about 3 km^2 of new oceanic lithosphere is made every year, which is distributed along the constructive plate boundaries that are found at mid-ocean ridges. Although these statistics imply that mid-ocean ridges are important components of the Earth's dynamic systems, direct observations of this prodigious activity have lagged behind observations of geological processes on land, for obvious reasons. It is only since the 1970s that submersible craft have been able to take cameras, other instruments and, importantly, people, to the dark ocean depths of the mid-ocean ridges. There, some 2.5 km below sea level, very young lava flows and hot springs that gush water at up to 400°C show, together with the longer known background buzz of shallow-focus earthquakes, that mid-ocean ridges are hot-beds of geological activity (Figure 4.1). While Chapter 3 described the tectonics of constructive plate boundaries and their role in driving sea-floor spreading, this chapter describes the processes that generate oceanic lithosphere from the mantle that underlies the mid-ocean ridges. The overall aim of this chapter, therefore, is to discover what happens, why it happens, and where it happens.



(a)



(b)

Figure 4.1 Signs of activity at mid-ocean ridges. (a) The aftermath of a volcanic eruption on the East Pacific Rise ($9^\circ 50.6' \text{ N}$) in 1991 shows the corrugated surface of lava littered with fragments of organisms killed by the eruption. (b) The plume of black 'smoke' issuing from a chimney of metallic sulfide minerals is a cloud of dark metal sulfide particles precipitated when hot (380°C) mineral-rich water jets out of the sea floor and mixes with cold (2°C) seawater. ((a) R.M. Haymon; (b) Dr Ken MacDonald/Science Photo Library)

Section 4.1 reviews the structure of the oceanic lithosphere – the finished product of constructive plate boundary processes. The later sections then show how that finished product is generated, beginning in Section 4.2 with the starting material – mantle rock. Subsequent sections deal with the processes that cause the mantle to melt and produce magma (Sections 4.3 to 4.5), the history of that magma as it makes its way to the surface (Sections 4.6 to 4.7), and the interaction between igneous oceanic crust and the oceans (Section 4.8).

4.1 The structure and composition of oceanic lithosphere

From the point of view of plate tectonics, oceanic lithosphere is the relatively cold, strong, outer part of the Earth in the ocean basins. Far from mid-ocean ridges, this layer has a fairly uniform thickness of about 120 km. To understand how the Earth generates new oceanic lithosphere, it is helpful to know its composition and structure. But because the oceanic lithosphere is largely hidden from view and inaccessible, finding out about these rather fundamental aspects of the Earth's outer layer has required ingenuity and the piecing together of different types of information. Apart from the areas at and near mid-ocean ridges, the seabed is covered in very fine mud. Drilling into this sediment layer shows that it can be up to several hundred metres thick and that it is underlain by volcanic rocks of basaltic composition. The crests of mid-ocean ridges, on the other hand, are devoid of any obscuring sediment but display signs of recent volcanic activity (Figure 4.1a). These observations are a key factor in suggesting that the mid-ocean ridges are the youngest parts of the ocean basins and that they have an igneous origin. At some mid-ocean ridge sites, plumes of hot, metal-rich brines rise from the sea floor (Figure 4.1b). Both of these manifestations of heat loss from the Earth's interior signal that, multiplied along the 64 000 km length of the mid-ocean ridge system, constructive plate boundaries play a significant role in the slow but continuous cooling of the Earth.

To see the interior of the oceanic crust it is necessary to visit areas where faults have sliced through the oceanic lithosphere to expose deeper layers, as if cutting through a cake to see what lies beneath the outer layer of icing. One such opportunity is provided by oceanic fracture zones that have cut cross-sections through oceanic crust. These can be investigated from submersible craft or by dredging samples from the rock face. The rocks found in the fracture zones are basalt lava flows, basalt dykes, gabbro and peridotite (Box 4.1) arranged in a layered structure (Figure 4.4).

Box 4.1 Basalt, gabbro and peridotite

The rock types basalt, gabbro and peridotite are defined according to the minerals they contain and/or the rock's texture. In these rocks, the most common minerals are olivine (Figure 4.2a), pyroxene (Figure 4.2b) or plagioclase feldspar (Figure 4.2c). Olivine is a greenish-coloured, magnesium–iron silicate with the chemical composition $(\text{Mg,Fe})_2\text{SiO}_4$, where the brackets in the mineral formula signify that for every four oxygen atoms there are a total of two **cations** comprising a mixture of magnesium and iron. Pyroxene is dark-green to black, with the composition $(\text{Ca,Mg,Fe})_2\text{Si}_2\text{O}_6$. Plagioclase feldspar is a whitish silicate mineral containing aluminium, calcium, sodium and a very small amount of potassium with the composition $(\text{KSi,NaSi,CaAl})\text{AlSi}_2\text{O}_8$.

The chemical compositions of basalt, gabbro and peridotite vary within bounds according to the exact

composition of the minerals present and their proportions. However, given the general mineralogical make-up of these rocks, peridotite is poorer in Si, Al, Na and Ca, but richer in Mg than either gabbro or basalt. Gabbro and basalt have the same composition (referred to as basaltic); the differences between these two rock types are textural, with basalt having smaller and fewer crystals.

- Basalt (Figure 4.3a) is a fine-grained volcanic rock that typically contains small crystals of pyroxene, plagioclase and, in some cases, olivine.
- Gabbro (Figure 4.3b) is a coarse-grained rock composed of pyroxene and plagioclase and, in some cases, olivine.
- Peridotite (Figure 4.3c) is a coarse-grained rock composed of at least 90% olivine and pyroxene, with at least 40% olivine.

Basalt forms by relatively fast cooling of basaltic magma, either as a volcanic rock or in dykes or other minor intrusions. Gabbro forms when basaltic magma cools so slowly that the entire magma crystallises.



(a)



(b)

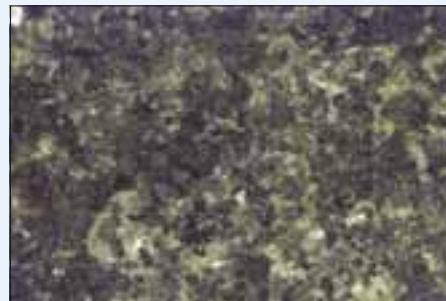


(c)

Figure 4.2 Photographs of the minerals (a) olivine (crystals ~2 cm across), (b) pyroxene (crystals ~1 cm long; the pale flecks are fragments of other minerals on the surface of the pyroxene crystals), and (c) plagioclase feldspar (crystals ~5 mm long).



(a)



(b)



(c)

Figure 4.3 Photographs of the rocks (a) basalt with white plagioclase and green olivine crystals (specimen is about 10 cm across), (b) gabbro (width of view about 5 cm), and (c) blocks of green peridotite up to 15 cm across encased in brown basalt.

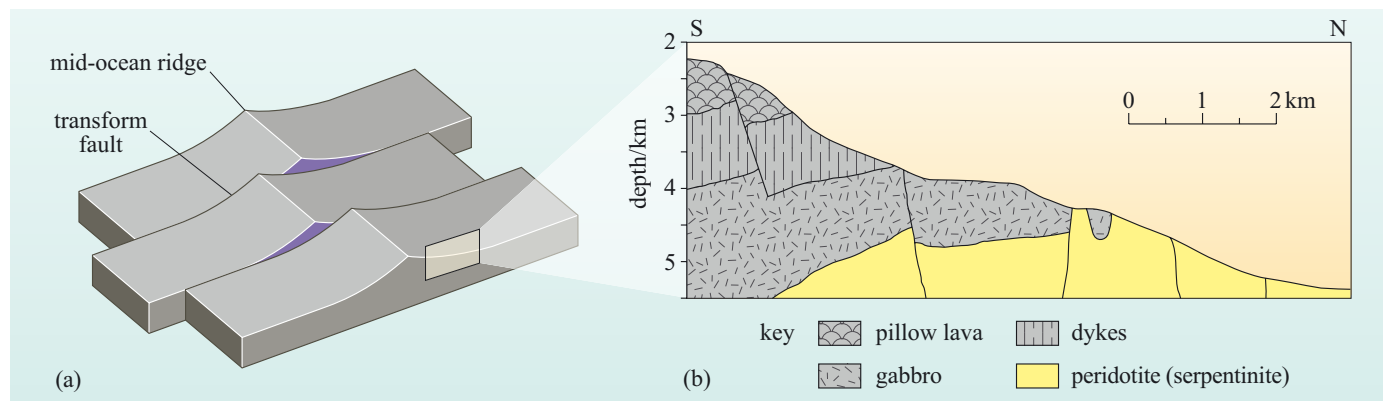


Figure 4.4 (a) Block diagram showing how transform faults at oceanic fracture zones expose the upper parts of oceanic crust. (b) Geological structure of the Vema Fracture Zone that offsets the Mid-Atlantic Ridge (around 10.75° N, 42° W) shows layers offset by faults and exposed by erosion. ((b) Adapted from Auzende et al., 1989)

Evidence for an igneous, layered oceanic crust with the composition of basalt also comes from, of all places, continental settings where convergent plate motions have caused segments of oceanic lithosphere to be tectonically emplaced onto continental and arc crust in a process called **obduction** (in contrast to subduction that removes oceanic lithosphere from the Earth's surface down into the mantle). The rocks that make up these obducted fragments are deep-sea sediments, basaltic pillow lavas, basaltic dykes (**sheeted dykes**), gabbro and peridotite (Figure 4.5). This association of rock types was first recognised in the early 19th century, when it was given the name ophiolite. However, it was not until the 1960s and the acceptance of plate tectonics that it was realised that ophiolites were stranded blocks of oceanic lithosphere trapped between colliding continents (Figure 4.6). Examples of ophiolites that teams of geologists have studied in detail occur in the Troodos Mountains of Cyprus, and in Oman, where the barren landscape gives almost complete exposure of rocks that once formed a complete section of oceanic crust (Figure 4.5).

Figure 4.5 (a) Photograph of pillow lavas in the Oman ophiolite formed by the flow of basalt lava under water. (b) Photograph of a cliff face composed entirely of near-vertical dykes of basalt in the Oman ophiolite. Each dyke has intruded rock made entirely of other dykes, resulting in a sheeted dyke complex.

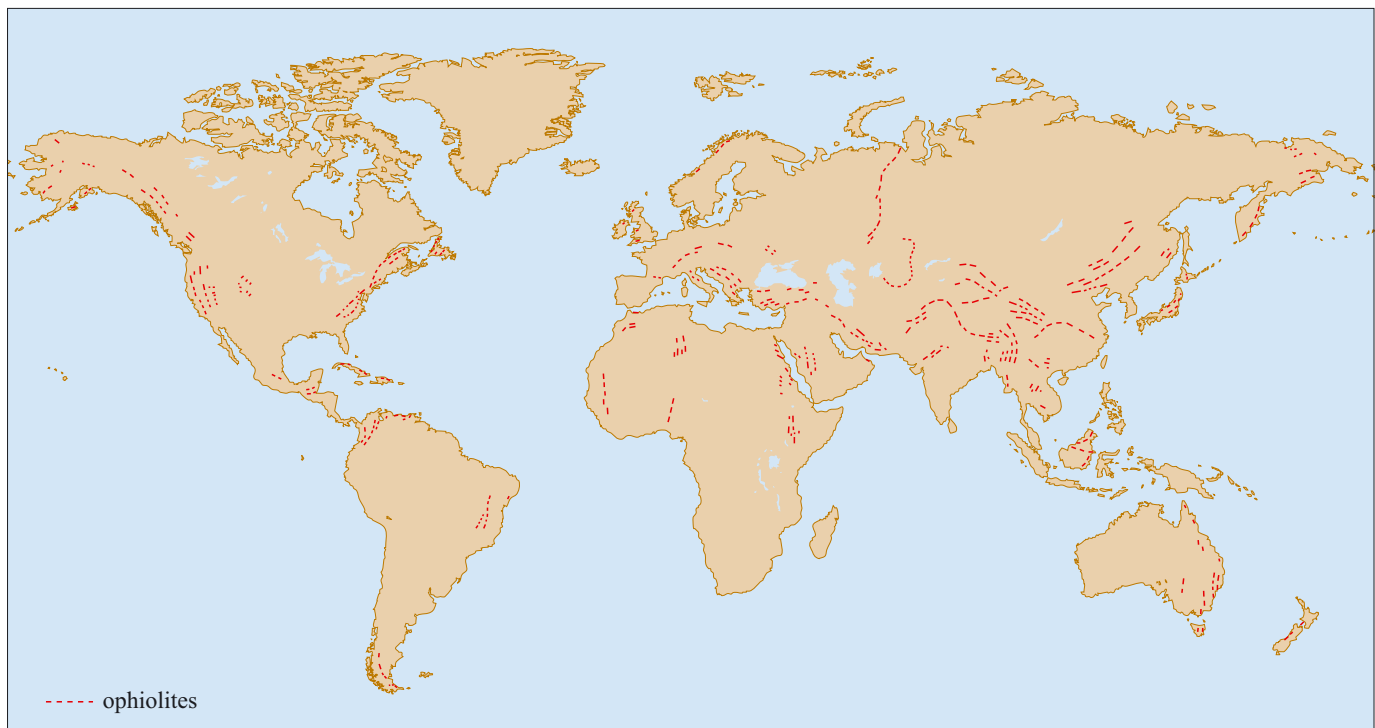
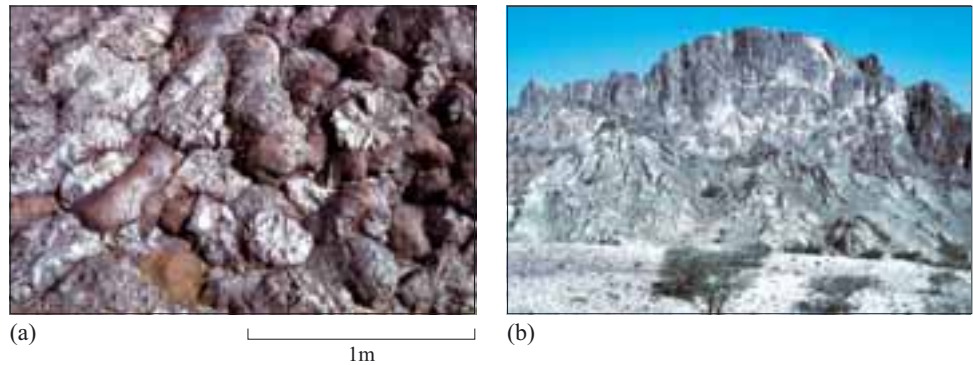


Figure 4.6 Map showing the global occurrence of ophiolites (marked as red dashed lines) illustrating how they often occur in long chains that were formed when they were obducted during mountain-building phases of continental collision.

The geologist's view of oceanic crust, based on studies of ophiolites and deep-sea rocks, can be compared with the results of geophysical investigations of the physical properties of oceanic crust. Foremost among these is the velocity of P-waves in rocks (v_p), because this depends on the mineral composition, making seismology a useful way of probing the geological composition and structure of the subsurface. The results of some seismic refraction experiments at several locations in the Atlantic Ocean off the coast of Florida are shown in Figure 4.7. These indicate that P-wave velocity increases with depth in the oceanic crust. The different profiles from this area share some common features: with increasing depth, v_p increases rather irregularly over the top 2 km. Below this depth, v_p increases much more gradually over a further 6 km until, at a depth of about 8 km, it increases sharply to a value close to 8 km s^{-1} , which is typical of the P-wave velocity in peridotite. The layer of nearly constant v_p above this boundary is interpreted as being made of gabbro, whereas the shallower layer most probably represents a series of pillow lavas that grade downwards into sheeted dykes.

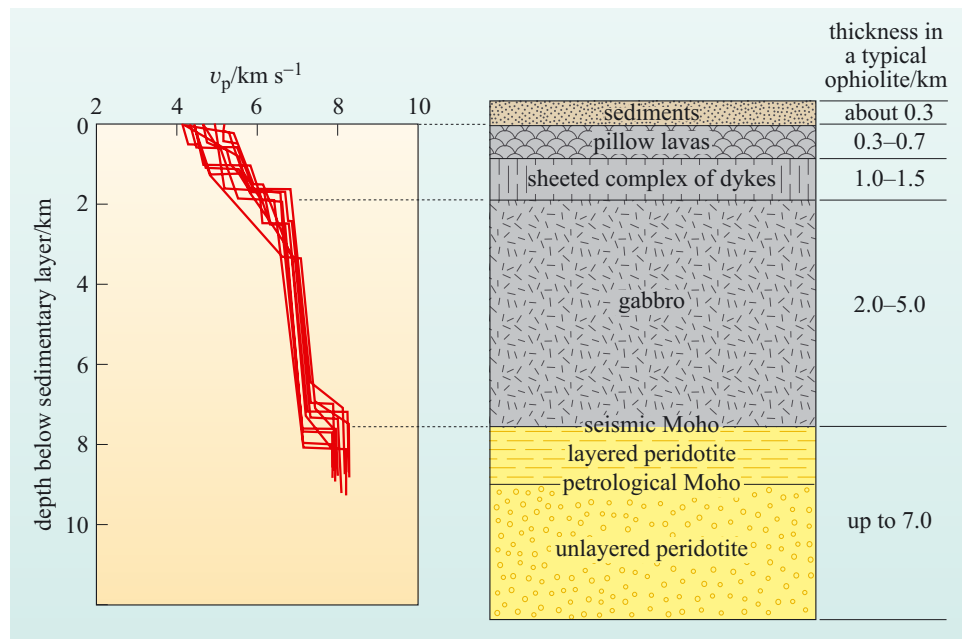


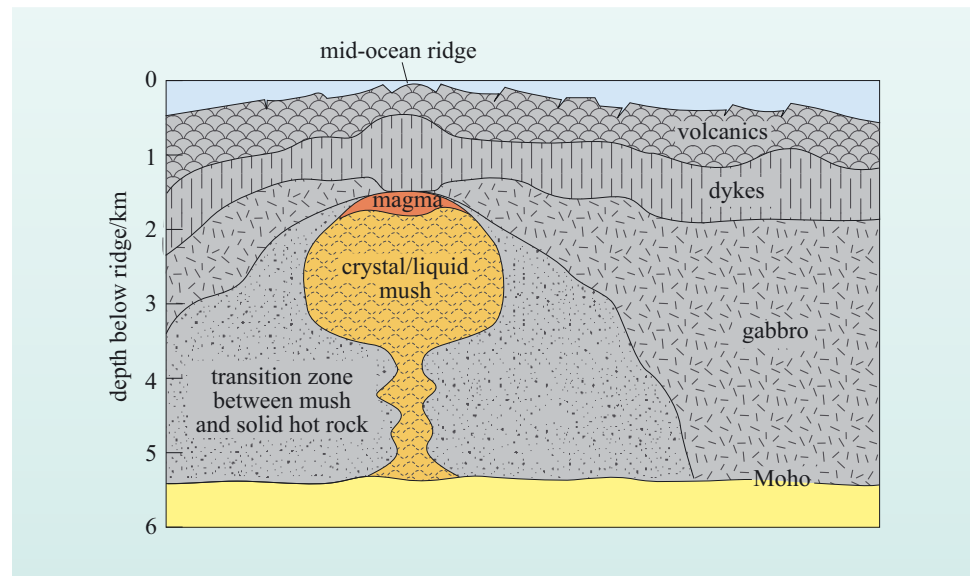
Figure 4.7 Seismic profile of 142 Ma-old oceanic crust in the Atlantic Ocean off the coast of Florida that lies beneath the surface layer of sediment. It shows a layered crustal structure that can be correlated with the layered geology seen in ophiolites. (Seismic profiles from White et al., 1992)

The peridotites found in ophiolites can be classed into two types based on their physical appearance: layered and unlayered. These are distinguished in Figure 4.7, but as they have the same P-wave velocities they cannot be distinguished seismically. The seismically defined Moho (i.e. the **seismic Moho**) is therefore interpreted as marking the shallowest occurrence of peridotite. Layered peridotites are thought to originate by crystallisation of olivine in magma chambers at constructive plate boundaries within oceanic crust, so the top of the mantle is best defined as being the contact between the layered and unlayered peridotite. This interface is known as the **petrological Moho** (Figure 4.7), **petrology** being the study of the composition, texture and structure of rocks. However, as the thickness of the layered peridotites is measured in hundreds of metres and the total thickness of the crust is measured in kilometres, the depth of the seismic Moho is a perfectly good estimate of the thickness of oceanic crust. On average, the igneous part of oceanic crust is 7 km thick. The sedimentary layer is much thinner and variable (between 0 km and 3 km, with an average thickness of 0.4 km).

Nearly the entire thickness of the oceanic crust has an igneous origin and is made from the solidified remains of the stored, transported and eruptive components of a basaltic magma system:

- gabbro that formed by solidification of slowly crystallising basaltic magma;
- sheeted dykes that were formed by solidification of basalt magma intruded into crust that was composed of older basalt dykes;
- lavas that were emplaced when the dykes broke through onto the sea floor (Figure 4.8).

Figure 4.8 Schematic cross-section of a mid-ocean ridge showing how the components of the system that produce the basaltic oceanic crust may be arranged. (Sinton and Detrick, 1992)



Basaltic magma is the raw material from which oceanic crust is constructed by sea-floor spreading, but why, when seismology shows that the mantle is normally solid, should magma be available beneath mid-ocean ridges? In other words, why should parts of the mantle be melting beneath mid-ocean ridges? To answer these questions, you will need to learn more about the material that the mantle is made from (Section 4.2) and then to find out what conditions are required for that material to melt (Section 4.3).

4.2 Rocks and minerals in the upper mantle

Parts of the Earth's mantle can be seen at the surface in ophiolites and oceanic fracture zones, but these are not the only source of mantle samples available to geologists. Lumps of mantle rock are occasionally picked up and carried by rapidly rising magma as it bursts upwards to the surface. These accidental but very special rock fragments, known as mantle xenoliths, can be found at some basaltic volcanoes such as Lanzarote in the Canary Islands, and Kilbourne Hole in New Mexico (Figure 4.3c). A rare type of CO₂-rich magma known as **kimberlite** (named after the Kimberley area of South Africa) also carries a cargo of xenoliths and crystals from the mantle, including diamond, which is the high-pressure form of carbon that forms at depths in excess of 150 km. Before looking more closely at the types of rock found in the upper mantle, the next section gathers together information about the minerals that are found in these (and other) rocks.

4.2.1 Minerals in the upper mantle

There are only a few different minerals that are commonly found in mantle rocks, and these are listed in Table 4.1.

Table 4.1 Chemical formulae of minerals found in peridotite.

Mineral	Formula
olivine	$(\text{Mg,Fe})_2\text{SiO}_4$
pyroxene	$(\text{Ca,Mg,Fe})_2\text{Si}_2\text{O}_6$
garnet	$(\text{Ca,Mg,Fe})_3\text{Al}_2\text{Si}_3\text{O}_{12}$
plagioclase feldspar	$(\text{KSi,NaSi,CaAl})\text{AlSi}_2\text{O}_8$
spinel	$(\text{Mg,Fe}^{2+})(\text{Al,Cr,Fe}^{3+})_2\text{O}_4$

Question 4.1

The minerals in Table 4.1 can be classified in various ways. Which minerals contain the following?

- (a) silicon and oxygen
- (b) *no* silicon
- (c) aluminium
- (d) iron and magnesium.

Most mantle minerals are silicates that also contain iron and magnesium. These are examples of **ferromagnesian silicate minerals**. They are also known as **mafic** minerals because of their high content of magnesium and iron (particularly Fe(II) iron, Fe^{2+}).

On the atomic scale, the silicate minerals are constructed from building units that contain one Si^{4+} ion and four O^{2-} ions arranged in a pyramid-shape known as a tetrahedron (Figure 4.9). Each tetrahedron has the Si^{4+} ion in the centre of the pyramid and one O^{2-} at each corner; this is the **silicate tetrahedron**.

Question 4.2

What is the overall electronic charge of a silicate tetrahedron?

Silicate minerals are made up from silicate tetrahedra linked to each other, or occurring separately, with the intervening spaces filled by cations of the appropriate size and charge. This means that the chemical formula and atomic structure of each mineral are related. The main features of the minerals are as follows.

Olivine

The general formula for olivine is $(\text{Mg,Fe})_2\text{SiO}_4$. The brackets indicate that a range of compositions is possible. For example, olivine can be Fe_2SiO_4 , $\text{Fe}_{0.2}\text{Mg}_{1.8}\text{SiO}_4$, $\text{Fe}_{1.1}\text{Mg}_{0.9}\text{SiO}_4$ or any other combination of Fe and Mg as long as the atomic proportions of Fe and Mg atoms add up to 2 (for every four O atoms).

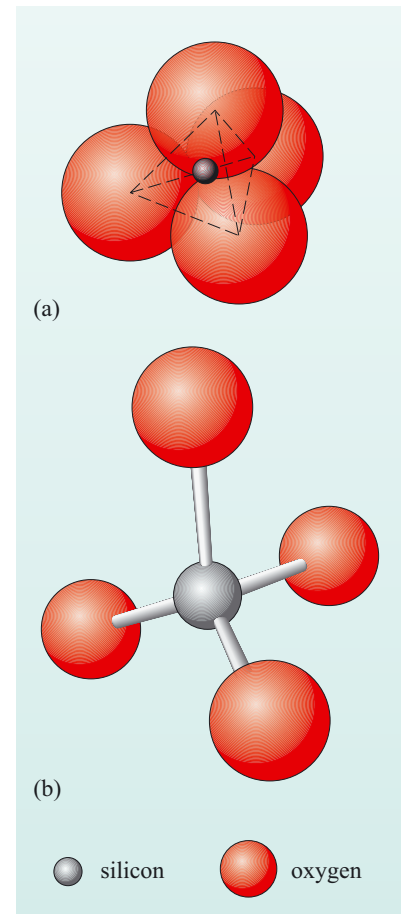


Figure 4.9 (a) One $(\text{SiO}_4)^{4-}$ tetrahedral unit (silicate tetrahedron) with the four atoms of oxygen and the one atom of silicon shown to true relative scale. (b) The atomic structure of the $(\text{SiO}_4)^{4-}$ unit showing how the silicon and oxygen atoms are linked.

You would be right in thinking that olivine contains individual $(\text{SiO}_4)^{4-}$ tetrahedra because SiO_4 appears in the formula. In the case of pure magnesium-rich olivine, the charges from two Mg^{2+} ions ($4+$) are required to balance the charge of one silicate tetrahedron ($4-$), which is why the chemical formula is Mg_2SiO_4 . The magnesium cations and silicate tetrahedra are arranged in a particular geometric pattern that allows the sizes and charges of the cations and tetrahedra to stay in stable positions, giving a crystal a specific shape and symmetry (Figure 4.2a).

But what happens if some of the Mg^{2+} ions are replaced with Fe^{2+} ions? The radius of an Mg^{2+} ion is 66 pm and the ionic radius of an Fe^{2+} ion is 74 pm. Because of their similar size and identical charge, Fe^{2+} ions can take the place of Mg^{2+} ions without distorting the overall mineral structure of olivine. The Mg and Fe atoms can substitute for each other and, in olivine, all compositional variants between Fe_2SiO_4 (**fayalite**) and Mg_2SiO_4 (**forsterite**) are possible because **complete ionic substitution** is allowed because of the similarity in size of the Fe^{2+} and Mg^{2+} ions. These variants are referred to as members of a **solid-solution series**, and the endmembers of the olivine solid-solution series are fayalite and forsterite. Fayalite and forsterite have identical atomic structures and are said to exhibit **isomorphism**, a term which means ‘equal form’.

Can Ca^{2+} ions enter the olivine structure? Calcium forms a much larger cation (radius = 99 pm) and cannot be accommodated easily in the structure of olivine. Consequently, only very limited ionic substitution of Ca^{2+} for Fe^{2+} or Mg^{2+} occurs, and it is unusual for olivine to contain much more than 1% Ca.

As a general rule, one ion may completely replace another to form a solid-solution series if the difference in their ionic radii *does not exceed 15%* of the radius of the smaller ion. If the size difference is greater than this, then **partial ionic substitution** will occur, and the compositional variation cannot extend continuously between the endmembers.

Pyroxene

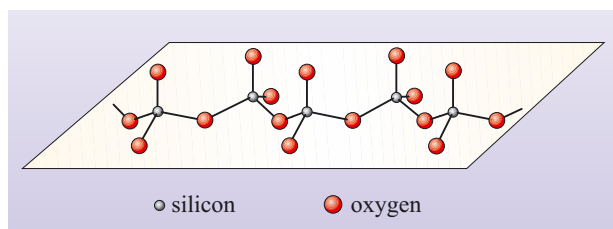


Figure 4.10 A chain of linked silicate tetrahedra, as found in pyroxene.

In the pyroxenes, the silicate tetrahedra are linked together by sharing two of their oxygens with two neighbouring tetrahedra, forming a chain of tetrahedra (Figure 4.10). The pyroxene structure has six oxygens for every two silicons, giving units of $(\text{Si}_2\text{O}_6)^{4-}$ that are charge balanced by two divalent cations – a mixture of Mg^{2+} and Fe^{2+} with some Ca^{2+} , hence the general pyroxene formula $(\text{Ca},\text{Mg},\text{Fe})_2\text{Si}_2\text{O}_6$, which is sometimes written in simplified form as $(\text{Ca},\text{Mg},\text{Fe})\text{SiO}_3$.

- Would you expect to find complete ionic substitution between $\text{Mg}_2\text{Si}_2\text{O}_6$ and $\text{Ca}_2\text{Si}_2\text{O}_6$?
- No. The ‘15% rule’ applies and, because the ionic radius of Ca^{2+} is greater than that of Mg^{2+} by 50%, only partial ionic substitution occurs.

Because Ca^{2+} is considerably larger than Mg^{2+} and Fe^{2+} , the atomic structure becomes increasingly distorted as the Ca content increases. When Ca makes up less than about 10% of the cations in pyroxene, the silicate chains and intervening cations form crystals with three symmetry axes at right angles (i.e. orthogonal), which is identical to the symmetry of a cereal box. These are the **orthopyroxenes**.

1 picometre (pm) = 10^{-12} m

(so-called because of their characteristic orthogonal symmetry axes). The simplest orthopyroxenes form a solid-solution series between the endmembers **enstatite** ($\text{Mg}_2\text{Si}_2\text{O}_6$) and **ferrosilite** ($\text{Fe}_2\text{Si}_2\text{O}_6$).

Pyroxenes with more than about 10% Ca have an atomic structure with only two symmetry axes because of the distortions required to fit the large Ca^{2+} ions into the spaces between the chains of silicate tetrahedra. This group of pyroxenes are called **clinopyroxenes**. Their chemical composition can be expressed as mixtures of the endmembers **diopside** ($(\text{Ca,Mg})_2\text{Si}_2\text{O}_6$), **hedenbergite** ($(\text{Ca,Fe})_2\text{Si}_2\text{O}_6$), enstatite ($\text{Mg}_2\text{Si}_2\text{O}_6$) and ferrosilite ($\text{Fe}_2\text{Si}_2\text{O}_6$). Note that diopside is a chemical mixture of $\text{Mg}_2\text{Si}_2\text{O}_6$ and $\text{Ca}_2\text{Si}_2\text{O}_6$. Similarly hedenbergite is a chemical mixture of $\text{Fe}_2\text{Si}_2\text{O}_6$ and $\text{Ca}_2\text{Si}_2\text{O}_6$.

The mineral with the formula $\text{Ca}_2\text{Si}_2\text{O}_6$ is **wollastonite**, but it is not a pyroxene because the structural adjustments needed to fit in all the Ca^{2+} ions are so great that it no longer has the same structure of pyroxenes; hence wollastonite is recognised as a different mineral in its own right. Nonetheless, wollastonite is often considered as an endmember when describing the chemical compositions of pyroxenes. For example, diopside can be described as 50% enstatite, 50% wollastonite.

Pyroxene from mantle rocks also contains small amounts of Al, Ti, Cr and Na.

Garnet

Like olivine, garnet contains individual $(\text{SiO}_4)^{4-}$ silicate tetrahedra. Three of these are required to account for the Si_3O_{12} part of the mineral's chemical formula $((\text{Ca,Mg,Fe})_3\text{Al}_2\text{Si}_3\text{O}_{12})$. Unlike olivine, garnet contains trivalent ions, notably Al^{3+} and, in some cases, Fe^{3+} . A comparatively large range of cations can fit into garnet (it might even be called a 'chemical dustbin'), so there are many endmembers to the garnet **isomorphous series**. Here, we mention only pyrope ($\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}$), almandine ($\text{Fe}^{2+}_3\text{Al}_2\text{Si}_3\text{O}_{12}$) and grossular ($\text{Ca}_3\text{Al}_2\text{Si}_3\text{O}_{12}$). Garnet found in mantle peridotites is rich in the pyrope (Mg) component.

Feldspar

As noted earlier, in the pyroxenes the corners of two silicate tetrahedra are shared with neighbours to produce a chain structure (Figure 4.10). In feldspar all four corners are shared, resulting in a three-dimensional framework. The simplest silicate mineral with a framework structure is quartz (which is *not* a feldspar), which has the simple formula SiO_2 because each Si^{4+} ion can lay claim to only half of each of the four O^{2-} ions that surround it. Quartz is a common mineral in the continental crust but it is absent from mantle peridotite. In feldspar, either one or two of the Si^{4+} ions in the framework are replaced by Al^{3+} ions.

Where *one* out of every four silicons is replaced by aluminium, charge balance is re-established by introducing one Na^+ or one K^+ , and this yields the **alkali feldspars**: **albite** ($\text{NaAlSi}_3\text{O}_8$) and **orthoclase** (KAlSi_3O_8).

- Would you expect complete ionic substitution between albite and orthoclase, given ionic radii of 97 pm for Na^+ and 133 pm for K^+ ?
- The radius of K^+ is $\frac{(133 - 97)}{97} \times 100\% = 37\%$ greater than that of Na^+ .

Because this is higher than the 15% limit, complete ionic substitution between albite and orthoclase does not occur.

When Al^{3+} replaces Si^{4+} in *two* out of every four tetrahedra, the divalent cation Ca^{2+} restores the neutral charge, giving **anorthite** ($\text{CaAl}_2\text{Si}_2\text{O}_8$).

- Will there be a solid-solution series between anorthite and albite?
- The similarity of ionic radii for Ca^{2+} and Na^+ (99 pm and 97 pm respectively) allows complete ionic substitution to occur between albite and anorthite to form a solid-solution series.

This is known as the plagioclase feldspar series. As Na^+ is replaced by Ca^{2+} in going from albite ($\text{NaAlSi}_3\text{O}_8$) to anorthite ($\text{CaAl}_2\text{Si}_2\text{O}_8$), so a second Si^{4+} (in every four tetrahedra) is replaced by Al^{3+} . In other words, this solid-solution series involves substitution of CaAl for NaSi to keep the structure electrically balanced, and this is an example of **coupled substitution**. A small amount of K is also present in most plagioclase feldspars because K can substitute for Na to a limited extent.

Spinel group

This is a group of non-silicate minerals loosely described by the chemical formula MOR_2O_3 , where M is a divalent cation (e.g. Mg^{2+}) and R is a trivalent cation (e.g. Al^{3+}). The formula of spinels can also be written as MR_2O_4 . Important endmembers within this rather diverse group are spinel (MgAl_2O_4), magnetite ($\text{Fe}^{2+}(\text{Fe}^{3+})_2\text{O}_4$) and chromite ($\text{Fe}^{2+}\text{Cr}_2\text{O}_4$).

Mineral compositions

Most of the minerals in mantle rocks are ferromagnesian minerals with solid solution between Mg and Fe endmembers, but we have not said which endmembers dominate. In all cases, the mineral composition can be expressed in terms of the proportion of the Mg endmember (relative to the total amount of the Mg and Fe^{2+} endmembers) using the ratio $\text{Mg}/(\text{Mg} + \text{Fe}^{2+})$, where Mg and Fe^{2+} are the number of Mg and Fe^{2+} atoms in the mineral formula. This ratio can be multiplied by 100 to express it as a percentage. The value, either of the ratio or of the percentage, is called the **magnesium number** and is often abbreviated as **Mg-number** or **Mg#**.

Question 4.3

- (a) An olivine crystal has the composition $\text{Mg}_{1.8}\text{Fe}_{0.2}\text{SiO}_4$. What is its Mg# as a ratio and as a percentage?
- (b) A clinopyroxene crystal has the composition $\text{Mg}_{1.2}\text{Fe}_{0.4}\text{Ca}_{0.4}\text{Si}_2\text{O}_6$. What is its Mg# as a ratio and as a percentage?

Chemical analysis of the ferromagnesian minerals in mantle rocks shows that they are much richer in the Mg endmember than the Fe endmember. Olivines in peridotite xenoliths have Mg# in the range 88 to 92. In other words, they have between 88% and 92 % forsterite, and between 12% and 8% fayalite. Using the conventional abbreviations Fo for forsterite and Fa for fayalite, these compositions are written more concisely as $\text{Fo}_{88}\text{Fa}_{12}$ to $\text{Fo}_{92}\text{Fa}_8$. The Mg#s of orthopyroxenes and clinopyroxenes cover a similar range to those of olivine but extend to slightly higher values. Mantle-derived garnets are rich in Cr and Mg.

4.2.2 Rock types in the upper mantle

Rocks from the mantle, whether from the base of ophiolite sequences or xenoliths found in volcanic rocks, are dominated by olivine, orthopyroxene and, to a lesser extent, clinopyroxene. The content of the mafic minerals is so high that the rocks are termed **ultramafic rocks**; ultramafic rocks can be subdivided into different types depending on the relative proportions of the three minerals olivine, orthopyroxene and clinopyroxene.

To show the proportions of these three components in any mixture calls for a triangular shaped diagram, known as a **ternary diagram**, as explained in Box 4.2. The diagram used for classifying ultramafic rocks is shown in Figure 4.12.

Box 4.2 Plotting and reading ternary diagrams

A ternary diagram can be used to plot the composition of any substance or mixture in terms of three components. Examples include the chemical compositions of feldspar and pyroxene, and the proportions of different minerals in rocks. Consider a mixture of three components A, B and C. A ternary diagram consisting of a triangle with corners representing the three components (100% A, 100% B, and 100% C) is required to plot the composition of this mixture (Figure 4.11).

- A sample that contained only component A (i.e. 100% A) would plot at the corner labelled A.
- A sample that contained only a mixture of B and C (i.e. 0% A) would plot on the edge of the triangle joining the corners B and C; the greater the proportion of component C, the closer it would be to the C corner.

Figure 4.11a shows how the proportion of A is represented by lines parallel to the edge BC.

Suppose you wanted to plot the composition of a sample that contained 40% A, 30% B and 30% C (a shorthand way of writing this is $A_{40}B_{30}C_{30}$, or $A_{0.40}B_{0.30}C_{0.30}$). This is done as follows.

- 1 Determine on which 'A' line the composition must lie (as in Figure 4.11a).
- 2 Find the 'B' line.
- 3 The point where these two lines intersect marks the composition of the mixture.
- 4 As a final check, mark the 'C' line, which should pass through the same point (Figure 4.11b).

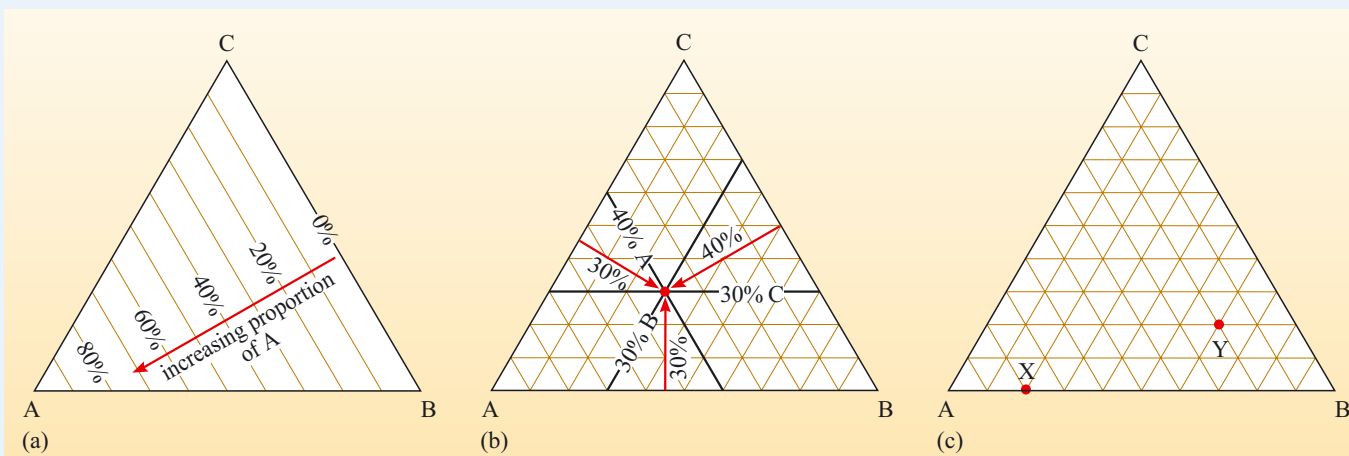


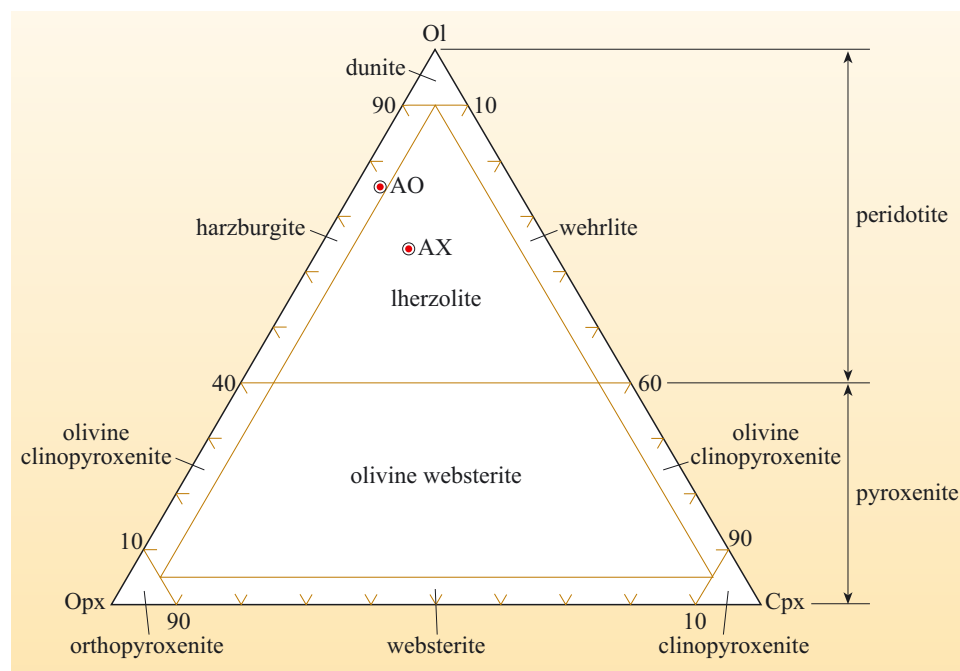
Figure 4.11 (a) A ternary diagram with three components A, B and C. (b) A ternary diagram with the composition $A_{40}B_{30}C_{30}$ plotted. The percentages of each component are indicated by the thick lines, which intersect at a point that shows the overall composition. (c) A ternary diagram for use with Question 4.4.

Question 4.4

Using the ternary diagram in Figure 4.11c, determine the compositions indicated by the two points X and Y.

Figure 4.12 is split into areas, or fields, that define the nomenclature for ultramafic rocks, which are themselves defined as containing more than 90% mafic minerals. Those that contain more than 40% olivine are defined as peridotites. These are further divided according to the proportions of olivine, orthopyroxene and clinopyroxene. The average composition of peridotite xenoliths from basalts and kimberlites is plotted on Figure 4.12 and falls in the lherzolite field. Lherzolite samples can be subdivided according to which Al-bearing mineral they contain. Kimberlites have mantle xenoliths that contain garnet, so these are called garnet lherzolite. Spinel lherzolites are found in some basalts from ocean islands (such as Lanzarote) and the continents (e.g. Kilbourne Hole crater, New Mexico) and in oceanic fracture zones. Plagioclase lherzolites are found rarely in oceanic settings.

Figure 4.12 Classification of ultramafic rocks based on the relative proportions of olivine (Ol), orthopyroxene (Opx) and clinopyroxene (Cpx). Average compositions of peridotites from ophiolites and oceanic fracture zones (AO) and of xenoliths from basalts and kimberlites (AX). (Dunite, harzburgite, lherzolite, and websterite are named after the places where the rock type was first recognised: Dun Mountain in New Zealand, the Harz Mountains in Germany, Etang de Lherz in the French Pyrenees, and Webster in North Carolina, USA; wehlrite is named after Adolf Wehrle, a 19th century Austrian Councillor of Mines.)



Question 4.5

- What type of ultramafic rock is represented by the average oceanic peridotite composition (labelled AO) in Figure 4.12?
- What are the approximate proportions of olivine, orthopyroxene and clinopyroxene in this average composition?

Compared to the average lherzolite, oceanic peridotites have a lower proportion of clinopyroxene and a higher proportion of olivine.

4.2.3 Peridotites and basalts

The chemical compositions of different rocks can be useful in deducing how the rocks formed. Table 4.2 lists typical chemical compositions of a harzburgite from an ophiolite, a lherzolite xenolith, and a mid-ocean ridge basalt.

Table 4.2 Chemical compositions (in weight per cent) of two peridotites and a mid-ocean ridge basalt. See also Box 4.3.

	Harzburgite	Lherzolite	Mid-ocean ridge basalt
SiO ₂	44.69	45.35	48.77
TiO ₂	0.02	0.16	1.15
Al ₂ O ₃	0.86	4.26	15.90
FeO	8.17	8.24	9.81
MnO	0.12	0.14	0.17
MgO	45.04	38.17	9.67
CaO	1.09	3.39	11.16
Na ₂ O	0.02	0.29	2.43
K ₂ O	0.01	0.03	0.08

Box 4.3 The chemical composition of rocks

Although virtually every chemical element can be found in most rocks, oxygen, together with ten other elements (Si, Ti, Al, Fe, Mn, Mg, Ca, Na, K, and P) account for nearly all of the matter in a given rock. These ten elements are therefore known as the **major elements**, as opposed to other elements that occur in much lower abundances, which are known as **trace elements**. Because most minerals contain O, with matching amounts of cations needed to balance the electronic charge, rock analyses are usually reported in terms of the weight percentage of the major element oxides SiO₂, TiO₂, etc., as shown in Table 4.2. Iron can occur in two oxidation states – Fe²⁺ and Fe³⁺, so some rock analyses give the separate concentrations of their oxides FeO and Fe₂O₃ whereas other analyses report the concentration of all of the iron in terms of either FeO or Fe₂O₃. Some rocks also contain a few per cent water. When the percentages of all the oxides are added up, the total rarely comes to 100%. This is because of (i) analytical errors and (ii) components such as water and/or some trace elements that are actually quite abundant in a given rock were not analysed.

- In what way do the chemical compositions of lherzolite and harzburgite in Table 4.2 reflect the composition of the minerals found in those rocks?
- Olivine and pyroxene from peridotites have high Mg#s so, because these minerals are the main constituent of peridotite, the rocks also have much higher MgO contents than FeO contents.

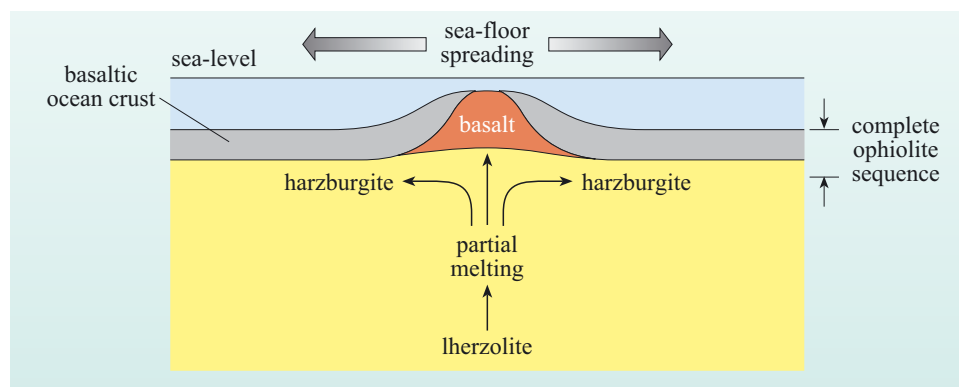
The composition of basalt is starkly different from that of lherzolite or harzburgite, having much less magnesium, slightly higher silica and iron, and much higher concentrations of other oxides. Clearly, basalt is not a molten version of peridotite – the chemical compositions are far too different for them to represent the same material, so basalt must have another origin.

A clue that there is a relationship between the three rock compositions in Table 4.2 comes from noticing that, for every oxide, the composition of the lherzolite falls between that of harzburgite and basalt. It is possible to split lherzolite into two compositionally different portions – a mobile basalt liquid and a harzburgite residue. Basalt, therefore, forms by the partial melting of lherzolite:



This process is fundamental in processing lherzolite from within the mantle and turning it into oceanic lithosphere comprising basaltic crust that is underlain by harzburgite, as seen in ophiolite sequences (Figure 4.13).

Figure 4.13 Cartoon of a constructive plate boundary, seen in cross-section, showing partial melting of lherzolite to produce basalt liquid and solid residual harzburgite.



The explanations of why partial melting happens, and exactly where within the mantle it happens, are the topics of the next section.

4.3 Experiments with peridotite

The conditions in the mantle where peridotite begins to melt must involve higher temperature (T) and higher pressure (P) than are encountered on the Earth's surface. But how high must T and P be for melting to happen? Taking a sample of peridotite and subjecting it to high T and P in the laboratory and seeing whether it melts would be an ideal way of answering this question and is, in fact, the approach taken. While the ability to replicate the conditions in the upper mantle in the laboratory is more than fifty years old, the techniques and hardware involved continue to improve. A variety of apparatus exists that has different capabilities in terms of the temperatures and pressures that can be achieved. Whatever the exact design of the equipment, a small sample of powdered rock is subjected to a given pressure and temperature for several hours or even days. Over a long enough time period the sample reaches chemical equilibrium. The temperature and pressure are then quickly returned to room conditions so that the material has no time to change back into its initial state. Instead, the material that was present at high pressure and high temperature becomes quenched, with any liquid that was present being turned to glass. Chemical analysis and microscope studies are then used to identify any minerals present.

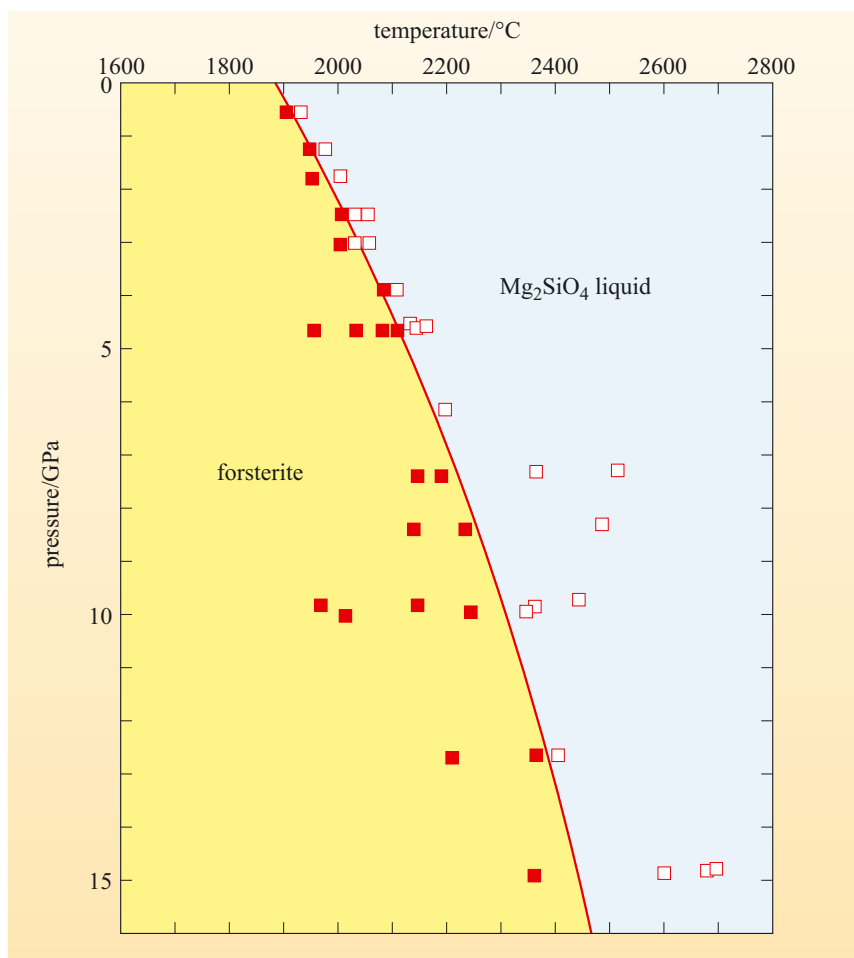
To understand the results from these types of experiment you need to understand the terms ‘phase’ and ‘phase diagram’. A **phase** is a substance with a particular chemical composition and molecular structure. Olivine and quartz are examples of two phases because each has its own composition and structure. Basalt liquid and olivine are different phases, as are SiO_2 liquid and quartz (solid SiO_2). Graphite and diamond are both forms of solid carbon but have different structures (it is their structures that make graphite soft and diamond hard). These two phases are both solid. **Phase diagrams** generally illustrate the conditions of temperature, pressure and chemical composition under which different phases exist. Because the processes that form igneous rocks are fundamentally based on reactions between minerals and liquids that either involve melting a crystalline rock to produce a liquid, or crystallising a liquid to produce crystals, phase diagrams that define the relationships between minerals and liquids at different temperatures and pressures are a cornerstone for understanding how igneous processes operate.

4.3.1 The phase diagram of Mg_2SiO_4

The phase diagram of Mg_2SiO_4 (the chemical composition of forsterite, the magnesium endmember of olivine) is shown in Figure 4.14 and introduces some of the main aspects of using phase diagrams. In this example, the chemical composition is fixed (it is Mg_2SiO_4) and the diagram shows the results of experiments in which Mg_2SiO_4 has been held at various high temperatures and high pressures. The outcome at each P and T can be shown using different symbols, in this case a filled square denotes solid crystals of forsterite whereas an open square denotes liquid Mg_2SiO_4 . As you might expect, forsterite melts if the temperature is sufficiently high.

The conditions under which forsterite or liquid Mg_2SiO_4 is stable are separated by a line that divides the phase diagram into two areas. Each area is known as a **stability field**, and the boundary between them is a **phase boundary**. Both phases coexist at conditions that plot exactly on the phase boundary, but on either side of it only one phase is stable.

Figure 4.14 Pressure–temperature phase diagram of Mg_2SiO_4 . Filled squares represent crystalline Mg_2SiO_4 , i.e. forsterite. Open squares represent liquid Mg_2SiO_4 . Note: pressure increases down the vertical axis. (Adapted from Herzberg, 1987)



- Does the melting temperature of forsterite depend on pressure?
- Yes. As pressure increases, the melting temperature increases.

The reason that the melting temperature increases with pressure is that more energy is required to liberate atoms from the ordered atomic structure of a mineral when the mineral's atoms are densely packed together as a result of being compressed under high pressure.

4.3.2 The phase diagram of lherzolite

Although forsterite is the main component of mantle olivine, and olivine is the main component of mantle rocks, the phase diagram of lherzolite is needed to investigate the mineralogy and melting behaviour of the mantle. To this end, the Japanese researcher Takahashi subjected samples of lherzolite xenolith from the Kilbourne Hole crater in New Mexico to pressures as high as 3 GPa, which is equivalent to the pressure at around 100 km below the Earth's surface (Box 4.4), and to high temperatures. The results of his experiments define the phase diagram shown in Figure 4.15, which shows two important features:

- depending on pressure and temperature, the peridotite either remained as a solid, produced a mixture of solid and liquid, or was completely liquid
- depending on pressure, and to a lesser extent temperature, the identity of the mineral that contains most of the aluminium in the rock changes. So, whether plagioclase lherzolite, spinel lherzolite or garnet lherzolite is stable depends on pressure, and hence on depth within the mantle.

Box 4.4 Pressure and depth

The pressure at any point on or inside the Earth depends on the weight of the overlying materials. On the surface, the weight of the atmosphere gives rise to atmospheric pressure. An underwater swimmer experiences a hydrostatic pressure that increases with the depth of water. A rock inside the Earth is subjected to **lithostatic pressure** due to the weight of overlying rock.

The pressure due to an overlying layer of material depends on the density, ρ , and thickness, h , of the layer:

$$P = \rho gh \quad (4.1)$$

where g is the acceleration due to gravity (approximately 9.8 m s^{-2}) and units are ρ : kg m^{-3} , g : m s^{-2} , and h : m, pressure P : $\text{kg m}^{-1} \text{ s}^{-2}$, which is equivalent to newtons per square metre (N m^{-2}) or pascals (abbreviated Pa).

$$1 \text{ Pa} = 1 \text{ N m}^{-2} = 1 \text{ kg m}^{-1} \text{ s}^{-2}$$

Atmospheric pressure has a value of 10^5 Pa (which is equal to 0.1 MPa). Lithostatic pressure increases by about 30 MPa for every kilometre depth beneath the surface of the Earth.

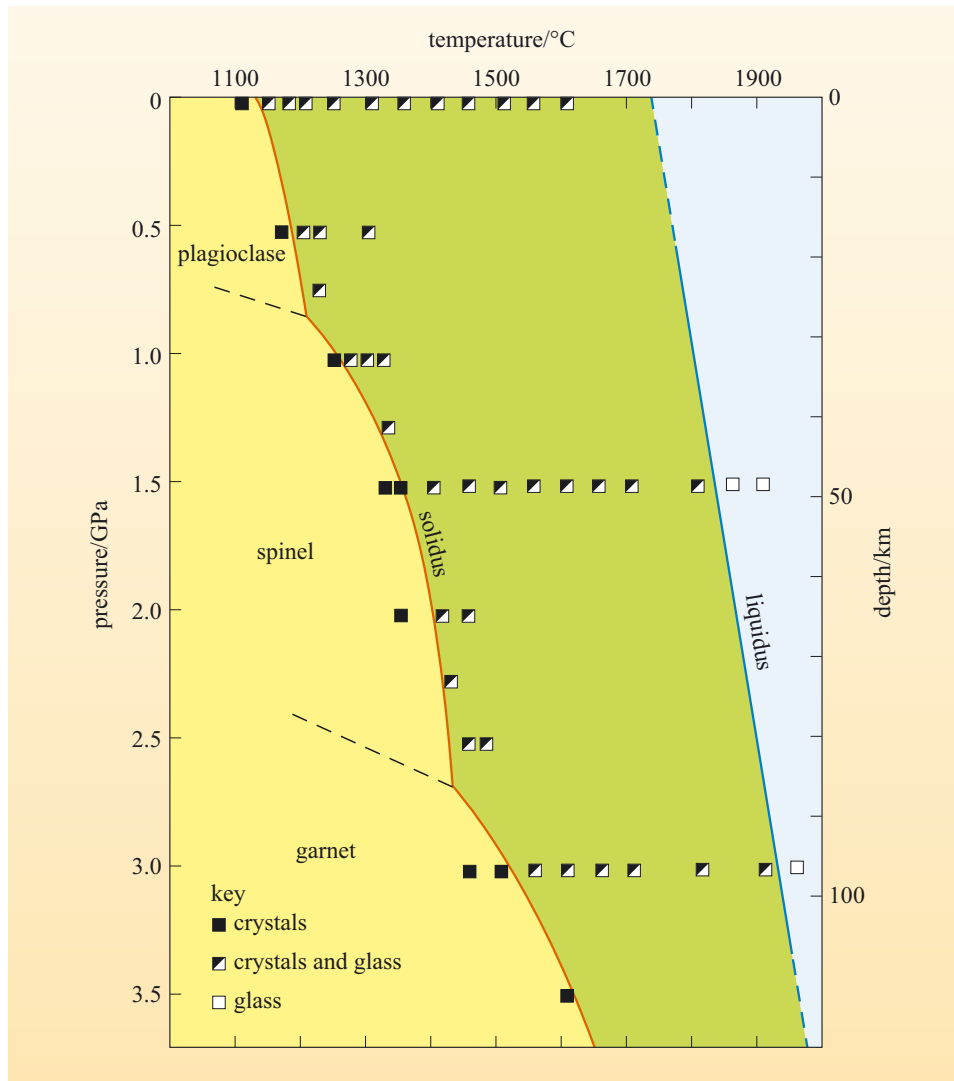


Figure 4.15 Phase diagram of lherzolite sample KLB-1 determined by Takahashi (1986). Open squares represent experiments that produced only glass (quenched liquid), half-filled squares represent a mixture of crystals and glass, and filled squares represent crystals only. Note that pressure increases down the vertical axis and the pressure is proportional to depth (labelled on the right-hand vertical axis). (Adapted from Takahashi, 1986)

Question 4.6

The sample used by Takahashi (Figure 4.15) was a spinel peridotite. From what range of depths could this sample have come from?

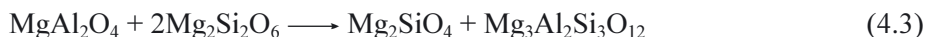
One reason for expecting mineralogy to change with pressure is that the structure of a mineral is sensitive to pressure. When pressure is increased, a mineral becomes slightly denser as its atoms are compressed together. At some point, the amount of compression becomes too much for the mineral to sustain its structure, forcing a major readjustment of the positions of the atoms to make a much more tightly packed structure. This is what happens when graphite changes to diamond during the manufacture of industrial diamonds by compressing carbon to very high pressure. Diamond and graphite are **polymorphs** (meaning ‘in many forms’) of carbon, i.e. they are different forms of the same substance. Diamond has a higher density (3500 kg m^{-3}) than graphite (2000 kg m^{-3}), reflecting the tighter packing of carbon atoms in the high-pressure polymorph.

In other cases, the effect of increasing pressure is to cause two or more minerals to react chemically, forming a new assemblage of minerals with an overall higher density, a process that happens during the metamorphism of rocks in the crust. The reactions that occur in lherzolite with increasing pressure (simplified by referring only to the magnesium endmembers of the minerals) are:



forsterite anorthite spinel enstatite diopside
(olivine) (plagioclase) (orthopyroxene)(clinopyroxene)

and



spinel enstatite forsterite pyrope
(orthopyroxene) (olivine) (garnet)

Question 4.7

What does the mineralogy of the different types of lherzolite xenoliths found in basalts and in kimberlites tell you about the relative depths from which these two magma types come? (*Hint*: refer to Section 4.2.2.)

Turning now to the melting behaviour of lherzolite, use Figure 4.15 to answer the following question.

Question 4.8

At a pressure of 1.5 GPa (equivalent to a depth of almost 50 km), what is the maximum temperature at which KLB-1 lherzolite is still solid, and what is the minimum temperature at which it is totally liquid?

Constrained by the experimental data points, the line labelled '**solidus**' in Figure 4.15 is the curve below which the system is entirely solid (where 'below' means 'at lower temperature'). This area, which is below the solidus, is known as the **subsolidus** field. At higher temperature, the **liquidus** is the curve above which the system is completely liquid. Thus, we can say that at 1.5 GPa the solidus is crossed at a temperature of just above 1350 °C and the liquidus is crossed at about 1830 °C (see answer to Question 4.8). Between the solidus and liquidus lies the zone of partial melting, so it is within this region of pressures and temperature that basalts are generated.

Experiments on KLB-1 and other lherzolite samples reveal that with increasing temperature (at a given pressure) there are systematic changes in the degree of melting, the composition of the liquid produced, the identity of the minerals that are still present, and the chemical composition of those minerals. Some of these results are summarised in Table 4.3 and Figure 4.16, and they are explored in Question 4.9.

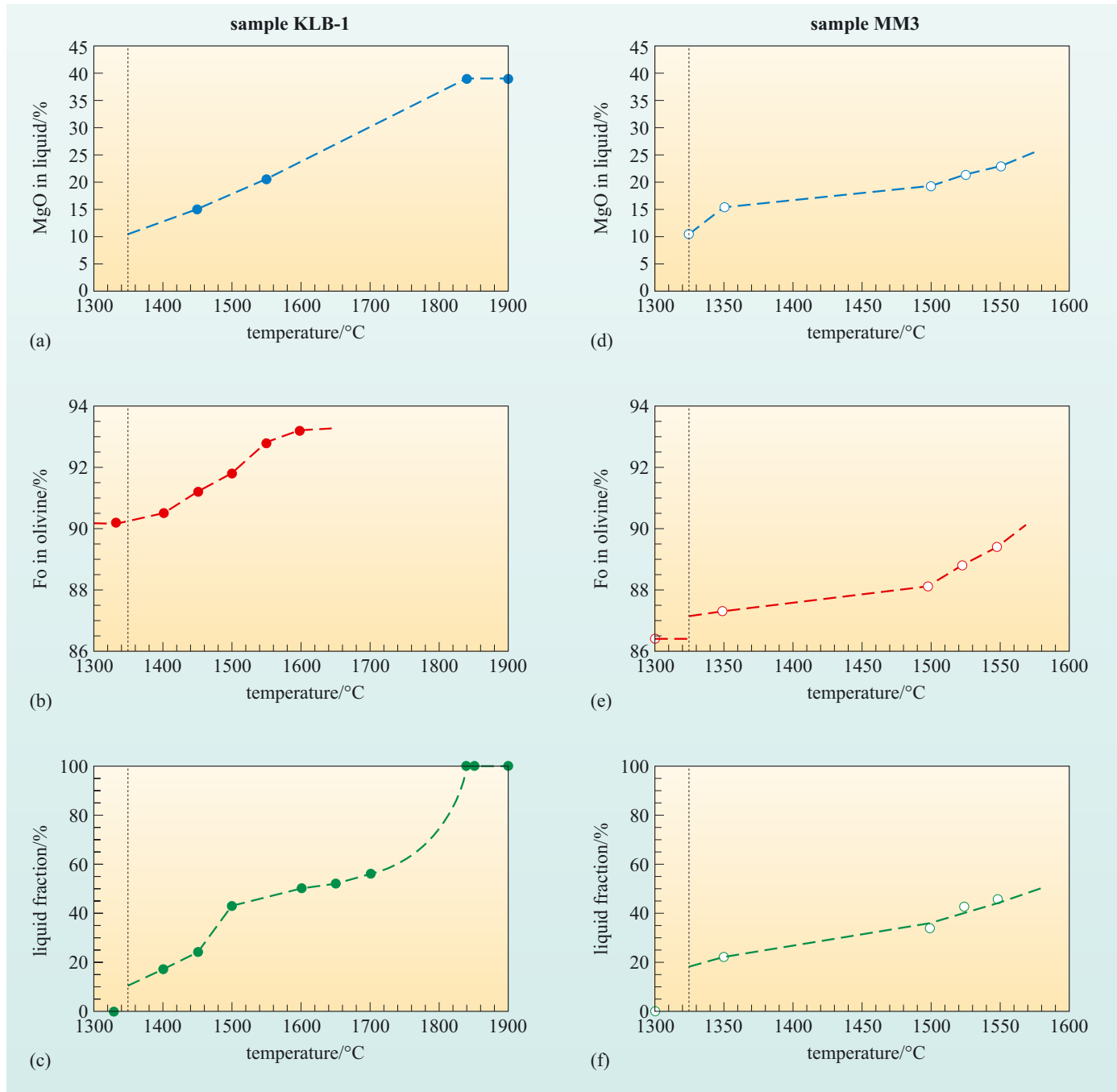


Figure 4.16 Melt fraction, MgO content of liquid, and Fo content of olivine versus temperature for two examples of lherzolite at a pressure of 1.5 GPa: (a–c) KLB-1 and (d–f) MM3. The appropriate solidus temperature is indicated by the vertical dotted line. ((a–c) adapted from Takahashi et al., 1993; (d–f) adapted from Falloon et al., 1999)

Table 4.3 Phases present in KLB-1 at a pressure of 1.5 GPa (Takahashi, 1986).

Temperature/°C	Olivine	Orthopyroxene	Clinopyroxene	Spinel	Liquid
1325	✓	✓	✓	✓	
1350	✓	✓	✓	✓	
1400	✓	✓	✓		✓
1450	✓	✓			✓
1500	✓				✓
1600	✓				✓
1700	✓				✓
1800	✓				✓
1850					✓
1900					✓

Question 4.9

Use Table 4.3 and Figure 4.16 to decide which of the following statements describe correctly the melting behaviour of spinel lherzolite at 1.5 GPa. For those statements that are wrong, rewrite them so that they are correct.

- A The solidus temperature lies between 1300 °C and 1400 °C.
- B The first phase to disappear with increasing temperature is orthopyroxene.
- C Spinel disappears at a lower temperature than clinopyroxene.
- D Orthopyroxene is the mineral phase that survives to the highest temperature.
- E Clinopyroxene disappears at a lower temperature than olivine.
- F With increasing temperature the amount of partial melting increases.
- G The amount of melt produced at the solidus is between 0% and 20%.
- H The Fo content of olivine becomes lower with increasing amount of melting.
- I The MgO content of the liquid decreases with increasing temperature.
- J The MgO content of the liquid produced at temperatures just above the solidus is lower than the MgO content of the peridotite.

The next two sections develop the idea of partial melting by investigating why it is that some materials undergo partial melting rather than melting completely at a single temperature.

4.4 Why does partial melting happen?

To understand the reasons why peridotite does not melt at a single temperature, but undergoes partial melting over a range of temperatures, it is useful to consider some much simpler systems that still have some similarity with

peridotite. The aim of this approach is that, by understanding how a simple system behaves, you can gain some insight into the behaviour of the more complex real system. So, because lherzolite is a mixture of olivine, orthopyroxene, clinopyroxene and a small amount of plagioclase, spinel or garnet, and because each of these minerals forms a solid-solution series, two simple versions are considered. These are a single mineral that forms a solid-solution series (forsterite–fayalite), and a mixture of two pure minerals (forsterite–diopside).

4.4.1 The phase diagram of forsterite–fayalite

As there are two components in this system (forsterite and fayalite, abbreviated to Fo and Fa), it is an example of a **binary system**. And as there is a compositional continuum between the two endmember phases, it is a binary system with a solid solution. The most useful phase diagram of such a system, in terms of understanding its melting behaviour, is a diagram that plots chemical composition against temperature (the pressure is fixed) as shown in Figure 4.17.

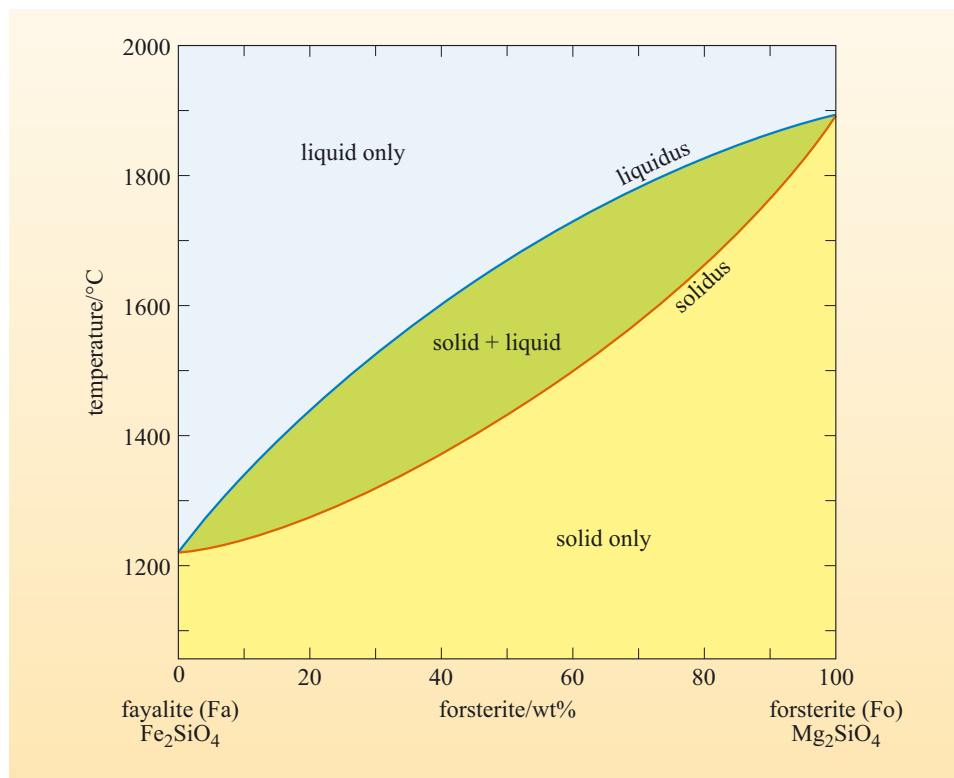


Figure 4.17 Experimentally determined phase diagram of temperature against composition (at atmospheric pressure) for the olivine system.

Composition is plotted on the horizontal axis of Figure 4.17 as the percentage (by weight) of forsterite in olivine. Any olivine composition can be plotted or read off the scale: Fe-rich examples plot to the left and Mg-rich examples plot to the right. The phase diagram is split into three areas by two curves:

- a sample whose composition and temperature plot in the top field (above the liquidus) will be completely molten
- the middle area contains samples in which liquid and solid coexist
- the lowest area (below the solidus) contains samples that are totally solid.

The olivine phase diagram shows how the composition and temperature of liquids and solids are related. For instance, olivine that is 100% forsterite ($\text{Fo}_{100}\text{Fa}_0$) has a liquidus and solidus that coincide at 1890 °C. This is what you would expect from Figure 4.14, which shows that pure forsterite melts at a single temperature of 1890 °C. Pure fayalite ($\text{Fo}_0\text{Fa}_{100}$) also melts at a single temperature, but one that is much less than that of $\text{Fo}_{100}\text{Fa}_0$.

But more interestingly, imagine a number of experiments in which samples of olivine with the composition $\text{Fo}_{50}\text{Fa}_{50}$ are heated to different temperatures. At 1150 °C this composition plots in the solid-only field of the phase diagram (Figure 4.18).

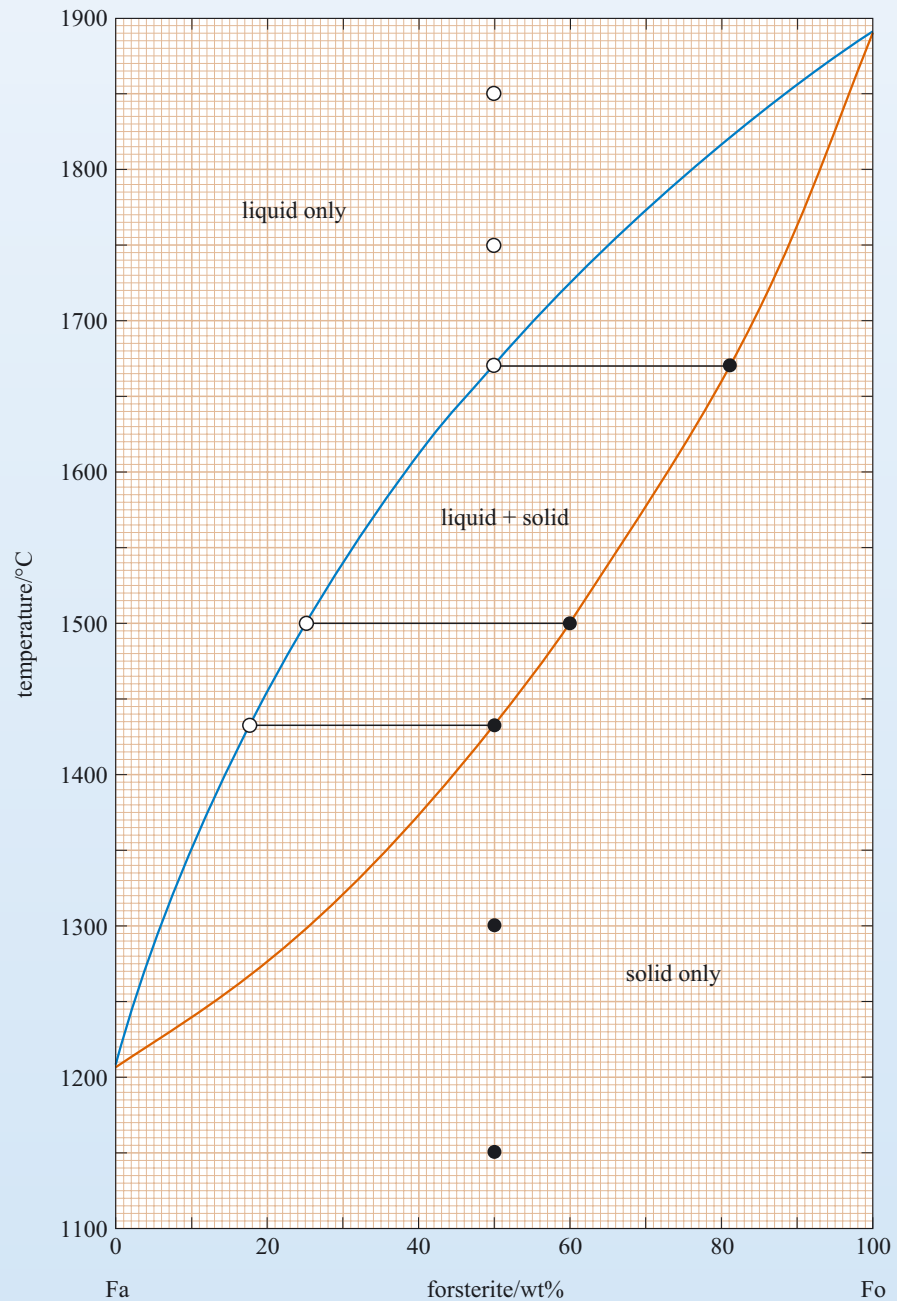


Figure 4.18 Phase diagram of temperature against composition (at atmospheric pressure) for olivine, showing the results of heating $\text{Fo}_{50}\text{Fa}_{50}$. Filled circles give the composition and temperature of solid olivine. Open circles refer to liquid. Horizontal tie lines join the compositions of solid and liquid phases that coexist at particular temperatures.

Likewise, at 1300 °C it is still solid. As the temperature is increased, the solidus is intercepted at 1433 °C; this is the lowest temperature at which solid and liquid can coexist in a system of composition $\text{Fo}_{50}\text{Fa}_{50}$.

What will be the composition of the first drop of liquid that appears at the solidus? The liquidus is the curve showing the temperature and composition of liquids that coexist with solid. And the solidus is the curve showing the temperature and composition of solids that coexist with liquid. Thus, when melting starts at 1433 °C, the liquid composition can be read off the phase diagram as $\text{Fo}_{19}\text{Fa}_{81}$ (to the nearest 1%). A **tie line** has been drawn between phases that coexist at any one time, so the tie line at 1433 °C shows that liquid $\text{Fo}_{19}\text{Fa}_{81}$ coexists with olivine crystals of $\text{Fo}_{50}\text{Fa}_{50}$ composition at this temperature. Because the overall composition is still $\text{Fo}_{50}\text{Fa}_{50}$, the proportion of liquid must be infinitesimally small.

At higher temperatures, the proportion of liquid increases at the expense of solid. At 1500 °C the bulk composition, $\text{Fo}_{50}\text{Fa}_{50}$, plots between the solidus and liquidus curves and so the sample is partially molten at this temperature. Drawing in the horizontal tie line identifies the liquid composition as $\text{Fo}_{25}\text{Fa}_{75}$ and this coexists with solid $\text{Fo}_{60}\text{Fa}_{40}$. In order for the bulk composition of this mixture to be $\text{Fo}_{50}\text{Fa}_{50}$ there must be more solid than liquid.

At 1670 °C the bulk composition, $\text{Fo}_{50}\text{Fa}_{50}$, plots on the liquidus.

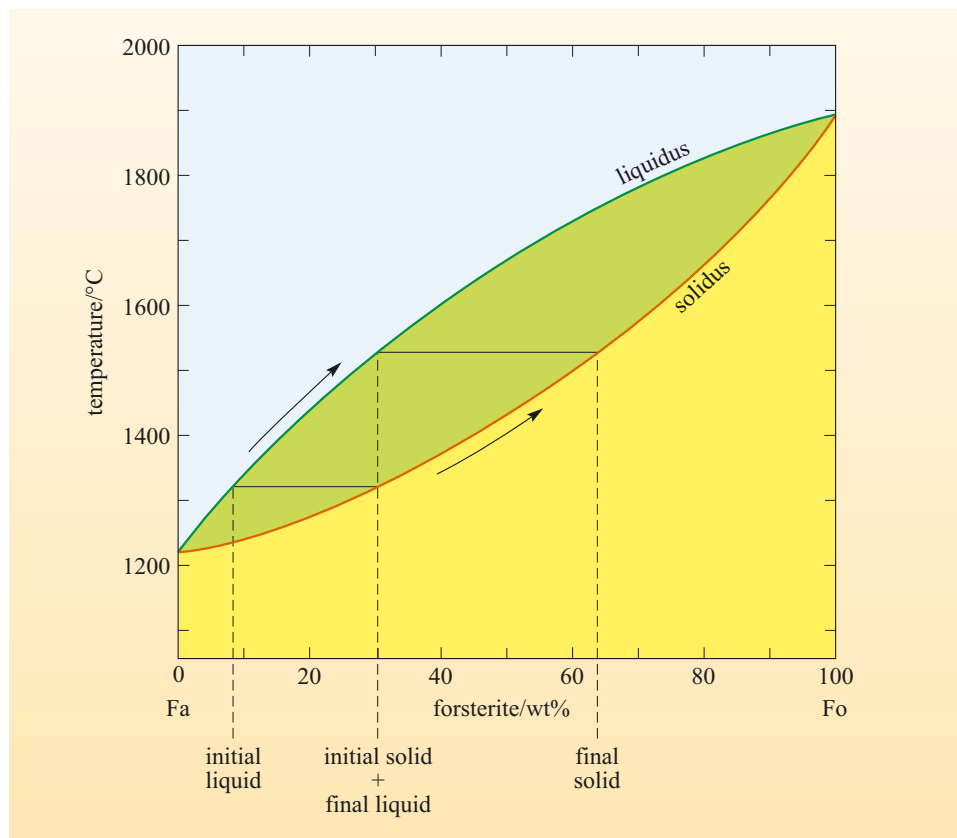
- What does this mean?
- This is the highest temperature at which solid can still exist within a system of bulk composition $\text{Fo}_{50}\text{Fa}_{50}$.

Here, the liquid has composition $\text{Fo}_{50}\text{Fa}_{50}$ and, as the position of the tie line indicates, it coexists with an infinitesimally small amount of solid $\text{Fo}_{81}\text{Fa}_{19}$ olivine. Liquid $\text{Fo}_{50}\text{Fa}_{50}$ is the only phase present at temperatures above 1670 °C (e.g. the points at 1750 °C and 1850 °C on Figure 4.18 are both in the liquid-only field).

The behaviour of a binary solid-solution system (e.g. olivine) can be summarised as follows.

- 1 Partial melting occurs in a system showing solid solution.
- 2 The amount of liquid produced increases continuously from 0% at the solidus to 100% at the liquidus. Strictly speaking, an infinitesimally small amount of liquid is present at the solidus and an infinitesimally small amount of crystals is present at the liquidus.
- 3 During partial melting the liquid contains a higher concentration, and the solid contains a lower concentration, of the low melting-temperature endmember of the solid-solution series. Thus, liquids formed by partial melting of olivine will always be richer in Fa, and the solids richer in Fo, than the bulk composition.
- 4 As the temperature increases, both the liquid and the solid become richer in the high melting-temperature component (forsterite in the case of olivine solid solution). This is illustrated in Figure 4.19 for a composition of $\text{Fo}_{30}\text{Fa}_{70}$.

Figure 4.19 The compositions of coexisting solid and liquid evolve along the solidus and liquidus respectively as temperature increases. The example shows the behaviour of a sample of composition $\text{Fo}_{30}\text{Fa}_{70}$.



Points 3 and 4 are important because they mean that partial melting produces liquids with different chemical compositions from that of the material being melted. Also, partial melting will leave a residual solid whose composition is different from that of the initial material because it will be depleted in the components that preferentially entered the melt.

4.4.2 The phase diagram of forsterite–diopside

Olivine is not the only mineral present in lherzolite, so an alternative to thinking about the mantle in terms of the solid solution Fo–Fa is to consider a mixture of two minerals, such as olivine and orthopyroxene, or olivine and clinopyroxene. To choose an appropriate mixture to help understand partial melting, recall the idea put forward in Section 4.2.3 that harzburgite is the solid residue left behind when lherzolite melts to produce basalt liquid. This involves a decrease in the amount of clinopyroxene as shown, for example, by the shift in composition on Figure 4.12 from AX (lherzolite) to AO (average harzburgite), away from the clinopyroxene corner. So, the melting behaviour of mixtures of olivine (the dominant mineral in peridotite) and clinopyroxene should be a useful model. Furthermore, the Mg endmembers forsterite (Mg_2SiO_4) and diopside ($\text{CaMgSi}_2\text{O}_6$) are appropriate simplifications for the compositions of these minerals in the mantle.

The phase diagram of the binary system forsterite–diopside (Fo–Di) at a pressure of 1.5 GPa is given in Figure 4.20. At this pressure pure Fo melts at 1965 °C (Figure 4.14) and pure diopside melts at 1590 °C.

At what temperature will melting start to happen when a mixture of forsterite and diopside crystals is heated?

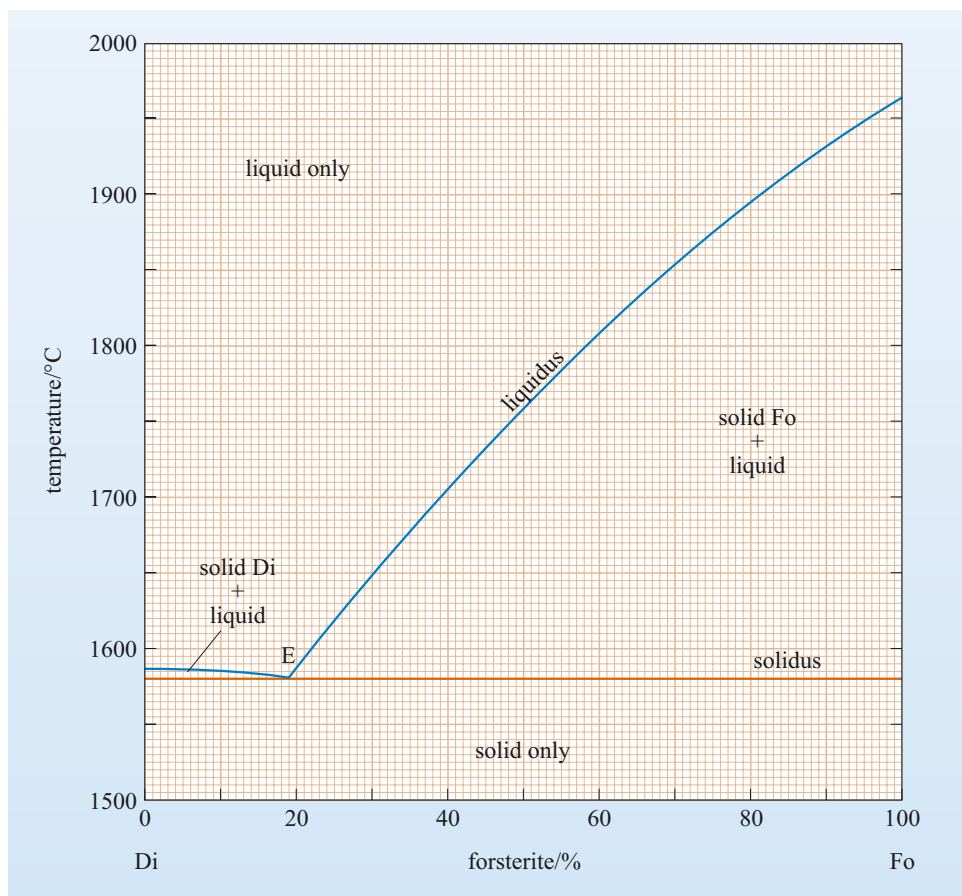


Figure 4.20 Phase diagram of temperature against composition for the system forsterite–diopside at 1.5 GPa pressure. Point E is the eutectic.

On the Fo–Di phase diagram (Figure 4.20), the solidus is a horizontal line at a temperature of 1580 °C, so any mixture of forsterite and diopside in whatever proportions will start to melt at 1580 °C (and 1.5 GPa), which is a lower temperature than the melting temperature of either pure endmember! The coexisting liquid at this temperature lies on the liquidus at 1580 °C, which is at the point labelled E. This point is known as the **eutectic point**. The eutectic point (or just eutectic) is the only location on the phase diagram where the three phases: solid diopside, solid forsterite and liquid coexist. The composition of this liquid is read from the phase diagram as Fo₁₉Di₈₁; this is the **eutectic composition**. To recap, on heating any mixture of diopside and forsterite at 1.5 GPa, melting will start at 1580 °C, the liquid will have a composition of Fo₁₉Di₈₁ and it will coexist with crystals of Fo₁₀₀ and Di₁₀₀.

What happens at higher temperatures depends on the bulk composition. Consider a sample with the composition Fo₆₀Di₄₀ at a temperature of 1700 °C. This plots in the solid forsterite + liquid field on Figure 4.20, so the sample is now partially molten. The compositions of the two phases present at this temperature are found by drawing the horizontal tie line at 1700 °C, which shows that solid forsterite will coexist with a liquid of composition Fo₃₉Di₆₁.

Question 4.10

What are the compositions of the coexisting phases when a sample of Fo₆₀Di₄₀ is at 1800 °C?

The sample $\text{Fo}_{60}\text{Di}_{40}$ plots on the liquidus at 1810°C . This is the temperature at which the last crystals of forsterite dissolve into the liquid; the sample is completely molten at all temperatures above 1810°C .

Question 4.11

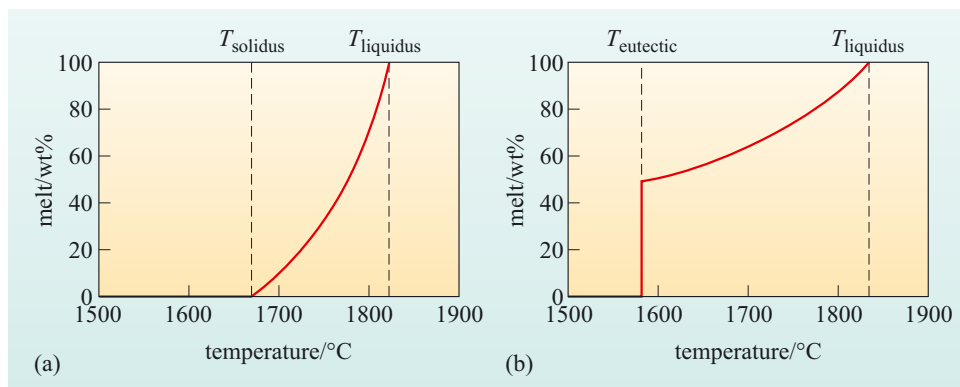
Consider a sample with the composition $\text{Fo}_{10}\text{Di}_{90}$ and use the Fo–Di phase diagram (Figure 4.20) to answer the following.

- When this sample is heated from 1500°C to 1700°C , at what temperature will a liquid first appear?
- Which mineral is consumed first?
- What is the composition of the first liquid that is generated?
- At what temperature does the sample become completely molten?

The answer to Question 4.11b reveals an important aspect of eutectic systems: the first mineral to be completely melted in a eutectic system depends on the relative compositions of the sample and the eutectic. Thus, although pure Di has a much lower melting point than pure Fo, it will be forsterite that disappears first if the sample contains less than 19% Fo.

The amount of liquid that is produced at the eutectic temperature depends on the bulk composition of the sample. In our example of $\text{Fo}_{60}\text{Di}_{40}$, all but an infinitesimally small amount of the diopside crystals is consumed at the eutectic, so it might be tempting to jump to the conclusion that 40% of the sample becomes liquid at the eutectic. This cannot be the case, however, because the eutectic liquid contains some forsterite (the eutectic composition is $\text{Fo}_{19}\text{Di}_{81}$), so more than 40% of the sample is molten at 1580°C . (In fact, in this example, 49% of the sample becomes liquid at the eutectic in order to allow the composition of the eutectic liquid to be derived from the bulk composition of the sample.) This specific example illustrates another important general principle – the amount of liquid produced at the onset of melting in a eutectic system can be large, whereas at the onset of melting in a solid-solution series the initial amount of liquid is infinitesimally small. In both of these model systems the proportion of liquid gradually increases with increasing temperature until it reaches 100% at the liquidus temperature (Figure 4.21).

Figure 4.21 Graphs of wt% melt against temperature illustrating the different behaviours of eutectic and solid-solution systems. (a) Binary solid solution. The curve is for $\text{Fo}_{80}\text{Fa}_{20}$ in the Fo–Fa system at 10^5 Pa (atmospheric pressure). (b) Binary eutectic system. The curve shows the example $\text{Fo}_{60}\text{Di}_{40}$ in the Fo–Di system at 1.5 GPa discussed in the text.



4.4.3 Partial melting of lherzolite and the generation of basalt

You are now in a position to understand some of the main results that were found in the experiments on peridotite at high pressure and high temperature (Section 4.3.2, Table 4.3, Figure 4.16). Because lherzolite contains different mineral phases, it has a similarity with eutectic systems (e.g. Fo–Di) and, because each of the mineral phases forms a solid-solution series, there are also similarities with systems such as Fo–Fa.

Question 4.12

For each of the following statements describing the partial melting of peridotite, decide whether it is explained by the way solid solutions and/or eutectic systems behave.

- A At a given pressure, partial melting occurs over a temperature range of several hundred degrees Celsius.
- B The olivine composition of the unmelted residue has a higher Fo content than the olivine in the original, unmelted peridotite.
- C The MgO content of the liquid increases as the degree of partial melting increases.
- D At least 10% of liquid is produced at temperatures just above the solidus.
- E Clinopyroxene disappears at a lower temperature than olivine.
- F Spinel disappears at a lower temperature than clinopyroxene.

Lherzolite xenoliths contain a much higher proportion of olivine than clinopyroxene, as seen from the way that peridotites plot near the olivine corner and near the olivine–orthopyroxene edge of the classification diagram for ultrabasic rocks (Figure 4.12). For instance, the sample KLB-1 whose phase diagram is shown in Figure 4.15 has 58% olivine, 25% orthopyroxene, 15% clinopyroxene, and 2% spinel. In experiments with this sample at 1.5 GPa, no clinopyroxene persists above about 1410 °C, but olivine is always present until the liquidus is reached at 1830 °C. Could this have been predicted from what is known about the Fo–Di phase diagram at 1.5 GPa (Figure 4.20)? In KLB-1, 73% of the rock is composed of olivine and clinopyroxene, so a simplified version of KLB-1 could contain

$$\frac{58\%}{73\%} \times 100\% = 79\% \text{ olivine and } \frac{15\%}{73\%} \times 100\% = 21\% \text{ clinopyroxene.}$$

Making another approximation by considering that olivine can be represented by forsterite, and clinopyroxene can be represented by diopside, then KLB-1 can be approximated by a mixture with the composition Fo₇₉Di₂₁.

- What is the melting behaviour of Fo₇₉Di₂₁?
- According to Figure 4.20, this mixture will melt at the eutectic at 1580 °C, with the disappearance of Di and the production of a eutectic liquid. At higher temperatures, as the melt fraction increases, the amount of solid Fo decreases and the liquid becomes richer in the Fo component. At the liquidus temperature of 1890 °C, all of the olivine has melted.

The model composition, therefore, mimics the real system in showing that clinopyroxene will melt before olivine. However, the model predicts a higher solidus temperature and a higher liquidus temperature. The reason for this is that these phase boundaries generally shift to lower temperatures as more chemical components are added. Thus at 1.5 GPa, pure Fo (Figure 4.14) melts at a higher temperature than a Fo–Di mixture (Figure 4.20), which in turn melts at a higher temperature than KLB-1 (Figure 4.15). When lherzolite melts, Al-bearing phases and then clinopyroxene melt completely at lower temperatures than orthopyroxene, while olivine melts at even higher temperatures.

4.4.4 Summary of Section 4.4

The partial melting of lherzolite (and other substances) can be understood using phase diagrams of composition versus temperature for solid-solution series and eutectic systems. Instructive systems that are simplified versions of lherzolite include the forsterite–fayalite (Fo–Fa) and forsterite–diopside (Fo–Di) systems. These show the following features.

The compositional phase diagram of a *binary solid-solution series* is characterised by a loop-shaped liquidus and solidus. This is exemplified by the olivine solid-solution series Fo–Fa (Figure 4.17). Here, partial melting of a given sample produces a liquid that is richer in Fa (the component with the lower melting temperature) than the overall bulk composition. The liquid forms at the expense of solid olivine that, therefore, becomes richer in Fo than the bulk composition. The compositions of coexisting solid and liquid phases plot on the solidus and liquidus curves at either end of a horizontal tie line. As melting progresses (as temperature increases), the compositions of the liquid and coexisting crystals become richer in the high-temperature endmember (Fo). In the melting interval between the solidus and liquidus temperatures, the proportion of melt increases steadily from 0% to 100%.

The partial melting behaviour of lherzolite shows features that are similar to the partial melting behaviour of olivine. There is a continuously smooth increase in melt fraction with temperature, and residual ferromagnesian minerals are richer in Mg (and poorer in Fe) than the same minerals in unmelted samples. The chemical composition of the partial melt also varies with the degree of melting, becoming more Mg rich as the melt fraction increases, analogous to the increase in Fo seen in the Fo–Fa system.

A *binary eutectic system* has a horizontal solidus on a compositional phase diagram (e.g. Figure 4.20). This means that partial melting will start at the same temperature irrespective of the bulk composition. The eutectic temperature is lower than the melting temperature of the system's endmembers. The eutectic point defines the composition and temperature of the first melt that forms. Partial melting at the eutectic produces a significant quantity of liquid, the exact amount of which depends on the composition of the bulk sample and the eutectic composition. This contrasts with the infinitesimally small amount of liquid generated at the solidus temperature in a solid-solution series.

In eutectic systems, the composition of the liquid reflects the amount of melting due to the changing proportions of the different minerals that contribute to the liquid. The phase that is less abundant in the bulk sample compared with the

eutectic composition will melt first at the eutectic temperature. This feature is observed when lherzolite melts, with Al-bearing phases and then clinopyroxene melting completely at lower temperatures than orthopyroxene and olivine melting at an even higher temperature. However, melting at the lherzolite solidus does not produce a large quantity of liquid, as in a eutectic system. This is because a solid solution in mantle minerals smoothes out the step-like increases in liquid fraction that characterise melting in eutectic systems.

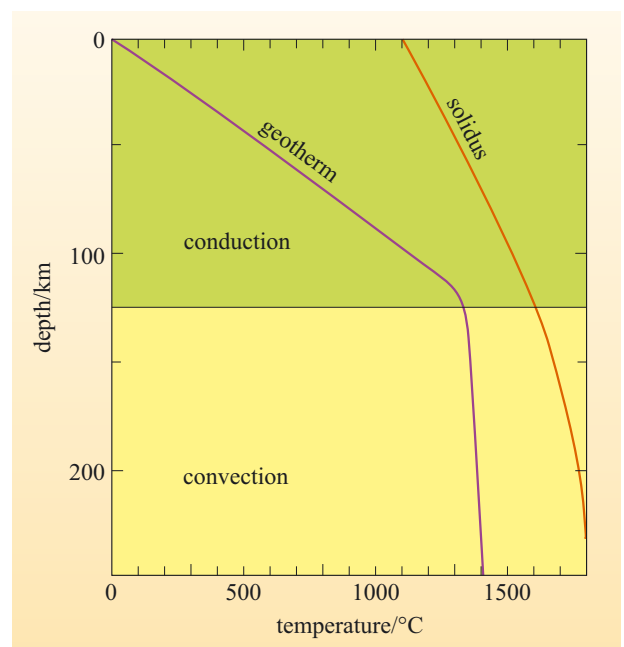
4.5 Magma generation at constructive plate boundaries

Pressures and temperatures in the mantle beneath mid-ocean ridges are large enough for the mantle to experience partial melting to produce basalt magma. This section starts by outlining the general mechanism for producing magmas at constructive plate boundaries (Section 4.5.1) and then shows how these conditions depend on mantle temperature (Section 4.5.2) so that observations of oceanic crust can be used to infer temperature conditions in the upper mantle (Section 4.5.3).

4.5.1 Crossing the solidus

The temperature structure beneath old oceanic lithosphere (with the exception of hot-spot volcanoes; a separate issue that we return to later in Chapter 7) is not capable of generating magmas. In areas of old oceanic lithosphere, the surface heat flux is $\leq 50 \text{ mW m}^{-2}$ (Box 3.2) and this heat comes partly from the decay of trace amounts of radioactive U, Th and K and partly from the cooling of the already hot interior. Convection currents in the asthenosphere move heat to the base of the lithosphere, and that heat is then conducted to the surface through the lithosphere, resulting in a geotherm like that shown in Figure 4.22. In the convecting mantle, descending currents of mantle become compressed, causing them to warm slightly as a result of their atoms being squashed together. The reverse happens when rising mantle is decompressed – it cools slightly. (The same effect accounts for the drop in atmospheric temperature with increasing altitude.) In these cases, a rising or falling segment of mantle does not lose or gain heat energy – the heat energy only leaves the convecting system across the steep temperature gradient at the lowest part of the conductive part of the geotherm. The geotherm in the convecting mantle is therefore said to be **adiabatic** (from the Greek for impassable – a reference to the fact that heat passes neither from nor to a substance undergoing an adiabatic process). The temperature gradient in the lithosphere is conductive and, in the underlying mantle where convection occurs, it is adiabatic. Furthermore, the geotherm always lies below the mantle solidus, which explains why the older parts of oceanic basins do not have any volcanic activity.

Figure 4.22 The geotherm and mantle solidus in old oceanic lithosphere. The solidus curve is an average taken from experiments on slightly different lherzolite samples.



At mid-ocean ridges the surface heat flux is very high (>100 up to $>300 \text{ mW m}^{-2}$), which is an immediate indication that very hot material lies close to the surface. At the ridge, oceanic lithosphere splits apart at rates of a few centimetres per year, so deep material must upwell at the same rate to fill the gap and then flow sideways in a pattern like the one shown in Figure 4.23a (see also Figure 4.13). This flow pattern underneath the plate boundary brings deep, and therefore hot, mantle material towards the surface. Its path is illustrated in Figure 4.23b.

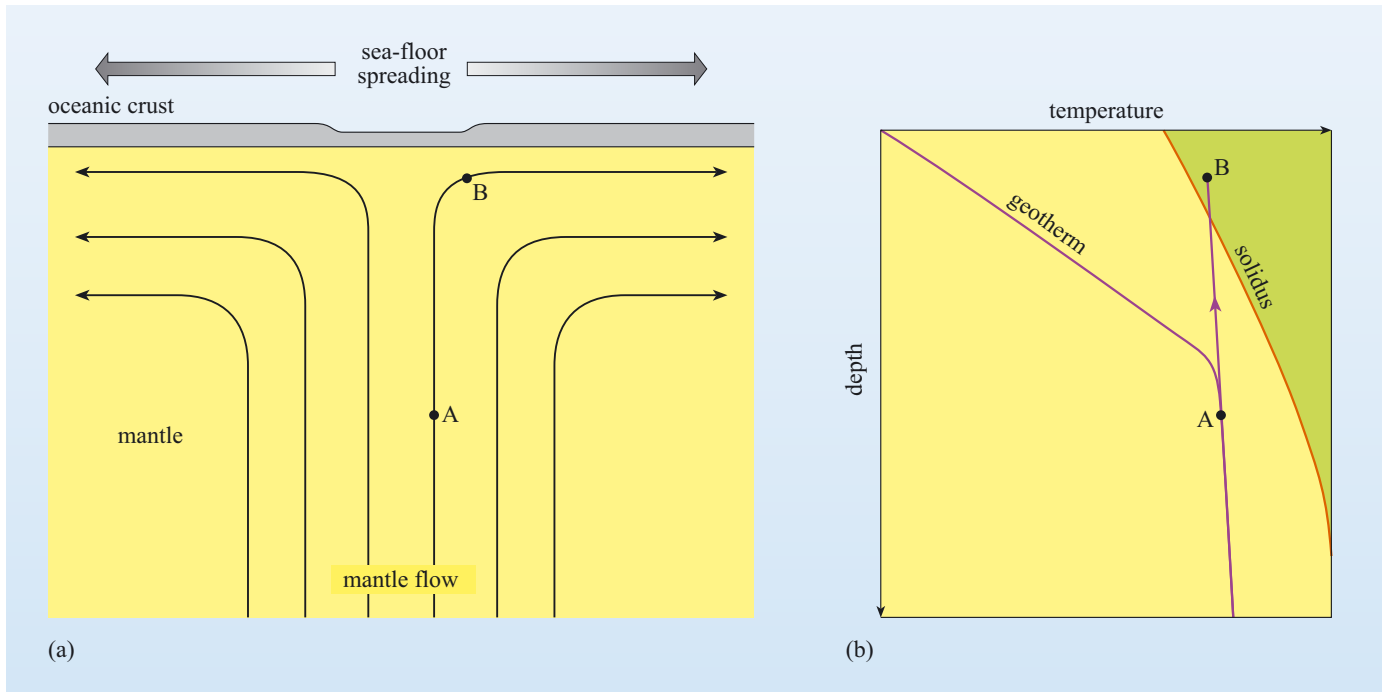


Figure 4.23 (a) Sketch cross-section of a constructive plate boundary. The crust moves apart, opening a gap that becomes filled with basalt escaping from the underlying mantle (see also Figure 4.13). The mantle rises below the plate boundary and then moves sideways to follow the divergent paths of the crust. A parcel of mantle that is originally at point A moves to point B. Its pressure and temperature decrease along the adiabatic gradient as shown in (b) and it crosses the mantle solidus during its journey. ((a) Adapted from Langmuir et al., 1993)

Deep beneath a constructive plate boundary, a region of mantle within the convecting asthenosphere (labelled point A) moves upwards in response to the separation of the plates at the surface. It follows an adiabatic path, so its temperature decreases slightly as it rises along an adiabatic gradient and by the time it has risen to point B it has crossed the solidus and will be partially molten – the basaltic magma needed to make oceanic crust has been created. The change from solid mantle to partially molten mantle has therefore been caused by a decrease in pressure, so it is referred to as **decompression melting**. In this case, partial melting is not caused by heating but by decompression caused by large-scale flow in the mantle that is induced by plate divergence.

Figure 4.23 shows how decompression melting can happen beneath a mid-ocean ridge, and this general concept can be used to link together two aspects of the Earth – the temperature of the upper mantle and the composition and structure of oceanic crust. This link comes about because of a chain of connections:

- 1 The amount of partial melting will determine how much basalt is generated, and this in turn will determine how thick the oceanic crust is.
- 2 The amount of partial melting depends on the pressures and temperatures reached within the zone of partial melting where the magma is generated (in other words, how far the decompressed mantle – point B in Figure 4.23 – lies above the solidus).
- 3 These P and T conditions depend on the initial temperature of the rising mantle and the amount by which it decompresses.

4.5.2 The temperature of the convecting mantle

The temperature in the convecting mantle varies along the adiabatic gradient, decreasing by about $1\text{ }^{\circ}\text{C}$ for every 3 km of ascent. To specify the temperature, a depth, or pressure, needs to be specified. The conventional way of doing this is to extrapolate the **adiabat** to the surface, which is at a pressure of 10^5 Pa , as shown in Figure 4.24a. This extrapolated temperature is known as the **potential temperature** (T_p).

If T_p is less than the solidus temperature at atmospheric pressure (about $1120\text{ }^{\circ}\text{C}$ from Figure 4.15), then decompression melting would not happen (Figure 4.24b). This means that T_p for the Earth's mantle must be greater than $1120\text{ }^{\circ}\text{C}$. When mantle with a higher T_p rises, it will cross the solidus at some pressure, P_s , where the temperature is T_s (Figure 4.24c). For the melting reaction:

solid phases \rightarrow liquid + residual solid phases

to happen, some energy is needed. This energy (the latent heat of fusion) is supplied by a change in temperature, causing the system to cool as it undergoes more and more melting. Because of the phase changes involved in partial melting, the actual temperature path that the mantle follows once it has crossed the solidus will be similar to the one shown in Figure 4.24c, where T_{act} is the actual temperature that the rising parcel of mantle would have if it rose all the way to the surface, and this is less than T_p .

Sea-floor spreading allows the mantle to come very close to the surface because the oceanic lithosphere at the plate boundary is extremely thin. In detail, the amount of melting achieved will increase as the mantle reaches shallower and shallower levels after it has crossed the solidus because the path it follows moves it farther away from the solidus. Figure 4.24d illustrates two hypothetical situations in which mantles with different potential temperatures undergo decompression melting that extend to the surface.

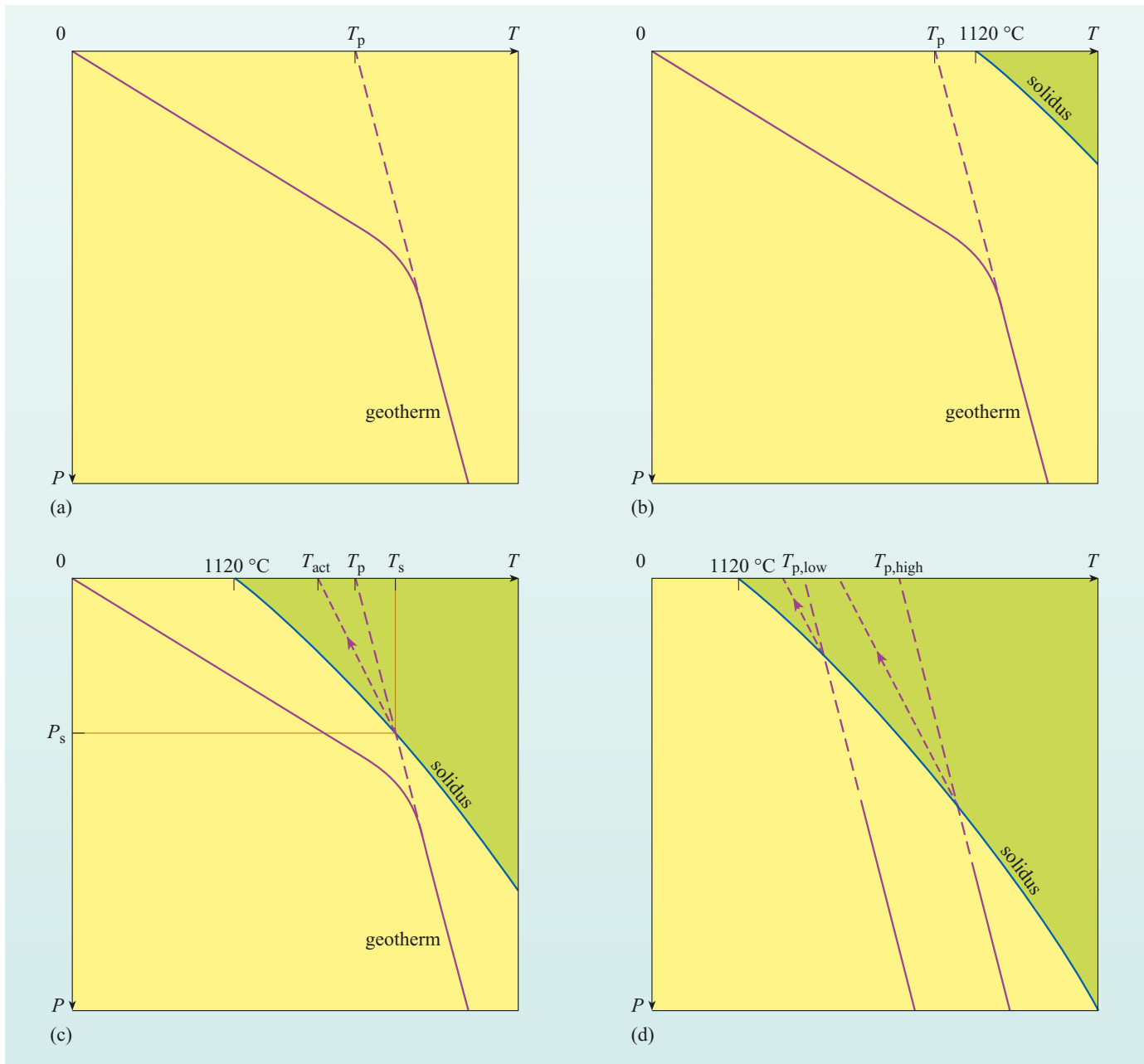


Figure 4.24 (a) Sketch showing how the extrapolation of temperature (dashed line) along the adiabatic gradient from any depth in the convecting mantle to the surface defines the potential temperature (T_p) of the convecting mantle. (b) Mantle with T_p less than about 1120°C cannot cross its solidus by decompressing. (c) Mantle with T_p greater than 1120°C crosses the solidus at (T_s, P_s) and continues to decompress along the path leading to $(T_{\text{act}}, 0)$. (d) Decompression paths of mantle with potential temperatures of $T_{p,\text{low}}$ and $T_{p,\text{high}}$ (see Question 4.13).

Question 4.13

Of the two cases ($T_{p,\text{low}}$ and $T_{p,\text{high}}$) shown in Figure 4.24d, which one will cause partial melting to start at the greater depth, and which will produce the larger amount of liquid?

The answers to Question 4.13 lead to the interesting conclusion that the amount of basaltic magma (and therefore the thickness of basaltic oceanic crust) that can be produced at a mid-ocean ridge will depend on the potential temperature of the mantle. Therefore a measurement of the thickness of oceanic crust can, in principle, be used as a ‘thermometer’ with which to estimate the temperature of the Earth’s upper mantle. This idea is followed to its logical conclusion in the next subsection.

4.5.3 The thickness of oceanic crust and the potential temperature of the mantle

To restate the conclusion of the last section, the amount of liquid produced by decompression melting of the mantle will depend on the potential temperature of the mantle and the lowest pressure reached by the rising mantle. To make this idea quantitative requires measurements of the amount of melting at different pressures and temperatures above the solidus of mantle peridotite. Among the first researchers to take this step were Dan McKenzie and Mike Bickle (Cambridge University, UK), who compiled data from high-pressure, high-temperature experiments and used them to predict the melting behaviour of mantle peridotite. Their study was published in 1988 and, although it continues to be embellished by further experimental data, it sets out the basic quantitative correlation between potential temperature and the extent of decompression melting, under the simplifying assumption of decompression to atmospheric pressure. For a given potential temperature, their model predicts the depth at which partial melting would start, the amount of basalt produced, and the maximum proportion of melting. Their results are summarised in Figure 4.25.

- The average thickness of oceanic crust is about 7 km. What does this imply about the mantle’s potential temperature?
- According to the model of Figure 4.25a, this implies a potential temperature of 1280 °C.

As well as helping to explain the basaltic product of partial melting, the model in Figure 4.25 helps to explain why harzburgite is found in the mantle sections of ophiolites and oceanic peridotites. According to Figure 4.25, a potential temperature of 1280 °C should result in up to 25% melting, and this is similar to the amount of melting at which clinopyroxene no longer remains in the mineral assemblage of partially molten peridotite in laboratory experiments. The model predicts also that partial melting will start at a depth of 42 km, which is in the spinel–peridotite field (Figure 4.15).

To summarise, the calculations of partial melting during adiabatic ascent of lherzolite beneath oceanic spreading ridges account for production of 7 km thick basaltic crust and complementary harzburgite if the mantle has a potential temperature of 1280 °C. Furthermore, the model predicts about 25% melting at the top of a zone of partial melting that extends to a depth of about 40 km.

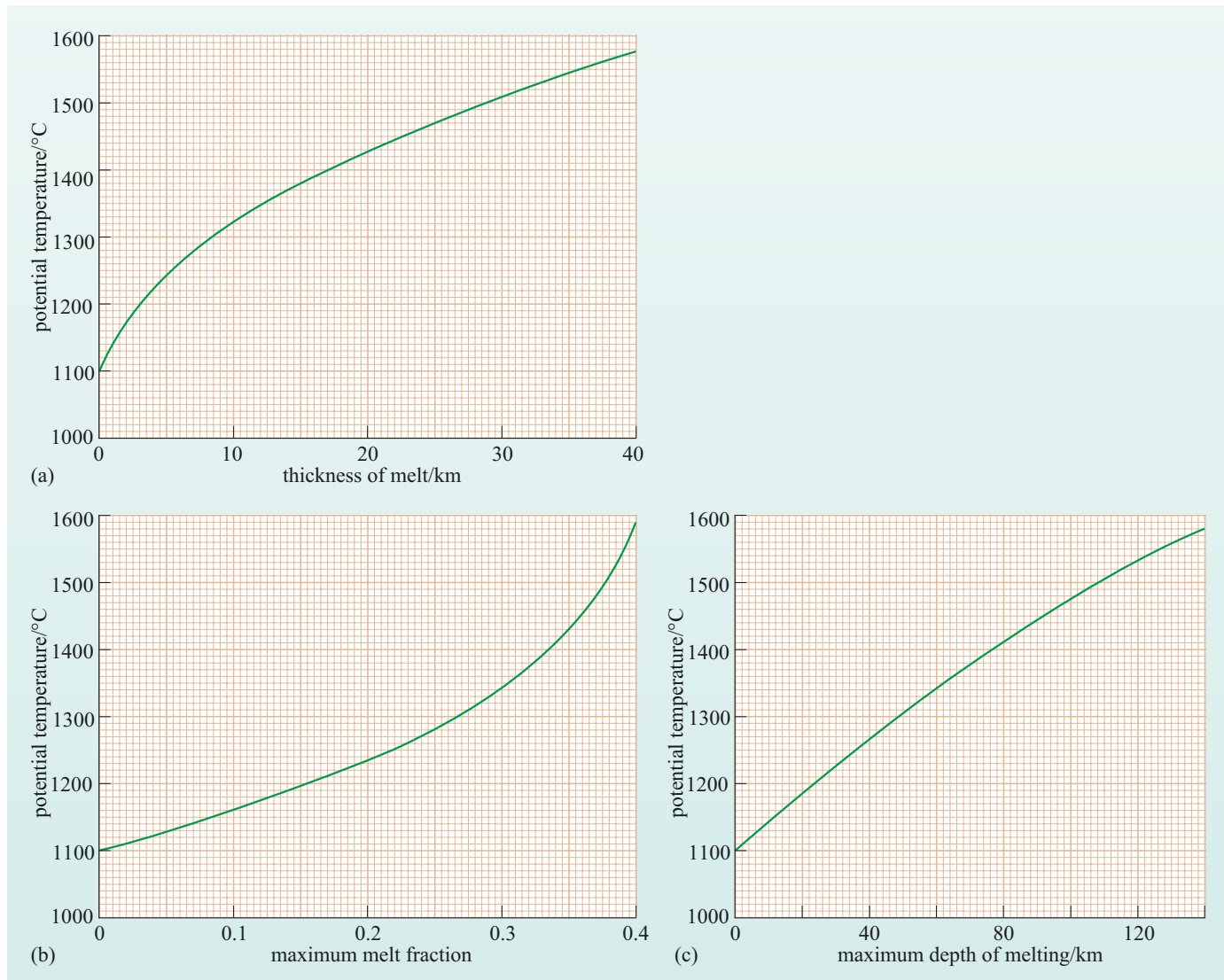


Figure 4.25 Calculated relationships between the potential temperature of the mantle and: (a) thickness of melt produced and extracted; (b) maximum mass fraction of partial melt in the melting zone; and (c) maximum depth at which melting of adiabatically ascending mantle will start to partially melt. The model assumes that decompression continues to atmospheric pressure (10^5 Pa). (Adapted from McKenzie and Bickle, 1988)

Question 4.14

Iceland lies on the Mid-Atlantic Ridge and is part of the constructive plate boundary between the Eurasian and North American Plates. Unusually, this part of the mid-ocean ridge system rises above sea level, and the Icelandic crust has an average thickness of 27 km. What is the potential temperature of the mantle beneath Iceland that is implied by this crustal thickness? Use Figure 4.25 to help you.

4.6 From mantle to crust

Decompression melting produces basaltic magma in the mantle beneath mid-ocean ridges. Being less dense than the residual harzburgite, the basalt liquid is buoyant and therefore migrates towards the surface where it forms oceanic crust, either by erupting onto the sea floor where it forms lava flows, or intruding and crystallising at depth to form the gabbro that makes up the lower oceanic crust. When lava erupts into the cold ocean water at a mid-ocean ridge, the hot lava spreads out from the volcanic vents. Its surface chills to a glassy skin that stretches, breaks and bursts open repeatedly, building up a bewildering accumulation of tubes, channels, ponds and mounds of lava. The chilled rind of these submarine lava flows is composed of dark, glassy rock containing a few crystals of olivine, plagioclase and sometimes clinopyroxene and it represents magma quenched from the state in which it erupted. The larger crystals can reach sizes of up to a few millimetres across and are referred to as **phenocrysts**. Because the glassy sample was quenched by water, the phenocrysts (e.g. the visible plagioclase and olivine crystals in the basalt of Figure 4.3a) must have been present before the lava became solid, and have been carried in the magma as it travelled to the surface. This section shows how the composition of lavas such as these can be used to discover the history of the magma after it has left its source region in the upper mantle. It starts by comparing the magma generated in the mantle with the magma that erupts on the sea floor.

4.6.1 Chemical compositions of partial melts and mid-ocean ridge basalt

If **mid-ocean ridge basalt** (often abbreviated as **MORB**) has come directly from the mantle, then it will be in equilibrium with olivine and the other residual solid phases in its peridotite source region. You can see from the olivine phase diagram (Section 4.4.1) that liquids that are rich in Mg and poor in Fe will be in equilibrium with olivine that is also rich in Mg and poor in Fe. In principle, given the composition of a liquid (such as a basalt lava from a mid-ocean ridge), this can be checked to see whether it is in equilibrium with the olivine found in the mantle source region of basalts, which is taken to be Fo₈₈₋₉₂ (Section 4.2.1). If the liquid lava is in equilibrium with mantle olivine, then it can be assumed that the magma has come directly from the mantle.

The Canadian geologists Peter Roeder and Ron Emslie were the first to demonstrate experimentally the systematic correlation between the compositions of coexisting basaltic liquids and olivine (Figure 4.26). Their results, which have been corroborated by several hundred more experiments carried out over a wide range of temperatures, pressures and liquid compositions, show that the ratio of the number of moles of Fe²⁺ atoms to Mg atoms in the olivine and in the liquid are related by the equation:

$$\left(\frac{\text{Fe}^{2+}}{\text{Mg}} \right)_{\text{olivine}} = 0.3 \left(\frac{\text{Fe}^{2+}}{\text{Mg}} \right)_{\text{liquid}} \quad (4.4)$$

An important point here is that the iron in a magma is present as a mixture of Fe^{2+} and Fe^{3+} . In most basalts between about 5% and 20% of the total iron is present as Fe^{3+} , whereas the iron in olivine is exclusively Fe^{2+} . This is why care has been taken in specifying in Equation 4.4 that the iron that is relevant is Fe^{2+} .

Equation 4.4 can also be expressed in terms of the distribution coefficient (denoted K_D), which is defined by the gradient of the straight line passing through the data in Figure 4.26:

$$K_D = \frac{\left(\frac{\text{Fe}^{2+}}{\text{Mg}}\right)_{\text{olivine}}}{\left(\frac{\text{Fe}^{2+}}{\text{Mg}}\right)_{\text{liquid}}} = 0.3 \quad (4.5)$$

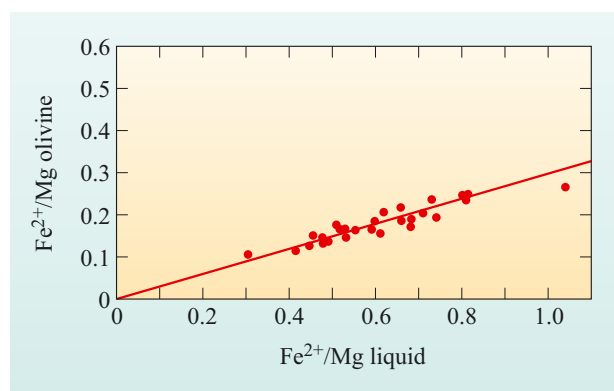


Figure 4.26 The results of experiments by Roeder and Emslie (1970), demonstrating that the ratio of the number of Fe^{2+} to Mg atoms in olivine is proportional to the Fe^{2+}/Mg ratio in the liquid with which it is in equilibrium. (Adapted from Roeder and Emslie, 1970)

This is a very useful result because it means that, without having to know the temperature or pressure, it is possible to test whether a given olivine composition and a given liquid composition are in chemical equilibrium. The practicalities of this approach are demonstrated by considering a specific example.

First, note that the chemical equilibrium between olivine and liquid is based on a balance between the number of Mg and Fe^{2+} atoms in the olivine and in the liquid. However, conventional chemical analyses of rocks (e.g. Table 4.2) express the chemical composition in terms of the weight per cent of the oxides MgO and FeO, and *not* the proportions of atoms. The atomic proportions are calculated from the wt% oxides by dividing the wt% by the relative molecular mass. The relative molecular mass of MgO is 40 (i.e. the sum of the relative atomic masses of Mg and O, which is $24 + 16 = 40$), while the relative atomic mass of FeO is 72. Therefore, a basalt containing 12 wt% MgO and 8.0 wt% FeO will have

$$\frac{\text{Fe}^{2+}}{\text{Mg}} = \frac{\left(\frac{8.0}{72}\right)}{\left(\frac{12}{40}\right)} = 0.37 \text{ (to 2 sig. figs).}$$

- What is the Fe^{2+}/Mg ratio of the olivine in equilibrium with this basalt?
- Applying Equation 4.4, this liquid should be in equilibrium with an olivine with $\text{Fe}^{2+}/\text{Mg} = 0.3 \times 0.37 = 0.11$.

Until now, we have considered olivine compositions in terms of the percentages of forsterite (Fo) and fayalite (Fa). The percentage Fo is simply the percentage of Mg in the $(\text{Mg},\text{Fe})_2$ part of the olivine formula:

$$\%Fo = 100\% \times \frac{Mg}{Mg + Fe^{2+}}$$

This ratio was introduced towards the end of Section 4.2.1 as the Mg-number, often abbreviated as Mg#

$$Mg\# = 100\% \times \frac{Mg}{Mg + Fe^{2+}} \quad (4.6)$$

Therefore, an olivine with composition $Fo_{80}Fa_{20}$ has an Mg# of 80 (see also Section 4.2.1). But how does this relate to the Fe^{2+}/Mg ratio of the olivine?

To convert from Mg# to the atomic ratio Fe^{2+}/Mg or vice versa, Equation 4.6 can be rewritten by dividing the top and bottom lines by Mg:

$$Mg\# = 100\% \times \frac{\left(\frac{Mg}{Mg}\right)}{\left[\left(\frac{Mg}{Mg}\right) + \left(\frac{Fe^{2+}}{Mg}\right)\right]} = 100\% \times \frac{1}{1 + \left(\frac{Fe^{2+}}{Mg}\right)} \quad (4.7)$$

and this can be rearranged to make Fe^{2+}/Mg the subject:

$$\frac{Fe^{2+}}{Mg} = \frac{100 - Mg\#}{Mg\#} \quad (4.8)$$

- What is the Fe^{2+}/Mg ratio of $Fo_{80}Fa_{20}$?
- Applying Equation 4.8 and substituting an Mg# of 80:

$$\frac{Fe^{2+}}{Mg} = \frac{100 - 80}{80} = 0.25.$$

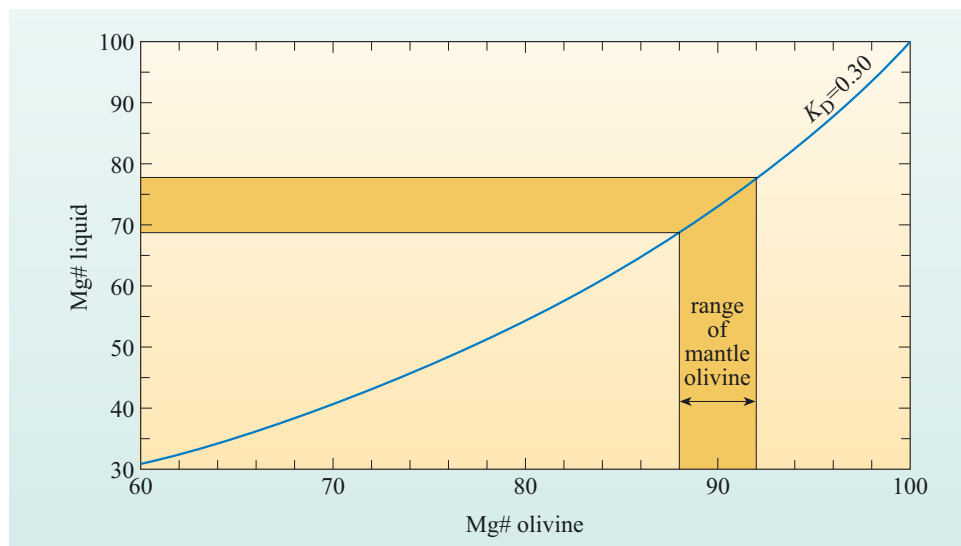
- An earlier example established that basalt with $Fe^{2+}/Mg = 0.37$ would be in equilibrium with olivine with $Fe^{2+}/Mg = 0.11$. What are their Mg#s?
- Applying Equation 4.7 in each case gives a liquid Mg# of 73 and olivine Mg# of 90.

The equilibrium relationship between olivine and liquid (Equation 4.4) can be rewritten in terms of the Mg#s of olivine and liquid:

$$Mg\#_{liquid} = \frac{30 Mg\#_{olivine}}{100 - 0.7 Mg\#_{olivine}} \quad (4.9)$$

and this equation is shown graphically in Figure 4.27. It illustrates what we found out from the Fo–Fa phase diagram, namely that olivine has a higher Mg# than the coexisting liquid.

Figure 4.27 Equilibrium relationship between the Mg#s of olivine and liquid for $K_D = 0.3$. The heavily shaded bands show the ranges in composition of mantle olivines and the liquids with which they are in equilibrium.

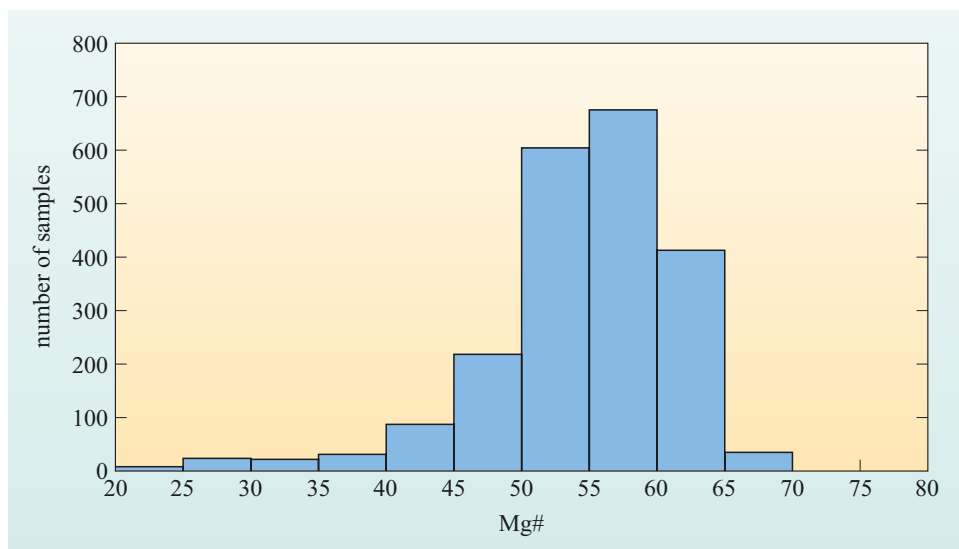


The Mg#s of mantle olivine are in the range 88 to 92 so, using Equation 4.9 or Figure 4.27, these olivines will be in equilibrium with basaltic liquids with Mg#s in the range 69 to 78. We now need to compare the actual compositions of basalts erupted at mid-ocean ridges with the compositions produced in the mantle.

The Mg#s of more than 2000 samples of MORB are summarised in Figure 4.28, from which two conclusions can be drawn:

- very few mid-ocean ridge basalts have Mg#s of 69 to 78, therefore very few mid-ocean ridge basalts are in equilibrium with mantle olivine
- the erupted lavas have a wide range of compositions, with most lavas having Mg#s in the range 45 to 65.

Figure 4.28 Histogram of Mg#s of mid-ocean ridge basalts. (Adapted from Sinton and Detrick, 1992)



It is very rare, therefore, for magma erupted at mid-ocean ridges to have come straight from its mantle source region unchanged. Between leaving the mantle and arriving at the surface, the magma has changed composition. The liquid produced by partial melting is referred to as a **primary magma**, and will have a suitably high Mg# reflecting equilibrium with mantle olivine. The origin of oceanic

crust, therefore, must start with a narrow range of primary magmas (based on the narrow range of $Mg\#_{\text{olivine}}$) and end with a wide range of magmas with lower $Mg\#$ s (see Figure 4.28). The transformations that take place between leaving the mantle and arriving at the sea floor are the subject of the next sections.

4.6.2 The liquidus temperatures of mid-ocean ridge basalt

Sections 4.4 and 4.5 showed how the origin of basalt by partial melting of peridotite could be understood in terms of the pressure–temperature phase diagram of peridotite and the P – T path followed by hot mantle as it decompressed. Can phase diagrams be used to help us think about how hot basalt liquid will behave as it moves through cold crust to the surface? We can make an initial stab at what happens by revisiting the Fo–Fa phase diagram. According to Figure 4.18, what will happen to a liquid of composition $Fo_{70}Fa_{30}$ as it cools from 1800 °C?

The liquidus curve is intercepted at 1775 °C, so at this temperature olivine crystals start to grow. The composition of this olivine, read from the solidus at 1775 °C, will be $Fo_{91}Fa_9$. With further cooling, the liquid tracks down the liquidus curve and the solid tracks down the solidus curve. So, as this magma cools, Mg-rich olivine crystals grow from the liquid, causing the concentration of Mg in the liquid to fall. As the temperature continues to decrease, more olivine crystallises, and the liquid becomes poorer in the Fo component, as does the coexisting olivine crystals. This is the reverse of the situation during partial melting that we looked at in Figure 4.19. The phase diagram shows that crystallisation is caused by cooling, and that there are correlations between the composition of the liquid, the composition of the crystals and the temperature.

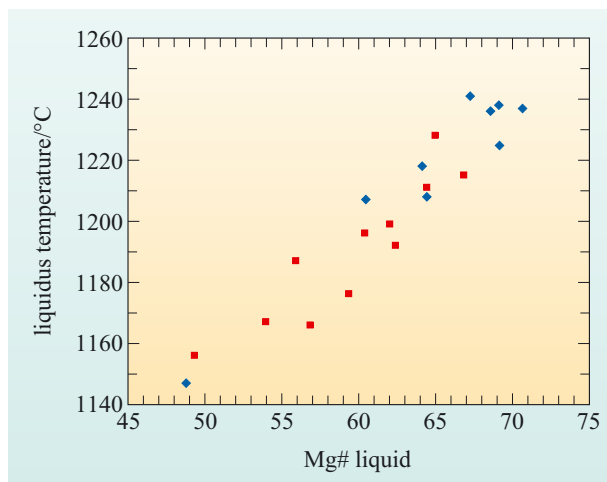
- In this model of crystallisation during cooling, does the $Mg\#$ of the liquid increase, decrease, or stay the same?
- With decreasing temperature, the liquid becomes poorer in the Mg endmember, and richer in the Fe endmember, so the $Mg\#$ of the liquid decreases.

In principle, the relatively low $Mg\#$ of mid-ocean ridge basalt glasses (i.e. low relative to the values expected for liquids in equilibrium with mantle olivine) indicates that MORB liquids may have been produced by crystallisation of some olivine from a magma of initially higher $Mg\#$ and higher temperature.

Turning to laboratory experiments with real mid-ocean ridge basalts, samples with different compositions (and therefore different $Mg\#$ s) are found to have different liquidus temperatures. If a single sample is taken and held at different temperatures, mixtures of crystals and liquids of different composition are produced. The correlation between the compositions of liquids and their liquidus temperatures is shown in Figure 4.29 and it mimics the pattern anticipated from the Fo–Fa phase diagram: high $Mg\#$ compositions have higher liquidus temperatures than low $Mg\#$ compositions.

The temperature at which the global population of mid-ocean ridge basalts is erupted can be estimated from their chemical compositions, specifically their $Mg\#$ s, using Figures 4.28 and 4.29.

Figure 4.29 Results of melting experiments on mid-ocean ridge basalts done at atmospheric pressure. Diamonds represent the liquidus temperatures of different samples. The squares represent the temperatures and liquid compositions of liquids produced in different experiments on the same sample. Both sets of data overlap on this diagram. (Data from Grove and Bryan, 1983; Tormey et al., 1987; Grove et al., 1990 and 1992; Yang et al., 1996; Juster et al., 1989)



Question 4.15

The quenched liquids (glass) in many mid-ocean ridge lavas have Mg#s in the range 50–65%. What range of liquidus temperatures do these liquids have?

The answer to Question 4.15 is important because it shows that mid-ocean ridge magmas cool by up to 100 °C or more between leaving their mantle source and reaching the sea floor.

4.6.3 Crystallisation of mid-ocean ridge basalt

When basalt magma cools, crystallisation occurs and drives the composition of the remaining liquid to a lower Mg#. But because the erupted lavas have lower Mg#s than the primary magmas, the majority of the crystals that formed after the primary magma had left the region of partial melting in the mantle must have been removed from the magma before it erupted. It is likely that this is achieved where magma accumulates at depth in the crust and slowly cools. It is within such magma chambers that crystals can start to grow. Being denser than the liquid, crystals are likely to sink through the magma, accumulating on the floor of the chamber in layers. The continuous physical separation of the crystals as they grow from the liquid is known as **fractional crystallisation**. This term is used because the sample becomes physically divided into separate fractions, i.e. liquid and crystals. The accumulated crystals are referred to as **cumulate crystals**, or just **cumulates**.

A second way in which the magma can behave is where the crystals remain suspended within the magma, forming phenocrysts. In this case, there are a further two possibilities: either the growing crystals continuously re-equilibrate with the magma, or previously formed crystals become isolated from the liquid by becoming coated by new layers of crystal growth. In the first of these two cases, the phenocrysts will have constant composition, whereas in the second case only the edges of the phenocrysts are in equilibrium with the liquid and each crystal will be compositionally zoned with, for example, the cores of olivine crystals being richer in Fo than the edges of the crystals.

Fractional crystallisation produces a series of liquids whose compositions reflect the amount and composition of removed (cumulate) crystals, but what determines which minerals will crystallise from the cooling magma?

Question 4.16

Results of crystallisation experiments on a Mid-Atlantic Ridge basalt are shown in Table 4.4. What was the sequence of mineral growth as lower temperatures were reached? (You may neglect the small amount of spinel and just concentrate on the silicate minerals.)

Table 4.4 The results of experiments with a sample of basalt when held at different temperatures, showing whether liquid, olivine, plagioclase or clinopyroxene were found. (✓* indicates that a small amount of spinel was also present.) (Data from Grove and Bryan, 1983)

Temperature/°C	Liquid	Olivine	Plagioclase	Clinopyroxene
1245	✓			
1228	✓	✓*		
1215	✓	✓*		
1211	✓	✓*		
1199	✓	✓*	✓	
1196	✓	✓*	✓	
1192	✓	✓*	✓	
1187	✓	✓*	✓	✓
1176	✓	✓	✓	✓
1167	✓	✓	✓	✓
1166	✓	✓	✓	✓
1156	✓	✓	✓	✓

As well as using experiments to discover the order in which different minerals crystallise during cooling, samples of rock brought up from the sea floor can also reveal the order of crystallisation. Each rock is a mixture of glass together with a few phenocrysts, and the identity of the phenocrysts depends on the composition of the glass (i.e. the composition of the liquid in equilibrium with those crystals). Glasses with lower Mg#s represent liquids that have been generated at lower temperatures, and therefore by greater amounts of fractional crystallisation, than glasses with higher Mg#s. The sequence in which different phenocrysts appear with decreasing Mg# (and decreasing MgO content) of the glass thus represents the crystallisation sequence of the minerals in mid-ocean ridge basalt magma. The commonest sequence is olivine, then olivine + plagioclase, and then olivine + plagioclase + clinopyroxene (see Question 4.16).

To summarise:

- during partial melting of peridotite it was found that, with increasing temperature, some minerals were consumed in the melting reactions (plagioclase, spinel, garnet and clinopyroxene) whereas others persisted to higher temperatures (orthopyroxene and olivine)
- in the case of progressive crystallisation due to cooling, the reverse phenomenon occurs: certain minerals do not start to crystallise until the temperature has become sufficiently low (and the liquid composition has sufficiently low Mg#).

4.6.4 Variation diagrams

Mid-ocean ridge glasses show a range in composition reflecting the progressive removal of olivine, plagioclase and clinopyroxene. These different minerals have different compositions. For example, relative to basalt:

- olivine is rich in MgO and poor in CaO
 - plagioclase is poor in MgO but rich in CaO
 - clinopyroxene is rich in both MgO and CaO.
- When olivine is the only crystallising mineral, what will happen to the MgO and CaO concentrations in the liquid?
- MgO will decrease and CaO will increase.

This effect can be seen on a plot known as a **chemical variation diagram** (or just **variation diagram**). Take a look at Figure 4.30, which plots the CaO and MgO wt% contents of liquids and olivine. Starting with an initial liquid (L_1), crystallising some olivine causes the liquid composition to move farther away from the composition of the olivine (Figure 4.30a) until L_2 is reached. Crystallising still more olivine changes the liquid composition to L_3 . Given the array of liquids L_1 to L_3 , their alignment on this diagram links them to fractional crystallisation involving olivine removal. In other words, fractional crystallisation causes the composition of the liquid to change, or evolve. Magmas which have undergone only a small amount of fractional crystallisation and are still fairly close in composition to that of a primary magma are said to be **primitive magmas**, whereas those that are further removed are said to be **evolved magmas** (but note these terms are relative). Evolved magmas may also be described as **derivative magmas** or **fractionated magmas**. They can also be called **differentiated magmas** because the production of evolved magmas from a single starting magma is called differentiation.

If at some point plagioclase starts to crystallise alongside olivine, then the liquid composition will evolve in a direction that moves it away from the average composition of the crystallising mixture of olivine and plagioclase. This average composition lies on the straight line joining the compositions of the two minerals and depends on the proportions of each mineral. This is shown schematically on Figure 4.30b.

If a mixture of olivine, plagioclase and clinopyroxene is crystallising, then the compositional path followed by the liquid will move away from the composition of the three-phase mixture (Figure 4.30c). In this case, the overall composition of the crystals plots within the triangle defined by drawing straight lines between each of the three minerals.

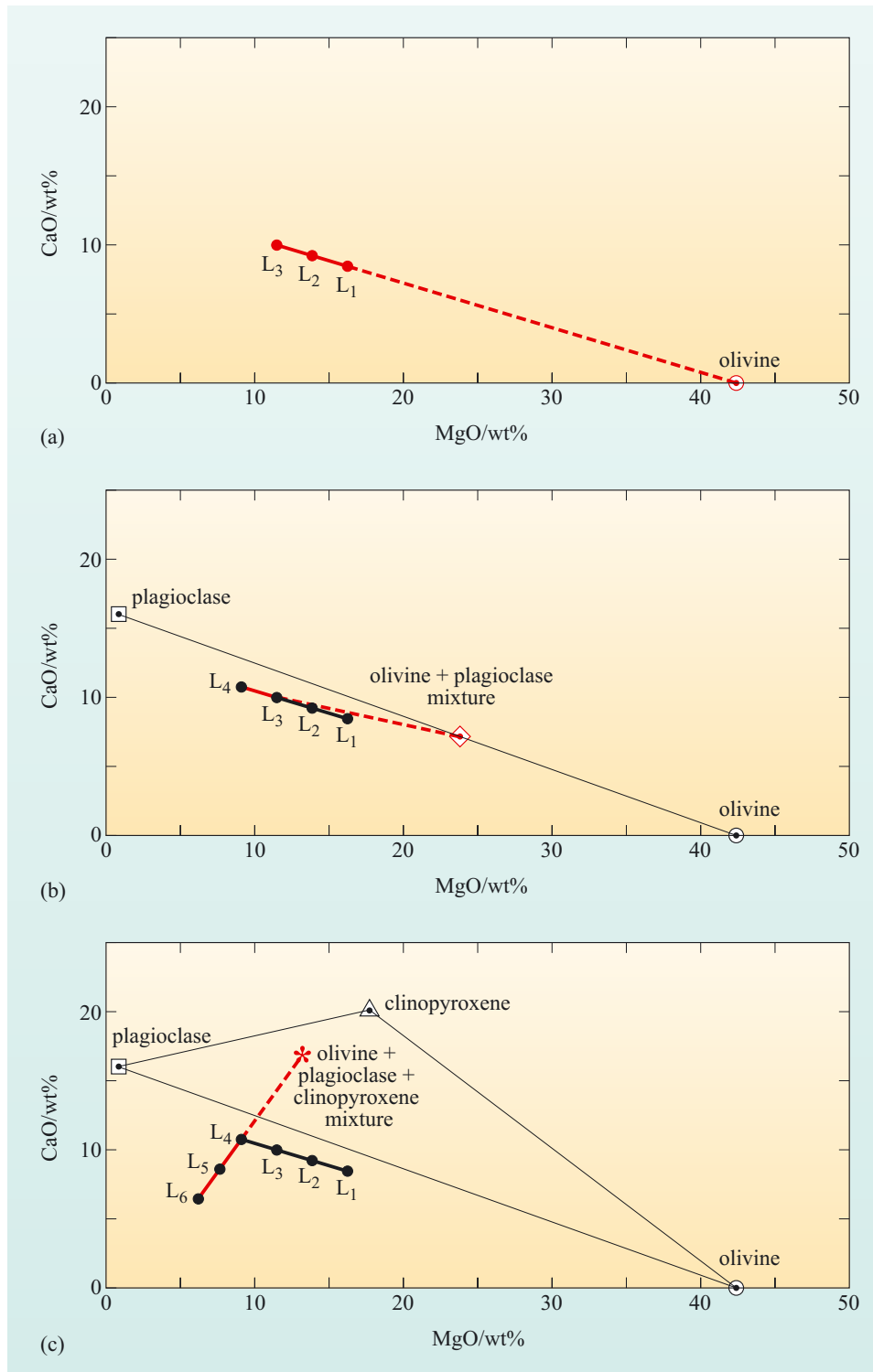


Figure 4.30 Schematic variation diagrams showing the trajectory of liquid composition of basalt undergoing fractional crystallisation of (a) olivine (L₁ to L₃), (b) a mixture of plagioclase and olivine (L₃ to L₄), and (c) a mixture of olivine, plagioclase and clinopyroxene (L₄ to L₆). For simplicity, mineral compositions are assumed to remain constant, whereas solid-solution effects will cause them to change as a result of changing liquid composition.

Variation diagrams such as those in Figure 4.30 are useful for a number of reasons. First, they allow the compositions of many samples to be compared without having to scrutinise tables of data. Second, and more importantly, the shapes of the trends can be used to infer the processes that have controlled the evolution of the magmas. For instance, a trend of decreasing MgO and slightly increasing CaO can be due to fractional crystallisation involving the removal of olivine or a mixture of olivine and plagioclase. A trend of decreasing MgO and decreasing CaO reflects fractional crystallisation of a mixture of olivine, plagioclase and clinopyroxene.

- Figure 4.31 shows the chemical compositions of glasses from pillow lavas collected from the East Pacific Rise. Describe the trend and suggest how it was formed.
- The correlation shows decreasing MgO and CaO. This trend can be explained if different samples have been formed by different amounts of fractional crystallisation of olivine, plagioclase and clinopyroxene (as in the trend from L₄ to L₆ in Figure 4.30c).

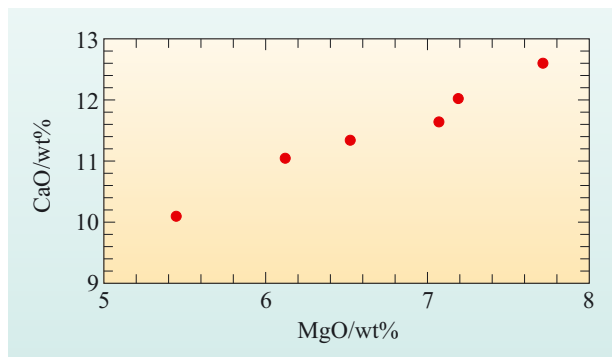


Figure 4.31 Variation diagram of wt% CaO against MgO for basalt glasses from 9° 30' N on the East Pacific Rise. (Data from Pan and Batiza, 2003)

The interpretation of Figure 4.31 is that magma became lodged in the magma chamber(s) beneath the East Pacific Rise, and cooled sufficiently for olivine, plagioclase and clinopyroxene to crystallise, and that these crystals became separated from the magma, forming cumulates. By this process, fractional crystallisation produced a range of liquid compositions, dependent on the amount of crystallisation: the larger the amount of crystals removed, the lower the MgO content of the liquid.

Depending on the stage of magma evolution, the removed cumulate crystals will be composed of olivine and a mixture of olivine and plagioclase, or a mixture of olivine, plagioclase and clinopyroxene. Rocks made of these minerals are dunites (olivine only) or gabbros. So, dunite and gabbro cumulates are

the expected by-products of the fractional crystallisation of basalt, as well as the slowly crystallised equivalents of basalt. Thus, the igneous oceanic crust is composed of evolved liquids quenched as sea-floor lavas or dykes, the cumulate complement of those liquids, and slowly cooled basalt magmas.

4.6.5 The depths of magma chambers beneath mid-ocean ridges

Between the depths where magmas are generated and the sea floor where they erupt, mid-ocean ridge magmas undergo fractional crystallisation to yield the evolved liquids that erupt and the cumulate gabbros that remain in the crust. But what about the depths at which fractional crystallisation occurs?

Studies of the passage of seismic waves through the crust beneath mid-ocean ridges have been successful in locating the tops of magma chambers at between 1.2 km and 2.4 km below the East Pacific Rise (Detrick et al., 1987) and about 3 km below the Valu Fa Ridge (situated in the Lau Basin, North of New

Zealand). But in other cases, no magma chambers have been detected by this method. This could be because either the chambers simply are not there or the chambers are too small or too deep to be detected amongst the noise in the seismic data. Seismology has been only partly successful in detecting magma chambers because of limitations with the method, so other ways of pinpointing the depths at which magmas may be stored or may fractionate need to be used.

Because pressure can have an influence on which minerals crystallise from a magma, there is the potential to use the compositions of the liquids or the crystals in the erupted lavas to infer the pressure (and therefore depth) at which fractional crystallisation occurred. This can be illustrated using a simple model system involving the first mineral to crystallise from primitive mid-ocean ridge basalt (i.e. olivine) and the last mineral to crystallise (i.e. clinopyroxene).

Both of these minerals form solid-solution series, so the first step is to focus on the pure endmembers. The Mg endmembers are usually chosen because the olivine and clinopyroxene phenocrysts in basalts are richer in Mg than in Fe.

- What are the pure Mg endmembers of olivine and clinopyroxene called and what are their chemical compositions?
- Forsterite (Mg_2SiO_4) and diopside ($\text{CaMgSi}_2\text{O}_6$) respectively.

Figure 4.32 shows the phase diagram of Fo–Di at two pressures: 10^5 Pa (atmospheric pressure) and 2 GPa. In this system, for a starting composition that is rich in Fo, forsterite will be the first mineral to crystallise (which is analogous to the initial crystallisation of olivine when primitive basalt magma cools).

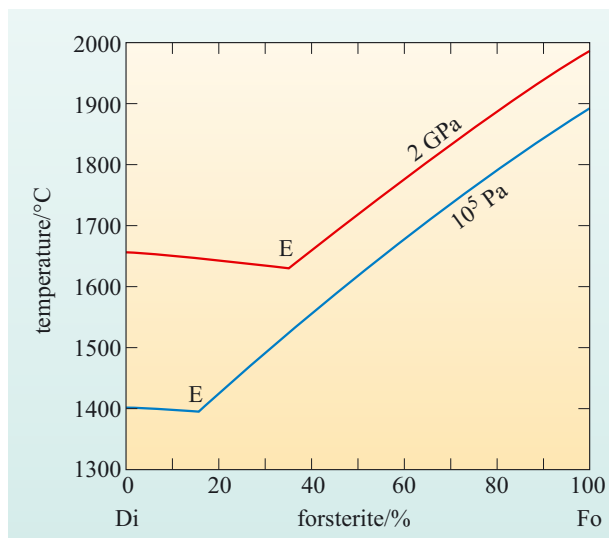


Figure 4.32 Liquidus curves in the system Fo–Di phase at pressures of 2 GPa and 10^5 Pa. E is the eutectic point.

- According to Figure 4.32, what is the composition of the liquid when Di starts to crystallise at 2 GPa, after fractional crystallisation of Fo?
- The composition of the liquid follows the liquidus curve with falling temperature until it reaches the eutectic, at which point Di starts to crystallise. At 2 GPa the eutectic liquid has a composition of about $\text{Fo}_{35}\text{Di}_{65}$.

- What is the composition of the liquid that first crystallises Di at 10^5 Pa, and how does this differ from the composition at 2 GPa?
- At 10^5 Pa, the eutectic composition is at about $\text{Fo}_{16}\text{Di}_{84}$. This is closer to the composition of pure Di.

The Fo–Di phase diagram illustrates that the amount of Fo that must crystallise out of the magma before Di starts to crystallise depends on the pressure. The lower the pressure, the greater the amount of Fo needed to crystallise before Di appears. Fo has a higher MgO content than Di (57.2 wt% versus 18.6 wt% MgO) so, in terms of the MgO content of the evolving liquid, low-pressure fractional crystallisation of Fo will drive the liquid to a lower MgO content before Di starts to crystallise than will fractional crystallisation at high pressure.

The idea that reducing the pressure delays the start of Di crystallisation can now be applied to the chemically more complex mixture of natural basalt, and a variation diagram can be used to contrast the liquid compositions caused by fractional crystallisation of primitive MORB at high and low pressures. The plot of MgO wt% versus CaO wt% in Figure 4.33 illustrates the compositional evolution of the magma at different pressures:

- at high pressure, fractional crystallisation of a primitive magma with composition A would first involve olivine and then olivine + plagioclase until reaching point B, when clinopyroxene starts to crystallise and compositions evolve along the path of decreasing CaO wt% from B to C. The magmas along the line BC crystallise olivine + plagioclase + clinopyroxene;
- at low pressure, the first appearance of clinopyroxene is delayed and does not appear until the magma has evolved from A to D, after which point olivine + plagioclase + clinopyroxene crystallisation drives the composition along the path D to E.

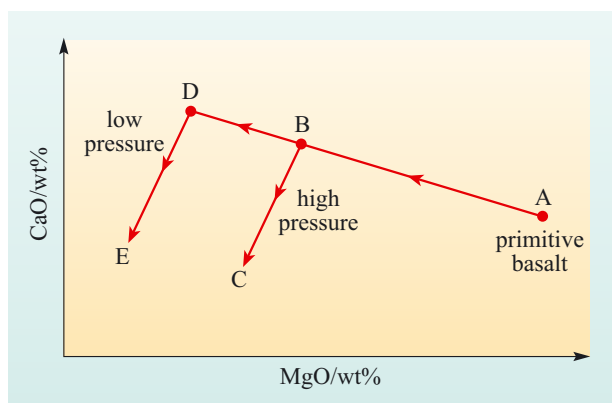


Figure 4.33 Hypothetical variation diagram showing how the compositional evolution of a primitive basalt (A) depends on the point at which clinopyroxene starts to crystallise, and how that point depends on pressure. See text for explanation.

According to Figure 4.33, liquids that are in equilibrium with crystals of olivine, plagioclase and clinopyroxene at low pressure will have different compositions from those at high pressure. For a given MgO content, liquids produced at lower pressure will have a higher CaO content than those produced by fractional crystallisation at higher pressure. Experiments on actual MORB samples bear this out, as seen in Figure 4.34a, which contrasts the compositions of liquids produced in experiments done at atmospheric pressure (10^5 Pa) and 800 MPa. To put these pressures into context, answer the following question.

Question 4.17

What is the pressure (a) on the sea floor below 3 km of sea water (density $1.0 \times 10^3 \text{ kg m}^{-3}$) and (b) at 30 km below the sea floor (assuming an average rock density of $3 \times 10^3 \text{ kg m}^{-3}$) and a water depth of 3 km? In both cases, assume that $g = 9.8 \text{ m s}^{-2}$ and give your answer to the correct number of significant figures. (Use Equation 4.1 in Box 4.4.)

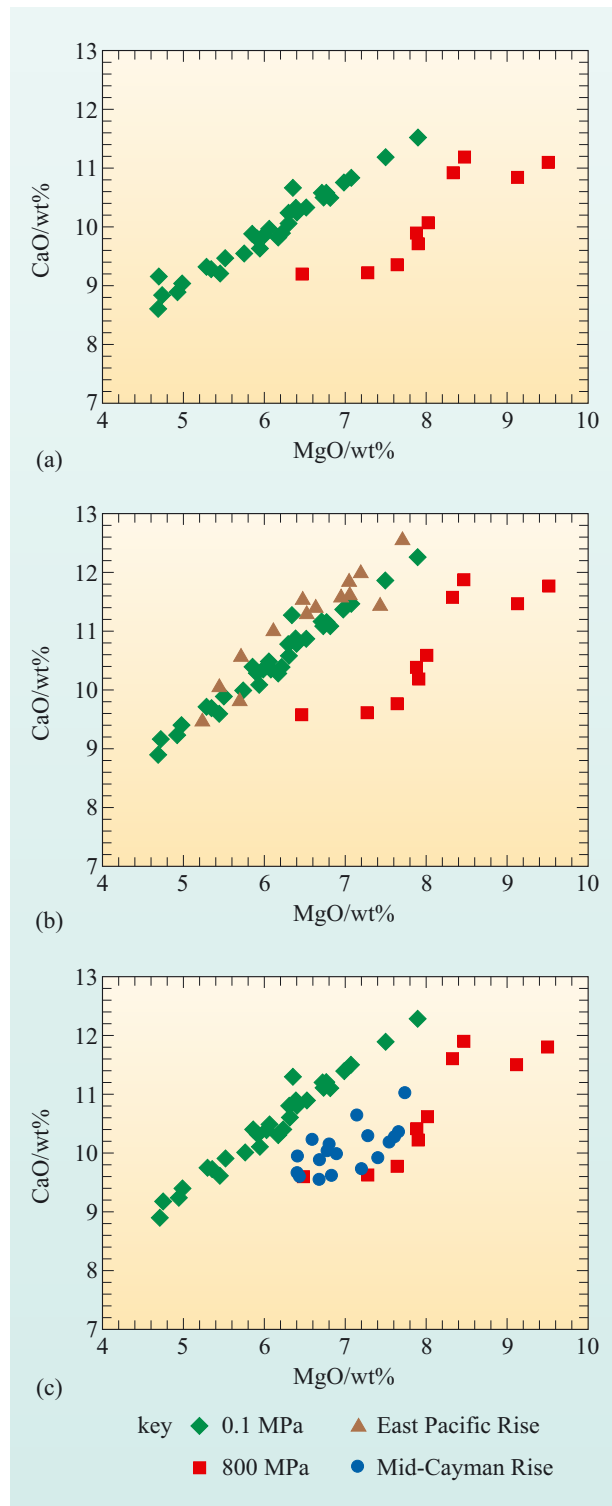
Figure 4.34 (a) Compositions of liquids in equilibrium with olivine + plagioclase + clinopyroxene produced in experiments on mid-ocean ridge basalts at 10^5 Pa. (b) Compositions of glasses from the East Pacific Rise at $9^\circ 30' \text{ N}$, in comparison with experimental glasses from (a). (c) Compositions of glasses from the Mid-Cayman Rise, in comparison with experimental glasses from (a). (Data for (a) from Tormey et al., 1987; Juster et al., 1989; Grove et al., 1990, 1992; Yang et al., 1996; for (b) from Pan and Batiza, 2003; for (c) from Elthon et al., 1995)

On Figure 4.34, liquids at atmospheric pressure fall on a narrow, well-defined trend, whereas the data at 800 MPa are much more scattered, but the points are still clearly separated from those at atmospheric pressure. The experimental compositions produced under known pressure, can be compared with the compositions of natural lavas produced by fractional crystallisation at some unknown pressure, in order to estimate the pressures at which the mid-ocean ridge magmas evolved.

The East Pacific Rise and the Mid-Cayman Rise in the northern Caribbean Sea represent a fast-spreading ridge ($11\text{--}12 \text{ cm y}^{-1}$ full spreading rate) with seismic evidence for the roof of a magma chamber at between 1.6 km and 2.2 km depth, and a slow spreading ridge (2.0 cm y^{-1} full spreading rate) respectively. Lavas from these two ridges show trends of decreasing CaO with decreasing MgO (Figure 4.34b–c), indicative of fractional crystallisation of olivine + plagioclase + clinopyroxene (see Figure 4.30c).

The East Pacific Rise lavas plot very close to the liquids from experiments at atmospheric pressure, so we infer that the East Pacific Rise magmas were produced by olivine + plagioclase + clinopyroxene fractional crystallisation in a magma chamber situated at shallow depths in the oceanic crust. On the other hand, the Mid-Cayman Rise lavas plot in a broader area to the right of the atmospheric pressure experiments but to the left of, and overlapping with, the area occupied by the liquids produced at 800 MPa. The crystallisation pressure of these magmas has been calculated to be about 600 MPa, and this pressure is found at a depth of about 20 km below the sea floor.

This indicates that the fast-spreading East Pacific Rise has a shallower magma chamber than the slow-spreading Mid-Cayman Rise. A possible reason for this difference is that slowly spreading ridges do not allow magmas to escape from depth as easily as at fast-spreading ridges, so the magma becomes trapped at greater depth.



4.7 Interactions with the oceans

So far in this chapter, the main themes have been the processes of partial melting and fractional crystallisation that generate oceanic crust at constructive plate boundaries. The material that forms the Earth's oceanic crust originates in the high- P , high- T environment of the upper mantle, but it ends up as part of the low- P , low- T environment of the Earth's surface where it is laid bare to the chemical, physical and biological processes of the hydrosphere and biosphere. Once formed, the oceanic crust continues to evolve, and this is the topic of the final section of this chapter.

Sea-floor spreading and volcanism at mid-ocean ridges are driven by the rise of hot mantle towards the surface. As a result, mid-ocean ridges sustain the highest surface heat fluxes on Earth, with half of the Earth's heat output coming from ocean lithosphere that is less than 65 million years old. Given this high heat flux and the fact that the upper ocean crust is a permeable rock mass soaking in seawater, it is perhaps not surprising that hot springs akin to those found at volcanic regions on land are found at mid-ocean ridges (Figure 4.1b). The first inkling that mid-ocean ridges sustained hydrothermal activity and vigorous submarine hot springs was the discovery, in the mid-1960s, of brine solutions at 40 °C to 60 °C and metal-rich sediment at the constructive plate boundary separating Arabia and Africa on the floor of the Red Sea. Scientists became eager to find more evidence of deep-sea hydrothermal systems, so an expedition was proposed to the Galapagos Ridge in the eastern Pacific Ocean at 86° W, where small thermal anomalies in the otherwise uniformly cold sea-floor ocean water had already been detected by remotely towed instruments. Using the submersible craft *Alvin*, which had been purpose built for deep-sea research, the first views of monotonous, barren, black basalt lava and the occasional clam shell at the mid-ocean ridge were seen in February and March 1977. As the expedition progressed, sensors detected temperatures above the normal 2.0 °C. *Alvin*'s lights lit up the

otherwise total darkness of the deep to reveal shimmering warm water rising out of cracks in the sea-floor basalt lava; large mussels, clams, red shrimp, white crabs, limpets and anemones populated the 17 °C water (Figure 4.35 and Box 4.5). Samples of water returned to the surface gave off the rotten-egg stench of hydrogen sulfide (H_2S).



Figure 4.35 A view of the East Pacific Rise, 2600 m below sea level, showing a community of crabs, bivalves, tube worms (the two white 'stalks' attached to the rocky sea floor left of lower centre) and fish. (Dr Ken MacDonald/Science Photo Library)

Box 4.5 Life in the dark

Dwelling on the surface of the Earth, we take it for granted that life's food chain is based on plants and micro-organisms that get their energy from sunlight. These are the photosynthetic organisms. However, no sunlight penetrates to the bottom of the ocean, so what sustains the communities of organisms that are found around mid-ocean ridge vents? At the bottom of the food chain are colonies of micro-organisms that rely on chemical energy rather than energy from sunlight. The main metabolic process in many of these micro-organisms is the oxidation of reduced sulfur. For example, reduced sulfur in H_2S gas is used to take CO_2 gas dissolved in seawater and generate organic carbon by the chemical reaction:



where $[\text{CH}_2\text{O}]$ denotes organic carbon. Organisms that live in this way are called **chemosynthetic**, in contrast to photosynthetic organisms. These chemosynthetic organisms tolerate high temperatures – and some require temperatures well above 100°C . The chemosynthetic organisms and the macrofauna that consume them form deep-sea vent communities that are alien to our normal concept of life, but their existence provides a new impetus to understand the intertwined geological and biological processes on Earth, and even on other planets.

Exploration gathered pace, and in the spring of 1979, diving on the East Pacific Rise at 21°N , *Alvin*'s crew encountered two types of hot-water vent. Not only did they find hot water seeping from cracks on a biologically infested sea floor (similar to the Galapagos Ridge sites found in 1977), they also discovered water at 380°C spouting from chimneys made of copper and iron sulfides. As the hot fluids rose through the water column they mixed with the ambient seawater, precipitating clouds of billowing black 'smoke' (Figure 4.1b). Here, in these spectacular 'black smoker' vents, was yet more evidence of sulfur-rich hot fluids gushing from the sea floor at mid-ocean ridges. At least fifty such areas of hydrothermal activity are now known from most of the mid-ocean ridge system, either as a result of planned exploration or serendipitous discovery of thermal or chemical anomalies in the oceans.

The underlying processes at hydrothermal areas are clear: cold seawater is drawn into the permeable oceanic crust, becomes heated and reacts with the rocks through which it flows before hot mineral-rich solutions rise to the surface, mixing and cooling to various degrees before flowing back out of the system into the ocean. As a result, the chemical composition of the oceanic crust becomes modified, as does the chemical composition of the ocean, and chemicals are supplied to mid-ocean ridge ecosystems. Two lines of evidence that give clues as to what happens in these hydrothermal systems are:

- the differences between pristine lava erupted on the ocean floor and rocks that have been chemically altered as a result of chemical reactions with hot seawater;
- the differences between normal cold seawater that enters hydrothermal systems and the hot effluent that comes out of the hydrothermal systems.

4.7.1 Evidence from rocks

Section 4.6 dealt with the chemical composition of fresh lava that had been quenched on eruption and sampled before any later alteration processes had occurred. The chemical variations were due to the igneous processes that generated the magma and led to its differentiation. Samples that have undergone various degrees of chemical and mineralogical alteration after the magma solidified are needed to investigate hydrothermal alteration processes. Samples of altered ocean crust come from tectonically disrupted sections of sea floor (such as fracture zones), cores recovered by deep sea drilling, and dissected oceanic crust exposed in ophiolites.

These altered rocks are typically pale or dull green, in contrast to the black colour of freshly erupted glassy basalt, so are generally called **greenstones**. The familiar igneous ferromagnesian minerals olivine and clinopyroxene are absent, but minerals such as chlorite (a green mineral with a layered structure similar to mica), epidote (a pale-green mineral) and actinolite (a green Ca-bearing amphibole rich in iron) are present. Another contrast with fresh basalt is that the plagioclase feldspar found in most mid-ocean ridge basalts is relatively rich in the Ca endmember anorthite (abbreviated as An) and poor in the Na endmember albite (Ab), ranging from $An_{85}Ab_{15}$ to $An_{60}Ab_{40}$, whereas in the altered rocks, plagioclase is much richer in albite, ranging from $An_{35}Ab_{65}$ to An_5Ab_{95} . All of these minerals (i.e. chlorite, actinolite, epidote and plagioclase) form solid-solution series so their compositions are variable; examples of their compositions are given in Table 4.5. Chemical analyses of a relatively unaltered sample from the centre of a pillow lava and two more altered samples of greenstones from the edge of the same pillow are given in Table 4.6.

Table 4.5 Chemical analyses (reported as weight per cent oxides) of minerals from altered basalts recovered from the Mid-Atlantic Ridge. Analyses give the total iron (Fe^{2+} and Fe^{3+}) in terms of FeO and is denoted FeO_t . The shortfall between the reported total oxides and the ideal total of 100% reflects analytical uncertainties and components that were not analysed for, particularly water in the case of chlorite, epidote and actinolite. (Data from Humphris and Thompson, 1978)

	chlorite	chlorite	epidote	epidote	actinolite	actinolite	feldspar	feldspar
SiO ₂	28.34	28.42	37.50	37.75	49.75	51.01	67.68	63.50
Al ₂ O ₃	17.00	17.54	24.83	24.36	6.30	2.43	20.46	22.18
FeO _t	21.30	23.33	10.45	11.75	17.40	13.32	0.04	0.00
MnO	0.16	0.14	0.03	0.16	0.12	0.14	0.00	0.00
MgO	20.19	17.41	0.03	0.27	12.66	15.16	0.00	0.00
CaO	0.07	0.13	23.4	22.83	11.78	12.63	1.03	2.87
Na ₂ O	0.07	0.01	0.01	0.04	0.55	0.33	11.39	10.64
K ₂ O	0.03	0.03	0.03	0.00	0.06	0.07	0.00	0.00
total	87.16	87.01	96.28	97.16	98.62	95.09	100.60	99.19

Question 4.18

- (a) Which of the following statements are true?
- A Chlorite, epidote and actinolite are silicates.
 - B Chlorite, epidote and actinolite are hydrous minerals.
 - C Plagioclase, olivine and clinopyroxene are hydrous minerals.
- (b) Apart from silica, what are the main constituent oxides of (i) albite, (ii) chlorite, (iii) epidote, and (iv) actinolite?
- (c) What is the range of water content of the greenstones in Table 4.6?

Table 4.6 Chemical analyses (wt%) of greenstones from parts of a pillow lava from the Mid-Atlantic Ridge. Samples A and C are from the outermost and innermost parts of a single pillow respectively. Sample B is from a position between A and C. (Data from Humpbris and Thompson, 1978)

	A strongly altered	B moderately altered	C slightly altered
SiO ₂	39.31	47.70	52.30
TiO ₂	2.25	1.75	1.63
Al ₂ O ₃	17.59	14.15	14.79
FeO _t	12.71	9.50	8.17
MnO	0.11	0.10	0.14
MgO	12.67	9.70	6.03
CaO	3.35	8.40	9.83
Na ₂ O	2.98	2.97	3.22
K ₂ O	0.14	0.13	0.15
P ₂ O ₅	0.15	0.14	0.16
H ₂ O	6.53	3.26	1.85
Total	97.79	97.80	98.27

One of the main chemical differences between greenstones and fresh mid-ocean ridge basalt is that basalt has less than 0.1% H₂O, whereas greenstones can have more than 6% water. This is reflected in the abundance of hydrous minerals in greenstones and is a clear indication that seawater is incorporated into basalt during hydrothermal alteration. Simply adding water will dilute the concentrations of all other elements, in which case the chemical compositions of the greenstones would be intermediate between those of fresh basalt and those of seawater. On a variation diagram they would plot on a straight line joining the basalt and seawater compositions. But is this what the chemical compositions of the rocks show?

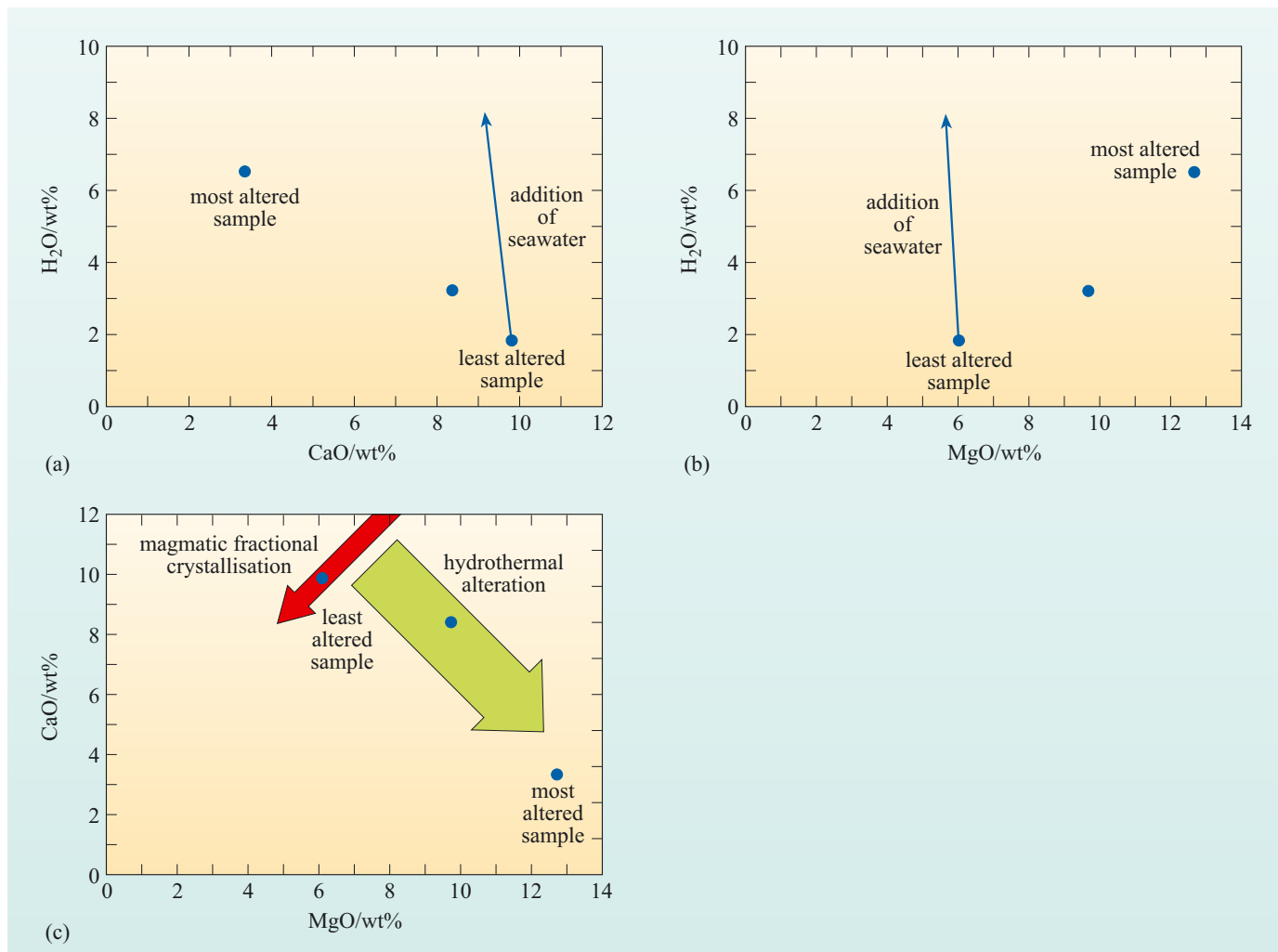


Figure 4.36 Compositions of the three altered lava samples from the Mid-Atlantic Ridge given in Table 4.6 plotted on variation diagrams of wt% (a) CaO versus H₂O, (b) MgO versus H₂O, and (c) MgO versus CaO. The arrows in (a) and (b) show the trend that would be produced by adding seawater to the least-altered sample. The broad arrows in (c) indicate the general effects that fractional crystallisation and subsequent hydrothermal alteration will have.

- Compare the array of compositions in Figure 4.36 with the trend expected from just adding water to the least-altered sample. How do you account for the composition of these rock samples?
- Because CaO decreases with increasing H₂O more sharply than by dilution, CaO must have been removed from the original rocks. MgO does not follow the expected trend of dilution by seawater either. Instead, it increases in concentration as the percentage of water increases, indicating that water and MgO have both been added to the rock.

Although the greenstones have formed by incorporation of water, their origin also involves the removal of some elements (notably Ca and Si) and the addition of others (notably Mg) and this gives them compositions that are distinct from those produced by igneous processes (Figure 4.36c). Pristine basalts show a trend of

decreasing MgO and CaO due to the removal of olivine, plagioclase and clinopyroxene crystals during fractional crystallisation in a cooling magma chamber (Section 4.6.4). The greenstones, on the other hand, have an array of compositions that reflect the addition of MgO and the removal of CaO to the original igneous composition.

Where does the added MgO come from, and where does the CaO that has been removed end up? The answer to this question comes from comparing the composition of seawater with the composition of the water that comes out of mid-ocean ridge hydrothermal systems. Hydrothermal water contains no Mg, whereas normal seawater has about 1290 ppm; hydrothermal waters have up to five times more Ca than seawater does (400 ppm). In overall terms, mid-ocean ridge hydrothermal systems remove Mg from the oceans (adding it to the crust) and add Ca to the oceans (removing it from the crust).

4.7.2 Mid-ocean ridge hydrothermal systems

Hydrothermal systems occupy large volumes of the crust at mid-ocean ridges, and can be divided into different parts according to the processes occurring (Figure 4.37). The broad area on either side of the plate boundary is where seawater penetrates the fractured oceanic crust and is drawn down into deeper, hotter crust; this is known as the **recharge zone** of the hydrothermal system. The hottest, deepest part of the system is where most of the chemical changes that determine the composition of the hydrothermal effluent occur, and this is known as the **reaction zone**. Further reactions occur as the hot fluid rises from the reaction zone to the surface, and this part of the system is known as the **discharge zone**.

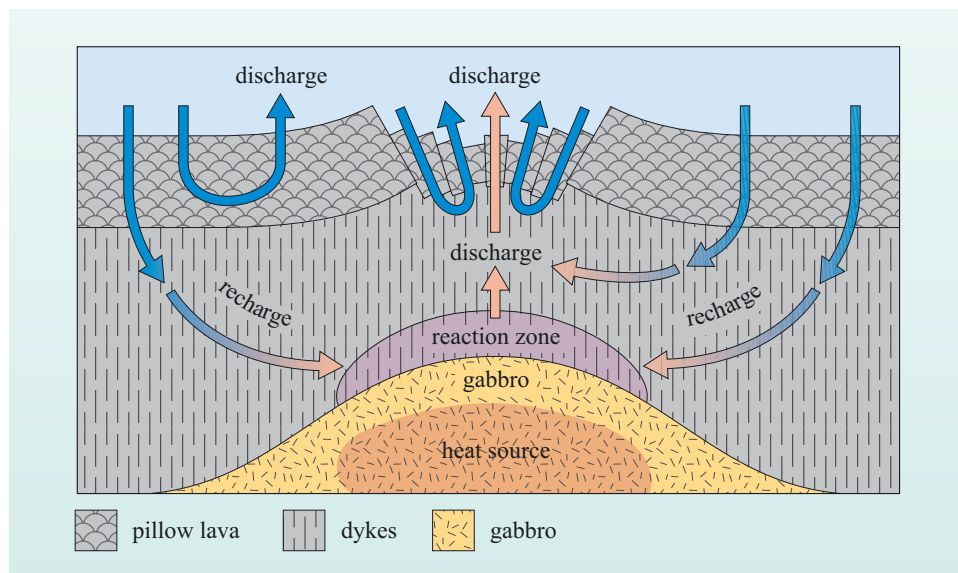


Figure 4.37 Schematic cross-section of oceanic crust at a mid-ocean ridge showing the flow paths of water through the recharge, reaction and discharge zones of the hydrothermal system. (Adapted from Alt, 1995)

Processes in the hydrothermal system have to be inferred from the composition of hydrothermal waters, investigations of hydrothermally altered rocks from ophiolites and ocean drilling, and theoretical modelling of chemical reactions at high temperature and pressure. Combinations of these approaches have led to an overall understanding of what happens in each of the zones.

1 Heating of seawater in the recharge zone

Direct sampling of water from recharge zones has not yet been achieved, although fluids from shallow off-axis regions have been obtained on the flanks of the Juan de Fuca Ridge (in the NE Pacific Ocean) and may be analogous to those from low-temperature regions of the recharge zone. These water samples, together with consideration of the behaviour of seawater when it becomes heated at crustal pressure, provide some ideas of the processes occurring in the recharge zone. The sampled fluids have less than 10% of the Mg contained in seawater, indicating uptake of Mg by the warm rocks through which the water passes. This happens by incorporation of Mg and water into basaltic rocks and minerals to form new, hydrous, Mg-rich minerals such as clay and chlorite. Other changes come about as a result of the change in the solubility of the dissolved components in seawater when it is heated. Once seawater is above about 150–200 °C (depending on pressure), it becomes saturated in CaSO_4 (the mineral anhydrite) such that all of the Ca and about two-thirds of the SO_4^{2-} is removed from the water by the crystallisation of anhydrite in pore spaces and veins within the rock. As well as removing Ca from the solutions that penetrate into deeper regions of the hydrothermal system, removal of SO_4^{2-} in the recharge zone limits the amount of sulfur taken to deeper levels.

2 Reaction of water with hot rock in the reaction zone

Fluids reaching the reaction zone have compositions that have been changed considerably by their passage through the recharge zone: most Mg has been removed, the remaining sulfur is largely in the form of sulfide ions rather than sulfate ions, and the pH has decreased. High-temperature reactions between this fluid and the hot rock (such as recently solidified intrusive rock or the rocks immediately above a magma chamber) yield new minerals and modify the water composition further. Fe, Mn, Cu, Zn and other metals are soluble in such hot acidic fluids and are therefore leached from the rock along with sulfur. The original igneous minerals (olivine, plagioclase and clinopyroxene) and glass are consumed in reactions that yield mineral assemblages involving albite and hydrous minerals that include chlorite, actinolite and epidote found in greenstones.

3 Precipitation of sulfides and sulfates, and mixing with seawater in the discharge zone

The hot mineralised solutions within the reaction zone are extremely buoyant with respect to cooler waters and therefore rise rapidly to the sea floor, undergoing decompression, cooling, and mixing with cooler compositionally different solutions. Phase changes that occur can involve the switch from a liquid into a mixture of vapour and a concentrated brine solution. Cooling can lead to deposition of metal sulfides in veins that give rise to mineral deposits. These mineral deposits can be found in ophiolites, notably in Cyprus and Oman, and have also been discovered by deep research drilling in the ocean crust. For instance, chemical analyses of a 1300 m long core that passes through pillow lavas and the underlying sheeted dyke complex of 5.9 million year old crust in the Nazca Plate (in the eastern Pacific Ocean) show the concentration of metal sulfide mineralisation is preferentially concentrated in the region where the dyke complex grades into the lavas. This is probably the location of heterogeneities in the rock structure that favoured mineral deposition.

4 Black and white smokers and hydrothermal plumes in the ocean

The most dramatic manifestation of hydrothermal systems at mid-ocean ridges are the billowing plumes of hot fluid and particles known as black smokers that flow from tall chimneys and mounds on the sea floor (Figure 4.1b). The black ‘smoke’ particles and the chimney are formed when hot, H_2S -rich, O_2 -poor, low-pH (acidic) hydrothermal fluid mixes with cold, sulfate-rich, O_2 -rich seawater. Within the first second of meeting seawater, mixing of the hottest fluids (between 290 °C and 400 °C) precipitates tiny particles of Cu, Fe and Zn sulfides and oxides, and these are carried up in the buoyant plume of water some 100–200 m before spreading out as a broad umbrella cloud. Some particles sink to the sea floor and become hydrothermal sediment that is rich in Fe and Mn.

- As well as the effect of seawater being mixed into hydrothermal fluid, the hydrothermal fluid also heats seawater. Can this lead to any minerals being precipitated?
- Yes. Heating seawater leads to precipitation of anhydrite.

Hot seawater, as well as mixtures of seawater and hydrothermal fluid, precipitates anhydrite where the hydrothermal fluid enters the sea, building a pipe that channels the hydrothermal fluid upwards. This pipe grows in size through a combination of sulfide and sulfate (anhydrite) precipitation, and matures in shape to form a chimney. Temperature gradients across the chimney wall, from the hot hydrothermal fluid flowing up the chimney to the cold seawater outside it, and mixing of fluids in the porous walls of the chimney, generate a compositional structure within the chimney as it grows, with the outside being composed of anhydrite and the inner walls of Cu–Fe sulfide (chalcopyrite). Individual chimneys can have an internal diameter of about 10 cm and can grow to 45 m in height, though most are much shorter than this.

Hydrothermal discharges between 150 °C and 290 °C produce white smokers, the white colour being due to pale-coloured minerals and precipitates, mainly of anhydrite, barite (barium sulfate), and amorphous (non-crystalline) silica.

Over time, mineral growth and dissolution in hydrothermal systems can clog the pathways used by the fluids and discharge may wane or cease, only to start up elsewhere. Once discharge has ceased, a further set of reactions takes place between seawater and the now cold mineral-rich deposits (the sulfide and sulfate-rich chimneys and sediment). Most important is the gradual dissolution of anhydrite into cold seawater. Because anhydrite cements parts of the chimneys together, this process leads to the weakening and collapse of chimneys, eventually reducing them to piles of sulfide-rich rubble.

4.7.3 Hydrothermal systems and the composition of seawater

Hydrothermal systems alter cold seawater by heating it, causing it to react with hot basalt and gabbro of the oceanic crust, and then rise to the surface. Because normal seawater enters the hydrothermal system but hot, mineral-rich water exits the system, the overall process results in a change to the original igneous composition of the crust. Equally, the input of hydrothermal fluid from black

smokers and other discharging vents adds components to the ocean water. Through these effects, the oceanic crust becomes hydrated whereas the ocean becomes depleted in some elements (particularly Mg) and enriched in others (such as Mn and Fe). Exact estimates of the strength of these sinks and sources are difficult to make, principally because hydrothermal fluids are very variable in composition and the rate of discharge is known in only a few isolated cases and for a very short period of time. A further complication is that Mn and Fe in hydrothermal fluids become removed from the ocean by precipitation of sulfides. So, the slow evolution of the chemical composition of ocean water by inputs from rivers and interactions with oceanic crust and hydrothermal processes is easier to conceptualise than it is to quantify.

Summary of Chapter 4

This chapter has taken a path that started in the upper mantle and followed the fate of hot peridotite as it rises towards the surface beneath a mid-ocean ridge as sea-floor spreading pulls the lithosphere apart. The geological processes that progressively change the composition of the mantle and crust (partial melting, fractional crystallisation, and alteration) have been studied largely in terms of the chemical changes that are caused when Earth materials experience changes in pressure, temperature or the chemical environment.

Decompression partial melting of peridotite produces basaltic magma that then cools and undergoes fractional crystallisation at shallow (crustal to uppermost mantle) levels to produce the range of volcanic and intrusive rocks that comprise oceanic crust. High-temperature hydrothermal systems at the mid-ocean ridge then modify large sections of the igneous crust, generating a hydrous greenstone crust, sulfide mineral deposits, and moderate the chemical composition of the oceans. High heat and chemical fluxes sustain sulfur-oxidising bacteria that are the primary producers in a food chain of deep-sea biota.

Learning outcomes for Chapter 4

You should now be able to demonstrate a knowledge and understanding of:

- 4.1 The layered structure of the oceanic lithosphere.
- 4.2 The different rocks and minerals that make up the mantle and oceanic crustal rocks.
- 4.3 Variations in the mineralogy and melting temperature of lherzolite with increasing pressure.
- 4.4 How the magmatic processes of partial melting and fractional crystallisation account for differences in the composition and evolution of mantle and oceanic crustal rocks.
- 4.5 The processes involved in decompression melting beneath oceanic ridges and how the percentage of melt produced is directly related to the potential temperature of the upper mantle, the depth of melting and the thickness of the oceanic crust.
- 4.6 The interactions and reactions that occur between seawater and the oceanic crust within hydrothermal systems beneath oceanic ridges.

Answers to questions

Question 2.1

(a) From Equation 2.2:

$$\Delta T = \frac{mv^2}{2MC}$$

where $m = 10^{15}$ kg (given), $v = 10^4$ m s⁻¹ (given), $C = 7.5 \times 10^2$ J kg⁻¹ K⁻¹ (given), and $M = 6.0 \times 10^{24}$ kg (given).

$$\begin{aligned}\Delta T &= \frac{10^{15} \text{ kg} \times (10^4 \text{ m s}^{-1})^2}{2 \times (6.0 \times 10^{24} \text{ kg}) \times (7.5 \times 10^2 \text{ J kg}^{-1} \text{ K}^{-1})} \\ &= 1.1 \times 10^{-5} \text{ K}\end{aligned}$$

(b) The number of planetesimals of mass 10^{15} kg each required to construct the Earth is:

$$\frac{6 \times 10^{24} \text{ kg}}{10^{15} \text{ kg}} = 6 \times 10^9$$

If each of these produced the temperature rise in part (a), the total temperature rise in the Earth would be $(6 \times 10^9) \times (1.1 \times 10^{-5} \text{ K}) = 66\,000 \text{ K}$.

This is a very crude, but very interesting, calculation. The temperature rise given by part (a) was based on the impact of a 10^{15} kg planetesimal on an Earth of current size. In practice, most of the planetesimals going into the construction of the Earth would have impacted a *growing* Earth (i.e. one smaller than the present Earth). As Equation 2.2 shows, if M were smaller, ΔT would be bigger. In other words, the 66 000 K temperature rise obtained above can be regarded as a *minimum* temperature rise – and it is already well over the vaporisation temperature of refractories (see Figure 1.16).

Question 2.2

The number of seconds in a year = 3.2×10^7 s y⁻¹ (to 2 sig. figs).

So the rate of heating is:

$$\frac{3.0 \times 10^{19} \text{ J y}^{-1}}{3.2 \times 10^7 \text{ s y}^{-1}} = 9.375 \times 10^{11} \text{ J s}^{-1} = 9 \times 10^{11} \text{ W (to 1 sig. fig)}$$

The rate of tidal heating generated within the Earth is approximately:

$$\begin{aligned}\frac{9 \times 10^{11} \text{ W}}{6 \times 10^{24} \text{ kg}} &= 1.5 \times 10^{-13} \text{ W kg}^{-1} \\ &= 2 \times 10^{-13} \text{ W kg}^{-1} \text{ (to 1 sig. fig.)}\end{aligned}$$

The data in Question 2.2 is to 1 significant figure, therefore your answer should be to 1 significant figure.

Question 2.3

- (a) The age of the Earth is 4.6×10^9 y. To calculate the proportion of the original radiogenic isotope remaining today, you first need to determine how many half-lives have occurred since the Earth was formed:

$$\text{number of half-lives} = \frac{\text{age of Earth}}{\text{isotope half-life}}$$

The proportion remaining after one half-life is $(\frac{1}{2})^1 = 0.5$, after two half-lives it is $(\frac{1}{2})^2 = 0.25$, and after three half-lives it is $(\frac{1}{2})^3 = 0.125$.

For ^{40}K , the number of half-lives =

$$\frac{4.6 \times 10^9 \text{ y}}{1.28 \times 10^9 \text{ y}} = 3.59 \text{ half-lives (to 3 sig. figs)}$$

So the proportion remaining today is $(\frac{1}{2})^{3.59} = 0.083$ (to 2 sig. figs).

Note: on many calculators this function can be performed by the key labelled x^y , which is often operated by the SHIFT key. Your calculator may differ in detail, but for many common calculators this calculation can be performed by keying:

$$0.5 \text{ [SHIFT] [x}^y\text{] } 3.59 \text{ [=]}$$

For ^{232}Th , the number of half-lives

$$\frac{4.6 \times 10^9 \text{ y}}{1.39 \times 10^{10} \text{ y}} = 0.331 \text{ half-lives (to 3 sig. figs)}$$

So the proportion remaining today is $(\frac{1}{2})^{0.331} = 0.79$ (to 2 sig. figs).

The data in Question 2.3a are to 2 significant figures, therefore your answer should be to 2 significant figures.

- (b) The amount of radiogenic heating is determined by utilising the proportion of the different radiogenic isotopes remaining today (you need to calculate the values for ^{235}U and ^{238}U as above) to calculate the original heat budget.

If only 0.0860 of the original ^{40}K remains today and its present rate of heat generation is $2.8 \times 10^{-12} \text{ W kg}^{-1}$ (from Table 2.1), its original heat budget is:

$$\frac{2.8 \times 10^{-12} \text{ W kg}^{-1}}{0.083} = 3.37 \times 10^{-11} \text{ W kg}^{-1} \text{ (or } 34 \times 10^{-12} \text{ W kg}^{-1} \text{ to 2 sig. figs)}$$

Note: the lowest level of accuracy is 2 sig. figs so the answer should be given to 2 sig. figs.

Likewise:

$$\text{initial heating by } ^{232}\text{Th decay} = \frac{1.04 \times 10^{-12} \text{ W kg}^{-1}}{0.79} = 1.3 \times 10^{-12} \text{ W kg}^{-1} \text{ (to 2 sig. figs)}$$

$$\text{initial heating by } ^{235}\text{U decay} = \frac{0.04 \times 10^{-12} \text{ W kg}^{-1}}{0.0113} = 3.6 \times 10^{-12} \text{ W kg}^{-1} \text{ (to 2 sig. figs)}$$

$$\text{initial heating by } ^{238}\text{U decay} = \frac{0.96 \times 10^{-12} \text{ W kg}^{-1}}{0.492} = 2.0 \times 10^{-12} \text{ W kg}^{-1} \text{ (to 2 sig. figs)}$$

Note: the lowest level of accuracy is 2 sig. figs so the answer should be given to 2 sig. figs.

By adding the four answers, the total initial heating was therefore about $4 \times 10^{-11} \text{ W kg}^{-1}$ compared with a modern total of about $5 \times 10^{-12} \text{ W kg}^{-1}$ (obtained by adding the four present-day heating rates given in Table 2.1), in other words it was at least eight times greater (we have not accounted for other short-lived isotopes, such as ^{26}Al , and the heat from their decay). This initial decay would have originally provided a major source of internal heating even in relatively small planetary bodies, and so would have contributed significantly to early planetary differentiation.

- (c) By summing the four values quoted in Table 2.1, the present rate of radiogenic heating is about $5 \times 10^{-12} \text{ W kg}^{-1}$, which is about 30 times greater than that attributable to estimates of tidal heating, and so represents the present major source of the Earth's internal heating.

Question 2.4

From Equation 2.13:

$$k = \frac{a^2 \Phi}{24\pi} \text{ m}^2$$

where $a = 10^{-3}$ ($0.001 = 1 \text{ mm}$), $\Phi = 0.1$

$$\begin{aligned} k &= \frac{(0.001 \text{ m})^2 \times 0.1}{24\pi} \\ &= \frac{1 \times 10^{-7} \text{ m}^2}{75.398} = 1.33 \times 10^{-9} \text{ m}^2 \end{aligned}$$

Using this estimate of permeability (k) in Equation 2.12:

$$v = \frac{k}{\eta} \Delta \rho g$$

where $\eta = 0.005 \text{ Pa s}$ ($1 \text{ Pa s} = 1 \text{ kg m}^{-1} \text{ s}^{-1}$), $\Delta \rho = 3500 \text{ kg m}^{-3}$, and $g = 9.8 \text{ m s}^{-2}$

$$\begin{aligned} v &= \frac{1.33 \times 10^{-9} \text{ m}^2}{0.005 \text{ Pa s}} \times 3500 \text{ kg m}^{-3} \times 9.8 \text{ m s}^{-2} \\ &= \frac{1.33 \times 10^{-9} \text{ m}^2}{0.005 \text{ kg m}^{-1} \text{ s}^{-1}} \times 3500 \text{ kg m}^{-3} \times 9.8 \text{ m s}^{-2} \\ &= \frac{2.66 \times 10^{-7} \text{ m}^3}{\text{kg s}^{-1}} \times 34\,300 \text{ kg m}^{-2} \text{ s}^{-2} \\ &= 9.1 \times 10^{-3} \text{ m s}^{-1} \end{aligned}$$

There are 3.2×10^7 seconds in a year and 10^3 metres in a kilometre.

$$\begin{aligned} \text{So the migration velocity} &= \frac{(9.1 \times 10^{-3} \text{ m s}^{-1}) \times (3.2 \times 10^7 \text{ s y}^{-1})}{1000 \text{ m}} \\ &= 2.912 \text{ km y}^{-1} \\ &= 290 \text{ km y}^{-1} \text{ to 2 sig. figs} \end{aligned}$$

The distance between the Earth's surface and the outer liquid core is 2900 km, which suggests that it would take about 10 years for metallic melt to sink to the core. During accretion, the Earth (and core) would have been growing, and gravitational acceleration would also be changing. If gravitational acceleration were smaller then this would reduce the migration velocity of the melt. For example, for a Mars-sized body with $g = 3.7 \text{ m s}^{-2}$, the migration velocity would be about 110 km y^{-1} , but the depth to the core would be less.

Question 2.5

The $\epsilon^{182}\text{W}$ value of the meteorite is:

$$\begin{aligned}\epsilon^{182}\text{W} &= \left[\left(\frac{0.864523}{0.864696} \right) - 1 \right] \times 10^4 \\ &= (0.999799 - 1) \times 10^4 \\ &= -0.000200 \times 10^4 \\ \epsilon^{182}\text{W} &= -2.0 \text{ (to 2 sig. figs)}\end{aligned}$$

Question 2.6

Compared with chondritic composition, the Earth's mantle is depleted in Rb, K and Na (Figure 1.22). This cannot be explained in terms of segregation into the core, as was the case for Fe and Ni depletion, since the core does not contain these elements. Instead, these depletions can be best explained using a similar argument to that offered to explain the Moon's depletion in volatiles. During the final stages of the assembly of Earth, the amalgamated planetary embryos were beginning to differentiate into layered bodies with dense cores, and with elements such as Rb, K and Na segregating toward the surface. Continuing impacts would have partially vaporised the surface layers, and some of the more volatile elements, including Rb, K and Na, (and probably Zn) may have been lost into space. Later in Earth history, the segregation of material to form oceanic and continental crust would have further augmented removal of these elements from the mantle.

Question 3.1

The coalfields formed from plants that grew in the tropical and sub-tropical climate belts. They must have drifted to their present positions from these latitudes either south or north of the Equator (i.e. 23° S – 23° N).

This represents a latitude drift of between 23° S to 55° N (i.e. 78°) and 23° N to 55° N (i.e. 32°).

If the Earth's radius is 6370 km, its circumference must be $2\pi \times 6370 \text{ km} = 40\,024 \text{ km}$.

1° of latitude (assuming the Earth is a sphere) is therefore $\frac{40\,024 \text{ km}}{360} = 111 \text{ km}$

Therefore the *minimum* distance that Britain can have drifted since the late Carboniferous is $32 \times 111 = 3552$ km (i.e. about 3500 km at the level of accuracy of this information), which gives a rate of:

$$\frac{3500 \times 10^6 \text{ mm y}^{-1}}{300 \times 10^6} \approx 12 \text{ mm y}^{-1} \text{ or } 10 \text{ mm y}^{-1} \text{ to 1 sig. fig.}$$

The *maximum* distance is at least 8672 km (i.e. about 8700 km), which gives a rate of:

$$\frac{8700 \times 10^6 \text{ mm y}^{-1}}{300 \times 10^6} \approx 29 \text{ mm y}^{-1} \text{ or } 30 \text{ mm y}^{-1} \text{ to 1 sig. fig.}$$

Both these values could be larger if Britain drifted in terms of longitude as well – in other words, if its course was not in a straight line.

Question 3.2

The APW parts of Europe and Siberia are the same as far back as the Triassic, but before this time the Siberian pole was to the west of the European pole. This indicates that the two regions were part of different land masses until the Triassic; at this time they must have collided, and afterwards they continued to move as a single unit.

Question 3.3

For the South Atlantic Ocean:

$$\text{half spreading rate} = \frac{38 \text{ mm y}^{-1}}{2} = 19 \text{ mm y}^{-1} \text{ or } 19 \text{ km Ma}^{-1}$$

The age of ocean crust (t) adjacent to the South American continental shelf is derived from:

$$\text{age} = \frac{\text{distance}}{\text{half spreading rate}} = \frac{3100 \text{ km}}{19 \text{ km y}^{-1}} = 163 \text{ Ma (or 160 Ma to 2 sig. figs)}$$

The age of ocean crust adjacent to the southern African continental shelf

$$= \frac{2700 \text{ km}}{19 \text{ km y}^{-1}} = 142 \text{ Ma or } 140 \text{ Ma to 2 sig. figs.}$$

Question 3.4

(a) (i) Rearranging Equation 3.1:

$$t = \left(\frac{d - 2500}{350} \right)^2 \text{ Ma} = \left(\frac{4700 - 2500}{350} \right)^2 \text{ Ma} \\ = 6.285^2 \text{ Ma} = 39.5 \text{ Ma} = 40 \text{ Ma (to 2 sig. figs)}$$

The age of the crust at this location is 40 Ma.

(a) (ii) The mean spreading rate is given by distance divided by time: $v = \frac{D}{t}$

$$\text{Therefore } v = \frac{1600 \times 10^6 \text{ m}}{40 \times 10^6 \text{ mm y}^{-1}} = \frac{1600}{40 \text{ mm y}^{-1}} = 40 \text{ mm y}^{-1}$$

(b) This is a half spreading rate because it refers to one side of the constructive margin only.

Question 3.5

- (a) Ocean depth increases with the square root of lithosphere age so the correlation is positive.
- (b) The departure from linearity occurs between 8 and 9 on the horizontal scale, which is equivalent to an age of 64–81 Ma.
- (c) For crust older than 64 Ma, ocean depth is less than predicted, i.e. the ocean floor is shallower, implying that it is warmer than predicted by the boundary-layer model.

Question 3.6

Compare your answer with Table 3.5.

Table 3.5 Completed Table 3.1.

	Relative plate motion/mm y ⁻¹	Direction of plate motion
American Plate fixed	40	east
Australian Plate fixed	40	west

If the South American Plate is stationary, then the spreading rate across the Mid-Atlantic Ridge means that the African Plate is moving at a rate of 40 mm y⁻¹ to the east, because the ridge is migrating east at the half spreading rate while the African Plate is moving at the full spreading rate. Conversely, if the Australian Plate is held stationary, then the Carlsberg Ridge migrates to the west at a rate of 20 mm y⁻¹ and the African Plate moves to the west at a rate of 40 mm y⁻¹.

Question 3.7

You should recall that:

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

but the graph is plotted as time against distance, so the gradient of the graph as it

is plotted is $\frac{\text{time}}{\text{distance}}$. To calculate the gradient, and hence the speed, choose two points on the best-fit straight line that are some distance apart and find their coordinates. For example, at a distance of 5500 km the age is 63 Ma, and at a distance of 1000 km the age is 11 Ma. The gradient of the line is:

$$\frac{63 \text{ Ma} - 11 \text{ Ma}}{5500 \text{ km} - 1000 \text{ km}} = \frac{52 \text{ Ma}}{4500 \text{ km}} = 0.0116 \text{ Ma km}^{-1}$$

The speed is therefore $\frac{1}{0.0116 \text{ Ma km}^{-1}}$
 $= 86.5 \text{ Ma km}^{-1} = 87 \text{ Ma km}^{-1}$ (to 2 sig. figs)

This is equivalent to 87 mm y^{-1} . The current overall direction of migration of volcanism is towards the southeast, therefore the plate is moving to the northwest. However, prior to about 43–45 Ma volcanism migrated more or less due south, therefore the Pacific Plate prior to 43–45 Ma was moving to the north.

Question 3.8

Table 3.6 Completed Table 3.2.

Force	Acts as a driving force	Acts as a resistive force	Might act as <i>either</i> a driving force <i>or</i> a resistive force
oceanic drag			✓
continental drag		✓	✓
ridge-push	✓		
transform fault		✓	
slab-pull	✓		
slab resistance		✓	
trench suction	✓	✓	

Question 3.9

Table 3.7 Completed Table 3.4.

	Good positive correlation	Poor correlation	Good negative correlation
Total plate area		✓	
Total continental area			✓
Effective length of ridge boundary		✓	
Percentage circumference of plate attached to descending slab	✓		
Percentage circumference of plate that is a transform fault		✓	

The positive correlation between absolute plate speed with the percentage of circumference attached to a descending slab implies that the slab-pull force (F_{SP}) is a major plate driving force.

Question 3.10

The spreading rate is equivalent to an ocean width increase of 30 km Ma^{-1} . For the continents to meet on the opposite side of the globe, the ocean will need to have opened to half the circumference of the Earth less the original width of the continent (i.e. $20\,000 \text{ km} - 5000 \text{ km} = 15\,000 \text{ km}$). The time taken to achieve this will be:

$$\text{time} = \frac{\text{distance}}{\text{speed}} = \frac{15000 \text{ km}}{30 \text{ km Ma}^{-1}} = 500 \text{ Ma}$$

Question 4.1

- (a) Olivine, pyroxene, garnet and plagioclase feldspar.
- (b) Spinel.
- (c) Garnet, plagioclase feldspar and spinel.
- (d) Olivine, pyroxene, garnet and spinel. (*Note:* pyroxene and garnet also contain calcium.)

Question 4.2

The total charge of one Si^{4+} ion (4+) and four O^{2-} ions (8-) is 4-. The formula for the tetrahedron can be written as $(\text{SiO}_4)^{4-}$.

Question 4.3

- (a) $\text{Mg\#} = \frac{\text{Mg}}{\text{Mg} + \text{Fe}^{2+}} = \frac{1.8}{1.8 + 0.2} = \frac{1.8}{2.0} = 0.90 \text{ or } 90\%$
- (b) $\text{Mg\#} = \frac{\text{Mg}}{\text{Mg} + \text{Fe}^{2+}} = \frac{1.2}{1.2 + 0.4} = \frac{1.2}{1.6} = 0.75 \text{ or } 75\%$

Question 4.4

The composition represented by X is 80% A and 20% B ($\text{A}_{80}\text{B}_{20}$). The composition represented by point Y is 20% A, 60% B and 20% C ($\text{A}_{20}\text{B}_{60}\text{C}_{20}$).

Question 4.5

- (a) The average oceanic peridotite has the composition of a harzburgite.
- (b) Reading from Figure 4.12, the composition AO has roughly 75% olivine, 21% orthopyroxene and 4% clinopyroxene.

Question 4.6

According to the phase diagram for KLB-1 (Figure 4.15), spinel lherzolite is stable at pressures equivalent to between about 25 km and 85 km. (*Note:* given that the sample came from below the Moho and that the crust is typically about 35 km thick, then the minimum depth would actually be about 35 km.)

Question 4.7

Spinel lherzolites are typical of mantle xenoliths found in basalts, whereas garnet lherzolites are typical of kimberlites (see Section 4.2.2). So, according to the phase diagram in Figure 4.15, kimberlites must come from within the garnet stability field, whereas basalts are more likely to come from the more shallow spinel lherzolite field. The presence of diamonds within some kimberlites is also consistent with these magmas coming from great depth.

Question 4.8

Reading from Figure 4.15, KLB-1 is solid at 1.5 GPa below about 1350 °C and liquid above about 1830 °C. KLB-1 is a mixture of crystals and liquid between these temperatures.

Question 4.9

Statements A, C, E, F, G and J are correct.

A is correct because, according to Table 4.3 and Figure 4.16, the per cent liquid changes from 0 to greater than 0 between 1300 °C and 1400 °C. The phase diagram in Figure 4.15 shows the solidus close to 1350 °C.

C and E are correct according to Table 4.3.

F and G are correct according to Figure 4.16.

J is correct because peridotite KLB-1 has an MgO content of about 40% (the composition of totally molten KLB-1) whereas the liquid produced nearer the solidus has a lower MgO content. At the solidus of MM3 the liquid has about 10% MgO, and extrapolating the results from KLB-1 to its solidus temperature implies a similar MgO content in that case too.

The wrong statements should read (modifications in *italics*):

B The first phase to disappear with increasing temperature is *spinel*.

D *Olivine* is the mineral phase that survives to the highest temperature.

H The Fo content of olivine becomes *higher* with increasing amount of melting.

I The MgO content of the liquid *increases* with increasing temperature.

Question 4.10

This sample plots in the solid forsterite + liquid field. The composition of the liquid coexisting with solid Fo is defined by the liquidus at 1800 °C, i.e. Fo₅₈Di₄₂.

Question 4.11

(a) Melting first occurs at the eutectic temperature of 1580 °C.

(b) At temperatures just above the eutectic temperature the sample will plot in the solid diopside + liquid field, meaning that forsterite is the first mineral to be consumed. This is because the bulk sample has a composition that falls on the Di-rich side of the eutectic composition.

(c) The first liquid has the composition of the eutectic, i.e. Fo₁₉Di₈₁.

(d) The sample becomes completely molten at the liquidus temperature for Fo₁₀Di₉₀. This is read from the phase diagram (with care, because the liquidus curve has a shallow slope) as 1585 °C.

Question 4.12

The behaviours described in statements D, E and F are due to there being more than one mineral present and are therefore similar to the behaviour of a eutectic system.

The behaviour described in statement B is caused by solid solution between Fo and Fa in olivine.

The behaviours described in statements A and C are encountered in both eutectic and solid-solution systems. In the case of statement C, MgO increases because the composition of peridotite is MgO rich, and Mg-rich olivine is the last mineral to melt completely.

Question 4.13

The mantle with the higher T_p crosses the peridotite solidus at a greater depth and reaches a final temperature that is farther above the solidus than the mantle with the lower T_p . The hotter mantle is therefore the one that yields the greater amount of liquid.

Question 4.14

From Figure 4.25a, a crustal thickness of 27 km implies a potential temperature of about 1480 °C.

Question 4.15

From Figure 4.29, the liquidus of mid-ocean ridge basalt with Mg# = 65% is about 1200–1230 °C and for Mg# = 50% about 1140–1160 °C, so typical eruption temperatures are in the range 1140–1230 °C.

Question 4.16

The order of crystallisation is: olivine; olivine + plagioclase; olivine + plagioclase + clinopyroxene.

Question 4.17

(a) Using Equation 4.1 (see Box 4.4) the pressure due to the weight of 3 km of seawater = $(1.0 \times 10^3) \text{ kg m}^{-3} \times 9.8 \text{ m s}^{-2} \times (3 \times 10^3) \text{ m} = 3 \times 10^7 \text{ Pa}$
(to 1 sig. fig.) = 30 MPa.

(b) The pressure 30 km below the sea floor is the sum of the pressure due to the water and the rocks. The pressure due to the water and rocks is
 $30 \text{ MPa} + (3 \times 10^3) \text{ kg m}^{-3} \times 9.8 \text{ m s}^{-2} \times (30 \times 10^3) \text{ m} = 9 \times 10^8 \text{ Pa}$ (to 1 sig. fig.) = 900 MPa.

Question 4.18

(a) Statements A and B are true. Statement C is false – none of these minerals is hydrous.

(b) (i) albite: Al_2O_3 and Na_2O

(ii) chlorite: Al_2O_3 , FeO , MgO

(iii) epidote: Al_2O_3 , FeO , CaO

(iv) actinolite: FeO , MgO , CaO and moderate amounts of Al_2O_3 .

(c) Water ranges from less than 2% to more than 6 %.

The elements

Table A1 The chemical elements and their abundances. (Data from Lodders, 2003)

Atomic number, Z	Name	Chemical symbol	Relative atomic mass, A_r	Solar System abundance		CI Chondrite abundance by mass/ppm
				Solar photosphere by number relative to Si = 1×10^6	CI chondrites by number relative to Si = 1×10^6	
1	hydrogen	H	1.01	2.88×10^{10}	5.5×10^6	21 000
2	helium	He	4.00	2.29×10^9	0.604	9×10^{-3}
3	lithium	Li	6.94	0.363	55.5	1.46
4	beryllium	Be	9.01	0.407	0.74	0.025
5	boron	B	10.81	14.5	17.3	0.71
6	carbon	C	12.01	7.7×10^5	7.1×10^6	35 000
7	nitrogen	N	14.01	1.95×10^6	5.54×10^4	2940
8	oxygen	O	16.00	1.41×10^7	7.55×10^6	458 000
9	fluorine	F	19.00	1047	841	60
10	neon	Ne	20.18	2.15×10^6	2.36×10^{-3}	1.8×10^{-4}
11	sodium	Na	22.99	5.75×10^4	5.75×10^4	5010
12	magnesium	Mg	24.31	1.00×10^6	1.04×10^6	95 900
13	aluminium	Al	26.98	8.51×10^4	8.31×10^4	8500
14	silicon	Si	28.09	1.0×10^6	1.0×10^6	107 000
15	phosphorus	P	30.97	8913	7833	920
16	sulfur	S	32.07	4.63×10^5	4.50×10^5	54 100
17	chlorine	Cl	35.45	9120	5240	704
18	argon	Ar	39.95	1.03×10^5	9.06×10^{-3}	1.33×10^{-3}
19	potassium	K	39.10	3800	3580	530
20	calcium	Ca	40.08	6.61×10^4	5.97×10^4	9070
21	scandium	Sc	44.96	42.7	34.2	5.83
22	titanium	Ti	47.88	3020	2422	440
23	vanadium	V	50.94	288.4	288.4	55.7
24	chromium	Cr	52.00	1.26×10^4	1.31×10^4	2590
25	manganese	Mn	54.94	7079	9170	1910
26	iron	Fe	55.85	8.13×10^5	8.63×10^5	183 000
27	cobalt	Co	58.93	2400	2250	502
28	nickel	Ni	58.69	4.78×10^4	4.78×10^4	10 600
29	copper	Cu	63.55	468	527	127
30	zinc	Zn	65.39	1202	1250	310
31	gallium	Ga	69.72	21.9	36	9.51
32	germanium	Ge	72.61	110	121	33.2

Atomic number, Z	Name	Chemical symbol	Relative atomic mass, A_r	Solar System abundance		CI Chondrite abundance by mass/ppm
				Solar photosphere by number relative to Si = 1×10^6	CI chondrites by number relative to Si = 1×10^6	
33	arsenic	As	74.92	—	6.09	1.73
34	selenium	Se	78.96	—	65.8	19.7
35	bromine	Br	79.90	—	11.3	3.43
36	krypton	Kr	83.80	55.2	1.64×10^{-4}	5.22×10^{-5}
37	rubidium	Rb	85.47	11.5	6.57	2.13
38	strontium	Sr	87.62	24.0	23.3	7.74
39	yttrium	Y	88.91	4.68	4.54	1.53
40	zirconium	Zr	91.22	11.2	11.5	3.96
41	niobium	Nb	92.91	0.759	0.752	0.265
42	molybdenum	Mo	95.94	2.40	2.80	1.02
43	technetium	Tc ^a	98.91	— ^b	— ^c	— ^c
44	ruthenium	Ru	101.07	2.00	1.81	0.692
45	rhodium	Rh	102.91	0.38	0.36	0.141
46	palladium	Pd	106.42	1.41	1.46	0.588
47	silver	Ag	107.87	(0.25)	0.49	0.201
48	cadmium	Cd	112.41	1.70	1.58	0.675
49	indium	In	114.82	1.05	0.18	0.079
50	tin	Sn	118.71	2.88	3.73	1.68
51	antimony	Sb	121.76	0.29	0.33	0.152
52	tellurium	Te	127.60	—	4.82	2.33
53	iodine	I	126.90	—	1.00	0.48
54	xenon	Xe	131.29	5.39	3.5×10^{-4}	1.74×10^{-4}
55	caesium	Cs	132.91	—	0.367	0.185
56	barium	Ba	137.33	4.27	4.44	2.31
57	lanthanum	La	138.91	0.389	0.441	0.232
58	cerium	Ce	140.12	1.10	1.169	0.621
59	praseodymium	Pr	140.91	0.148	0.174	0.093
60	neodymium	Nd	144.24	0.912	0.836	0.457
61	promethium	Pm ^a	146.92	— ^c	— ^c	— ^c
62	samarium	Sm	150.36	0.2818	0.254	0.145
63	europium	Eu	151.96	0.096	0.095	0.0546
64	gadolinium	Gd	157.25	0.380	0.332	0.198
65	terbium	Tb	158.93	0.055	0.059	0.036
66	dysprosium	Dy	162.50	0.398	0.386	0.238
67	holmium	Ho	164.93	0.098	0.090	0.056
68	erbium	Er	167.26	0.246	0.255	0.162
69	thulium	Tm	168.93	(0.029)	0.037	0.024

Atomic number, Z	Name	Chemical symbol	Relative atomic mass, A_r	Solar System abundance		CI Chondrite abundance by mass/ppm
				Solar photosphere by number relative to Si = 1×10^6	CI chondrites by number relative to Si = 1×10^6	
70	ytterbium	Yb	170.04	0.347	0.248	0.163
71	lutetium	Lu	174.97	0.033	0.036	0.024
72	hafnium	Hf	178.49	0.219	0.170	0.115
73	tantalum	Ta	180.95	—	0.021	0.014
74	tungsten	W	183.85	(0.372)	0.128	0.089
75	rhenium	Re	186.21	—	0.053	0.037
76	osmium	Os	190.2	0.813	0.674	0.486
77	iridium	Ir	192.22	0.692	0.645	0.470
78	platinum	Pt	195.08	1.59	1.36	1.00
79	gold	Au	196.97	(0.30)	0.196	0.146
80	mercury	Hg	200.59	—	0.413	0.314
81	thallium	Tl	204.38	(<0.36)	0.185	0.143
82	lead	Pb	207.2	2.88	3.26	2.56
83	bismuth	Bi	208.98	-	0.14	0.110
84	polonium	Po ^a	209.98	— ^c	— ^c	— ^c
85	astatine	At ^a	209.99	— ^c	— ^c	— ^c
86	radon	Rn ^a	222.02	— ^c	— ^c	— ^c
87	francium	Fr ^a	223.02	— ^c	— ^c	— ^c
88	radium	Ra ^a	226.03	— ^c	— ^c	— ^c
89	actinium	Ac ^a	227.03	— ^c	— ^c	— ^c
90	thorium	Th ^a	232.04	—	0.035	0.0309
91	protoactinium	Pa ^a	231.04	— ^c	— ^c	— ^c
92	uranium	U ^a	238.03	<0.01	0.0093	0.0084

^a No stable isotopes.

^b Detected in spectra of some rare evolved stars.

^c No naturally occurring long-lived isotopes.

Appendix B

SI fundamental and derived units

Table B1 SI fundamental and derived units.

Quantity	Unit	Abbreviation	Equivalent units
mass	kilogram	kg	
length	metre	m	
time	second	s	
temperature	kelvin	K	
angle	radian	rad	
area	square metre	m ²	
volume	cubic metre	m ³	
speed, velocity	metre per second	m s ⁻¹	
acceleration	metre per second squared	m s ⁻²	
density	kilogram per cubic metre	kg m ⁻³	
frequency	hertz	Hz	(cycles) s ⁻¹
force	newton	N	kg m s ⁻²
pressure	pascal	Pa	kg m ⁻¹ s ⁻² , N m ⁻²
energy	joule	J	kg m ² s ⁻²
power	watt	W	kg m ² s ⁻³ , J s ⁻¹
specific heat capacity	joule per kilogram kelvin	J kg ⁻¹ K ⁻¹	m ² s ⁻² K ⁻¹
thermal conductivity	watt per metre kelvin	W m ⁻¹ K ⁻¹	m kg s ⁻³ K ⁻¹

The Greek alphabet

Table C1 The Greek alphabet.

Name	Lower case	Upper case	Name	Lower case	Upper case
Alpha	α	A	Nu (new)	ν	N
Beta (bee-ta)	β	B	Xi (cs-eye)	ξ	Ξ
Gamma	γ	Γ	Omicron	ο	O
Delta	δ	Δ	Pi (pie)	π	Π
Epsilon	ε	E	Rho (roe)	ρ	P
Zeta (zee-ta)	ζ	Z	Sigma	σ	Σ
Eta (ee-ta)	η	H	Tau (torr)	τ	T
Theta (thee-ta; ‘th’ as in theatre)	θ	Θ	Upsilon	υ	Υ
Iota (eye-owe-ta)	ι	I	Phi (fie)	φ	Φ
Kappa	κ	K	Chi (kie)	χ	X
Lambda (lam-da)	λ	Λ	Psi (ps-eye)	ψ	Ψ
Mu (mew)	μ	M	Omega (owe-me-ga)	ω	Ω

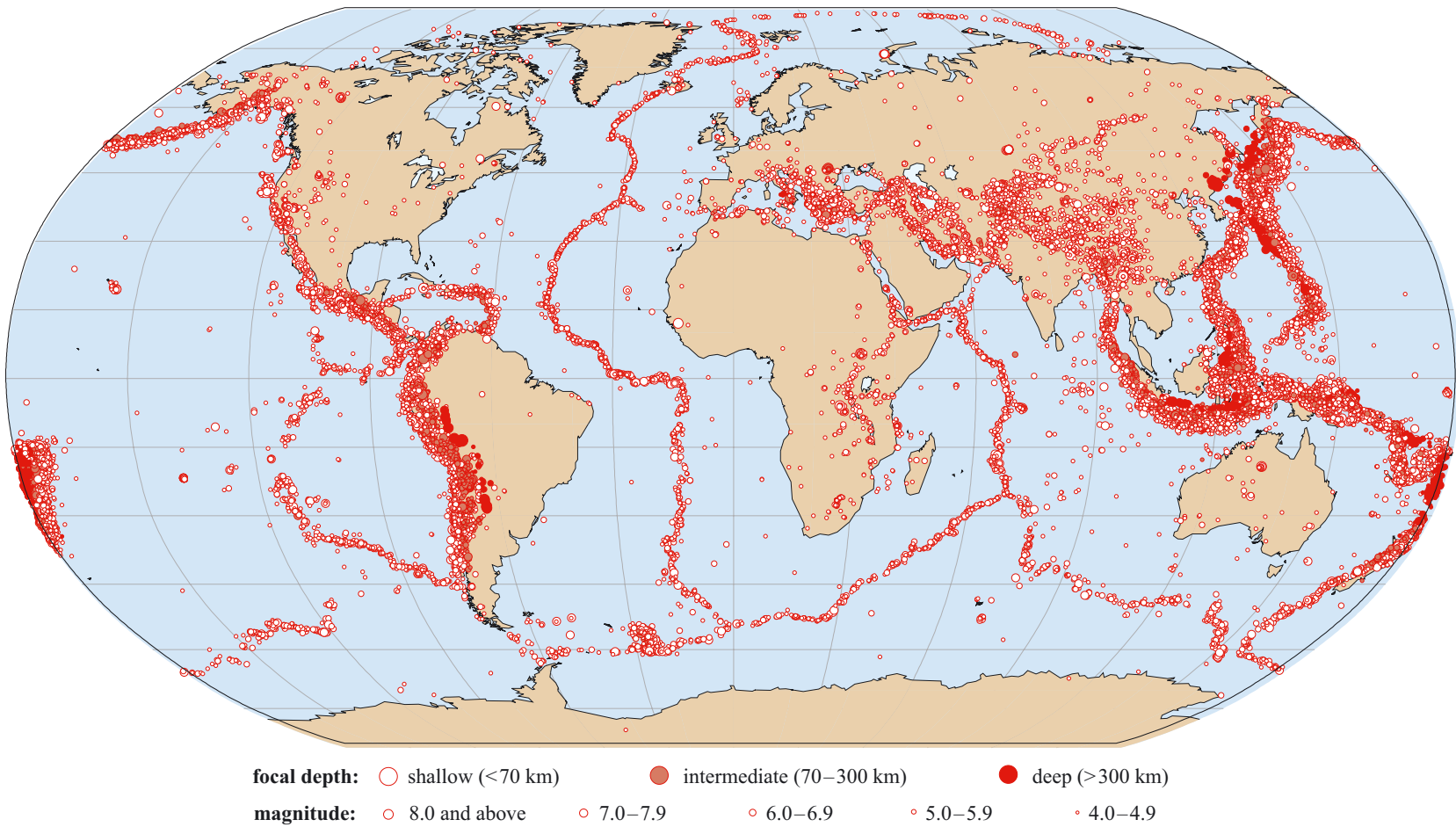


Figure 3.9(a) Global earthquake epicentres between 1980 and 1996. Only earthquakes of magnitude 4 and above are included. (BGS)

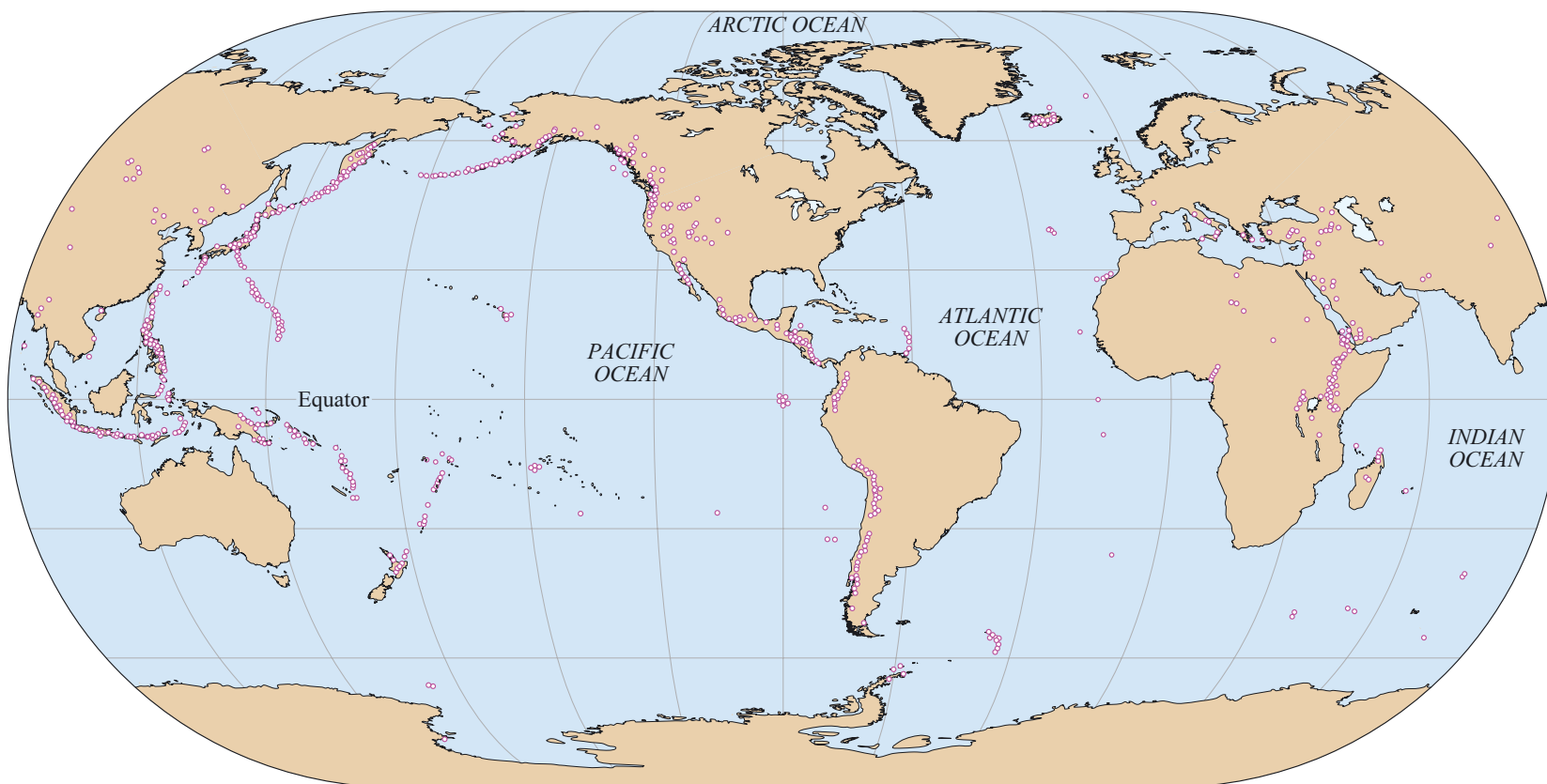


Figure 3.9(b) Map showing locations of active, sub-aerial volcanoes. (Adapted from Johnson, 1993)

Further reading

Chapter 1

Fowler, C.M.R. (2005) 'The Solid Earth' Cambridge University Press, pp. 685.

Drake, M.J. and Richter, K. (2002) 'Determining the composition of the Earth', *Nature*, vol. 416, pp. 39–44.

McDonough, W.F. and Sun, S.-s. (1995) 'The composition of the Earth', *Chemical Geology*, vol. 120, pp. 223–253.

Watts, A.B. (2001) 'Isostasy and Flexure of the lithosphere', Cambridge University Press, pp. 458. (Chapter 1 gives a good introduction to the principles behind isostasy.)

Chapter 2

Canup, R.M. and Richter, K. (eds) (2000) *Origin of the Earth and Moon*, Tuscon, University of Arizona.

Wood, B.J., Walter, M.J. and Wade, J. (2006) Accretion of the Earth and segregation of its core, *Nature*, vol. 441, pp. 825–833.

Chapter 3

Fowler, C.M.R. (2005) 'The Solid Earth' Cambridge University Press, pp. 685

Mussett, A.E. and Khan, M.A. (2000) *Looking into the Earth*, Cambridge, Cambridge University Press.

Chapter 4

Nicolas, A. (1990) *The Mid-Oceanic Ridges*, Berlin, Springer-Verlag.

Acknowledgements

Grateful acknowledgement is made to the following sources for permission to reproduce material in this book:

Cover photo Dirk Wiersma/Science Photo Library.

Figures 1.1a, 1.1b and 1.1c USGS/Cascades Volcano Observatory; *Figures 1.2a, 1.2b, 1.2d, 1.2e, 1.4a, 1.13* NASA; *Figure 1.2c* United States Geological Survey; *Figure 1.3* NOAA/NASA; *Figure 1.4b* NASA Jet Propulsion Laboratory (NASA-JPL); *Figure 1.5* Watts, A.B. (2001) *Isostasy and Flexure of the Lithosphere*, Cambridge, Cambridge University Press; *Figure 1.11* Earle, P.S. and Shearer, P.M. (1994) 'Characterization of global seismograms using an automated-picking algorithm', *Bulletin of the Seismology Society of America*, vol. 84, p. 366; *Figure 1.15* Gill, R.C.O. (1982) *Chemical Fundamentals of Geology*, Chapman and Hall; *Figure 1.16* Morgan, J.W. and Anders, E. (1980) 'The chemical composition of Earth, Venus and Mercury', *Proceedings of the National Academy of Science*, vol. 77, p. 6973; *Figure 1.17* Copyright © Natural History Museum, London; *Figure 1.19 and 1.20* McSween, H.Y. (1987) *Meteorites and their Parent Planets*, Cambridge, Cambridge University Press.

Figure 2.5 Rushmer, T., Minarik, W.G. and Taylor, G.J. (2000) 'Physical processes of core formation' in Canup, R.M. and Righter, K. (eds) *Origin of the Earth and Moon*, Tucson, University of Arizona; *Figure 2.7* Walter, M.J. et al. (2000) 'Siderophile elements in the Earth and Moon: metal/silicate partitioning and implications for core formation' in Canup, R.M. and Righter, K. (eds) *Origin of the Earth and Moon*, Tucson, University of Arizona Press; *Figure 2.11* Snyder, G.A. et al. (2000) 'Chronology and Isotopic Constraints on Lunar Evolution' in Canup, R.M. and Righter, K. (eds) *Origin of the Earth and Moon*, Tucson, University of Arizona Press; *Figure 2.15* Courtesy of Steve Moorpath; *Figure 2.16* Image courtesy of Keiko Kubo; *Figure 2.17* Reprinted with permission from Watson, E.B. and Harrison, T.M. (2005) 'Zircon thermometer reveals minimum melting conditions on earliest Earth', *Science Magazine*, vol. 308, May 2005. Copyright 2005 AAAS; *Figure 2.18* Bill Bachman/Science Photo Library; *Figure 2.19* Adapted from Abe, Y. et al. (2000) 'Water in the early Earth' in Canup, R.M. and Righter, K. (eds) *Origin of the Earth and Moon*, Tucson, University of Arizona Press.

Figure 3.1a Christian Darkin/Science Photo Library; *Figure 3.1b* M-Sat Ltd/Science Photo Library; *Figure 3.3* Hallam, A. (1975) *Alfred Wegener and the hypothesis of continental drift*, Scientific American Inc; *Figure 3.4* Creer, K.M. (1965) 'Palaeomagnetic data from gondwanic continents' in Blackett, P.M.S. et al. (eds) *A Symposium on Continental Drift*, The Royal Society; *Figures 3.5a, b and c* Mussett, A.E. and Khan, M.A. (2000) *Looking into the Earth: an introduction to geological geophysics*, Cambridge, Cambridge University Press; *Figure 3.6* Hirtzler, J. R. et al. (1966) 'Magnetic anomalies over the Reykjanes Ridge', *Deep Sea Research*, vol.13, Elsevier Science Limited; *Figures 3.8 and 3.28* Adapted from Bott, M.H.P. (1982) *The Interior of the Earth, Its Structure, Constitution and Evolution*, Edward Arnold;

Figure 3.9a Adapted from the British Geological Survey World Seismicity Database, Global Seismology and Geomagnetism Group, Edinburgh; *Figure 3.9b* Adapted from Johnson, R.W. (1993) *AGSO Issues Paper No. 1, Volcanic Eruptions and Atmospheric Change*, Australian Geological Survey Organisation; *Figure 3.11* Parsons, B. and Sclater, G. (1977) 'An analysis of the variation of ocean floor bathymetry and heat flow with age', *Journal of Geophysical Research*, vol. 82, no 5, February 1977. © 1977 American Geophysical Union; *Figures 3.13 and 3.14* Fowler, C.M.R. (2005) *The Solid Earth*, Cambridge, Cambridge University Press; *Figure 3.16* Sykes, R. (1966) 'The seismicity and deep structure of island arcs', *Journal of Geophysical Research*, vol. 71, no 12. © American Geophysical Union; *Figure 3.17* Watts, A.B. (2001) *Isostasy and flexure of the lithosphere*, Cambridge, Cambridge University Press; *Figure 3.18* Westbrook, G.K. (1982) 'The Barbados ridge complex', Special Publication of *The Geological Society of London*, Blackwell, © The Geological Society; *Figure 3.21* Dalrymple, G.B. et al. (1973) 'The origin of the Hawaiian islands', *American Scientist*, vol. 61, Scientific Research Society (Sigma XI); *Figure 3.22* Adapted from Clague, D. and Dalrymple, G.B. (1987) USGS Professional Paper 1350, United States Geological Survey; *Figure 3.30* Adapted from Livermore, R. et al. (2005) 'Paleogene opening of the Drake Passage', *Earth and Planetary Science Letters*, vol. 236, pp.459–470.

Figure 4.1a Professor R.M. Haymon; *Figures 4.1b and 4.35* Dr Ken MacDonald/Science Photo Library; *Figure 4.2* Andy Tindle (The Open University); *Figure 4.4* Adapted from Auzende, J.M. et al. (1989) 'Direct observation of a section through slow spreading oceanic crust', *Nature*, vol. 337. Copyright © Nature Publishing Group; *Figure 4.5* Nigel Harris (The Open University); *Figure 4.6* Gass, I.G. (1990) 'Ophiolites', *The Earth: Tectonics of Continents and Oceans*, W.H. Freeman and Co; *Figure 4.7a* White, R.S., McKenzie, D. and O'Nions, R.K. (1992) 'Oceanic crustal thickness from seismic measurements and rare earth element inversions', *Journal of Geophysical Research*, vol. 97, no. B13. The American Geophysical Union; *Figure 4.14* Herzberg, C. T. (1987) 'Magma density at the high pressure part 1: the effect of composition on the elastic properties of silicate liquids', *Magmatic Processes: Physiochemical Principles*, no. 1, The Geochemical Society; *Figure 4.15* Takahashi, E. (1986) 'Melting of mantle peridotite to 14 GPa', *Journal of Geophysical Research*, vol. 91. © American Geophysical Union; *Figure 4.26* Based on Roeder, P.L. and Emslie, R.F. (1970) 'Olivine-liquid equilibrium', *Contributions in Mineralogy and Petrology*, vol. 29. Springer-Verlag.

Sources of figures, data and tables

- Abe, Y. et al. (2000) 'Water in the early Earth' in Canup, R.M. and Righter, K. (eds) *Origin of the Earth and Moon*, Tucson, University of Arizona.
- Alt, J.C. (1995) 'Subseafloor processes in mid-ocean ridge hydrothermal systems' in Humphris, S.E., Zierenberg, R.A., Mullineaux, L.S. and Thompson, R.E. (eds) *Seafloor Hydrothermal Systems*. American Geophysical Union Geophysical Monograph 91, pp. 85–114.
- Anders, E. and Grevasse, N. (1989) Abundances of the elements: meteoritic and solar, *Geochimica et Cosmochimica Acta*, vol. 53, pp. 197–214
- Auzende, J.M., Bideau, D., Bonatti, E., Cannat, M., Honnorez, J., Lagabrielle, Y., Malavieille, J., Mamaloukas-Frangoulis, V. and Mevel, C. (1989) 'Direct observation of a section through slow-spreading oceanic crust', *Nature*, vol. 337, pp. 726–29.
- Boyett, M. and Carlson, R.W. (2005) '¹⁴²Nd Evidence for Early (>4.53 Ga) Global Differentiation of the Silicate Earth', *Science*, vol. 309, pp. 576–581.
- Caro, G. Bourdon, B. Birck, J.-L. and Moorbath, S. (2003) '¹⁴⁶Sm-¹⁴²Nd evidence from Isua metamorphosed sediments for early differentiation of the Earth's mantle', *Nature*, vol. 423, pp. 428–431.
- Detrick, R.S., Buhl, P., Mutter, J., Vera, E., Harding, A., Kent, G. and Orcutt, J. (1987) 'Multichannel seismic imaging of a crustal magma chamber along the East Pacific Rise', *Nature*, vol. 326, pp. 35–41.
- Drake, M.J. and Righter, K. (2002) 'Determining the composition of the Earth', *Nature*, vol. 416, pp. 39–44.
- Dziewonski, A.M. and Anderson, D.L. (1981) 'Preliminary Reference Earth Model', *Physics of the earth and Planetary Interiors*, vol. 25, pp. 297–356.
- Earle, P.S. and Shearer, P.M. (1994) 'Characterization of global seismograms using an automated-picking algorithm', *Bulletin of the Seismology Society of America*, vol. 84, pp. 366–76.
- Elthon, D., Ross, D.K. and Meen, J.K. (1995) 'Compositional variations of basaltic glasses from the Mid-Cayman Rise spreading center', *Journal of Geophysical Research*, vol. 100, pp. 12 497–512.
- Falloon, T.J., Green, D.H., Danyushevsky, L.V. and Faul, U.H. (1999) 'Peridotite melting at 1.0 and 1.5 GPa: an experimental evaluation of techniques using diamond aggregates and mineral mixes for determination of near-solidus melts', *Journal of Petrology*, vol. 40, pp. 1343–1375.
- Foley, C.N., Wadha, W. Borg, L. E., Janney, P. E., Hines, R. and Grove, T.L. (2005) 'The early differentiation of Mars from ¹⁸²W–¹⁴²Nd isotope systematics in SNC meteorites', *Geochimica et Cosmochimica Acta*, vol. 69, pp. 4557–4571.
- Gill, R.C.O. (1982) *Chemical fundamentals of Geology*, Chapman and Hall.
- Grove, T.L. and Bryan, W.B. (1983) 'Fractionation of pyroxene–phyric MORB at low pressure: an experimental study', *Contributions to Mineralogy and Petrology*, vol. 84, pp. 293–309.

- Grove, T. L., Kinzler, R.J. and Bryan, W.B. (1990) 'Natural and experimental phase relations of lavas from Serocki Volcano' in Detrick, R. et al. (eds) *Proceedings of the Ocean Drilling Program Scientific Results*, vol. 106–109, pp. 9–17.
- Grove, T.L., Kinzler, R.J. and Bryan, W.B. (1992) 'Fractionation of mid-ocean ridge basalt' in Phipps Morgan, J., Blackman, D.K. and Sinton, J.M. (eds) 'Mantle flow and melt generation at mid-ocean ridges', *American Geophysical Union Geophysical Monograph*, vol. 71, pp. 281–310.
- Herzberg, C.T. (1987) 'Magma density at high pressure, Part 1: the effect of composition on the elastic properties of silicate liquids' in Mysen, B.O. (ed.) *Magmatic Processes: Physicochemical Principles*, Geochemical Society Special Publication 1, pp. 25–46.
- Humphris, S.E. and Thompson, G. (1978) 'Hydrothermal alteration of oceanic basalts by seawater', *Geochimica et Cosmochimica Acta*, vol. 42, pp. 107–25.
- Juster, T.C., Grove, T.L. and Perfit, M.R. (1989) 'Experimental constraints on the generation of FeTi basalts, andesites, and rhyodacites at the Galapagos spreading center, 85° W and 95° W', *Journal of Geophysical Research*, vol. 94, pp. 9251–74.
- Kleine, T., Munker, C., Mezger, K. and Palme, H. (2002) 'Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry', *Nature*, vol. 418, pp. 952–955.
- Langmuir, C.H., Klein, E.M. and Plank, T. (1992) 'Petrological systematics of mid-ocean ridge basalts: constraints on melt generation beneath ocean ridges', in Phipps Morgan, J. (ed.) 'Mantle Flow and Melt Generation at Mid-ocean Ridges', *American Geophysical Union Monograph*, vol. 71 pp. 183–280, D.K. Blackman & J.M. Sinton.
- Lee, D-C., Halliday, A.N., Leya, I., Weiler, R. and Wichert, U. (2002) 'Cosmogenic tungsten and the origin and earliest differentiation of the Moon', *Earth and Planetary Science Letters*, vol. 198, pp. 267–274.
- Livermore, R. et al. (2005) 'Palaeogene opening of the Drake Passage', *Earth and Planetary Science letter*, vol. 236, pp. 459–470.
- Lodders, K. (2003) 'Solar System abundances and condensation temperatures of the elements', *The Astrophysical Journal*, vol. 591, pp. 1220–1247.
- McDonough W. F. and Sun, S-S. (1995) 'The composition of the Earth', *Chemical Geology*, vol. 120, pp. 223–253.
- McKenzie, D. and Bickle, M.J. (1988) 'The volume and composition of melt generated by extension of the lithosphere', *Journal of Petrology*, vol. 29, pp. 625–79.
- McSween, H.Y. (1987) *Meteorites and their Parent Planets*, Cambridge, Cambridge University Press.
- Minster, J. F. and Allègre, C. J. (1980) '⁸⁷Rb–⁸⁷Sr dating of LL chondrites', *Earth and Planetary Science Letters*, vol. 56, pp. 89–106.
- Morgan, J.W. and Anders, E. (1980) 'The chemical composition of Earth, Venus and Mercury', *Proceedings of the National Academy of Science*, vol. 77, pp. 6973.

- Murty, V. R. and Patterson, C. C. (1962) 'Primary isochron of zero age for meteorites and the Earth', *Journal of Geophysical Research*, vol. 67, pp. 1161–1167.
- Pan, Y. and Batiza, R. (2003) 'Magmatic processes under mid-ocean ridges: a detailed mineralogic study of lavas from East Pacific Rise 9° 30' N, 10° 30' N, and 11° 20' N', *Geochemistry Geophysics Geosystems*, vol. 4, no. 11, doi: 10.1029/2002GC000309.
- Roeder, P.L. and Emslie, R.F. (1970) 'Olivine–liquid equilibrium', *Contributions to mineralogy and petrology*, vol. 29, pp. 281.
- Rushmer, T., Minarik, W.G. and Taylor, G.J. (2000) 'Physical processes of core formation' in Canup, R.M. and Righter, K. (eds) *Origin of the Earth and Moon*, Tucson, University of Arizona.
- Sherstein, A., Elliott, T., Hawkesworth, C., Russell S. and Masarik, J. (2006) 'Hf–W evidence for rapid differentiation of iron meteorite parent bodies', *Earth and Planetary Science Letters*, vol. 241, pp. 530–542.
- Sinton, J.M. and Detrick, R.S. (1992) 'Mid-ocean ridge magma chambers', *Journal of Geophysical Research*, vol. 97, pp. 197–216.
- Takahashi, E. (1986) 'Melting of a dry peridotite KLB-1 up to 14 GPa: implications on the origin of peridotitic upper mantle', *Journal of Geophysical Research*, vol. 91, pp. 9367–82.
- Takahashi, E., Shimazaki, T., Tsuzaki, Y. and Yoshida, H. (1993) 'Melting study of a peridotite KLB-1 to 6.5 GPa, and the origin of basaltic magmas', *Philosophical Transactions of the Royal Society of London, Series A*, vol. 342, pp. 105–20.
- Tormey, D.R., Grove, T.L. and Bryan, W.B. (1987) 'Experimental petrology of normal MORB near the Kane Fracture Zone: 22°–25° N, mid-Atlantic ridge', *Contributions to Mineralogy and Petrology*, vol. 96, pp. 121–39.
- Watts, A.B. (2001) *Isostasy and Flexure of the Lithosphere*, Cambridge, Cambridge University Press.
- White, R.S., McKenzie, D. and O'Nions, R.K. (1992) 'Oceanic crustal thickness from seismic measurements and rare earth element inversions', *Journal of Geophysical Research*, vol. 97, pp. 19 683–715.
- Yang, H.J., Kinzler, R.J. and Grove, T.L. (1996) 'Experiments and models of anhydrous, basaltic olivine–plagioclase–augite saturated melts from 0.001 to 10 kbar', *Contributions to Mineralogy and Petrology*, vol. 124, pp. 1–18.
- Yin, Q., Jacobsen, S.B., Jamashita, K., Blichert-Toft, J., Telouk, P. and Albarede, F. (2002) 'A short timescale for terrestrial formation from Hf–W chronometry of meteorites', *Nature*, vol. 418, pp. 949–952.