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

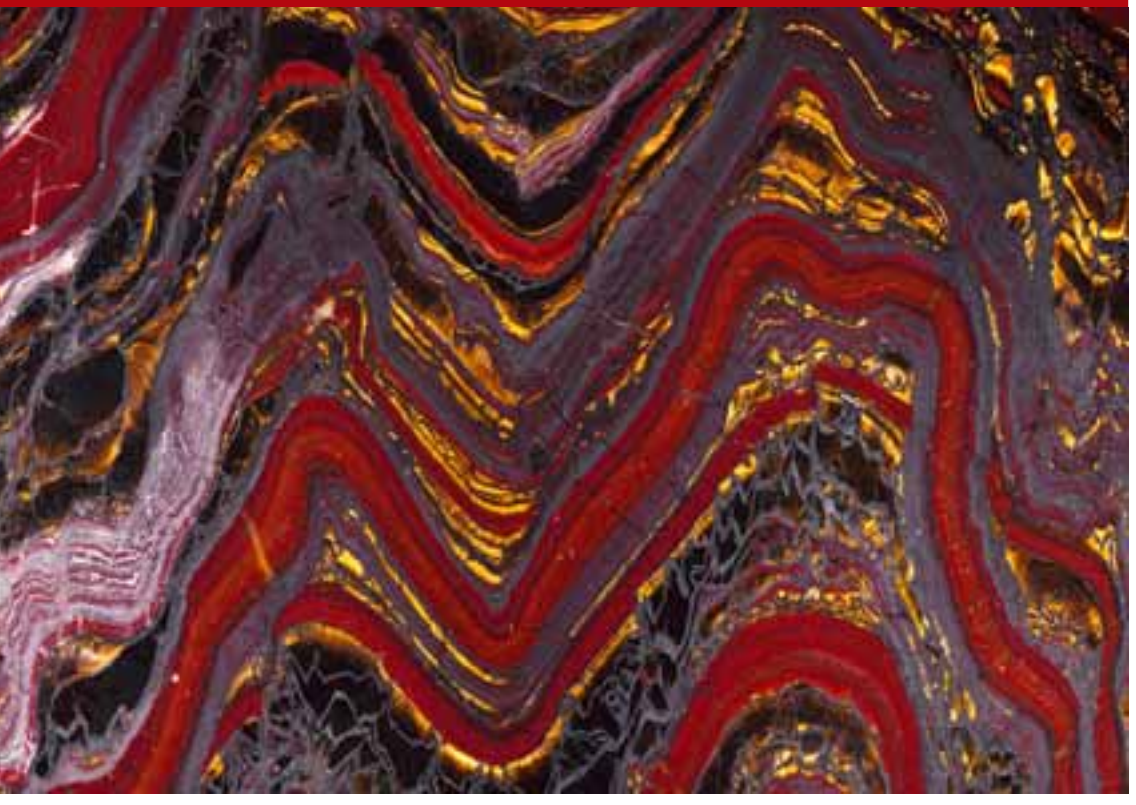
The Open University

Book 1 Part 2

Our Dynamic Planet

Edited by Nick Rogers

Authors: Nigel Harris, Ian Parkinson and Nick Rogers



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.

Edited, designed and typeset by The Open University.

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

ISBN 978 0 7492 1903 1

1.1

Contents

Processes at destructive plate boundaries

Chapter 5

Destructive plate boundaries are the sites where oceanic plates are recycled back into the mantle and, in terms of understanding the Earth as a system, they are the key locations for material flux back into the mantle. The modification of the subducted plate in the shallow mantle beneath a volcanic arc and the ultimate fate of a plate as it plunges into the deeper mantle play an important role in both the short-term and long-term recycling of chemical elements within the Earth. In addition to volcanism, the majority of major earthquakes, particularly deep earthquakes, are located at destructive plate boundaries (Figure 3.9), a fact brought into stark reality in 2004 with the 26 December (Boxing Day) tsunami, which was triggered by a shallow earthquake located on the Sumatran plate boundary where the Indo-Australian Plate is being subducted beneath the Eurasian Plate.

In detail, subduction zones are complicated by a wide variety of factors that can influence the style of the plate boundary (Figure 5.1). In this chapter, the key parameters are described and their influence on the structure, tectonics, seismicity and magmatism of subduction zones is discussed.

You should recall from Chapter 3 that there are three major types of destructive plate boundary:

- ocean–ocean boundaries, where oceanic lithosphere is subducted beneath oceanic lithosphere;
- ocean–continent boundaries, where oceanic lithosphere is subducted beneath continental lithosphere;
- continent–continent boundaries, where two continental plates collide.

The first two plate boundaries are the focus of this chapter and continent–continent collision is discussed in Chapter 6.

It is clear that the dominant material that is subducted is the oceanic lithosphere.

- What are the main components of oceanic lithosphere?
- You should recall from Chapter 4 that oceanic lithosphere is made up of oceanic crust, consisting of pillow basalts, sheeted dykes and gabbros, and the underlying residual mantle comprising harzburgite.

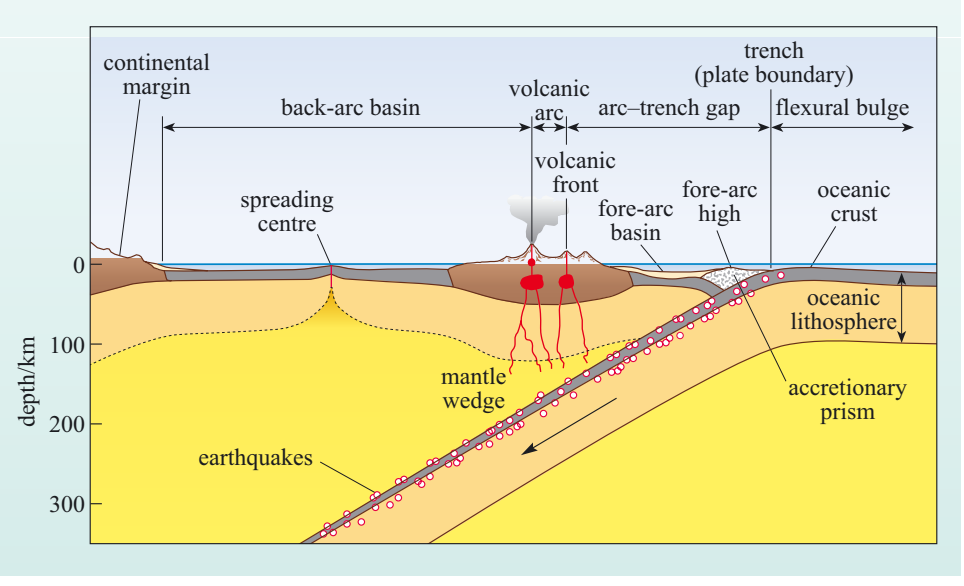


Figure 5.1 Generalised cross-section through oceanic and continental subduction zones. The terms used in this figure are explained in Section 5.1.

Importantly, hydrothermal fluids at the ridge crest chemically modify the oceanic crust. In addition, hydrothermal activity can extend deep within the upper part of the residual mantle to produce **serpentinites**. The oceanic plate, as it moves away from the ridge crest, cools and experiences low-temperature interaction with seawater; sediments gradually accumulate and, in some cases, further volcanic activity related to hot spots, known as intra-plate volcanism, may occur, producing seamounts and ocean islands. Sediment composition can vary widely depending on the latitude and proximity to continents and ocean islands. All of these effects contribute to the variable chemical composition of the oceanic plate, but one key factor is common to all plates: the addition of relatively large amounts of H_2O, both as pore water and locked up in hydrous minerals. The recycling of H_2O and other volatiles through subduction zones is one of the most important aspects of destructive plate boundaries and will be discussed with respect to volcanism in this chapter and to the inner workings of the planet in Chapter 7.

Although mid-ocean ridges, on average, produce more magma per year than volcanism associated with destructive plate boundaries, the majority of the world's explosive volcanism is found at destructive plate boundaries. This volcanism, the most hazardous and destructive, is consequently the one that captures the public's imagination. Mt St Helens, Pinatubo, Krakatau, Montserrat and Fuji are volcanoes that many people could name because of their histories of violent eruptions and all of these volcanoes are located at destructive plate boundaries.

From a geological point of view, destructive plate boundaries play a major role in the evolution of the planet, as subduction zones are thought to be the location of modern-day crustal growth. Furthermore, it is generally agreed that ancient subduction zones were the sites where a significant proportion of the continental crust was produced in the geological past. An understanding of the physical and chemical behaviour of modern-day subduction zones provides an essential framework for unravelling ancient crustal growth mechanisms.

5.1 Anatomy of a subduction zone

There is a large amount of nomenclature associated with subduction zones, much of which is illustrated in Figure 5.1. The **subducted plate** is often known as the **down-going plate**, and at depth is usually called the **subducted slab**. This mechanically strong slab is equivalent to the oceanic lithosphere discussed in Chapters 3 and 4. It is the location of most major earthquakes associated with destructive plate boundaries and the zone of earthquakes is called the Wadati–Benioff zone. The plate that lies above the subduction zone is termed the **overriding plate**, below which is a triangle of asthenosphere known as the **mantle wedge** overlain by the lithosphere of the volcanic arc itself. The **arc lithosphere**, which includes the crust and the lithospheric mantle, can be of variable thickness and composition and plays an important role in controlling the composition of the erupted lavas. In the oceans, arc crust is very thin (7 km) and the lithosphere thickness depends on the age of the plate. With time, the arc crust may thicken because of continued volcanism. The overriding plate at continental subduction zones often comprises pre-existing continental crust (30–50 km thick) and a much thicker lithosphere. The volcanoes form a **volcanic arc**, although the term island arc is commonly used in oceanic arcs because the volcanic arc is

defined by discrete islands (e.g. Guam, Krakatau and Pinatubo). On a map, the line of the volcanic islands defines the **volcanic-front**. Between the plate boundary, which is usually defined by a **trench**, is the **fore-arc** region. This zone can be very complicated because it can be a region undergoing either extension or compression, depending of the geometry of the plates involved in the subduction zone. The fore-arc usually comprises a low-lying **fore-arc basin** that is filled with sediment and an outer **fore-arc high**. Depending on the tectonics, the fore-arc may also contain an accretionary prism, which is a wedge of sediment scraped off the down-going plate, or faulted regions containing serpentinised peridotites. In continental arcs, the fore-arc may also contain remnants of older arcs, continental crust and material accreted from the down-going plate. Some subduction zones have a second line of volcanoes, usually submarine, where a new spreading centre is underlain by the subducting slab. This region is known as a back-arc basin, or sometimes as a **marginal basin**, from which magmas are erupted that are affected by the subducting slab. This may seem like a lot of names, but it reflects the complexity of destructive plate boundaries.

A rapid review of the global map of plate boundaries (Figure 5.2) shows that the majority of destructive boundaries are found around the periphery of the Pacific Ocean. Consequently, the majority of the world's volcanic arcs are located around the Pacific, and this is commonly referred to as the **Pacific Ring of Fire**.

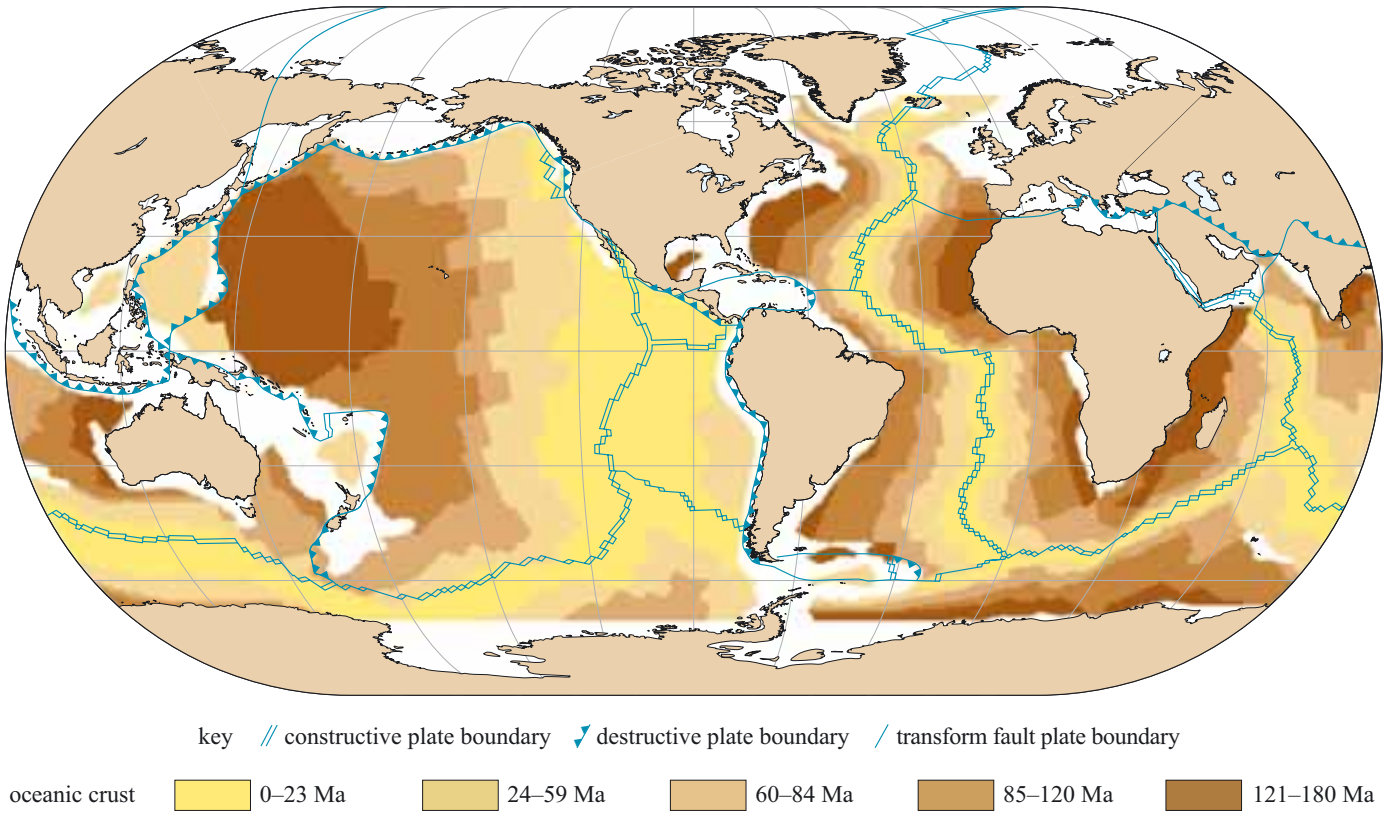


Figure 5.2 Global map showing the age distribution of the oceanic crust. (Adapted from Scotese et al., 1998)

- Using Figure 5.2 what can you say about the age of the subducting plate around the Pacific Ocean?

- Put simply, it varies hugely. The oceanic plate gets older the further away from a spreading centre. In the Pacific Ocean, the crust being subducted beneath the Mariana arc–trench system is the oldest because it is very distant from the East Pacific Rise where it formed. In places the crust is Jurassic in age. By contrast the Tonga–Kermadec system, being farther east, has Cretaceous crust being subducted. On the eastern margin of the Pacific, the average age of the subducting plate is younger. The Nazca Plate is <25 Ma in age, and even younger where it is being subducted under southern Chile. The Cocos Plate and particularly the Juan De Fuca Plate being subducted beneath the Cascades are both very young (<10 Ma) because of their proximity to a spreading centre (Figure 3.25).

- Can you recall another property of the subducting slab that varies between different destructive plate boundaries?

- Its convergence rate.

You discovered in Chapter 3 that subduction is driven by the gravitational descent of old, cold lithosphere into the mantle, and once a plate starts subducting it pulls younger lithosphere behind it. Thus, with time, lithosphere of almost any age can be subducted. Moreover, since the plate is being driven by slab-pull, which is the most effective of the plate driving forces, all plates converge on subduction zones at relatively high speeds that vary between 65 and 106 mm y^{-1}.

- What will be the effect of subducting old versus young lithosphere on the composition and physical properties of the subducting slab?

- Old lithosphere is colder than young lithosphere and is more likely to have accumulated a thick layer of marine sediments.

Thus, although the subducting slab is invariably oceanic, its composition varies, as does its thermal characteristics, both of which contribute to the development of the complexity of arc structure.

5.2 Geophysical observations

Whereas Figure 5.1 gives an overall impression of the general structure of a subduction zone, more detailed images of the seismic activity across specific arcs provide greater insights into that structure. In Chapter 3 you were introduced to the classification of shallow, intermediate and deep earthquakes in relation to the Tonga arc. Figure 5.3 is a cross-section of the upper part of the Japanese arc and so all of the earthquakes in this diagram could be described as shallow (0–70 km) or intermediate (70–300 km).

It is easy to see from Figure 5.3 that there are abundant shallow earthquakes located in the volcanic arc itself that may relate to magma movement and storage in the arc crust, or possibly extension in the back-arc region. There is also a zone

of earthquake foci that mark out the upper boundary of the descending slab, i.e. the Wadati–Benioff zone. In addition, there are other earthquakes deeper within the slab that may reflect the way in which the slab is deforming as it descends into the mantle. These intermediate earthquakes are more often compressional and relate to the resistance that the slab experiences as it pushes through the mantle. Deep earthquakes are also located within the descending slab, and are now considered to be generated by mineral transformations as the slab slowly heats.

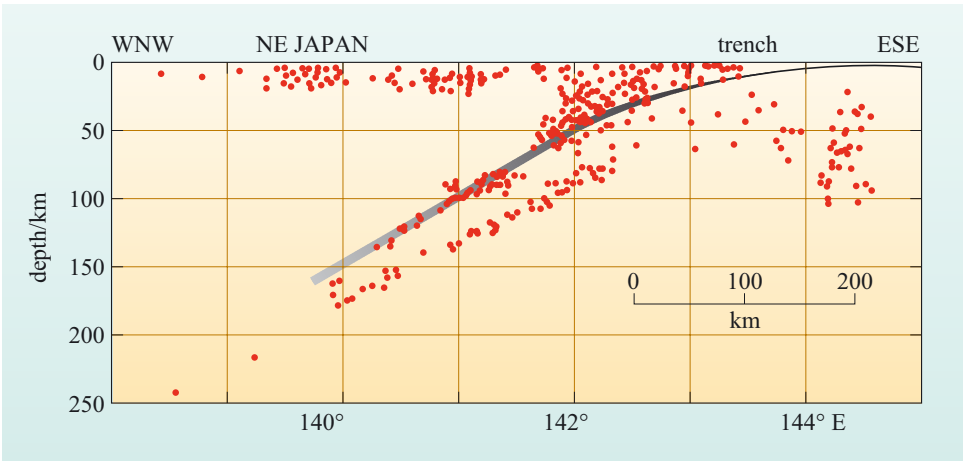


Figure 5.3 Simplified cross-section of the NE Japan subduction zone with the distribution of earthquake foci plotted as dots. The top of the subducting slab is marked with a black line.

Yet more information on the structure of subduction zones is derived from a technique known as seismic tomography (see Section 7.3). This technique compares the travel time of millions of individual seismic ray paths and compares them with the expected speed derived from an Earth reference model of the type described in Section 1.2.3. In essence, cold, dense material, such as subducted slab, allows seismic waves to travel faster, whereas warm, less-dense material makes them travel slower than the Earth reference model. For S-waves, the presence of partial melt has an additional slowing effect on the seismic waves. The results of seismic tomography are similar to an X-ray tomographic image of the human body but reveal regions of the Earth's interior where seismic velocities are either slower or faster than expected. To make the images easy to understand, seismically fast regions are coloured blue and seismically slow regions are coloured red. A collection of seismic tomographic images of some of the world's subduction zones is illustrated in Figure 5.4. From these diagrams the cold, dense subducting slab is well resolved, showing that the angle of subduction varies from arc to arc. A compilation of subduction angles is given in Table 5.1.

Table 5.1 Geometrical characteristics of some arc–trench systems.

Arc–trench system	Subduction angle	Curvature
Mariana	80–90°	highly arcuate
Lesser Antilles	45°	moderately arcuate
Tonga	45°	low curvature
Kurile	35–50°	moderately arcuate
Northern Chile	10–30°	low curvature

- Can you think of a reason why there is such a difference between the subduction angles beneath the Mariana and Chile arc–trench systems (Table 5.1)?
- You have already discovered that the age of the subducting crust is different for these two arcs. The steepness of the slab beneath the Marianas Islands is related to its dense, cold nature, which causes it to sink rapidly. This is the slab-pull force in action. Younger crust beneath Chile is more buoyant, and so will not sink into the mantle as easily.

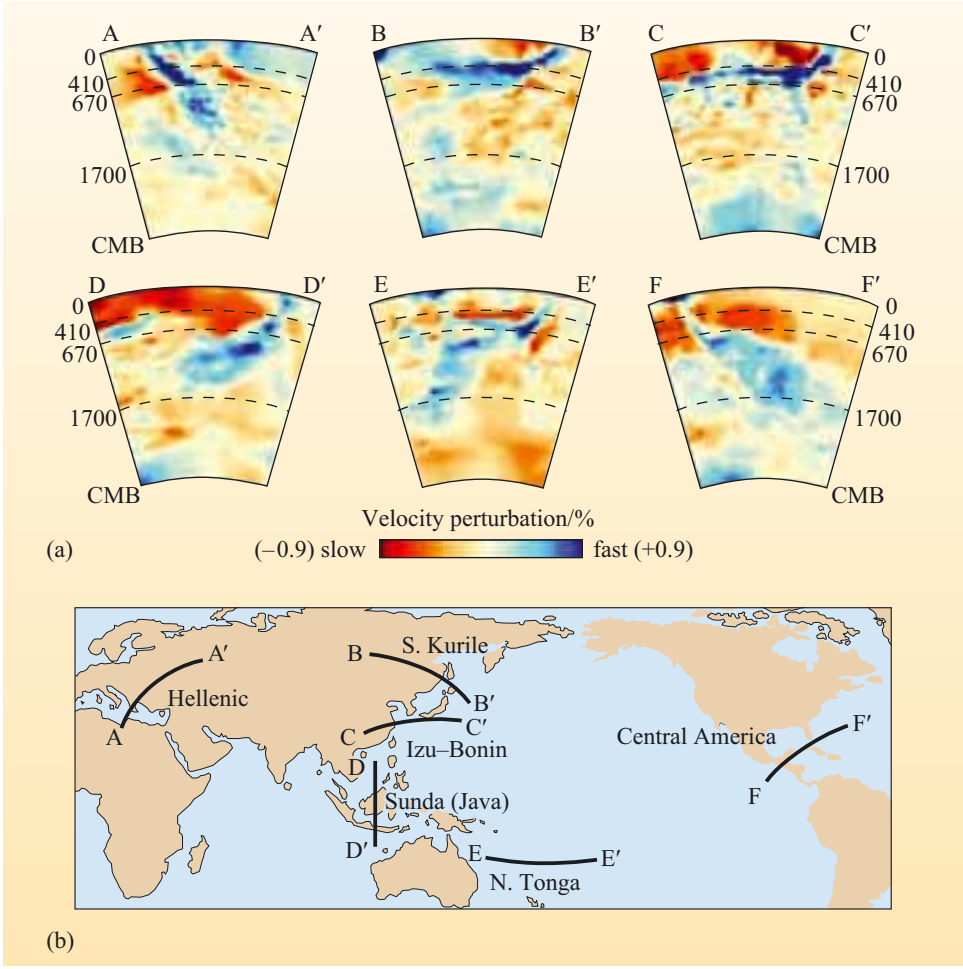


Figure 5.4 (a) Vertical seismic tomographic slices through the mantle at the localities shown in (b). Fast velocities indicate cold, dense anomalies, generally interpreted as subducting slabs or their remnants. Slow velocities imply warm, less-dense material. *Note*: CMB is the core–mantle boundary.

The seismic tomographic images also reveal that the slab does not break up and is not mixed into the upper mantle during its voyage down to 670 km into the mantle. It also remains cooler than the surrounding mantle. This is not a surprise, as a simple thermal calculation indicates that it would take several hundred million years for the slab to equilibrate thermally with the mantle at depth, whereas it takes a few tens of millions of years to subduct to 670 km depth. The fate of subduction zones at depths greater than 670 km will be discussed in Chapter 7.

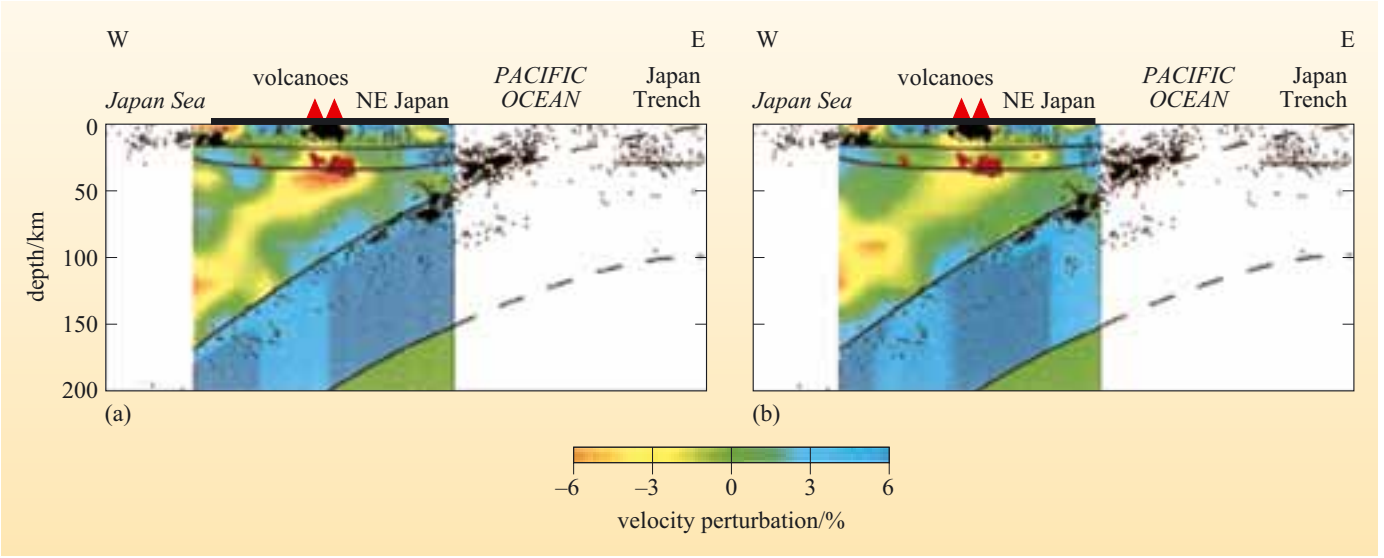


Figure 5.5 Cross-sections perpendicular to the volcanic front in NE Japan, at about 39° 50′ N running westwards from the Japan Trench, showing (a) P-wave and (b) S-wave velocity anomalies. Dots represent earthquake foci. Black lines beneath Japan represent seismic discontinuities. In order of increasing depth these are the Conrad discontinuity (upper/lower crust boundary), the Moho and the top and base of the subducting slab.

Figure 5.5 presents two detailed seismic tomographic images of the NE Japanese arc. Figure 5.5a uses only P-waves while Figure 5.5b uses only S-waves.

- What features can you observe in Figure 5.5 that are similar to and different from the less-detailed images in Figure 5.4?
- You should notice that the slab in Figure 5.5 is very well defined in blue tones (+3% to +6%), similar to the images of all the subducting slabs in Figure 5.4. The greater resolution of Figure 5.5 reveals a zone of seismically slow material that crosses the mantle wedge from 150 km depth (just above the slab–mantle wedge boundary) beneath the volcanic-arc crust. A zone of particularly slow velocity is located beneath the crust in Figure 5.5b.

- What do you think the zones of slower S-wave velocities might represent?
- One possibility is that they represent zones of hotter material. However, you should recall from Section 1.2.1 that S-waves do not propagate through liquids, so the S-wave signature both in the mantle wedge and beneath the arc crust may also indicate the presence of a small amount of melt.

Just how much melt this slow zone represents can be estimated by the decrease in S-wave velocity, and calculations indicate that a few per cent partial melt or fluid must be present. Therefore, the seismic tomography suggests that partial melting occurs within the mantle wedge and that this melt then traverses the wedge and perhaps ponds beneath the arc crust. As you will see in Section 5.4, these observations place major constraints on how melts can be produced within subduction zones.

Other geophysical observations shed further light on the tectonic processes that occur in subduction zones. Figure 5.6 illustrates the free-air gravity and heat flow

profile across a simple oceanic subduction zone. There is a small gravity high over the flexural bulge (where the down-going plate bulges just before the trench), a gravity low over the trench and a large gravity high over the fore-arc and volcanic arc. These data indicate that subduction zones are very dynamic and are not in isostatic equilibrium.

■ Can you recall from Chapter 3 what these anomalies reflect in the underlying structure?

■ The gravity low above the trench indicates that there must be either a downward force or a mass deficit in this region, and this may well be related to the subduction of H_2O-rich sediments. By contrast, the gravity high over the fore-arc and volcanic arc implies a mass excess at depth or, more likely, an upward acting force that could be related to the introduction of magma beneath the arc.

Turning to the heat flow data (Figure 5.6b), the fore-arc and trench are characterised by low heat flow, whereas the volcanic-arc has high heat flow.

■ What do you think are the causes of these variations in heat flow?

■ Low heat flow over the fore-arc and trench corresponds to the subduction of the cold slab, whereas high heat flow through the arc is consistent with the presence of magma at depth.

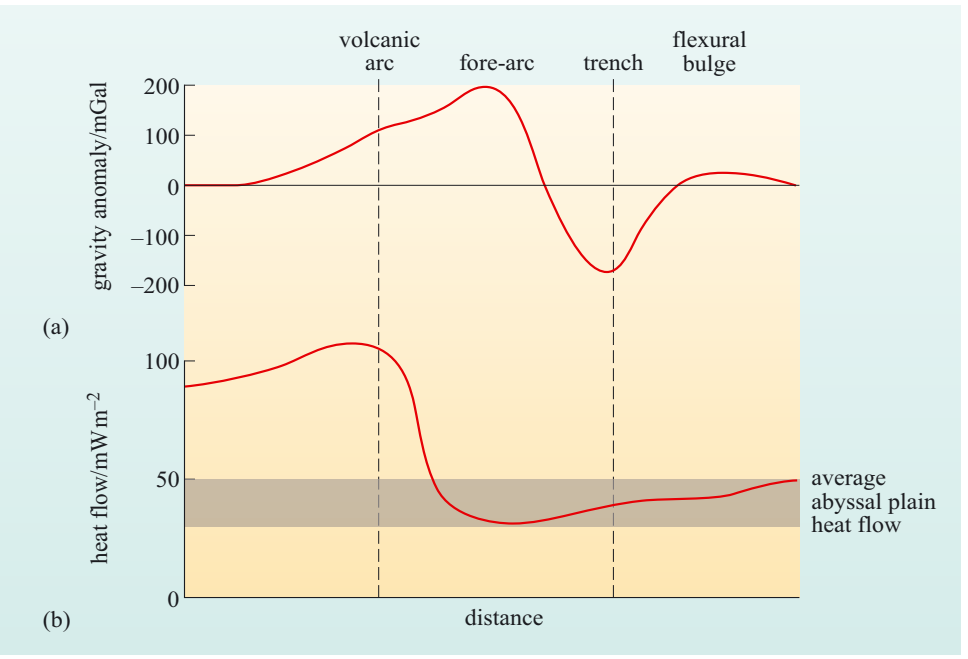


Figure 5.6 Sketch of (a) free-air gravity anomaly and (b) surface heat flow across a destructive plate boundary.

5.2.1 Summary of Section 5.2

You are now at a point where you have enough information to try to think quantitatively about melt generation in subduction zones.

- What are the most likely components in a subduction zone that may contribute to melt production?
- There are four components that could melt:
  (i) the mantle wedge, which is made of peridotite
  (ii) altered oceanic crust
  (iii) a wide variety of hydrated sediments
  (iv) serpentinised peridotites in the subducting slab.

Geophysical information further reveals that melt or fluid is present as deep as 150 km in the mantle wedge. These relatively simple observations allow some important conclusions to be drawn and a debate to be resolved that prevailed during the initial years of the plate tectonic revolution in the late 1960s and early 1970s. At that time scientists turned their attention to magma generation in subduction zones, and particularly to the problem of how subduction of cold material could produce hot magmas. Two competing models existed, one in which the oceanic crust of the subducted slab was heated up and melted and one that involved the melting of hydrous ('wet') peridotite in the mantle wedge. Initially, the slab-melting model was favoured, in part because many workers thought that andesites (which are more silica rich than basalts) were the primary melts found in most arcs, and because slab melts are also rich in silica. This perhaps reflects the fact that many experimental petrologists (people who perform high-pressure and/or high-temperature experiments on igneous rocks) had not visited many island arcs. There is still a perception that subduction zones dominantly generate andesites, whereas in reality modern oceanic arcs produce, on average, basaltic andesites and many of these arcs erupt primitive basaltic lavas (e.g. Vanuatu in the southwest Pacific Ocean, and Grenada in the Lesser Antilles). Therefore, it is now widely accepted that magmatism in the great majority of subduction zones is due to partial melting of the mantle wedge. However, there may be special circumstances where the slab does melt, and some arcs where melts are a mixture of slab melts and melting of the mantle wedge.

5.3 Thermal structure of subduction zones

The thermal structure of the Earth away from destructive plate boundaries is relatively simple, with the temperature increasing with depth along a geotherm.

- Would you expect the geotherm beneath a destructive plate boundary to be simple?
- No. The subduction of cold oceanic lithosphere into the hot mantle should modify the thermal structure of the subduction zone, producing a more complex geotherm.

  The geotherm across the mantle wedge first increases with depth to a maximum and then cools again as it approaches the subducted slab. This is known as an **inverse geotherm**.

You have already seen that seismic tomography can image differences in P- and S-wave velocities of up to 12% beneath active arcs that are primarily related to temperature differences in the down-going plate and the overlying mantle wedge. Specifically, the down-going plate remains relatively cold compared with the mantle wedge. Seismic tomography, therefore, provides a qualitative guide to the relative temperature differences in the subduction zone. But can it produce a more quantitative picture of the thermal structure of a subduction zone? It would be particularly useful to have a thermal map of a subduction zone in which key input parameters, such as the rate of subduction, the age of the subducted plate and the angle of subduction, could be varied. Such a model could be tested against modern-day subduction zones where the values for the input parameters are known, but could also be used to make predictions about ancient subduction. Moreover, predicted pressure and temperature conditions could be compared with phase diagrams that describe mineralogical and melting reactions to make predictions about the location and type of melt generation in a given subduction system.

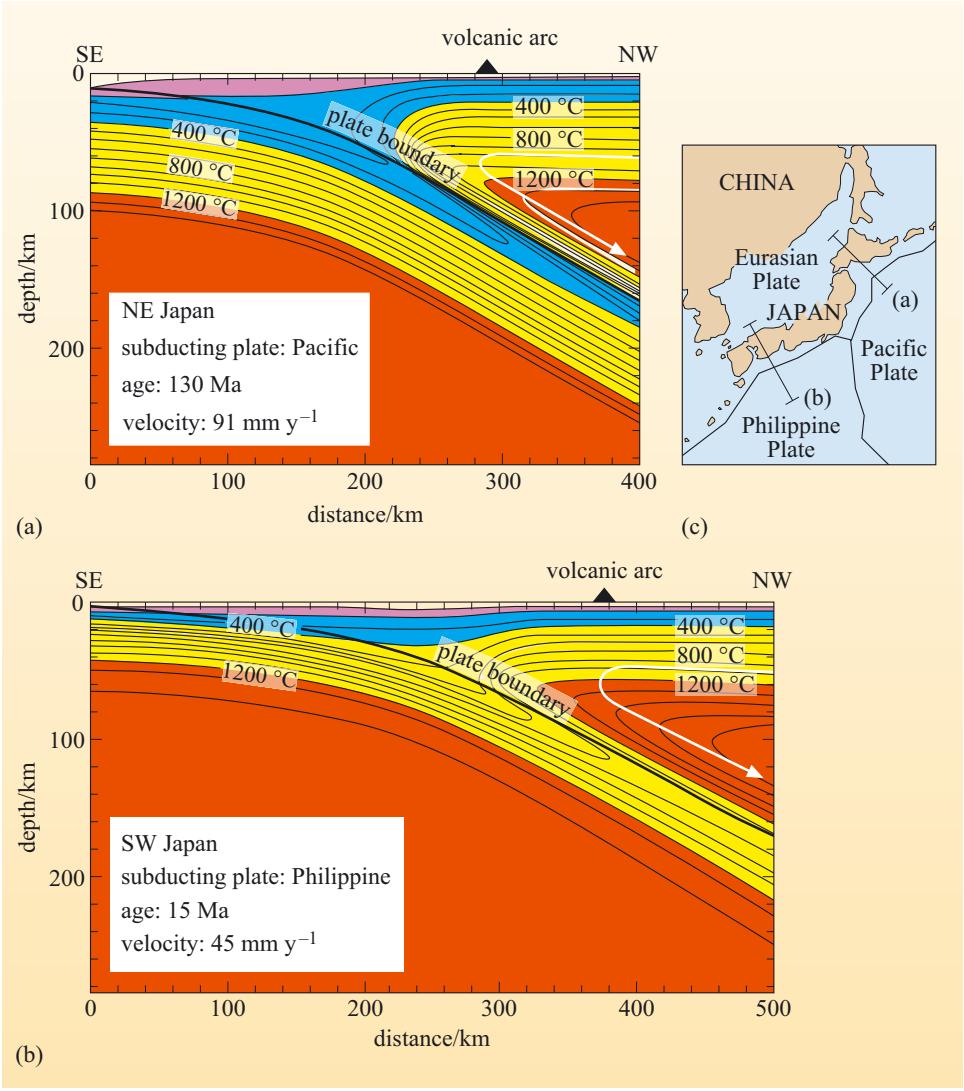


Figure 5.7 Models of the thermal structure of two different subduction zones: (a) an old slab being subducted rapidly; (b) a young slab being subducted relatively slowly. The isotherms are colour coded and the direction of corner flow is illustrated by the white arrow. (Peacock, 2000)

Mathematical models and computer simulations of the thermal structure of subduction zones have existed for many years, but modern computation power allows models of considerable sophistication, particularly with regard to how material flows within the mantle wedge. Some simple thermal models for a subduction zone are illustrated in Figure 5.7. This is a steady-state model because all of the input parameters have been held constant for long enough for the thermal structure to become stable. In all of the models, lines of equal temperature (**isotherms**) are depressed to greater depths in the vicinity of the down-going plate. This thermal perturbation produces a very different heat distribution from that found in a normal plate (as seen at the left-hand edges of all the models). Therefore two key thermal consequences of continuous subduction of cold oceanic lithosphere can be deduced:

- the Wadati–Benioff zone is cooler then the surrounding mantle
- melts generated in the mantle wedge migrate into, and thus heat, the arc lithosphere, particularly the arc crust.

The elevation of the isotherms in the arc lithosphere is hard to see in Figure 5.7 because of the scale, but is obvious from the surface heat flow (Figure 5.6b).

You might expect that continuous subduction would progressively cool the overlying mantle wedge, and this would be true were it not for an additional control on the thermal structure of the subduction zone. As the slab is subducted, it cools the immediately overlying mantle wedge. The resultant viscous drag causes the subducting plate to couple with the overlying mantle wedge, pulling hot mantle down parallel to the convergence direction. This results in **corner flow**, in which material dragged down by the slab is replaced at higher levels by asthenosphere from the back-arc region. The direction of corner flow is indicated on the thermal models in Figure 5.7 by the white arrow. Alternatively, **arc-parallel flow** in which asthenosphere flows parallel to the length of the arc, can also form as a consequence of variations in geometry along the length of the arc. This allows material to be sucked from the edges of the arc to its centre (e.g. Aleutian and Kurile arcs). This effect extends out of the plane of the thermal models illustrated in Figure 5.7.

For all models, it is important to assess how the different input parameters affect the thermal structure of the subduction zones.

■ What is the effect of the age of the subducting slab on the thermal structure of a subduction zone?

■ As the plate ages it becomes colder and denser and the lithosphere becomes thicker. Therefore, an old, cold slab will take longer to heat up and will depress the isotherms to a greater degree. By contrast, a very young slab will be relatively warm and will heat up more rapidly and depress the isotherms to a lesser degree.

■ What is the effect of the rate of subduction on the thermal structure of a subduction zone?

■ The rate of subduction is directly proportional to the depression of the isotherms. Rapid subduction will allow the cold slab to traverse to greater depths, thus depressing the isotherms to a greater extent. Slow subduction limits the depth to which the cold slab will go before it is heated up, which in turn reduces the extent to which it depresses the isotherms.

These simplified thermal models are a useful starting point, but the most up-to-date models are best illustrated with a specific example that is based on a real subduction zone. You have already encountered seismic tomographic images (Figure 5.5) of Honshu in NE Japan. This is a subduction zone where old oceanic crust is being subducted rapidly. Until recently, most models assumed the viscosity of the mantle was the same throughout the mantle wedge (called **isoviscous**). This was convenient as it simplified much of the mathematics of the modelling that readily produced corner flow. However, models that take into account the effect of H_2O (from the subducting slab) on the viscosity of the mantle wedge produce some rather different results. H_2O has the effect of reducing the viscosity of the mantle, and this allows the mantle to flow more easily. You will see later (Chapter 7) that viscosity is a fundamental parameter in determining how the Earth works, and H_2O has a very strong effect on modifying this property.

In a subduction zone, therefore, there is a strong **positive feedback** between subducting H_2O from the Earth's surface and the thermal and volcanic evolution of the subduction zone. In the Honshu model, the reduction in viscosity above the slab–wedge interface has the effect of increasing the flow of hot asthenosphere into the wedge that pulls down hotter isotherms closer to the slab. This produces a very strong thermal gradient across the top of the slab in this model. Previously, the only way to produce this effect in thermal models was to include an arbitrary amount of frictional heating between the slab and wedge that was not well constrained. Reassuringly, the more sophisticated models confirm the effects previously seen in the simple thermal models. Fast subduction of an old slab (Honshu) tends to keep the slab cold, and slow subduction of a young, hot slab keeps the slab hot.

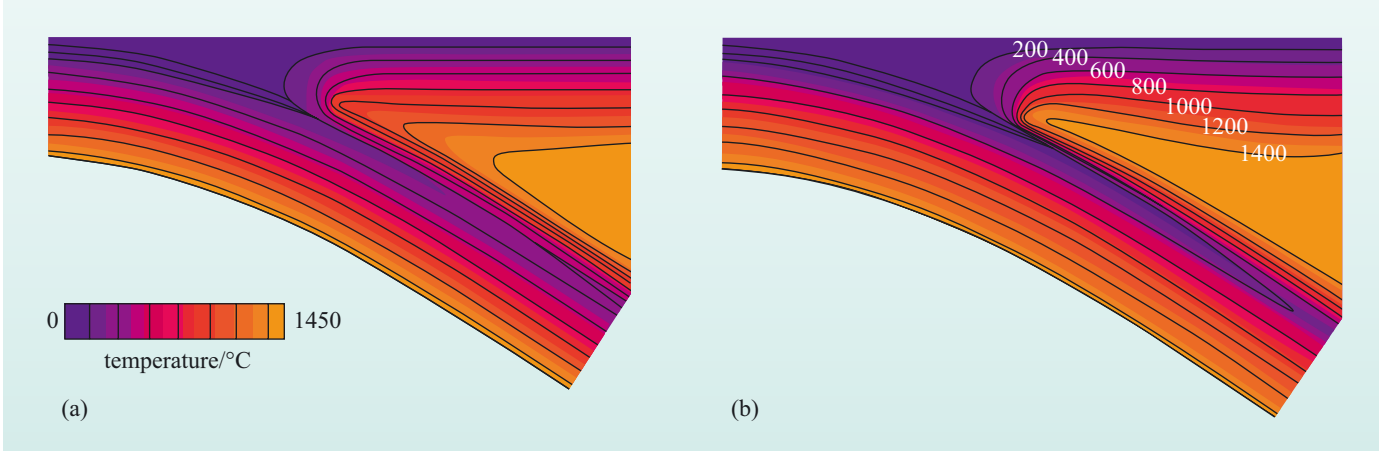


Figure 5.8 Models of the thermal structure of the Honshu subduction zone: (a) isoviscous model; (b) H_2O-modified viscosity. (van Keken et al., 2002)

5.4 Applying the thermal models to melt production

To understand whether any of the components that make up the source of arc lavas could actually melt or influence melt production, the information from the thermal models needs to be translated on to appropriate phase diagrams for each of the possible source compositions. You should now realise that evidence for the mantle wedge melting can be deduced from the observation that primitive basaltic melts are found in many arcs. H_2O influences this melting as indicated by the elevated H_2O contents of many arc lavas. Before considering what happens in the mantle wedge, the next section reviews phase diagrams for each of the three slab components, namely sediment, hydrated oceanic crust and the underlying hydrated peridotite. To help interpret what happens in the slab, a variety of pressure–temperature paths have been added to the phase diagrams. These correspond to thermal models where the age of the slab and the rate of subduction are varied to represent the range of probable conditions found in different subduction zones.

5.4.1 Basalt phase diagram

When geologists started to accept sea-floor spreading and plate tectonics it became obvious that subduction of oceanic basalts was a critical aspect of plate tectonics.

- What types of experiment do you think experimental petrologists attempted?
- The obvious experiment to attempt was to derive a phase diagram for the system MORB + H_2O because hydrothermally altered oceanic basalt is an important component of the subducted slab.

Figure 5.9 is a simplified phase diagram for a H_2O-saturated basalt.

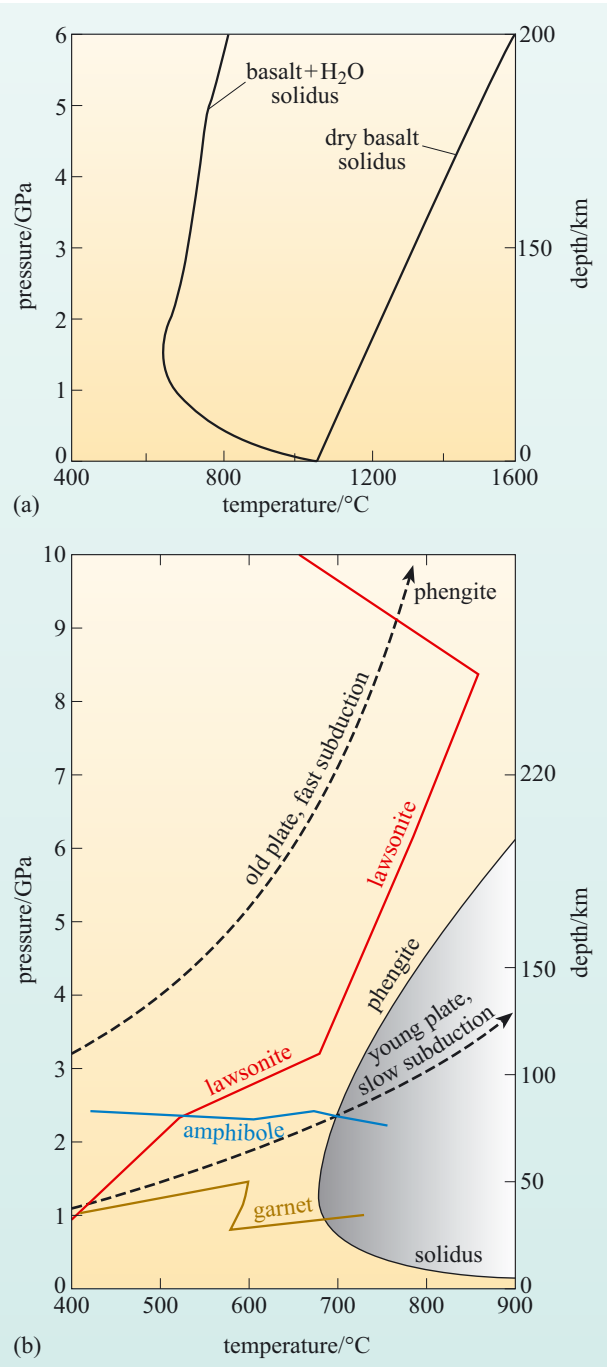


Figure 5.9 (a) Simplified phase diagram for a basalt illustrating the effects of H_2O on the melting temperature of the subducted basaltic slab. (b) Simplified phase diagram for H_2O-saturated basalt with *P–T* paths for the slab for two different subduction zone thermal regimes and the stable sub-solidus phases. Garnet is stable at all pressures above the lower brown line, whereas amphibole is only stable up to pressures below the blue line labelled amphibole. Lawsonite is stable to the left of the labelled red line whereas phengite is stable to the right of the red line labelled lawsonite up to high temperatures and pressures. Several key phases such as clinopyroxene have been excluded from the diagram for clarity reasons and/or because they have a wide stability range. *Note*: the depth increases upwards on this figure. Melt can be present in the shaded area to the right of the solidus. (Adapted from Schmidt, 2002)

- What is the most obvious effect that H_2O has on the basalt phase diagram?
- It reduces the solidus temperature over a range of pressures by as much as 800 °C relative to the dry solidus.

Below the solidus, however, H_2O also interacts with a basaltic rock to produce a great variety of hydrated or hydrous minerals. For example, as the slab gets deeper H_2O reacts with the original igneous minerals so that at depths of a few tens of kilometres the dominant minerals are amphibole, chlorite, clinopyroxene, **lawsonite** (a Ca–Al silicate mineral that contains up to 11 wt% H_2O) and a variety of mica minerals. These phases are not all on Figure 5.9 for reasons of clarity. The amphiboles are sodium rich and have a characteristic blue colour, which explains the name given to **blueschist facies metamorphic rocks**. As pressure and temperature increase, further reactions take place and new minerals appear. These reactions all involve the release of H_2O in what is called a **dehydration reaction**. In particular **garnet** becomes a progressively more important mineral and, by 70 km depth, the blue amphibole dehydrates and the rock is made up of dominantly a sodium-rich clinopyroxene (not shown on Figure 5.9) and garnet. Lawsonite and a type of mica called **phengite** are the remaining H_2O-bearing phases. This pyroxene–garnet bearing rock is known as an **eclogite** and is a characteristic low-temperature, high-pressure metamorphic rock. Overall, the reactions involve the hydrated basalt progressively losing H_2O until a depth of about 200 km, when only about 1 wt% H_2O is present in the rock.

- Study Figure 5.9. Do either of the thermal models predict that basaltic crust will melt during subduction?
- Basalt only melts for young plates subducted slowly.

For most thermal models, for example those illustrated in Figure 5.9, basalt does not melt, although for slow subduction of a young plate it is possible to produce some melt.

Melt produced from hydrated basalt is enriched in SiO_2 and sodium and broadly dacitic in composition, and because it is produced in the presence of garnet it has a characteristic trace-element composition. Some arcs do erupt lavas with compositions that are consistent with a component of slab melt, and such melts are found in arcs that subduct young slabs (e.g. the Aleutian Islands in the North Pacific Ocean, and South Chile). Overall, the role of the altered oceanic crust in most arcs is to give up H_2O progressively as the slab dehydrates under the fore-arc and volcanic front.

5.4.2 Sediment phase diagram

There are a wide range of sediments that could be subducted, which makes using a single phase diagram for a sediment slightly simplistic. However, experiments have been performed on various deep-sea clay compositions that are reasonably representative of pelagic sediment. A simplified phase diagram for such a composition is presented in Figure 5.10. At low pressures, sediments undergo a series of dehydration reactions that drive off some chemically bound H_2O in addition to any free pore fluids, and a large amount of H_2O is lost into the fore-arc region. This is most often expressed on the sea floor as mud volcanoes and the intrusion of masses of serpentinite.

The critical aspects of these experiments are:

1 The solidi temperatures of sediments vary from below that of the solidus of a 'wet' basalt, for clays that contain some quartz, to higher than the 'wet' basalt solidus for quartz-poor clays. This implies that, under certain conditions, sediment can melt without the underlying oceanic crust melting. The most recent thermal models (e.g. Figures 5.7 and 5.8) have a very high thermal gradient close to the sediment–mantle wedge interface, and this makes the melting of sediment more likely without recourse to melting the rest of the slab.

2 At temperatures above the solidus sediment a variety of hydrous phases break down as melting progresses. These are phengite, amphibole and biotite, which break down respectively with increasing temperature. The stability of these high-temperature hydrous phases means that subducted sediment behaves in a similar manner to the altered oceanic crust, in that it continuously dehydrates as the slab descends. The melt produced from sediment melting is essentially rhyolitic in composition, but it also contains high alkalis (Na_2O and K_2O) and in excess of 10 wt% H_2O.

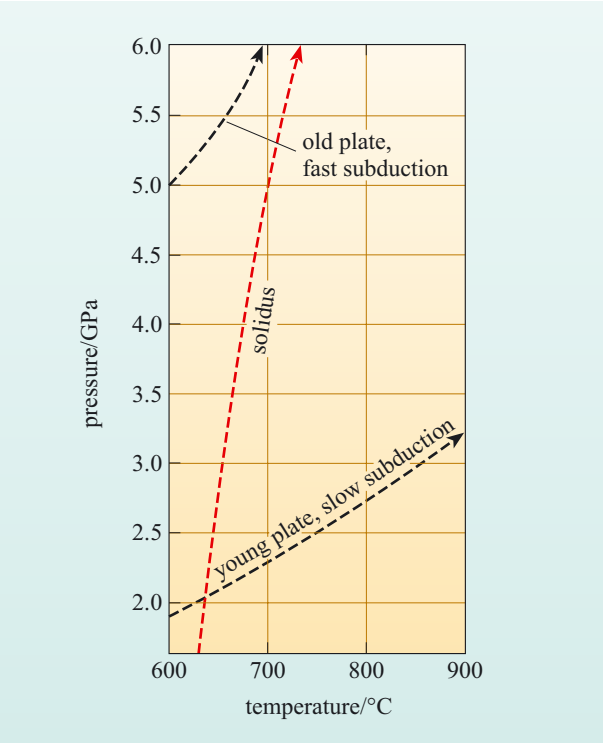


Figure 5.10 Simplified phase diagram for H_2O-saturated red clay with *P–T* paths for the slab from a variety of subduction thermal models.

5.4.3 Peridotite phase diagram

The peridotite phase diagram presented in Figure 5.11 is for an H_2O-saturated peridotite at low temperatures. At low pressures a whole host of hydrous minerals are present, including serpentine minerals, chlorite, talc and amphibole. As the temperature rises, the number of stable hydrous phases decreases, as does their potential for including structurally bound H_2O. Amphibole is stable to a pressure of ~3 GPa, but at higher pressures only chlorite and then a serpentine mineral called antigorite are present. Thermal models indicate that it is highly unlikely that the hydrous mantle in the slab ever melts, but it potentially will dehydrate and flush H_2O through the overlying basalt and sediment. Moreover, the possibility of stabilising hydrous phases at high pressures in a peridotite composition may allow the transport of H_2O into the deep mantle, a possibility explored in Chapter 7.

■ Is the peridotite–H_2O phase diagram appropriate only for the subducted slab?

■ No, because dehydration reactions in the subducted sediment and oceanic crust release H_2O into the overlying mantle wedge. The inverse geothermal gradient in the mantle wedge means that the mantle just above the slab is also quite cool, and so the same peridotite phase diagram is also useful for understanding the base of the mantle wedge.

Consider what happens to the base of the mantle wedge, which is viscously coupled to the slab. It will receive large amounts of H_2O and possibly some melt from the slab. This hydrated peridotite is then dragged downwards and at some point may release its H_2O to the mantle wedge and initiate melting. The movement of this hydrated amphibole-bearing peridotite in the mantle wedge is an important constraint on the location of partial melting in the mantle wedge. Therefore, it has been the focus of several models for melt generation in subduction zones, which usually involve a series of hydration and dehydration reactions within the peridotite that allow the transport of H_2O from the slab to a hotter part of the mantle wedge before eventually encountering conditions where the mantle melts. Regardless of the actual mechanism, melting in the mantle wedge above a subduction zone is most likely related to the stability of amphibole or other hydrous minerals.

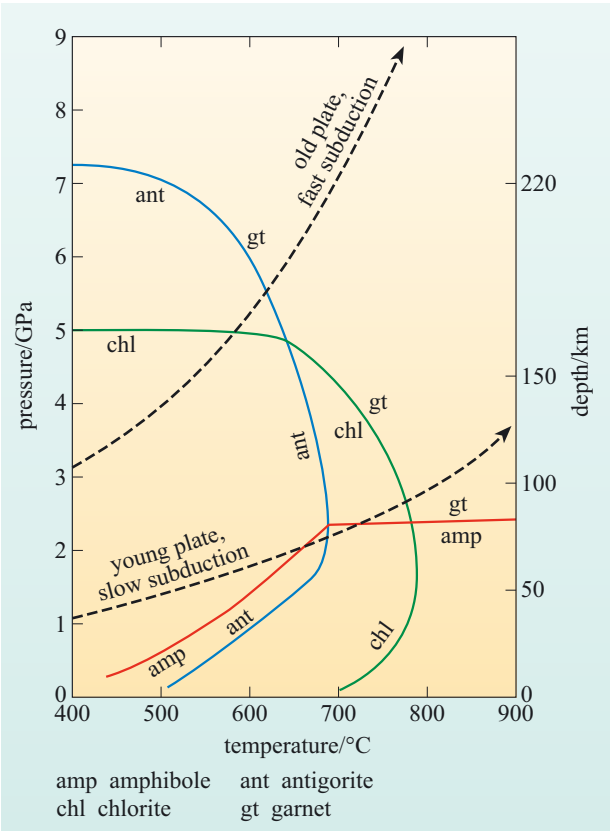


Figure 5.11 Simplified phase diagram for H_2O-saturated peridotite with *P–T* paths for the slab for two different subduction zone thermal regimes (labelled NE Japan and Cascades) and the stable sub-solidus phases. Peridotites stabilise many phases at high temperatures and pressures and on this figure they are simplified into regions that contain the key phases, amphibole (amp), chlorite (chl), the serpentine mineral antigorite (ant) and garnet (gt). Garnet is stable to the right or above any line it is labelled against, meaning it is a high-pressure phase. By contrast, amphibole is stable at low pressures below and to the right of its line. Chlorite and antigorite are stable to the left of their respective stability lines, with antigorite stable to higher pressures. There are regions were more than one of these key phases is stable (e.g. chlorite and antigorite). (Poli and Schmidt, 2002)

- Can you recall any other evidence that might provide more information about the distribution of melt in the mantle wedge?
- The seismic tomography beneath Japan (Figure 5.5) suggests that melt or fluid is present close to the slab at 150 km depth and that melt is distributed along a diagonal path across the mantle wedge.

5.4.4 Melting the mantle wedge

Figure 5.12 is a phase diagram for both a dry (anhydrous) and wet (hydrous) peridotite with two calculated geotherms for the mantle wedge. One geotherm assumes that the viscosity does not change and the other assumes a more realistic mantle wedge viscosity that takes into account the effects of H_2O on mantle viscosity. The geotherm has a distinct but expected shape, with its highest temperature located in the centre of the mantle wedge, away from the cooling influence of the Earth's surface and the subducting slab.

The hydrous peridotite phase diagram is relevant to understanding what happens in the mantle wedge (Figure 5.12). If enough H_2O is added to a mantle peridotite then

there is a major change in the shape of the peridotite solidus. The curved shape reflects the stability of amphibole, which is stable in hydrous peridotites below 3 GPa. The key observation is that the temperature that a hydrous peridotite will start to melt is several hundred degrees lower than for a dry peridotite. However, even small amounts of H_2O decrease the temperature of the solidus, without forming amphibole, as shown by the series of lines roughly parallel to the dry solidus in Figure 5.12.

- What key difference is there between the two mantle wedge geotherms?
- The isoviscous geotherm only crosses the H_2O-saturated mantle peridotite solidus, whereas the geotherms using H_2O-dependent viscosities cross both the anhydrous and all of the solidi with as little as 0.2% water. The geotherm also shows that there are parts of the mantle wedge where amphibole would not be stable.

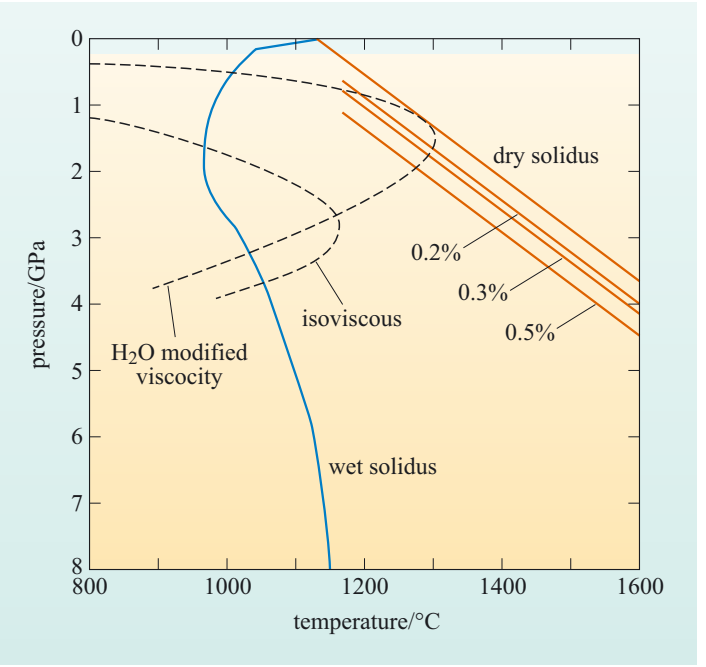


Figure 5.12 Simplified phase diagrams for anhydrous and H_2O-saturated peridotite. Also plotted are solidii curves for peridotite plus 0.1, 0.2 and 0.5 wt% H_2O. Also plotted are geotherms for the mantle wedge from two subduction thermal models (see Figure 5.8).

How do these model results compare with real observations on the composition of arc lavas? You should recall that one of the assumptions for earlier thermal models and experimental studies of magma generation in arcs was that andesites were the primary melts in arcs, but that more recent studies have discovered that basalts are very common in many oceanic arcs. Moreover, these basalts have high MgO contents (~10 wt%) and magnesium numbers consistent with being in equilibrium with the mantle (see Section 4.2.1). Experimental studies have also found that these arc basalts were produced at 1300–1400 °C at 2 GPa. However, these same melts have high H_2O contents, and this presents a dilemma. If these magnesian arc basalts are representative of primary magmas generated within the mantle wedge then these temperatures are too high for them to have been produced at the H_2O-saturated solidus. These observations, therefore, imply strongly that, even though hydrous melting is an important aspect of magma generation beneath arcs, melting in the mantle wedge is influenced by an additional process or processes.

Ideas about melting in the mantle wedge have changed substantially in the last 10 years. The observation that many volcanic arcs are located about 110 km above the subducting slab was used to argue that melting was associated with a pressure-related dehydration reaction, usually ascribed to the breakdown of amphibole in the slab, which triggered melting in the wedge. You should now realise that there are many dehydration reactions that occur in the slab, and these all have the potential to add H_2O to the overlying wedge over a range of pressures. Experiments in which hydrous peridotite is melted produce about 8% partial melt when amphibole breaks down. However, most arc lavas are thought to have been generated by at least as much partial melting as that which produces MORBs and possibly by as much as 30% (see Section 5.8).

Except when amphibole is present, H_2O behaves as an incompatible element during partial melting of the mantle and so is lost rapidly from the mantle into the melt phase. If H_2O is added to the mantle it will lower its melting temperature, but when this hydrated mantle is melted it will lose its H_2O and mantle melting will revert to an anhydrous solidus. Current models favour a mechanism in which H_2O is released into the wedge to produce melting at high temperatures, consistent with the high temperatures found for some magmas. Geochemical evidence not presented in this chapter has found that the degree of partial melting is directly related to the H_2O content of the source, and this also supports a flux melting model. There is also growing evidence that there may be some decompression melting.

Returning to some of the geophysical evidence presented earlier in the chapter, the deep melt generated in the Honshu arc could be related to hydrous peridotite being dragged down to 150 km and eventually releasing H_2O and generating some melt. The melt and H_2O then migrate through the wedge and induce high-temperature melting in the centre of the mantle wedge. Clearly, there is more research needed to understand melting in subduction zones, but hydrous melting of the mantle wedge at high temperatures is a model that currently fits most observations.

To illustrate the effects of varying the rate of subduction and the age of the subducting plate, representative (endmember) pressure–temperature paths have been added to the various slab phase diagrams. These paths are labelled ‘old plate, fast subduction’ and ‘young plate, slow subduction’. The first of these pressure–temperature paths is representative of, for example, NE Japan (see Figure 5.7a), whereas the second is similar to the thermal structure of SW Japan (Figure 5.7b) or the Cascades in western North America. The key is to note whether the pressure–temperature path crosses the solidus leading to melting or whether it just crosses boundaries between mineral reactions.

- What happens to the various components in the subducted slab (i.e. sediment, altered basaltic crust and hydrated peridotite) during subduction for (a) the old plate, fast subduction and (b) the young plate, slow subduction thermal models?

- In both models the slab is hot enough to melt the subducted pelagic sediment, although this may only happen at depth in model (a). Modelling using H_2O-modified viscosities produce steeper thermal gradients at the top of the slab and this will make melting of the subducted sediment more likely. By contrast, for the old plate, fast subduction model (a) the basaltic portion of the slab remains cold enough not to melt, but hot enough to lose H_2O by dehydration reactions. In the young plate, slow subduction model (b), the basaltic crust will dehydrate and the hydrous parts may also melt. Finally, the upper part of the mantle lithosphere will dehydrate in both models, but it will happen more rapidly in the young plate, slow subduction model (b).

Slab melts have been discovered in volcanic arcs such as the Cascades and are known as **adakites**. They have chemical compositions consistent with melts produced in hydrous basalt melting experiments.

Summary

It can be concluded from this discussion of the thermal structure of subduction zones that a series of complicated melting and dehydration reactions can occur both within the slab and the mantle wedge. These are controlled by the rate of subduction and the age of the subducted slab. Moreover, recent modelling strongly suggests that the very process of subduction provides a mechanism for lowering the viscosity of the mantle wedge by adding H_2O, and this may help to enhance the flow of hot (1300–1400 °C) mantle below the arc. Sediment melting may occur in most arcs, because there is a very high thermal gradient across the top of the slab due to the enhanced flow of the mantle wedge. The basaltic slab will only melt when young plates (<25 Ma) are subducted slowly, but this is a rare occurrence (e.g. the Aleutian Islands and the Cascades). More commonly, the slab will dehydrate by a series of progressive dehydration reactions. The actual method of melting the mantle wedge is still controversial, but hydrous fluxing of the wedge is a model that currently fits most observations.

5.5 The composition of island-arc magmas

Subduction-zone magmas represent one of the three important broad tectono-magmatic associations, with magmas produced at mid-ocean ridges and intra-plate magmas making up the other two. Given the differences in tectonic setting for these different associations you might expect significant differences in their chemical composition. Figure 5.13 is a plot of magmas from these three tectonic settings on a total alkali ($Na_2O + K_2O$) versus SiO_2 diagram, commonly known as a **TAS** diagram. This type of diagram is used to classify volcanic rocks using simple chemical criteria. Each suite of lavas illustrates the range of lava compositions found in each tectonic setting, and although only a representative fraction of the geochemical analyses available are plotted, some clear distinctions between the magma types can be seen. Mid-ocean ridge samples have a restricted compositional range and are predominantly basaltic in composition with only rare intermediate compositions. Intra-plate magmas span a much

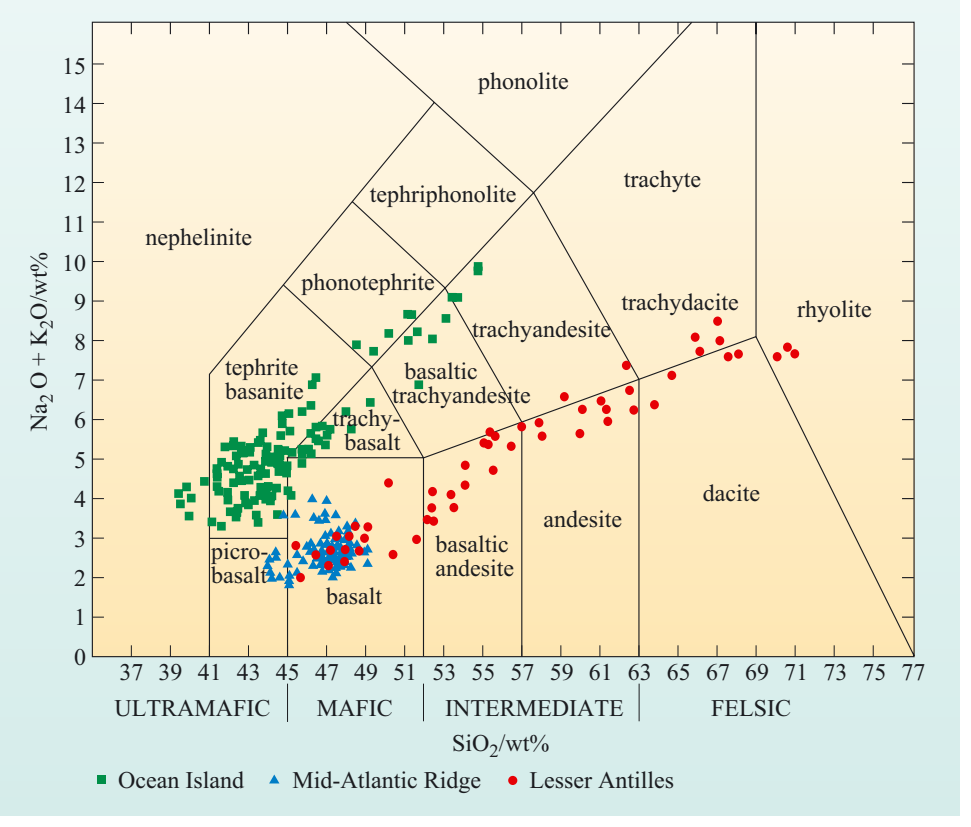


Figure 5.13 Plot of total alkalis ($Na_2O + K_2O$) versus SiO_2 showing fields of different rock names. Representative suites of data from an island-arc volcano (from the Lesser Antilles), an ocean island volcano and MORBs from the Mid-Atlantic Ridge are plotted. (Adapted from Le Bas et al., 1986)

wider SiO_2 range, and generally are enriched in the alkali elements, sodium and potassium. Subduction-zone magmas have the greatest range in SiO_2 contents and generally have lower total alkali contents for a given SiO_2 than intra-plate magmas. The predominant rock association found in subduction-zone settings is a **basalt–andesite–dacite–rhyolite association**.

The TAS diagram provides two important insights into magma generation in subduction zones. Magmas from all three tectonic settings include some basaltic compositions, and you should recall from Section 4.6 that such compositions, are generated by partial melting of peridotite in the mantle. Increasing SiO_2 contents in lavas are the result of fractional crystallisation, and the wide range and high SiO_2 contents of arc lavas strongly imply that fractional crystallisation can readily take place in the arc crust.

You will consider the difference in composition between MORBs and arc lavas in more detail later in this chapter, but there is one final difference between these two magmas worthy of mention. Magmas erupted subaerially (or in shallow water depths) tend to lose all or some of their volatiles, such as H_2O, CO_2 and SO_2 to the atmosphere. By contrast, those lavas that erupt beneath significant ocean depths, under pressures of a few hundred times that of the atmosphere, tend to retain volatile components. When MORBs and arc lavas are collected from the deep ocean floor, the glass in these submarine basalts will quench all of the volatiles that were originally dissolved in the magma. Comparison between primitive quenched glasses reveals that subduction-zone basalts contain significantly higher H_2O contents than MORBs and basalts from other tectonic settings. Primitive arc lavas can contain 3–6 wt% H_2O compared with generally less than 1 wt% in MORBs, and they are also enriched in the element chlorine (Cl).

- Where do you think the H_2O and Cl dissolved in arc lavas originated?
- The logical source of these volatiles is from the subducted slab, either from altered oceanic crust or from sediments. Ultimately, the H_2O and Cl come from seawater, providing a direct link between the hydrosphere and the mantle.

5.6 Evolution of arc magmas

The observation that arc lavas extend to higher SiO_2 compositions (Figure 5.13) is not just of academic interest, because high-SiO_2 lavas have some interesting but potentially hazardous characteristics. Increasing the SiO_2 content of a melt increases what is called the **polymerisation** of the melt. In effect, this means that atoms are more strongly bonded to each other (they form the equivalent of bigger molecules) and this makes the melt more viscous. The viscosity of a liquid reflects the ability of the atoms to flow past each other. As an analogy, you can imagine the viscosity of a basalt to be like water and the viscosity of a SiO_2-rich melt, such as a dacite or rhyolite, to be like very thick porridge. The combination of high viscosity and volatile content is a recipe for potential disaster. Consider the effects of heating a pan of water and a pan of porridge. As the water boils, bubbles of H_2O vapour form and rise easily through the water and burst gently at the surface. By contrast, any gas in the porridge forms much larger bubbles which tend to burst more violently at the surface and spatter hot porridge over the cooker! The analogy for what happens

during an eruption is not perfect, but it serves to illustrate that more viscous magmas tend to build up high gas pressures and then erupt explosively. The eruptions of Mt St Helens, Krakatau and Mt Pinatubo are examples of the effects of catastrophic degassing of magma, and contrast with a less viscous and less volatile-rich basaltic lava flow erupting on Hawaii. The eruption of Krakatau obliterated most of the island and ejected volcanic ash and gases high into the stratosphere. This affected atmospheric conditions across the globe and reduced global temperatures, whereas a Hawaiian eruption causes local difficulties related to acidic smog and the disruption of local infrastructure. Clearly, it is useful to be able to understand how and why SiO_2-rich melts are produced in subduction zones.

- Can you think of any reasons why basaltic arc lavas might evolve into such SiO_2-rich compositions?

- There are three possible reasons why arc lavas behave differently from those from other tectonic settings: (i) arc lavas have elevated H_2O contents and this may influence the crystallisation history of the magma; (ii) some volcanic arcs, particularly continental arcs, have thicker lithosphere and this may act to slow down the transport of magma and allow for more extensive crystallisation; (iii) magmas in continental arcs rise from their source through high-SiO_2 crust, which may act to contaminate them in some way.

Crystallisation has an important role to play, and it is necessary to define what happens when a magma crystallises. In Section 4.6 you used experimental data and phase diagrams to illustrate how a MORB crystallises. Crystallisation usually happens when a magma cools. In the case of a simple experiment where the crystals are continuously in contact with the melt, some simple rules are observed. This type of crystallisation is called **equilibrium crystallisation**, and because the crystals are not separated from the melt the bulk composition of the crystals and liquid added together do not change. Also, because chemical equilibrium is maintained at all times, the crystals are homogeneous (i.e. they do not vary in composition from the core to rim of the grain). For example in the binary olivine system, although the forsterite composition of the olivine crystals changes continuously during crystallisation it is always homogeneous at any point in the crystallisation history. The simple effect of equilibrium crystallisation is that, if you start with a basalt, although the liquid may change its composition during cooling, the bulk composition remains basaltic, which is not a particularly good start for making a wide range of magma compositions.

There is another type of crystallisation called **fractional crystallisation**, which involves the continuous or episodic removal of crystals from a magma system. This has a rather dramatic effect on the composition of magmas and can be illustrated using a simple binary phase diagram.

5.6.1 The plagioclase binary phase diagram

Plagioclase is an important mineral in the crystallisation of most lava associations because it has a wide range of compositions and is stable over a wide range of temperature. The plagioclase feldspar system involves a solid solution between anorthite ($CaAl_2Si_2O_8$) and albite ($NaAlSi_3O_8$). The solid solution in the plagioclase system is not as simple as in the olivine system because it involves the coupled

substitution of $Ca^{2+}Al^{3+}$ for $Na^{+}Si^{4+}$ (notice that the charge is balanced), but the shape of the binary diagram is very similar (Figure 5.14). Anorthite is the higher temperature feldspar and the mass proportion of anorthite is given as the An content (e.g. 50% anorthite is An_{50}). During equilibrium crystallisation, if a liquid of An_{50} composition cools it will start to crystallise An_{82} crystals at 1445 °C (points a and b Figure 5.14). If the liquid continues to cool to 1385 °C, when it is 50% crystallised (line c to d), it will consist of 50% An_{68} crystals and 50% An_{32} liquid. Below 1290 °C it will be completely crystallised and the crystals will have the same initial bulk composition of An_{50}. The same diagram can be used to illustrate the effect of fractional crystallisation, where some crystals are separated at a point during the cooling process. Starting with the same bulk composition and allowing it to undergo equilibrium crystallisation to 1385 °C, 50% An_{68} crystals are in equilibrium with liquid with a composition of 50% An_{32} (point c). At this point, if all of the crystals are removed then this leaves a liquid with a bulk composition of An_{32}. This composition (point c) can be treated as a new liquid and upon cooling to 1290 °C will be composed of 50% An_{49} crystals and 50% An_{15} liquid (points m and n) and, ultimately, will become completely crystallised with a bulk composition of An_{32} below 1205 °C (point p). This simplified version of fractional crystallisation illustrates that more sodium-rich compositions can be reached at lower temperatures from the same initial bulk composition by this mechanism than is possible by equilibrium crystallisation.

It is possible to envisage a mechanism that removes crystals more than once during crystallisation. This would drive the liquid composition to be very sodium rich and produce a set of crystals with different anorthite compositions. The process described

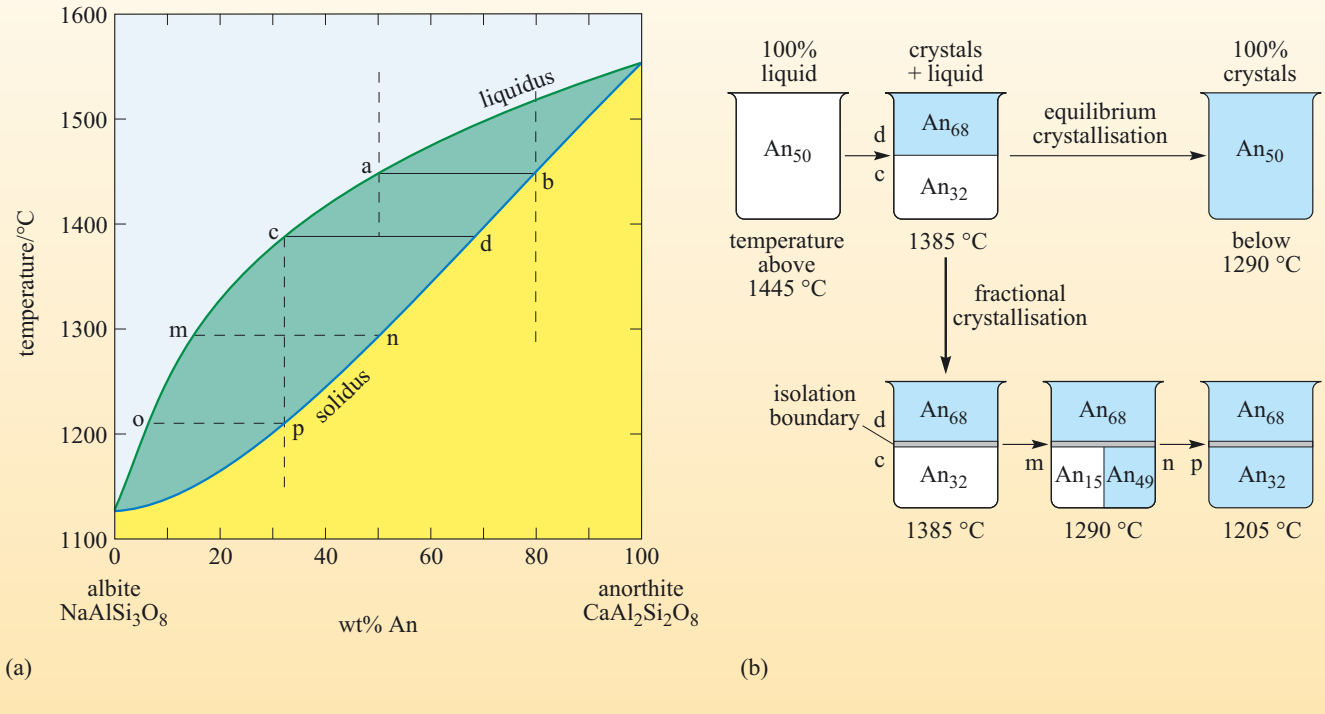


Figure 5.14 (a) Phase diagram of temperature against composition for the binary plagioclase system at 10^5 Pa. A sample of An_{50} crystallises under equilibrium conditions until 1385 °C, whereupon the crystals (point d) become isolated from the liquid (point c) and cool separately. The horizontal lines are tie-lines between coexisting crystals and liquid, before (solid) and after (dashed) the crystals were separated at 1385 °C. (b) Diagrammatic sketch illustrating the crystallisation of the sample of An_{50}.

is still an equilibrium process, in that each crystal that separated out has, up to that point, kept in equilibrium with the liquid and remains homogeneous. But what happens in nature? One of the most striking features of many arc lavas is that they contain plagioclase phenocrysts with complicated zoning. Such a **zoned crystal** is illustrated in Figure 5.15. Each of the concentric rings represents a growth of the crystal with a distinct anorthite content. Clearly, this crystal is not homogeneous and so could not have been formed by simple equilibrium crystallisation.

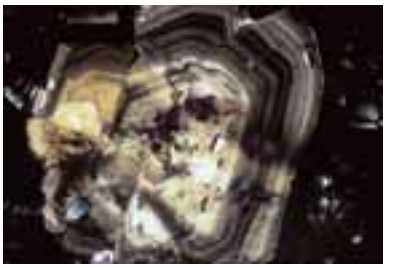


Figure 5.15 A zoned plagioclase phenocryst. The concentric lines represent boundaries between regions of different compositions.

- How might such a zoned crystal be formed?
- There are several possible explanations for generating such a crystal. First, each zone might represent a period of equilibrium crystal growth, and the crystal then separated out. In this case, each zone is like a tree ring that records some chemical information about the magma the crystal was growing in at that time. Second, only the outer layer (rim) kept in equilibrium with the melt. This indicates that there was not enough time for the ions to move (diffuse) across the crystal and to re-equilibrate the inner zone (core) with the magma before the crystal was physically separated from the melt.

Zoned crystals provide good evidence for fractional crystallisation and they potentially record information about changes in magma composition prior to eruption. But before fractional crystallisation can have an effect on magma composition the crystals need to separate from the melt. Crystals that do separate from the melt will accumulate on the sides and floors of magma chambers and conduits to form cumulate rocks, similar to those found in the lower part of the oceanic crust (Chapter 4). Arc cumulates have a distinct mineralogy that provides further information about the evolution of magmas in subduction zones.

5.6.2 The Lesser Antilles: fractional crystallisation in an arc

The Lesser Antilles islands in the eastern Caribbean are a well-studied example of an island arc associated with the westward subduction of the South American Plate beneath the Caribbean Plate. This volcanic arc contains several active volcanoes, including Montserrat, which erupted in the late 1990s, Mt Soufriere on the island of St Vincent that erupted in 1979 and Mt Pelée on Martinique which killed 30 000 people in 1902.

Two volcanoes, one from the island of Dominica and one from St Lucia, can be used to illustrate the chemical evolution of arc magma. These two volcanoes have been active in the last few million years and are located close to each other. However, one has erupted mainly basaltic rocks and the other produced andesites and dacites, and by studying these lavas it should be possible to discover whether the higher SiO_2-content andesites and dacites could be produced by fractional crystallisation of the basalts.

Major elements

Figures 5.16 and 5.17 show SiO_2 wt% plotted against CaO and MgO wt% respectively for a range of volcanic rocks from Dominica. The average compositions of four minerals, i.e. olivine, plagioclase, clinopyroxene and orthopyroxene commonly found as phenocrysts in the lavas are also plotted on the

figures. In both Figures 5.16 and 5.17, the volcanic rocks define smooth trends and the shape of the trends can be used to infer what minerals or groups of minerals have fractionated from these magmas. You may remember from Chapter 4 that if a mineral is removed from a magma, then the liquid moves away from the composition of the mineral on a chemical variation diagram. If two minerals are crystallising, then a point along the line connecting the two minerals defines the composition of the phases being extracted, conventionally referred to as the **extract**. If three or four minerals are crystallising, then the composition of the extract lies within a triangle or quadrilateral on the variation diagram defined by the three or four mineral compositions, respectively.

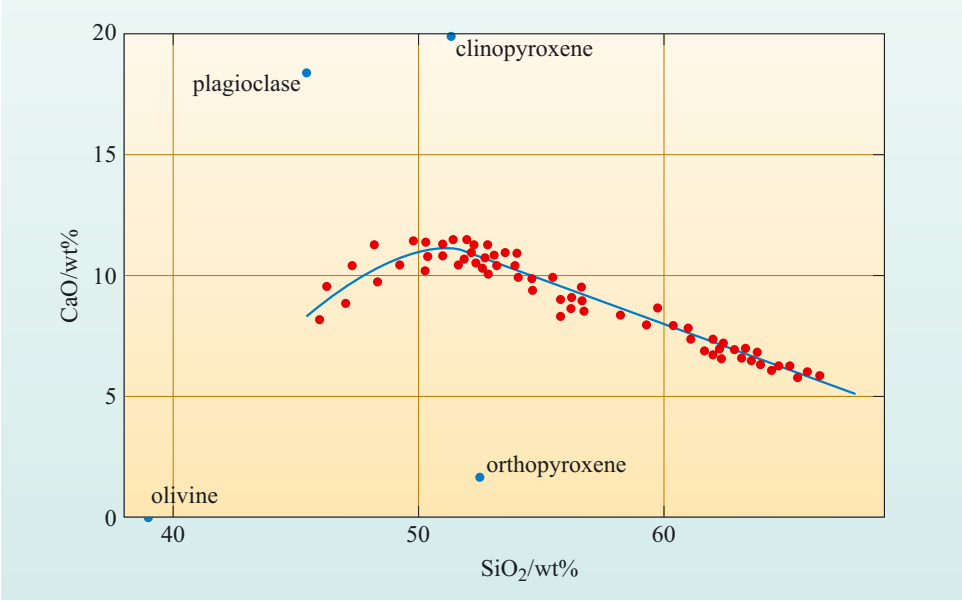


Figure 5.16 CaO versus SiO_2 concentrations from a suite of lavas and phenocrysts from two volcanoes on the island of Dominica. The blue line represents a smoothed compositional trend.

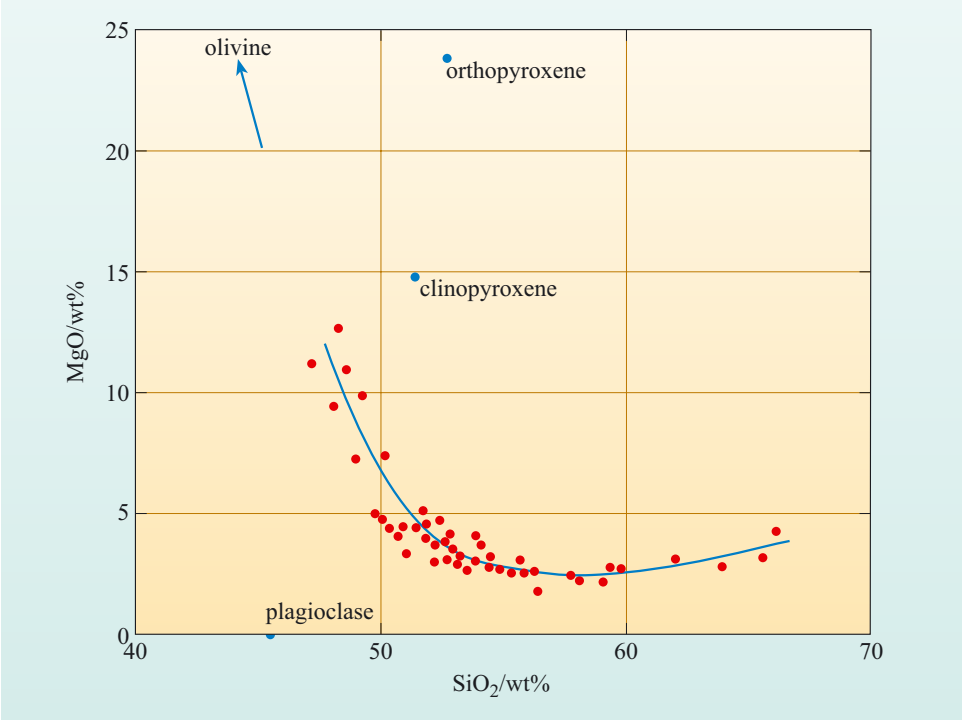


Figure 5.17 MgO versus SiO_2 concentrations from a suite of lavas and phenocrysts from two volcanoes on the island of Dominica. The blue line represents a smoothed compositional trend.

- Can you observe any kinks in the trends defined by the data in Figures 5.16 and 5.17?
- Yes, there is a strong kink in the trends at about 50–52 wt% SiO_2 particularly on Figure 5.16.

- If the observed variations in SiO_2, CaO and MgO were controlled by fractional crystallisation, what could you say about the composition of the crystals extracted (the extract) compared with that of the liquid? For samples with less than 50 wt% SiO_2, would the extract contain more or less of the following constituents than the liquid: (a) SiO_2, (b) CaO, and (c) MgO?
- The trends in the data for rocks with less than 50 wt% SiO_2 have an increase in CaO and SiO_2 and a decrease in MgO. The extract must, therefore, contain (a) less SiO_2, (b) less CaO and (c) more MgO than these liquids.

- What about the extract controlling liquids with 55–60 wt% SiO_2?
- The extract contains less SiO_2, but more CaO and just slightly more MgO than those liquids.

- Look at the trend in the data for rocks with less than 50 wt% SiO_2. Does the composition of any one phenocryst plot in such a way that crystallisation of that mineral alone could drive the composition of the liquid along the trend?
- Yes, the removal of just olivine would drive the composition of the liquids along a line of increasing CaO until it starts to flatten out at about 50 wt% SiO_2. Similarly, olivine has a high MgO content and, therefore, removal of olivine reduces the MgO contents of the liquids with less than 50 wt% SiO_2.

- For rocks with 52–60 wt% SiO_2, is it possible that just one of the phenocryst minerals could produce the trends observed in the data?
- The answer is no, because no one mineral plots along the trends of data for rocks with more than 52 wt% SiO_2. If this trend reflects fractional crystallisation, then it must be due to more than one phenocryst phase crystallising. The extract has to be some combination of a high-calcium phase (plagioclase and/or clinopyroxene) and a low-calcium phase (olivine and/or orthopyroxene).

The chemical variation diagrams give a guide to which minerals crystallise and their proportions but cannot give a unique answer. To find out which minerals actually crystallised together in the Dominica lavas involves an examination of the rocks themselves. Phenocrysts are the minerals that were crystallising at the time the liquid was quenched during eruption on the Earth's surface. Thus, identification of the phenocryst minerals and their proportions provides an estimate of the bulk composition of the aggregate minerals (the extract) that were crystallising together shortly before they were erupted. However, the phenocryst assemblage may not be completely representative of the mineral proportions that were extracted during a long period of fractional crystallisation.

By definition, fractional crystallisation involves the separation of crystals, and so earlier phenocryst minerals will now comprise the cumulates that reside at the base and sides of the underlying magma chamber.

A typical basalt from Dominica contains 10.5% olivine, 27.3% plagioclase and 6.2% clinopyroxene. The abundances represent the percentage the phenocrysts make up of the total rock, and so the relative proportion of the phenocryst assemblage, in this case, consists of 24% olivine, 62% plagioclase and 14% clinopyroxene.

- The aggregate phenocryst assemblage discussed above has a bulk composition of 14.2 wt% CaO and 44.5 wt% SiO_2. Where does this composition plot in relation to the compositional trend between 52–60 wt% in Figure 5.17? Could fractional crystallisation of this phenocryst assemblage be responsible for the variations in the lavas with 52–60 wt% SiO_2?
- The bulk composition of the aggregate phenocryst assemblage plots on an extension of the CaO–SiO_2 compositional trend, which could, therefore, be explained by extracting the phenocryst assemblage in the proportions found in the Dominica basalts. The bulk composition of this phenocryst assemblage plots in a position that could explain the trend not only between 52–60 wt% SiO_2, but also all the way up to 66 wt% SiO_2.

This is where other geochemical and geological information needs to be considered. Just using the SiO_2 versus CaO diagram could be misleading, as relying on one variation diagram limits the possible number of geometrical constraints. This is illustrated by the presence of an upturn in the SiO_2 versus MgO plot at about 60 wt% SiO_2 (Figure 5.17), suggesting a change in the phenocryst assemblage at his composition.

The phenocryst assemblage found in the basalts is not going to be the same one that is present in higher SiO_2 melts. High-SiO_2 rocks, such as rhyolites (the volcanic equivalent of a granite), consist largely of different minerals (quartz and feldspars) from those found in basalts (olivine, clinopyroxene and plagioclase). Magnesium-rich olivine and quartz are not stable together in the same rock, reacting together to produce orthopyroxene. Moreover, it has already been shown that during fractional crystallisation the composition of minerals that have a solid solution change their composition as the melt composition evolves. The plagioclase compositions plotted in Figures 5.16 and 5.17 are just average values. Analyses of plagioclase from a rock of higher SiO_2 content on Dominica gives 8.2 wt% CaO and 58.6 wt% SiO_2. This is consistent with the chemical formula of anorthite-poor plagioclase, because the decrease in Ca^{2+} is coupled with the substitution of Si^{4+}. If you plot this composition in Figure 5.17 you will find that it lies on the trend of the rock analyses, which suggests that the separation of anorthite-poor plagioclase alone could control the compositions of rocks with the highest SiO_2 content. The slight rise in MgO contents in the SiO_2-rich samples (Figure 5.17) is consistent with this interpretation, as plagioclase contains virtually zero MgO.

- Would you expect olivine to be crystallising from a melt with 3 wt% MgO and 60 wt% SiO_2 and if so what would be its Mg#?
- The experimental data for crystallising MORB (Section 4.6.3) indicates that olivine becomes a less important phase during extensive crystallisation and if it did crystallise, then it would produce a low Mg# olivine.

In Dominica, olivine, which is abundant in low-SiO_2 rocks, is rarely observed in rocks with more than 60 wt% SiO_2, indicating that the liquids have moved out of the olivine stability field. Instead, orthopyroxene starts to crystallise; and since, like olivine, it contains little CaO and significant amounts of MgO, the transition from olivine to orthopyroxene is not marked by any kink on the trends in Figures 5.16 and 5.17.

The gradual changes in the composition of the lavas from the two volcanoes on Dominica, therefore, indicate that they may be related by fractional crystallisation. From the phenocryst assemblages observed in the rocks and their compositions, a fractional crystallisation path capable of generating the chemical variations observed in the lavas can be deduced. This can be summarised as follows:

1 At less than 50 wt% SiO_2, only olivine crystallises.

2 At 50–60 wt% SiO_2, plagioclase and clinopyroxene crystallise together with olivine (at lower SiO_2) and orthopyroxene (at higher SiO_2).

3 At more than 60 wt% SiO_2, the amount of pyroxene crystallising decreases until the chemical variation in very SiO_2-rich liquids appears to be controlled by crystallisation of anorthite-poor plagioclase alone.

Thus, it can be concluded that the andesites on Dominica appear to have been derived from the fractional crystallisation of a basaltic magma that was produced by partial melting in the mantle wedge.

This section demonstrates that, as the liquid evolves, the phenocryst assemblage changes. Plagioclase feldspar is present through most of the crystallisation history and will produce a wide range of plagioclase compositions, often resulting in zoned crystals. Figure 5.18 presents the composition of plagioclase from a suite of lavas from St Lucia as a function of the composition of the lavas. The lavas vary from basaltic andesites (56 wt% SiO_2) to silicic andesites (61 wt% SiO_2). Plagioclase is zoned in all of the rocks. However, the plagioclase composition varies systematically with the bulk composition of the lava, with a general decrease in the anorthite content of both the rim and core compositions. The rim composition changes from An_{43} to An_{20} and tracks the cooling liquid composition expected from the binary phase diagram (Figure 5.14). By contrast, the core composition reflects the composition of a crystal that was crystallised from earlier magmas and has not separated completely from the evolving liquid. Where zoning is less marked (i.e. the core and rim compositions are more similar), the plagioclase phenocrysts have reacted more extensively with the magma during cooling.

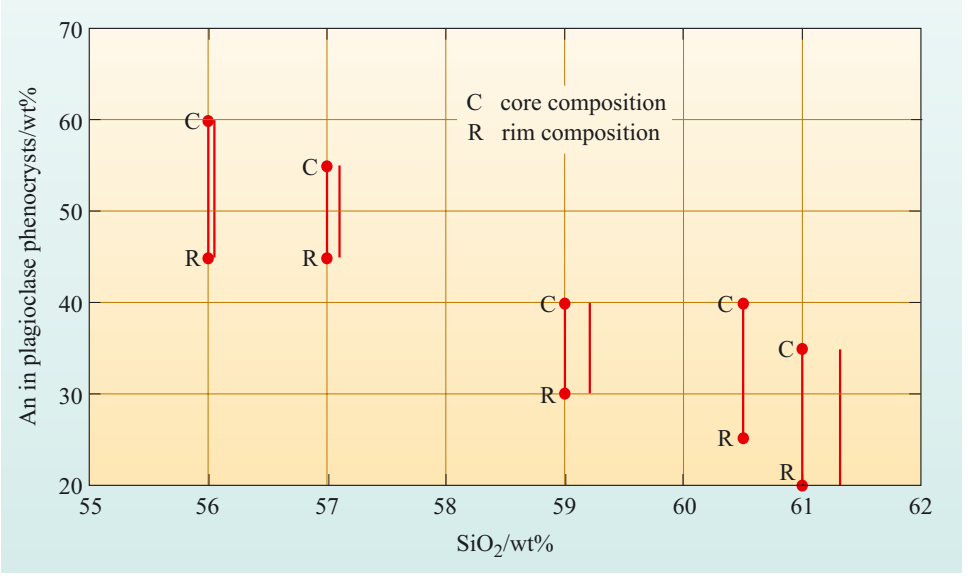


Figure 5.18 Variation of the composition of plagioclase phenocrysts with SiO_2 content of bulk sample from a suite of lavas from St Lucia.

One final comment needs to be made about the crystallisation history of these Lesser Antilles lavas. Plagioclase is an important phase throughout the crystallisation, but it is only stable up to pressures of 0.8 GPa, which equates to 30 km depth. In the Lesser Antilles, the crust is at least 30 km thick, so the presence of plagioclase confirms that crystallisation must take place within the crust rather than within the upper mantle.

Trace elements

Returning to the Dominica lavas, it has already been demonstrated that the major element geochemistry is consistent with crystallisation of olivine, clinopyroxene and plagioclase in varying proportions. Are the trace element variations in these lavas consistent with fractional crystallisation of these phases?

- Look at the partition coefficients for the elements yttrium (Y) and zirconium (Zr) in Table 5.2. What would you predict would happen to the Y and Zr concentrations during fractional crystallisation of olivine, clinopyroxene and plagioclase?
- The partition coefficients for these elements are less than one (except for Y in clinopyroxene). Y and Zr are therefore incompatible elements in olivine plus plagioclase ± clinopyroxene and their concentration should increase with increasing crystal fractionation.

Figure 5.19 is a plot of the Y and Zr variations in the Dominica lavas. Fractionation of basalts to basaltic andesites does indeed increase the concentration of these two elements, but there is a kink in the fractional crystallisation path. Something happens within the andesite field that produces a decrease in the Y content.

- What do you think happens at this point (the decrease in Y content) in the crystallisation history of the Dominica lavas?
- This must indicate where a new mineral crystallises which removes Y relative to Zr (i.e. Y behaves as a compatible element, $D>1$).

- From the partition coefficients listed in Table 5.2, which new minerals could cause the change in the Zr–Y trend?
- Only garnet and amphibole have Y partition coefficients high enough to produce the decrease in Y.

Table 5.2 Selected trace-element partition coefficients for minerals in a magma of andesitic composition.

	Rb	Sr	Ba	Zr	Y
olivine	0.001	0.001	0.001	0.01	0.01
orthopyroxene	0.003	0.05	0.001	0.08	0.1
clinopyroxene	0.001	0.07	0.001	0.25	1.0
garnet	0.01	0.02	0.02	0.5	11.0
plagioclase	0.04	4.4	0.3	0.03	0.06
amphibole	0.01	0.02	0.04	1.4	2.5

Whilst garnet is rarely found in andesitic lavas, amphibole is common in Dominica andesites and dacites (but not in the basalts). Amphibole is a hydrous mineral that is found in some evolved island-arc lavas, so it is not surprising that it crystallises from the more evolved magma compositions given the higher H_2O content of arc magmas relative to MORBs. However, amphibole stability also requires a reasonably high amount of Na_2O in the magma, which is why it is not usually found in arc basalts. The amphibole in the Dominica lavas has a major element composition very close to the clinopyroxene, and so its effect on the variation of CaO, MgO and SiO_2 contents is very similar. The Y–Zr plot (Figure 5.19) however demonstrates that, in addition to cpx, amphibole appears as a crystallising phase about halfway through the andesitic compositions causing Y to decrease. Meanwhile Y flattens off as Zr continues to increase in the dacite samples, indicating the percentage of amphibole crystallising decreases in these higher SiO_2 rocks. This has been confirmed by petrographic observation.

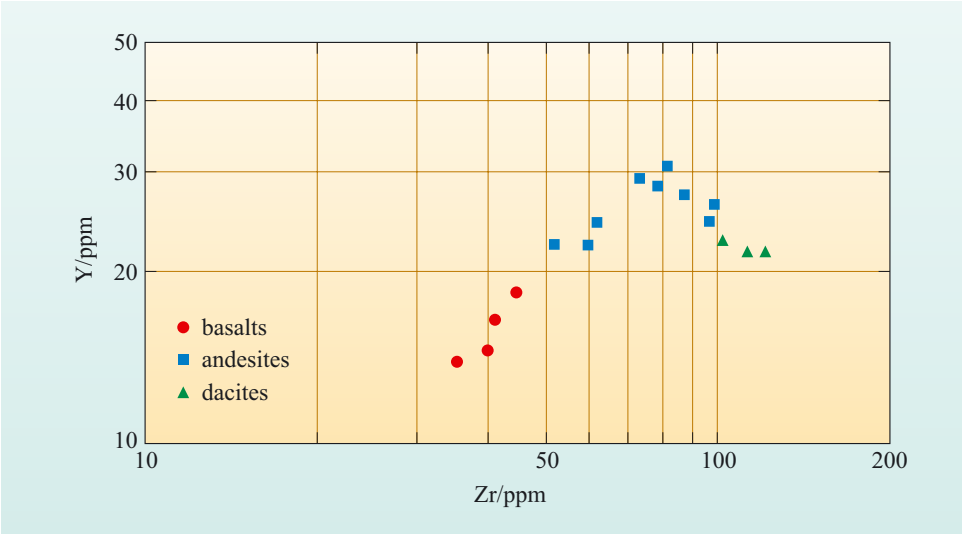


Figure 5.19 Plot of Zr versus Y for lavas from Dominica. Note scales of both axes are logarithmic.

5.6.3 Summary of arc magma evolution

To understand the evolution of an evolving magma, information from the mineralogy of the rocks and their major and trace elements composition are essential. In the case of the lavas from Dominica, fractional crystallisation of olivine, plagioclase and clinopyroxene produced andesitic lavas from a primary basaltic magma, while further crystallisation of plagioclase ± amphibole produced more SiO_2-rich compositions ranging from dacite to rhyolite. There are several factors governing why magmas at destructive plate boundaries evolve to more SiO_2-rich compositions than mid-ocean-ridge lavas. The rate of melt production from the mantle is generally less in arc systems than at mid-ocean ridges, so magmas can be stored in the crust for longer. High melt production island-arc volcanoes, such as Mt Fuji in Japan and Mt Klyuchevskaya in Russia, rarely produce magmas more evolved than basaltic andesites. Magma storage in the crust is also aided by the slightly thicker crust of many island arcs, particularly in continental arcs compared to that at mid-ocean ridges. Continental arcs are dominated by andesite, dacite and rhyolite with very few erupted basalts.

■ Can you think why it is difficult to erupt a basaltic lava in a continental arc?

■ Basaltic lavas contain dense ferromagnesian minerals and the melt itself is denser than the surrounding continental crust, even if it is considerably hotter. Therefore, it will not easily erupt. As the magmas fractionate, the liquids evolve towards less-dense compositions and these are easier to erupt in terms of buoyancy, if not viscosity.

It is no surprise, therefore, that andesitic lavas are common in the Andes, because the thick continental crust allows for extensive fractionation at depth. Moreover, the thickened crust in the Andes prevents some of the magma from erupting. As a result, rocks of andesitic to rhyolitic composition become trapped in the crust where they cool and crystallise more slowly, producing the coarse-grained, intrusive equivalents of dacite and rhyolite, which are granodiorite and granite respectively. If the new material added to the crust in this manner is derived solely by fractional crystallisation of a basaltic parent, then this represents a process whereby the crust grows, and most geoscientists are of the opinion that active crustal growth today occurs mostly above subduction zones. The mean composition of the continental crust is andesitic, and many modern arcs produce a significant quantity of basaltic andesitic to andesitic crust.

5.6.4 Some final thoughts on H_2O

There is one final aspect of magma generation in arcs to consider. H_2O is the component that makes subduction zones function, but what happens to H_2O and other volatile elements in arc magmas as they evolve? H_2O has a bulk partition coefficient similar to the highly incompatible rare earth element cerium (Ce).

- During fractional crystallisation, what will happen to the H_2O content of a melt?
- The H_2O content should increase exponentially as the melt evolves because it behaves like an incompatible element.

This is true in principle, although the ability of a melt to dissolve H_2O is a function of pressure. As long as the melt is stored at a depth of a few kilometres in the crust the H_2O will stay dissolved in the melt. Volcanic rocks from Dominica stabilise the hydrous mineral amphibole when they evolve to an andesitic composition, indicating that magmas do retain their H_2O, and some evolved arc lavas can retain up to 10 wt% H_2O both dissolved in the melt and bound in hydrous minerals. It was noted earlier in this chapter that SiO_2-rich melts become more viscous and that this stops volatiles, such as H_2O, degassing from magmas as they rise to shallower depths in the crust. When the confining pressure is released the vapour and silicate phases separate rapidly and generate an explosive eruption. This can have devastating effects, and the explosive eruptions of Mt Krakatau, Mt Pinatubo and Mt St Helens all resulted from the eruption of evolved volatile-charged magmas. These eruptions are so destructive because they inject ash and other debris into the atmosphere, but locally they also generated deadly ash flows. The eruption of Mt Pelée on Martinique killed many people because a turbulent flow of hot ash and gases at temperatures of 800 °C flowed at high speeds (~160 km h^{-1}) down a valley and flattened the town of St Pierre. Ash from such eruptions can be distributed around the globe and has the effect of cooling the Earth and disrupting air travel.

Other volatile elements also behave as incompatible elements, and sulfur is particularly concentrated in arc magmas because the melts are oxidised and, therefore, do not stabilise sulfide minerals that can crystallise and reduce the sulfur content of a melt. Sulfur is dissolved as a sulfate species in arc magmas and is readily partitioned into H_2O vapour as the melt degasses. This mechanism efficiently

transfers sulfur into the atmosphere. Sulfur from the Mt Pinatubo eruption in 1992 produced a global cooling of about 0.6 °C over the next year, and there are many instances in the geological record of such global cooling events. Sulfur concentration spikes in ice core records correlate with known eruptions (not always from arcs) and these correspond to historical records of global food shortages and cool summers. This is a good example where recycling of an element through a subduction zone has a profound effect on human society.

The longer term aspects of H_2O recycling will be considered further in Chapter 7, but a simplified diagram of the H_2O cycle for a subduction system is presented in Figure 5.20. Estimates of the H_2O fluxes are difficult to determine, but recent calculations suggest that although some H_2O is clearly taken deeper into the mantle, mainly in minerals in altered oceanic mantle, a significant amount is returned through the subduction zone system. This is good news, as it means the oceans are not going to disappear into the mantle!

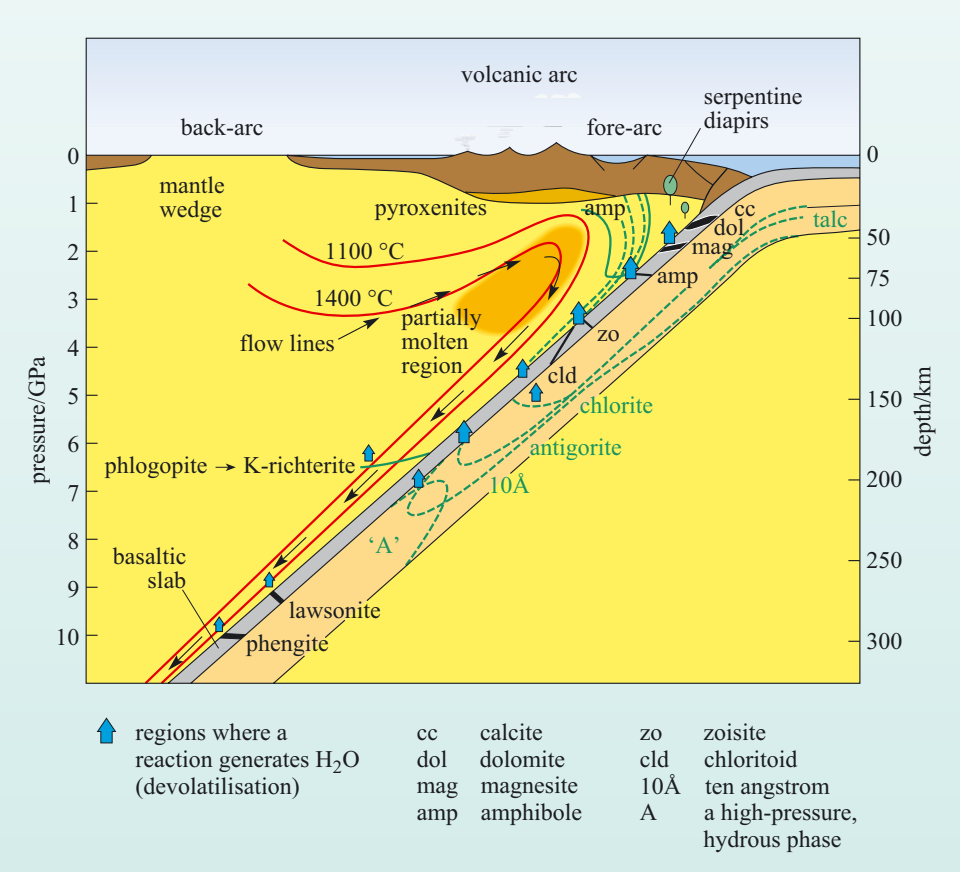


Figure 5.20 Schematic diagram of a subduction zone illustrating the key dehydration reactions that generate H_2O. Solid lines in the mantle wedge represent isotherms whereas the thin arrows are mantle flow lines. Phase boundaries in the basaltic slab are represented by thick lines and in the peridotite section of the slab by dotted lines. It is not important to know the exact position of all the phase boundaries, but that there are series of reactions that generate volatiles at increasing pressures (from calcite to phengite). The 'A' phase in the peridotite section of the slab carries water to depths much greater than the bottom of the figure (>300 km). Large arrows represent regions where a reaction generates H_2O. Mineral names involved in the reactions are labelled on the figure. (Poli and Schmidt, 2002)

5.7 Trace elements and magmagenesis

Before studying the trace element composition of subduction zone magmas in detail, it is important to consider how trace elements can be understood quantitatively. One of the reasons geochemists use trace elements is that they have a wide range of geochemical properties and there are some simple equations that can be used to model their behaviour during partial melting and crystallisation.

5.7.1 A quantitative approach

The distribution of a trace element between a mineral and melt is described quantitatively by the partition coefficient D_{min}, which is defined as

$$D_{min} = \frac{C_{min}}{C_{melt}} \quad (5.1)$$

where C_{min} is the concentration of an element in a mineral and C_{melt} is the concentration of the same element in a coexisting melt.

- Can you recall from Box 1.3 what is meant by the terms compatible and incompatible elements?
- A compatible element has a D_{min} value greater than one, and so preferentially partitions into a mineral phase. An incompatible element has a D_{min} of less than one, and so preferentially partitions into the melt.

The main controls on element substitution in crystal structures were discussed in Section 4.2.

- What are the two major factors that allow element substitution in minerals?
- You should recall that elements with similar ionic radii and charge tend to be able to partition into the same site within a mineral structure.

Nickel (Ni^{2+}), for example, easily substitutes for Fe^{2+} and Mg^{2+} in the olivine structure because of the similarity in their ionic radii, and it is a compatible element in olivine. However, Th^{4+} has a large ionic radius and is highly charged and, as a result, is highly incompatible ($D_{ol} \approx 0.000\ 01$). In a system in which more than one mineral is crystallising or melting it is possible to define a bulk partition coefficient, which is the sum of the proportion of each mineral multiplied by its respective partition coefficient. For example, if a melt is crystallising 20% olivine and 80% clinopyroxene and the partition for the element Y is 0.01 for olivine (ol) and 0.5 for clinopyroxene (cpx) then this mineral assemblage would have the following bulk partition coefficient:

$$D_{bulk} = 0.2D_{ol} + 0.8D_{cpx}$$

$$D_{bulk} = (0.2 \times 0.01) + (0.8 \times 0.5) = 0.002 + 0.4$$

$$D_{bulk} = 0.402$$

This can be written out in a more general form, such that if a mineral assemblage consists of phases A, B, C, D, … in mass proportions X_A, X_B, X_C, X_D, …, where

$$X_A + X_B + X_C + X_D ... = 1 \quad (5.2)$$

and the minerals have partition coefficients D_A, D_B, D_C, D_D, etc., then the bulk partition coefficient is

$$D_{bulk} = X_A D_A + X_B D_B + X_C D_C + X_D D_D ... \quad (5.3)$$

Having defined how elements are partitioned between minerals and melts their behaviour during melting and crystallisation can be quantified (Box 5.1).

Box 5.1 Equations that describe trace-element behaviour during partial melting and fractional crystallisation

During melting or crystallisation, trace elements partition between the melt and minerals in a predictable way. Consider a source rock with a concentration of a given element C_0 that melts to produce a liquid with a concentration C_L and some residual solid C_S. If the melt fraction is F, then the concentrations of the trace element in the source, melt and residue are related:

$$C_0 = FC_L + (1 - F)C_S \quad (5.4)$$

This is known as a mass balance equation.

With some rearranging the partial melting equation can be derived (Equation 5.5):

$$C_L = \frac{C_0}{F + (1 - F)D} \quad (5.5)$$

This can sometimes be found written in the alternative form:

$$C_L = \frac{C_0}{D + (1 - D)F} \quad (5.6)$$

Equations 5.5 and 5.6 relate the liquid composition to that of the source, the partition coefficient of the element of interest and the melt fraction. Equation 5.6 shows the partial melting equation in its simplest form because it assumes that the minerals on the solidus enter the melt phase in the same proportion that they are found in the rock. This is known as **modal melting**. You already know from melting of the mantle that a lherzolite containing clinopyroxene melts to produce a basalt and leaves a residual harzburgite (with essentially no clinopyroxene). Therefore in the mantle clinopyroxene must enter the melt at a much higher proportion than it is found in the original rock. When the phases that enter the melt do so in a different proportion from that in the original rock it is known as **non-modal melting**. The equation that takes this effect into account is not presented here as it is more complex, and for most elements the two equations give similar results for the composition of the liquid.

Equation 5.6 assumes that the melt does not separate from the mantle until melting is complete. In nature this is unlikely to happen, and melts can efficiently segregate from the mantle at porosities of less than 1% melt. These small melts eventually aggregate and form the final erupted lava. Again, the equations are not described here, but the composition of all of the aggregated small melt fractions is similar in composition to the liquid composition derived in Equation 5.6. However, the residual mantle is more affected by the differences in the melting model.

Equilibrium crystallisation can be described by Equations 5.5 or 5.6. In this case F represents the amount of liquid remaining, so the amount of crystallisation is $(1 - F)$.

Fractional crystallisation is described by a different equation. This assumes that each infinitesimal small crystal that is in equilibrium with the melt is then separated from the melt. This yields an equation containing an exponent:

$$C_L = C_0 F^{(D-1)} \quad (5.7)$$

Again, F represents the amount of liquid remaining, so the amount of crystallisation is $(1 - F)$.

Having defined the basic equations, it is useful to plot some simple results using these equations to illustrate how they control the concentrations of trace elements as a function of F and D. Figure 5.21 is a plot of the concentration of an element in a liquid relative to the starting composition. This ratio simply helps us understand whether a liquid is depleted or enriched in an element relative to the original composition. Curves are plotted for various bulk partition coefficients. For an element with $D = 1$, the melt does not change with degree of melting (changing F). For incompatible elements the initial melt has a very high concentration and this concentration decreases as partial melting increases (increasing F). The concentration is the highest for the most incompatible element and approaches $\frac{1}{F}$ for highly incompatible elements. For compatible elements the concentration of an element is always lower than the starting

composition. At low degrees of melting the concentration is lowest and it increases as the rock progressively melts. The concentrations are lowest for elements with the highest D values.

■ What is the concentration of Y in a melt if the mantle has 4 ppm Y, the bulk partition coefficient is 0.3 and 5% of the mantle has melted?

■ $C_0 = 4$, $D = 0.3$ and $F = 0.05$ therefore using Equation 5.6:

$$C_L = \frac{4}{0.3 + (1 - 0.3) \times 0.05} = 11.94 = 12 \text{ ppm to two significant figures.}$$

For crystallisation, Figure 5.21 is read in reverse, starting with 100% melt at $F = 1$. With increasing crystallisation (decreasing F) the concentrations of the incompatible elements increase and the concentrations of the compatible elements decrease in the liquid:

- for incompatible elements the increase in concentration is greatest for the elements with the lowest D value
- for compatible elements the decrease in concentration is greatest for elements with the highest D value.

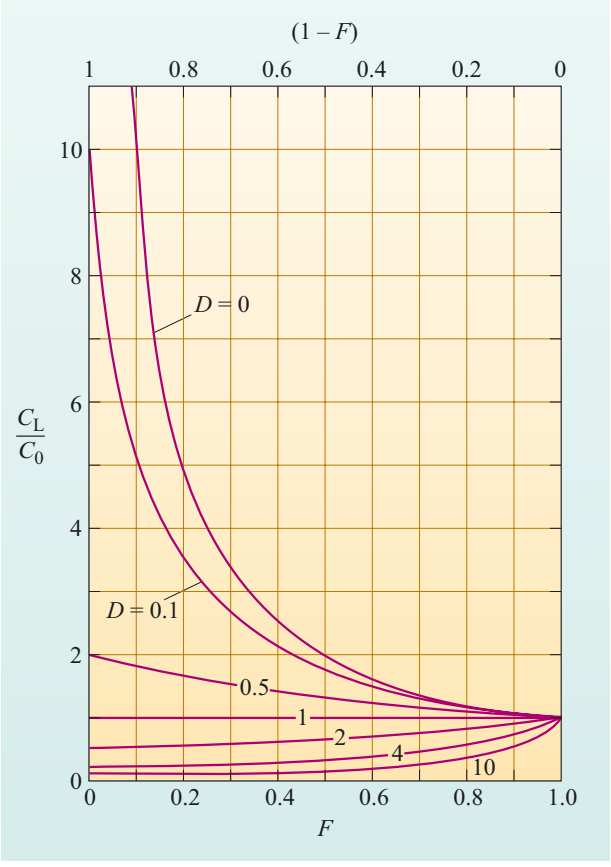


Figure 5.21 Plot of the ratio of the concentration in a liquid relative to the starting composition (C_L/C_0), as a function of the amount of melting F and the amount of crystallisation ($1 - F$), for different values of the partition coefficient D using the partial melting Equations 5.5 or 5.6. This diagram can also be used for equilibrium crystallisation. (Adapted from Cox et al., 1979)

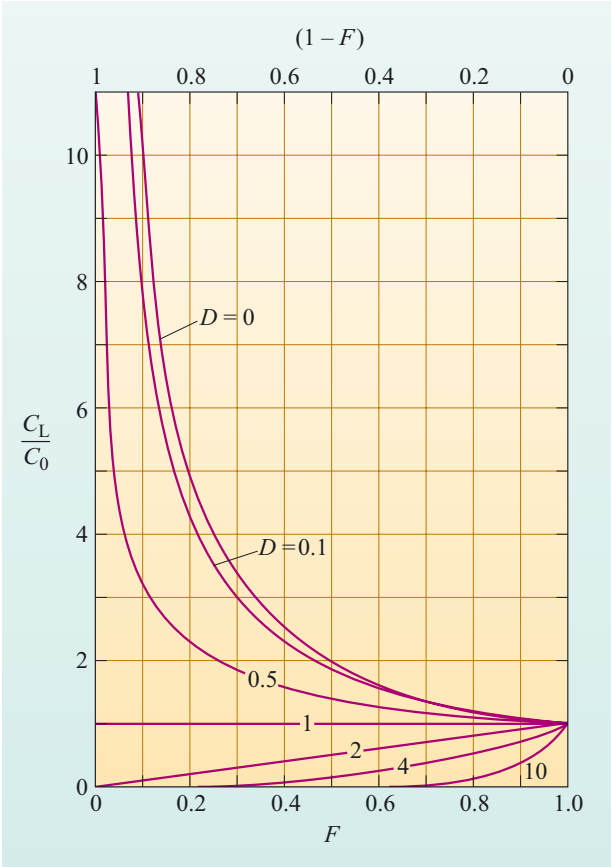


Figure 5.22 Plot of the ratio of the concentration in a liquid relative to the starting composition (C_L/C_0) as a function of F and the amount of crystallisation ($1 - F$), for different values of the partition coefficient D using the fractional crystallisation Equation 5.7. (Adapted from Cox et al., 1979)

- Sr has a partition coefficient of 4.4 in plagioclase. If plagioclase were the only crystallising phase, what would happen to the Sr concentration in the melt as plagioclase crystallised?
- As the melt crystallises, the Sr concentration will drop because it is concentrated into the plagioclase crystals.

Results using the fractionation crystallisation equation are illustrated in Figure 5.22. The curves are similar to those from the equilibrium crystallisation equation but the effects on concentration are more dramatic. Incompatible elements are more enriched and compatible elements are more depleted in the melts for a given F and D than for equilibrium crystallisation.

Question 5.1

Assuming the mantle contains 60% olivine, 20% orthopyroxene, 10% clinopyroxene and 10% garnet, what would be the bulk partition coefficient for Rb during partial melting of the mantle? Use the partition coefficients provided in Table 5.2.

Question 5.2

Calculate the concentration of Y in a melt, 80% of which has crystallised, for equilibrium and fractional crystallisation. Assume the original melt had 25 ppm Y and the bulk partition coefficient is 0.402.

5.8 The trace-element composition of island-arc magmas

In comparing the composition of **island-arc basalts (IABs)** with MORBs it has been established that primitive lavas from the two tectonic settings have similar major element compositions. One key geochemical contrast is that IABs contain higher H_2O contents, due to partial melting of the hydrous mantle wedge, and the H_2O is ultimately derived from the subducting slab. Given that the slab comprises altered oceanic crust and hydrated depleted peridotite, and may carry an array of exotic sediments into the mantle, it would be surprising if there were not more differences between IABs and MORBs.

Trace-element data are frequently presented in graphical form normalised to a reference composition. This is usually an estimate of the composition of either the Earth's primitive mantle or chondritic meteorites. Elements are ordered from left to right in terms of decreasing incompatibility during partial melting of the mantle. Thus, if a basalt is simply derived by partial melting of the mantle with no added components, then the resultant trace-element pattern will be a smooth curve. Figure 5.23 is a primitive mantle-normalised plot of a representative arc basalt and MORB that have roughly the same MgO contents.

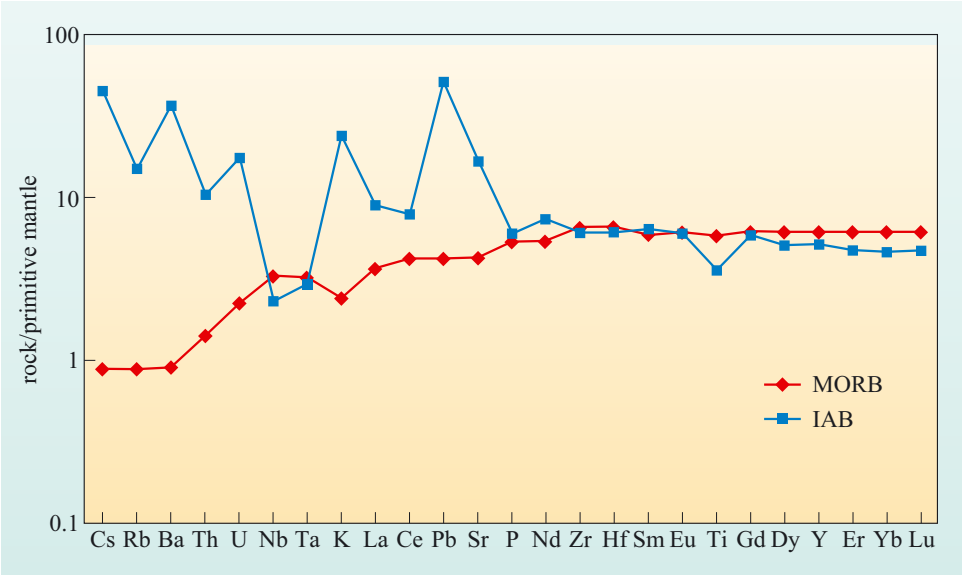


Figure 5.23 Primitive mantle-normalised multi-element patterns of a MORB and IAB with similar MgO contents. Element names and symbols are given in the Appendix.

■ Describe the differences between the mantle-normalised trace-element plots in Figure 5.23.

■ There are obviously big differences between the two trace-element patterns. The MORB pattern is very smooth with a steady decrease in the highly incompatible elements to the left of the figure.

This is because MORB represents moderate degrees of partial melting of a mantle that is slightly depleted in highly incompatible elements. The IAB plot has a very spiky pattern. The right-hand side of the pattern is broadly similar is shape to the MORB, although the IAB has slightly lower heavy rare earth elements (HREE) (Dy–Lu) contents. The IAB has lower Zr and Hf contents and large enrichments in K, Pb, Sr, U, Th, Ba, Rb and Cs and smaller enrichments of La and Ce. Of the highly incompatible elements to the left of the diagram, only Nb and Ta are depleted relative to MORB.

By choosing to compare two lavas with the same MgO content, the difference in the trace-element patterns is unlikely to result from differing degrees of crystallisation in a magma chamber. The elements to the right of the pattern such as the HREE and the element Y vary due to the amount of partial melting.

■ Y has bulk partition value of 0.3; what does the lower Y content of the IAB relative to MORB mean in terms of partial melting?

■ The partial melting equation dictates that the concentration of an incompatible element decreases with increasing degree of partial melting. Therefore, a first-order interpretation of the right-hand side of the normalised trace-element patterns is that IAB lavas are produced by higher degrees of partial melting. Alternatively, the source region of the IAB may have a lower concentration of Y than that of the MORB.

■ Th has a bulk partition coefficient of 0.0001; are the Th concentrations in the IAB and MORB consistent with the partial melting interpretation put forward above?

- Clearly, the Th concentration is much higher in the IAB than in the MORB, as are most of the other highly incompatible elements. This is not consistent with IABs being generated by higher degrees of partial melting than MORB, if they come from the same mantle source.

To explain the difference between the two trace-element patterns, there must be a difference in the composition of the material that is melted to produce an IAB. To understand a bit more about how to interpret these diagrams you need to know about the geochemical behaviour of various elements. Trace elements can be subdivided into those that are controlled by partitioning between a melt and a mineral and those that may also be controlled by partitioning between an aqueous fluid and a mineral. The elevated H_2O content of an IAB means that it is important to understand how some elements are partitioned into aqueous fluids. Elements with a low valency and large ionic radii tend to be soluble in aqueous fluids and are commonly known as **large ion lithophile elements (LILEs)**. Cs^{+}, Rb^{+}, K^{+}, Ba^{2+} and Sr^{2+} are all LILEs and have accordingly high solubilities. U^{6+} also is very soluble compared with U^{4+}. Many other higher valency ions, such as Y^{3+}, REE^{3+}, Th^{4+}, Zr^{4+}, Hf^{4+}, Ti^{4+}, Nb^{5+} and Ta^{5+}, all have smaller ionic radii and are insoluble in aqueous solutions.

- Do these differences in trace-element behaviour correspond to selective enrichments in the IAB relative to a MORB?
- In part, yes. Elements like Cs, Rb, Ba, Sr and U are highly enriched relative to their concentrations in MORBs, and this may be due to the transfer of these elements from the slab in an aqueous fluid.

The observations that some elements are added to the mantle wedge by an aqueous fluid is, not surprising, because IABs have elevated H_2O contents and it has already been demonstrated that dehydration reactions in the subducting slab release H_2O. However, this does not explain whether the fluid mobile elements come from the sediment or the oceanic crust, or why IABs are enriched in elements that are not fluid mobile, such as Th. To help understand this dilemma, the composition of an average subducted sediment (called global subducted sediment (GLOSS)) and the composition of an altered MORB are plotted on a primitive mantle-normalised diagram (Figure 5.24).

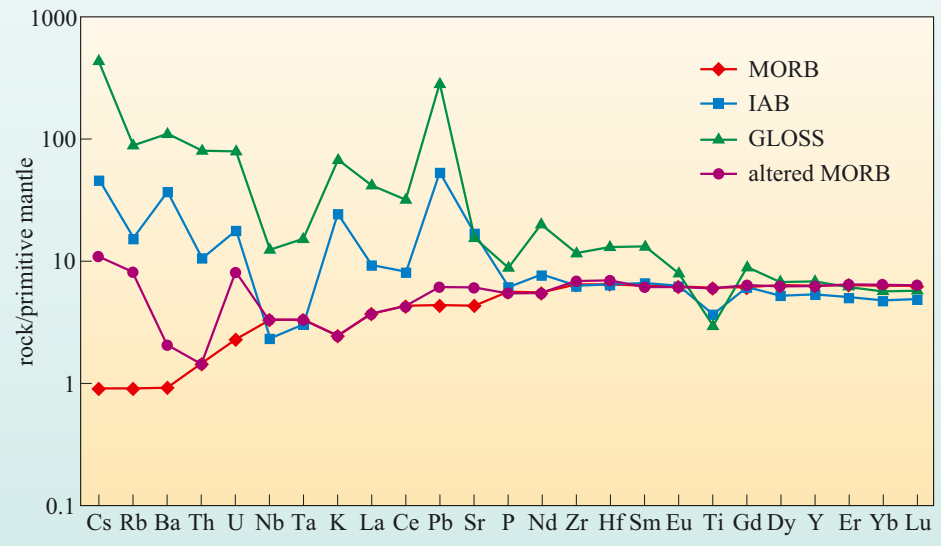


Figure 5.24 Primitive mantle-normalised multi-element patterns of GLOSS and an altered and fresh MORB.

- What compositional similarities exist between the GLOSS sediment, altered oceanic basalt, and IAB relative to fresh MORB?
- There are several clear similarities between the three patterns. In detail, GLOSS is enriched in all of the LILEs, as well as La, Ce, Pb and Th. It also has a marked depletion in Nb and Ta. By contrast, only the elements Cs, Rb, K and U are consistently enriched in the altered oceanic crust, whereas Ba, Pb and Sr are only slightly enriched relative to the fresh MORB.

- What process could have enriched IAB in the element Th which is insoluble in fluids?
- Th could be added in a melt derived from subducted sediment.

In principle, it is possible that sediment is physically removed from the slab and mixed into the mantle wedge. In practice, there is no real evidence for this happening, as the slab seems to remain relatively intact. It is more likely that the sediment partially melts, and this melt which is highly concentrated in these elements, is added to the mantle wedge. Simple mass balance arguments mean that only a small fraction of sediment melt is needed to enrich the mantle (about 0.2% sediment melt to 99.8% mantle). This explains why IABs can still have very low Nb and Ta concentrations, whilst being enriched in LILE.

The above discussion shows that there are some qualitative similarities between the material being subducted and the trace elements enriched in IAB lavas and in the case of elements that are not mobile in hydrous fluids there is a strong case for melted sediments. However, for the fluid-mobile LILEs, such as Rb and U that are enriched in both altered basalt and sediment, this proposal does not resolve whether these elements are added from either or both of the altered crust or sediments. This issue is the focus of the final section in Chapter 5.

5.9 Subduction zones and recycling

The trace-element patterns of subduction-related magmas are the product of (i) the different components that contribute to the magma source region, (ii) the processes that transport the components into the magma source and (iii) melt generation.

- What four source materials may contribute to the generaton of IAB?
- The mantle wedge; subducted sediments; subducted, altered, basaltic oceanic crust and the subducted, altered, depleted mantle peridotite.

- What processes can modify the source components?
- Melting and dehydration of both the subducted sediment and the basaltic slab could occur. Melting of the mantle wedge is probably the dominant process, along with dehydration of the altered subducted mantle. These processes are not mutually exclusive, and whether any, some or all happen depends on the thermal structure of the subduction zone.

As you have already discovered, the behaviour of an element during subduction processing depends on its geochemical properties, and in particular whether it is mobile in fluids or not and in how it partitions into a melt. When considering recycling it is also critical to assess how much of an element enters a subduction zone (its input flux) and how much leaves during the production of subduction zone magmas. Of the inputs to a subduction zone, altered oceanic crust probably exhibits the least compositional variation. By contrast, sediment input is highly variable both in amount and in terms of its composition, so it is important to assess the amount of sediment input into different subduction zones when considering the evolution and formation of IAB.

5.9.1 Sediment input

Sediments are a very important source for elements that can be subducted and these elements often originate in the continental crust. But can we be sure that sediments actually do get into the mantle and are incorporated into subduction-related magmatism? One of the more dramatic and least ambiguous results in this context has been the observation that the short-lived radioactive isotope ^{10}Be is present in amounts significantly above background levels in young lavas from island arcs (Box 5.2). The only way ^{10}Be can become incorporated in these lavas is if it has been transported into the mantle via subduction, released from the slab into the mantle wedge and then incorporated in the parental arc magma prior to eruption. Moreover, given that ^{10}Be has a half-life of 1.5 Ma, the complete cycle must be completed within about 7.5 Ma (Box 5.2)

Box 5.2 ^{10}Be: the smoking gun for sediment subduction

^{10}Be is an isotope of the element Beryllium, which is a Group 2 metal. ^{10}Be is produced continuously in the upper atmosphere by a process called **spallation**. This involves the bombardment of the nuclei of O and N atoms by cosmic rays. Hence, ^{10}Be is known as a **cosmogenic** isotope. ^{10}Be is radioactive and decays to ^{10}B with a half-life of 1.5 Ma. Be is carried by rainfall from the atmosphere on to the Earth's surface and into the oceans, where it becomes incorporated into pelagic sediments on the ocean floor. If these sediments are subducted, then ^{10}Be produced in the upper atmosphere can be transported into the upper mantle. If ^{10}Be is then transferred from pelagic sediments into the source region of subduction-zone magmas and is detected in erupted lavas, it provides an unambiguous tracer of element recycling in subduction zones.

As you might imagine, the amount of ^{10}Be produced in the atmosphere is not great, and with a half-life of 1.5 Ma the chances of finding measurable amounts of ^{10}Be in arc magmas might appear remote. The abundance of ^{10}Be in pelagic sediment is measured in units of atoms per gram, and there are about 5×10^9 atoms per gram of ^{10}Be in an average pelagic sediment. This is a tiny amount, but given that modern accelerator mass spectrometers can measure as low as ~10^6 atoms per gram of ^{10}Be, this is still about 1000 times more than the detection limit. ^{10}Be starts to decay effectively as soon as sediment is subducted, and after 10 half-lives (15 Ma) only $\frac{1}{1024}$ of the original parent isotope will be left making its concentration too low to measure precisely and be of any scientific use.

Question 5.3

If a plate is subducted at 60 mm y^{-1} and carries down pelagic sediment with recently acquired ^{10}Be, how long would it take to reach a depth of 100 km assuming the subduction zone dips at 45°? How long would it take to reach 100 km if the plate was only subducted at only 10 mm y^{-1}? What consequences does the difference in subduction rate have on whether a ^{10}Be signature can be detected in the arc lavas?

Figure 5.25 is a plot of ^{10}Be in MORBs, ocean island basalts (OIBs) and selected arcs. Both MORBs and OIBs have ~10^6 atoms per gram of ^{10}Be, which is indistinguishable from laboratory blanks (controls) that are nominally ^{10}Be free. By contrast, arc lavas can have up to (15–20) × 10^6 atoms per gram of ^{10}Be, with the Central American arc having the highest recorded ^{10}Be contents. The subduction rate along the Central American Trench is 80 mm y^{-1}, and so the most likely way in which the ^{10}Be was incorporated into these lavas is by recycling of ^{10}Be from subducted sediments.

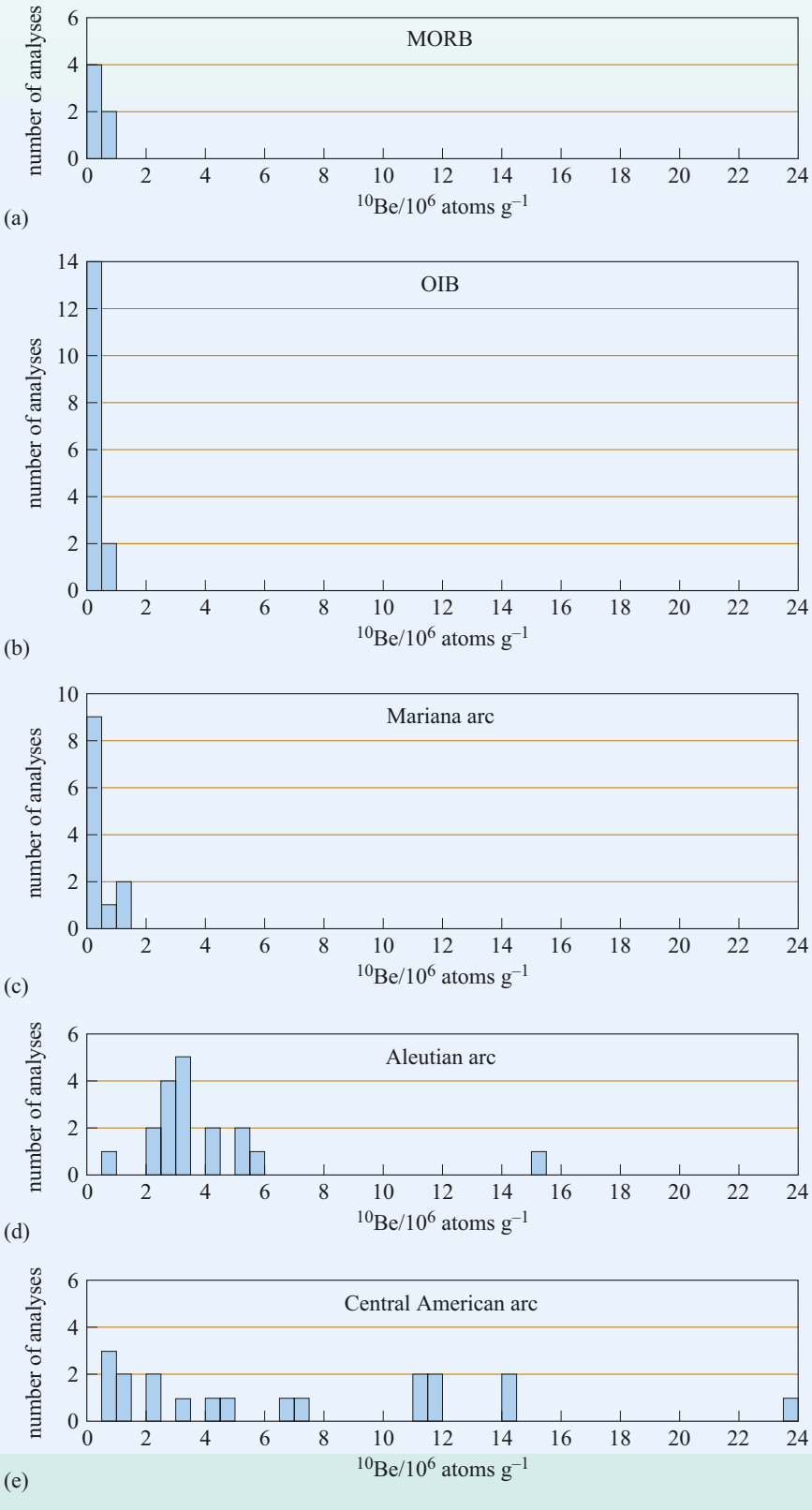


Figure 5.25 Histograms of ^{10}Be contents from (a) MORB; (b) OIB; (c) Mariana arc; (d) Aleutian arc and (e) Central American arc.

Despite the undoubted success of ^{10}Be for detecting some subducting sediments, not all sediments have the same composition, a similar number of arcs do not have a ^{10}Be signature, and in order to understand the link between ocean-floor sediments and subduction it is critical to understand the composition of oceanic sediments in more detail. Figure 5.26 is map of the distribution of oceanic sediment types on the ocean floor. The composition of the sediment varies depending on differences in ocean productivity, which primarily relate to latitude, and to ocean depth, which controls carbonate stability. The proximity to oceanic ridges and continents also influences the distribution of metalliferous sediments (at ridges) and **terriginuous input** (from the continents). Close to the trench some volcanic-derived material (**volcaniclastic sediment**) from the volcanic arc may also be subducted. Therefore, as a plate moves away from a spreading ridge, the sediment on the oceanic crust will get thicker and its composition will vary depending on the position of the plate on the Earth's surface. Figure 5.26 is a useful guide to the types of sediment being deposited at a given point on the Earth's surface, but a more representative assessment of the bulk composition of sediments being subducted can be gained by analysing cores through the sediment section close to the trench. Figure 5.27 illustrates simplified core sections from eight subduction zones where deep-sea drilling has sampled the sediments in enough detail to provide a reasonable estimate of the average sediment type being subducted. There are significant variations between trenches. For example, sediment entering the Guatemalan Trench is dominated by carbonate ooze, whereas the sediment being subducted beneath Vanuatu is mainly volcaniclastic.

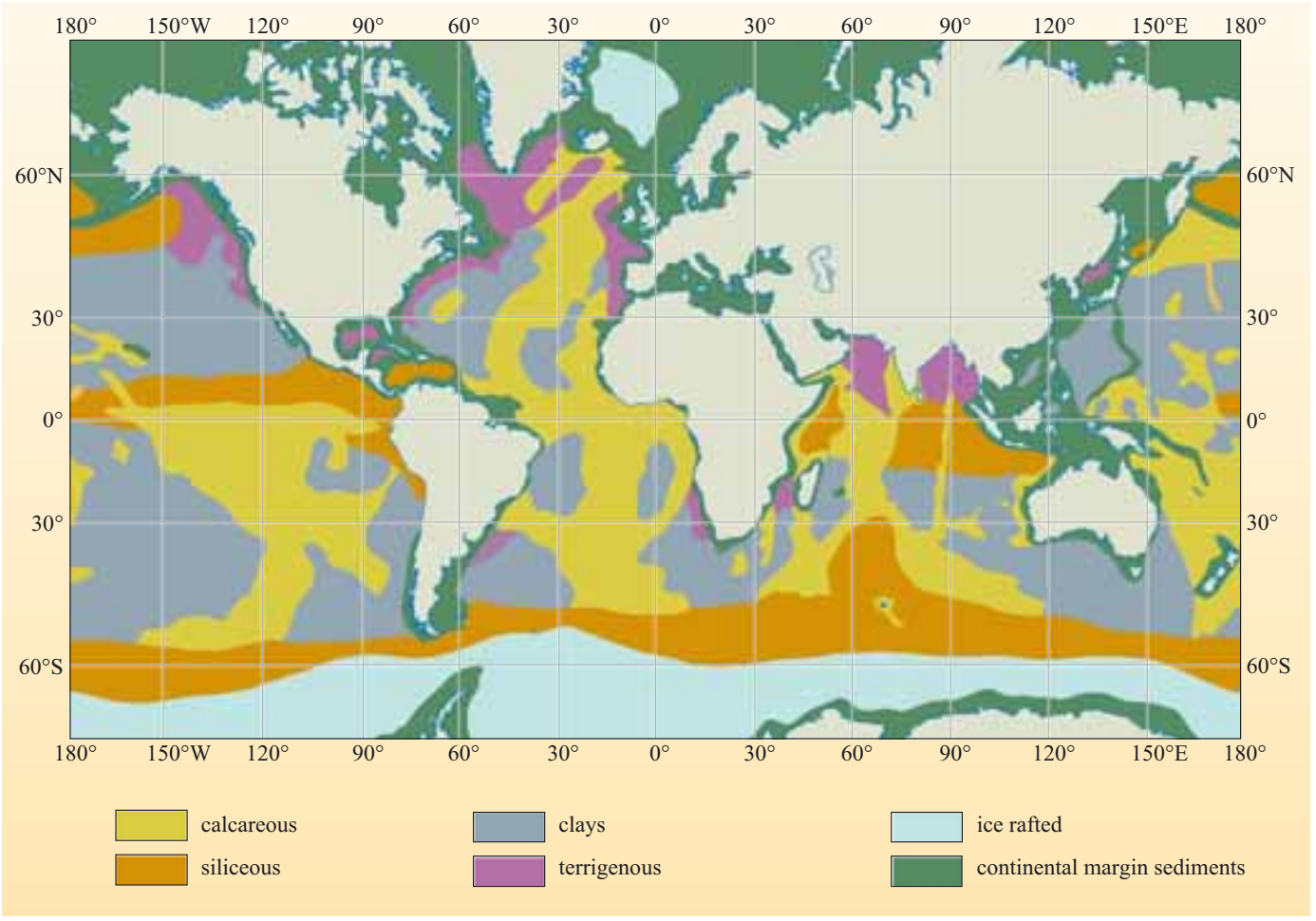


Figure 5.26 Map of the global distribution of sediment types currently being deposited on the ocean floor.

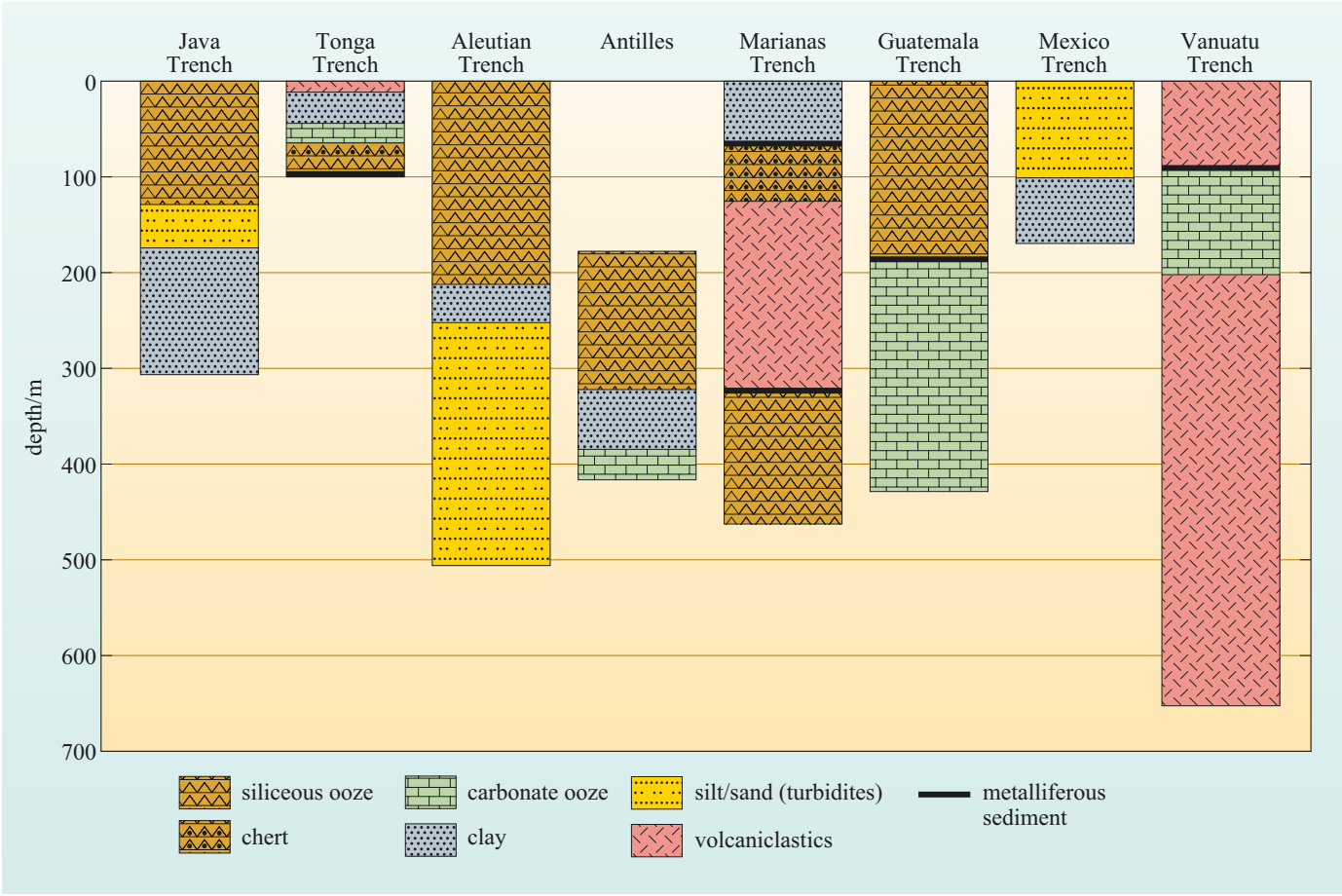


Figure 5.27 Average sedimentary columns being subducted at eight trenches. (Plank and Langmuir, 1993)

Two US geochemists, Terry Plank and Charlie Langmuir, analysed the sediment sequences and calculated an input for each element by taking the average chemical composition and multiplying it by the convergence rate and the mass of sediment. This yields an input for any given element in terms of grams per year per centimetre of arc length, which can then be compared with the composition of the lavas erupting in the adjacent arc.

Estimating the amount of a given element produced by arc magmatism for each subduction zones is rather tricky, not least because there are many submarine volcanoes that are not accessible and only the magma that is erupted can be easily sampled. So, rather than try to calculate the total input of an element it is easier to use a **proxy**. A proxy, in this case, is a geochemical indicator that varies in direct response to variations in the input. Having already established that some elements, such as Sr, Ba, Th, and U, are enriched in arc magmas relative to MORBs, the challenge is to quantify this enrichment and see whether the amount of enrichment in a given arc corresponds to the input of that element into the trench. However, a simple measure of the element concentration in a lava from a subduction zone is not directly related to its input.

■ What general processes affect the concentration of elements such as Ba in lavas?

■ Fractional crystallisation and partial melting.

Fractional crystallisation tends to increase the concentration of incompatible elements in magmatic rocks, while increasing degrees of partial melting decrease their concentration. Ideally, the effects of fractional crystallisation should be minimised by restricting attention to primary basaltic magmas, but as discussed earlier, these are generally rare in arc settings. So, to compare arcs on an equal basis, the effects of both fractional crystallisation and partial melting need to be evened out. Plank and Langmuir achieved this by taking data from a number of volcanoes from each arc, and for each volcano they plotted the variations of selected trace elements against MgO as an index of fractionation. The data define a linear or curved trend and the value of Ba at 6% MgO was used as a proxy for the composition of the primary melt. So, for Ba the corrected value becomes $Ba_{6.0}$. Plank and Langmuir also found that the Na_2O content at 6 wt% MgO provides a useful indicator of the degree of partial melting because Na is an incompatible element. High $Na_{6.0}$ contents indicate low degrees of partial melting, whereas low $Na_{6.0}$ contents indicate high degrees of partial melting. Therefore, dividing the trace-element value (e.g. $Ba_{6.0}$) by the sodium value at 6 wt% MgO (i.e. $Ba_{6.0}/Na_{6.0}$) gives a ratio that is a measure of the enrichment of a given element in the mantle source of the subduction zone melt. Figure 5.28 is a plot of input flux of Ba and Th versus their respected output flux ($Ba_{6.0}/Na_{6.0}$ and $Th_{6.0}/Na_{6.0}$) for seven different arcs. As you can see, there is a good correlation between the input sediment flux and the geochemical proxy for the output flux.

■ What does Figure 5.28 tell us about the relationship between the sediment subducted and the composition of arc lavas?

■ There is a direct relationship between the composition of the sediment subducted at a trench and the composition of the arc lavas erupted in the arc adjacent to the same trench. Therefore at least some of the sediment subducted must be recycled back into the source region of island-arc magmas.

Other elements, such as Sr, Rb, K and U, also reveal positive correlations between sediment input flux and (element)/$Na_{6.0}$ ratios. This partly answers a question raised earlier, which was whether these fluid mobile elements come from sediment or altered oceanic crust. The correlations suggest that they mainly come from sediment although they do not rule out an input from dehydration of the basaltic slab.

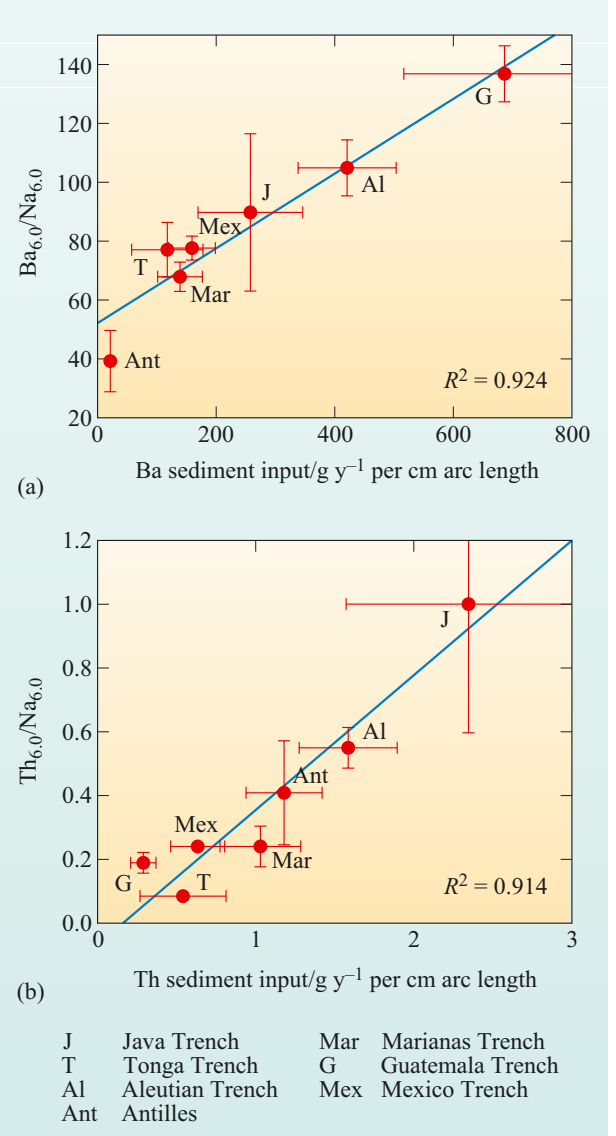


Figure 5.28 Correlations between trace element sediment inputs (in grams per year per centimetre arc length) and trace-element enrichment in arc basalts for: (a) Ba, and (b) Th. R^2 is the correlation coefficient between the two variables where a value > 0.9 indicates a significant correlation. (Plank and Langmuir, 1993)

Central American arc lavas have some of the most unambiguous evidence for sediment subduction and incorporation into the source region of arc magmas. They have high ^{10}Be signals and, because the sediment being subducted into the Central American trench is very distinct in composition, the trace-element composition of the sediments is recorded in the lava composition. For example, the sediment has a high Ba/Th ratio (Figure 5.28), and this high Ba/Th signature is found in the erupted lavas. Lavas from Nicaragua provide one last illustration of the link between sediment composition and volcanism in subduction zones. The uplift of the Panama peninsula cut off the flow of tropical waters from the Atlantic to Pacific Oceans 10 Ma ago. This dramatically changed the chemistry of the seawater in the Panama Basin, while the sediments being deposited in the basin changed from carbonate to siliceous ooze. More importantly, there was a factor of four increase in the U content of the sediment 10 Ma ago while the Th and Ba contents of the sediments do not change by more than 10%. Remarkably, this change in U/Th ratio is recorded in the composition of the lavas erupted in the Nicaraguan arc, in which lavas erupted after 7.5 Ma have elevated U/Th ratios compared with those erupted previously.

Question 5.4

What do these elevated U/Th ratios suggest about the Central American subduction zone? Is it consistent with other information about the Central American arc.

Understanding the details of element recycling through subduction zones is complicated, but by combining trace element and isotopic data there is strong evidence that sediments provide a significant flux of highly incompatible elements to most arc lavas. The oceanic crust may also contribute to the U and Sr budget (and H_2O is recycled from all parts of the subducted slab). Calculating absolute fluxes of elements is much more difficult, but, in the example from Central America given above, at least 75% of the sediment at the trench has to be subducted to balance the Th budget of the Nicaraguan lavas. Similarly, it has been calculated that all of the Be that is subducted is returned back to the crust by magmatism, illustrating that some element recycling is highly efficient within the upper 200 km of the mantle.

Summary of Chapter 5

Subduction zones are the major sites of recycling of surface material back into the mantle. The most dramatic effects of subduction are two of the Earth's major natural phenomena, namely earthquakes (particularly deep earthquakes) and explosive volcanism. H_2O is the key component in subduction zones because it influences chemical reactions in the subducting slab, the transport of elements from the slab, the viscosity of the overlying mantle wedge and the chemical evolution of arc magmas. This chapter has considered the recycling of material through the upper 200 km of the mantle; but, clearly, subducting slabs reach and penetrate the 670 km discontinuity, and many scientists think the slabs sink to the core–mantle boundary. Although some elements (Be, Th) are efficiently transported from the slab back into the mantle wedge, the extraction process, be it melting or dehydration, is never 100% efficient. Consequently, many elements are returned to the deep mantle, including some subducted H_2O. Chapter 7 explores what happens to the subducted slabs over geological time and how this influences the large-scale chemical and mechanical evolution of the mantle.

Learning outcomes for Chapter 5

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

5.1 The physical, internal and thermal structure of subduction zones.

5.2 How changes in the thermal structure of subduction zones influence location and physical and geochemical processes involved in the generation of melts.

5.3 How fractional crystallisation processes can account for the range in major element compositions, and the mineralogical and geochemical evolution of arc magmas.

5.4 How distribution (partition) coefficients of trace elements and mineral assemblages can be used to investigate the geochemical evolution of arc magmas.

5.5 The recycling of sediments and atmospheric isotopes through subduction zones and how they can provide insight into the timescale of subduction as well as where and how fluids and melts form within this tectonic setting.

Chapter 6

Processes during continental collision

We live on an unusual planet. Underlying the ocean basins, which cover two-thirds of the Earth's surface area, is a dense crust, about 7 km thick, forming the outer layer of the lithosphere. But what makes the Earth's structure extraordinary is the remaining third of the Earth's surface area, where the outer layer is a silica-rich, low-density crust, with an average thickness of about 40 km, which supports the continents on which we live. Such continental crust is almost certainly unique within our Solar System because, as you know from previous chapters, it requires both the subduction of tectonic plates and the presence of water locked up within mineral structures of subducted surface rocks for continental crust to form.

The rocks that make up the continental crust have, on average, a density of 2800 kg m^{-3}, which compares with a density of 3000 kg m^{-3} for oceanic crust. This is because much of the continental crust is made up of granitic rocks together with sediments such as sandstones, mudstones and conglomerates, all of which have a low density, largely because of the abundance of quartz and feldspar. Indeed, the average composition of the continental crust is similar to that of an intermediate igneous rock, such as andesite, which is significantly less dense than that of basalt or gabbro. As a consequence, continental lithosphere is too buoyant to descend far at a subduction zone. Hence, continents cannot be destroyed, and since their formation is irreversible they must grow through geological time. This is why the oldest rocks known from oceanic floor are a mere 180 Ma old, whereas rocks almost 4000 Ma old have been recovered from regions of stable continental crust in Canada and Greenland.

The constant movement of the Earth's tectonic plates causes the expanding continents to be shunted around like froth on the surface of a pond. As they grow through time, it is inevitable that sooner or later there will be collisions between two continental masses. Descent of the ocean crust into subduction zones at the edge of an ocean basin (Figure 6.1a) draws the continents either side of the ocean closer together (Figure 6.1b), ultimately closing the ocean basin completely and leading to continental collision (Figure 6.1c and d).

The consequences of continental collision are far reaching. Neither continent can be subducted back into the mantle due to the buoyancy of continental crust, so the forces that drive the plate movement prior to collision are brought to bear directly on the continental lithosphere itself. At this stage, further convergence of the plates must be taken up by deforming one or both of the plates of continental lithosphere. Under such strong compressive forces, the lithosphere contracts through folding and faulting of the rocks that make up the continental crust. In the upper part of the crust, the sedimentary sequences along the original continental margins are thickened, as sheets of rock are thrust under one another along low-angle faults (Figure 6.1d). The lower crust, being hotter, will deform in a more ductile fashion. Here, the geothermal gradient will steepen due to the effects of thick sequences of continental rocks with high concentrations of the heat-producing elements. Lower crustal rocks will also be subject to high pressures. These changes in pressure and temperature will result in mineral

recrystallisation (metamorphism) beneath collision zones. If the temperature rises above the solidi for common crustal rocks, then the crust will start to melt, resulting in a drop in mechanical strength of the lower crust and allowing further ductile deformation and thickening.

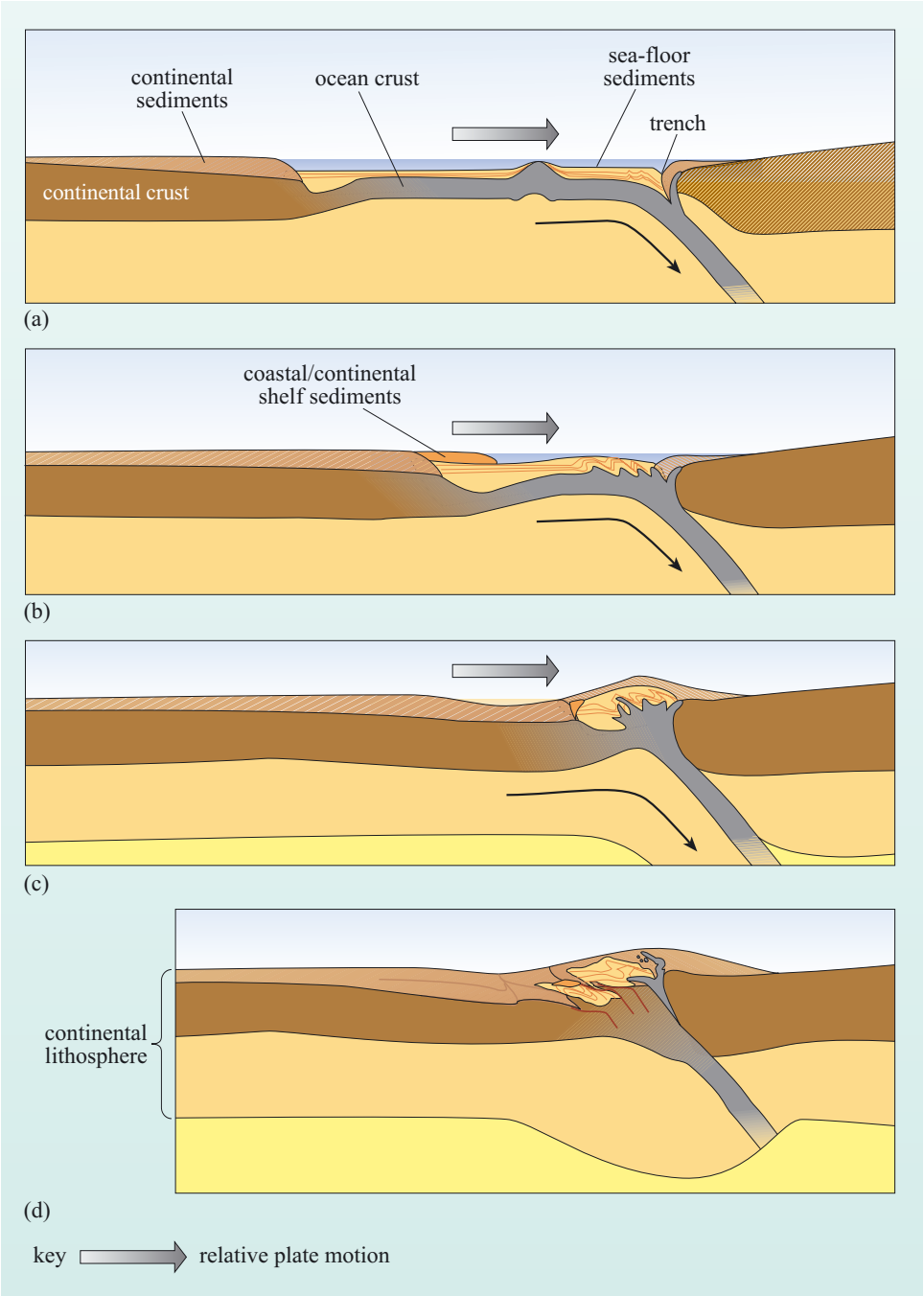


Figure 6.1 Schematic cross-sections showing the formation of a mountain range: (a) subduction at a destructive boundary (compare with Figure 4.3) causes (b) contraction of the ocean basin, which (c) leads to collision, and (d) thickening of the continental lithosphere.

- Bearing in mind what you read in Chapter 1 about the tendency for the lithosphere to 'float' in isostatic equilibrium, can you suggest what the consequences of thickening the crust would be?
- A thicker crust means a thicker lithosphere with a lower average density. Buoyancy forces in the mantle will cause such a lithosphere to rise, leading to land surface at high elevations.

The strong correlation between high surface elevation and a large crustal thickness (e.g. as seen in the case of the Andes or the Himalaya) is a clear indication of the effect of increased buoyancy forces on thickened crust.

This chapter will examine the tectonic processes that result from continental collision. In particular, it will look at the processes of metamorphism and partial melting that occur deep within the continental crust during continental collision as exemplified by the Himalayan orogeny.

6.1 Heating of the continental crust

All heat generated within the Earth results from radioactive decay. The principal isotopic systems responsible for internal heating stem from the decay of isotopes of three elements, K, U and Th, as was detailed in Chapter 2. All three of these elements are generally incompatible with respect to silicate mineral assemblages and so are concentrated in continental crust, which is initially made up of a high proportion of granitic rocks. Consequently, thick continental rocks sustain steep geothermal gradients, leading to metamorphism.

Metamorphism is a term used to describe the changes that affect existing rocks when they are subject to a change in pressure and/or temperature. It usually refers to chemical and physical reactions that take place in the *solid state*, though many metamorphic reactions are greatly assisted by the presence of any fluid that may be present along the grain boundaries or released by the reaction itself. Metamorphism covers a great range of processes, from subtle changes in the structure of clay minerals that occur in sediments after their burial, to the mineral reactions that occur deep in the crust at high pressures and temperatures. The field of metamorphism extends to the partial melting of rocks, so that, as you will see, there is overlap between the physical conditions under which igneous and metamorphic processes occur.

The simplest form of metamorphic reaction involves the change of a single mineral from one structure to another. If two or more minerals have the same composition, but contrasting structures, they are known as polymorphs. Carbon has two polymorphs, diamond and graphite, and their stability fields are plotted in Figure 6.2. Both polymorphs may coexist along the phase boundary that separates their stability fields.

Increasing the pressure of a rock will lead to reactions by which minerals of low density will be replaced by minerals of comparatively high density. Diamond is denser than graphite

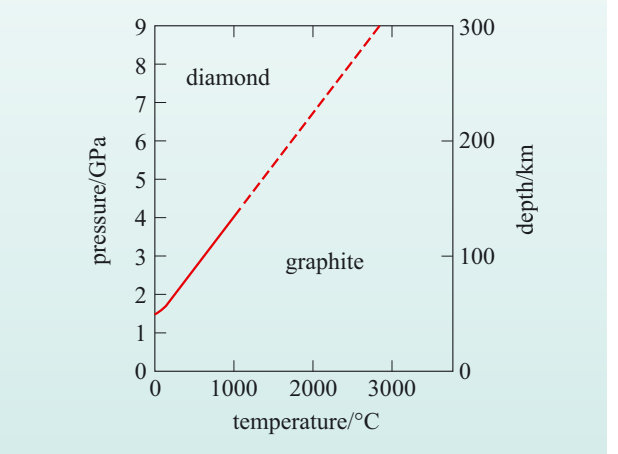


Figure 6.2 Stability fields of graphite and diamond in pressure–temperature space.

(a)

(b)

(c)

Figure 6.3 (a) Blue blades of kyanite growing within a quartz lens in mica schist (Glen Esk, Scotland). (b) Prismatic white andalusite growing within slate in the aureole of the Skiddaw Granite (Cumbria, England). (c) Lenses of fibrous white sillimanite within biotite gneiss (Langtang Valley, Nepal).

(3500 kg m^{-3} compared with 2000 kg m^{-3}), and so high pressures favour the graphite → diamond reaction. In other words, the transition from graphite to diamond results in a decrease in volume.

The graphite → diamond reaction is an example of a **geobarometer**, because the occurrence of either phase provides information on the pressure at which a rock has formed. For example, if graphite-bearing rocks are formed at a temperature of 1000 °C, then the maximum pressure the rock formed at was about 4 GPa, equivalent to a depth of about 120 km (Figure 6.2). It is not a very useful barometer though, because graphite is stable for all crustal pressures. Much more useful for petrologists who study crustal rocks are minerals that undergo metamorphic reactions under conditions commonly found within the crust. A particularly useful example is provided by three minerals that occur as minor components in many rocks that have aluminium-rich compositions: the **aluminosilicates**.

All three minerals are polymorphs with the same chemical formula, Al_2SiO_5, but their internal structures and external appearances are markedly different. The three polymorphs are called **kyanite**, **andalusite** and **sillimanite** (Figure 6.3). Each aluminosilicate is typical of a different range of pressure–temperature conditions, and their stability fields are shown in Figure 6.4. Two polymorphs can coexist along the phase boundaries that limit the stability field of each polymorph, and all three polymorphs can coexist at the point where the three phase boundaries converge. Such diagrams are the results of experiments performed in the laboratory, and the precise position of each boundary is subject to experimental uncertainty. This is particularly true for polymorphic reactions that have sluggish reaction rates.

- From Figure 6.4, estimate the relative densities of the three aluminosilicate minerals.
- For each phase boundary, the polymorph on the high-pressure side has the highest density. Thus, the order of decreasing density is kyanite, sillimanite, andalusite. (In fact, the densities are kyanite: 3600 kg m^{-3}, sillimanite: 3250 kg m^{-3} and andalusite: 3150 kg m^{-3}.)

The stability fields of the aluminosilicates (Figure 6.4) can be used to provide information of the physical conditions under which metamorphic rocks formed. For example, it is known the sillimanite-rich rock shown in Figure 6.3(c) could not have formed at temperatures less than about 500 °C (the minimum temperature for the sillimanite stability field). It is also known that the andalusite in Figure 6.3(b) could not have formed at

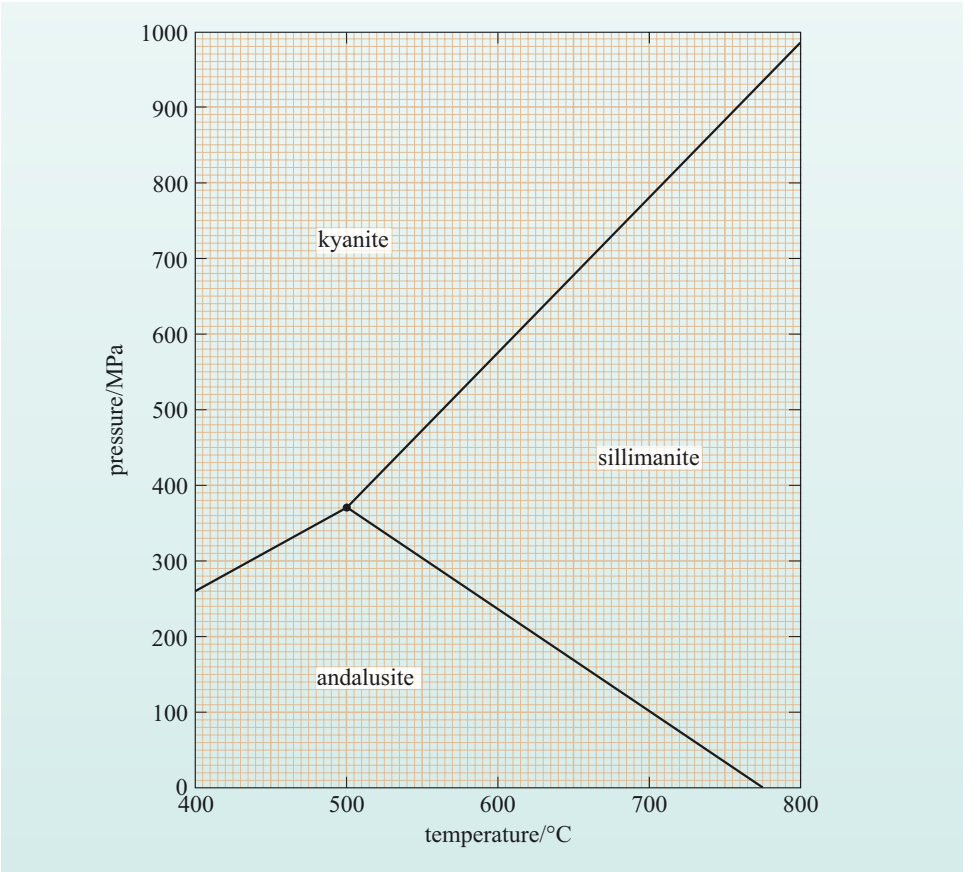


Figure 6.4 The stability fields of the three aluminosilicate minerals in pressure–temperature space. Note the pressure units are MPa. 1 GPa = 1000 MPa.

pressures greater than about 380 MPa (Figure 6.4). Since these andalusites are from baked sediments at the contact of a large granite in Cumbria, the granite must have been intruded at pressures of less than 380 MPa, equivalent to depths less than about 12 km. Thus, the aluminosilicates allow the pressure–temperature space of metamorphic rocks to be divided up into high-pressure (kyanite), low-pressure (andalusite) and high-temperature (sillimanite) regimes. Other minerals can be used in similar ways, and the conditions under which metamorphic rocks formed can sometimes be pinpointed quite accurately.

So far, this chapter has concentrated on the consequences of increasing pressure on the stability of single minerals and their polymorphs. In most natural rocks, several minerals take part in a metamorphic reaction in response to changing pressure *and* temperature. For example, if a limestone containing mostly calcite ($CaCO_3$) with minor quartz (SiO_2) is heated, a metamorphic reaction occurs:

$$\underset{\text{calcite}}{CaO_3} + \underset{\text{quartz}}{SiO_2} = \underset{\text{wollastonite}}{CaSiO_3} + \underset{\text{carbon dioxide}}{CO_2} \qquad (6.1)$$

In the metamorphosed limestone (known as a marble), small needles of the mineral wollastonite ($CaSiO_3$) are found along the grain boundaries of calcite and quartz (Figure 6.5).

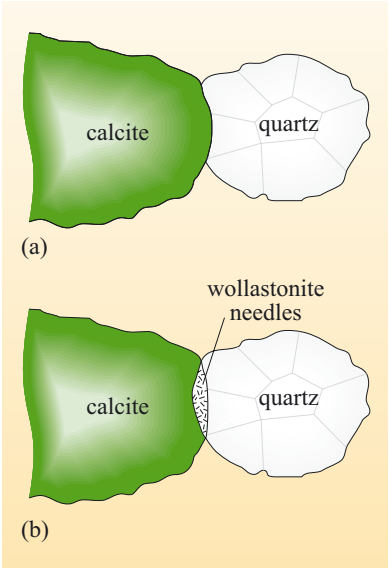


Figure 6.5 Sketch to show the formation of wollastonite ($CaSiO_3$) needles during metamorphism of limestone: (a) shows the minerals present before and (b) shows the mineral present after metamorphism.

The stability fields for calcite + quartz and for wollastonite + carbon dioxide (CO_2) are shown in Figure 6.6. Notice that the wollastonite + CO_2 assemblage is favoured by high temperatures.

Just as high pressures stabilise minerals with high densities, high temperatures stabilise minerals and gases with a high degree of disorder in their molecular structure. For polymorphs such as the aluminosilicates, the mineral with the more disordered state will always lie on the high-temperature side of the phase boundary. Thus, sillimanite has the most disordered molecular structure of the three Al_2SiO_5 polymorphs and kyanite has the lowest (Figure 6.4). Since the structures of gases are considerably more disordered than those of solids, any mineral reaction that releases a gas, such as given by Equation 6.1, will be triggered by an increase in temperature. In other words, the phase boundary for the reaction will be steep (at least at higher pressures), as we can see from Figure 6.6. The steep phase boundaries that describe reactions that produce CO_2 (**decarbonation reactions**) or H_2O (dehydration reactions) provide excellent **geothermometers**.

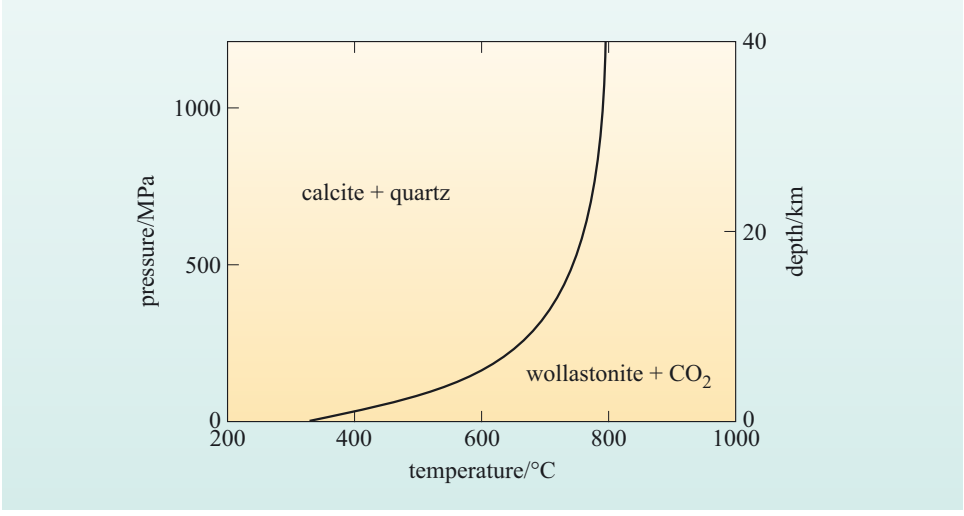


Figure 6.6 Boundary curve for the reaction calcite + quartz = wollastonite + CO_2.

- A limestone contains wollastonite, calcite and quartz. The adjacent rock is an aluminous sediment that contains kyanite. By considering both Figures 6.4 and 6.6, estimate the minimum pressure at which these rocks have formed.
- The minerals wollastonite calcite and quartz must lie on the boundary curve of the decarbonation reaction. This crosses the boundary between kyanite and sillimanite at a pressure of about 1000 MPa (Figure 6.4), which represents a minimum pressure for the assemblage.

Decarbonation reactions are particularly important for the Earth system as they ultimately release CO_2 into the atmosphere. Conditions for such reactions are primarily found within collision zones, and since CO_2 is a greenhouse gas, this is one example of continental collision contributing to climate change.

Different pressures and temperatures can change the minerals that are stable in a rock. It is also true that *different rock compositions will allow different minerals to form under the same pressure and temperature conditions*. Rocks that contain different aluminosilicates, for example, must have been rich in aluminium to form any aluminosilicates at all. Some compositions, like sandstone, made up of just quartz, are quite sterile in terms of mineral reactions because quartz is stable for all normal conditions encountered in the crust. In contrast, aluminium-rich sediments, known as mudstones, provide the raw materials for many mineral reactions. Such aluminous compositions are termed pelitic and their metamorphosed equivalents are referred to as **metapelites**. Mudstones are made up of clay minerals, which are prone to multiple changes during metamorphism on account of their chemical complexity. The following text explores the metamorphism of a mudstone under increasing pressure and temperature.

First, you need to consider the textural changes that are summarised in Table 6.1. The fine-grained clay minerals that make up a mudstone are deposited flat on the floor of the stream or lake. On compaction, the flakes are rotated to align at right angles to the compacting load, giving the rock a layered appearance. When this compaction is accompanied by heating, the clay minerals begin to recrystallise, and the process is promoted by hot, water-rich fluids being forced through the sediments caused by the compacting load of overlying sediments. In this way, new micaceous (mica-rich) minerals form. The micas will be either biotite, if iron rich, or muscovite, if aluminium rich. All micas are sheet-like or platy in shape and grow perpendicular to the direction of compression acting on the rock in which they grow. This stress is usually vertical due to the compacting load, but it may vary due to non-vertical stress induced by plate movements.

Slates are the result of a small increase in pressure and temperature acting on a mudstone. The minerals are still very fine grained but they are largely micaceous: the preferred orientation of the micas within the rock gives it a well-defined cleavage (exploited in the manufacture of roofing slates, for example). Further metamorphism of mudstones results in an increased grain size. Coarsely crystalline rocks with an aligned micaceous texture are known as **schists**. But micas are not the only minerals that form during metamorphism.

Table 6.1 Classification of metamorphosed sediments (metapelites).

Rock type	Appearance	Grain size
slate	very closely spaced, almost perfectly flat planes	fine
schist	moderately spaced, sub-parallel planes, characteristically with abundant mica	medium
gneiss	widely spaced layers with alternations between mica, amphibole or pyroxene-rich layers and quartz, and plagioclase-rich layers	coarse
migmatite	separation of granitic lenses within darker layers (usually of biotite) or amphibole	coarse

One of the aluminosilicate minerals can also result from a reaction involving micas. For example, a common dehydration reaction in schists is:

$$KAl_3Si_3O_{10}(OH)_2 + SiO_2 = KAlSi_3O_8 + Al_2SiO_5 + H_2O \quad (6.2)$$

muscovite — quartz — alkali feldspar — alumino silicate — water

Figure 6.7 Biotite gneiss from deformation of granite (from the Himalaya).

Higher temperatures and pressures will result in a segregation between quartz + feldspar and ferromagnesian minerals, such as biotite or amphibole. Since the former are light in colour and the latter are dark, this gives the rock a banded appearance. Such rock is known as a **gneiss** (which is pronounced 'nice'). Not only mudstones, but also other rock types such as granites and gabbros can form gneisses (Figure 6.7) when metamorphosed to sufficiently high temperatures and pressures. The dark layers may include micas, amphiboles or pyroxenes, depending on the degree of metamorphism and on the bulk composition of the rock.

It is under conditions close to the breakdown of muscovite that a gneiss may begin to melt. This is illustrated in Figure 6.8, where both the granite melting curve and the reaction boundary defined by Equation 6.2 are plotted. The granite melting curve shown is for melting in the presence of quartz, plagioclase, alkali feldspar and water. Note that when this reaction boundary crosses the 'wet' melting curve, muscovite and quartz combine with plagioclase (albite) to form a melt, facilitated by the water released from dehydrating the mica.

- If you follow the effects of increasing temperature on a typical metapelite at pressures of 500 MPa, what will happen when point X (Figure 6.8) is reached at about 650 °C?
- X lies on the water-present granite solidus, but no reaction will occur in the absence of H_2O. In order to form a granite magma at this point, the three minerals quartz, plagioclase, and alkali feldspar together with water must be present in the rock. Metamorphic rocks at temperatures of several hundred degrees Celsius rarely contain a free water phase and the melting of granite in the absence of water occurs at much higher temperatures. Also absent from many metamorphosed sediments is the mineral alkali feldspar.

- What will happen when point Y is reached on Figure 6.8?
- At point Y, the muscovite dehydration reaction is crossed. This will generate some melt, an aluminosilicate and alkali feldspar. Since quartz and plagioclase are common minerals in sediments, all four phases (plagioclase, alkali feldspar, quartz and H_2O) required for forming a granite minimum-melt composition are now present.

- Equation 6.2 also has Al_2SiO_5 as a product. Which aluminosilicate will be stable?
- From Figure 6.8, Y lies within the sillimanite field, and so sillimanite will form.

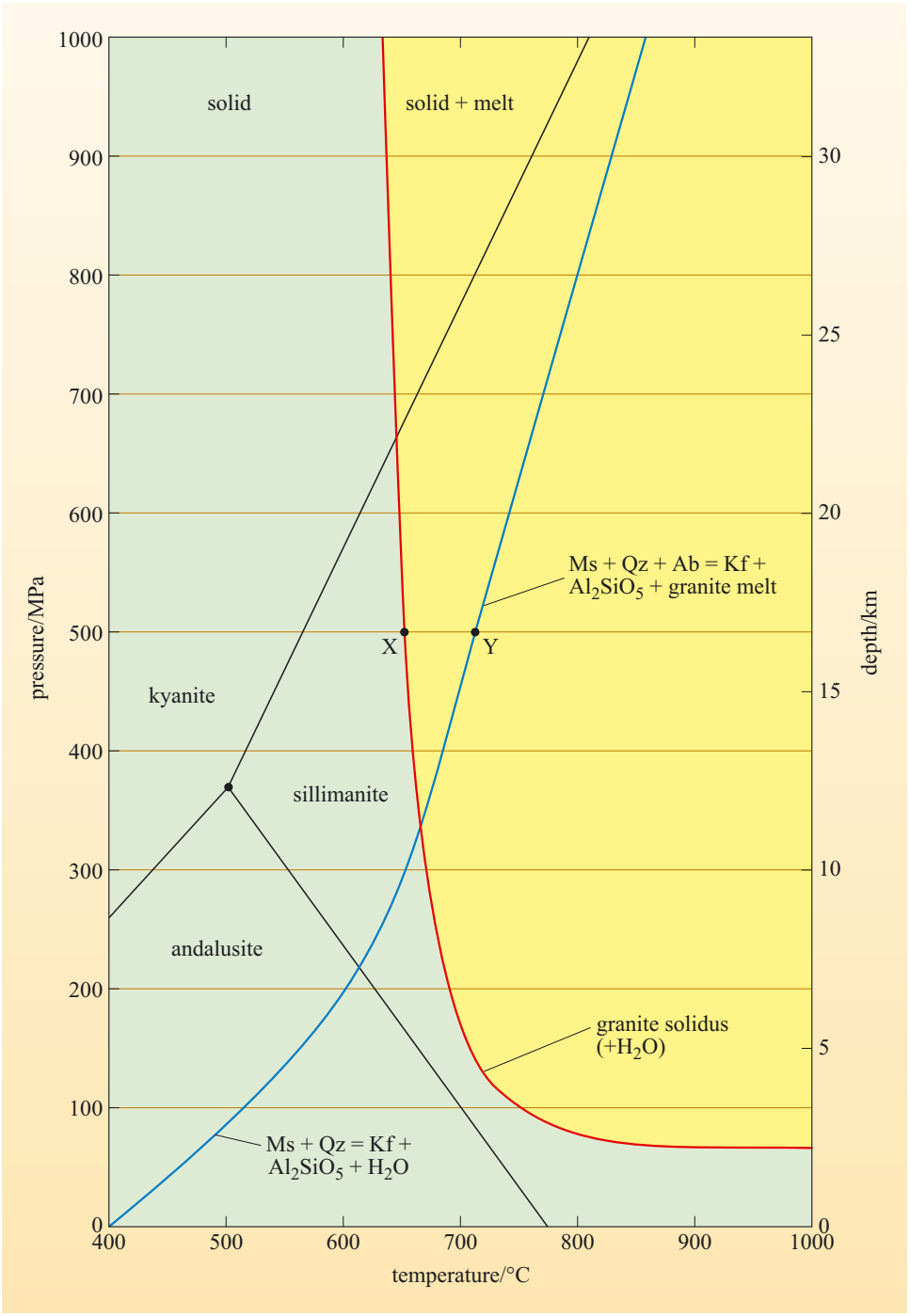


Figure 6.8 Solidus for melting of granite (in the presence of water) and the muscovite and quartz dehydration reaction (Equation 6.2). Also shown are the stability fields for the aluminosilicates. Ms = muscovite; Qz = quartz; Kf = alkali feldspar; Ab = albite.

If only a small proportion of melt is generated, then the granite magma will remain in the rock but be segregated from the micaceous layers. When these melts crystallise, they will form lenses of quartz, plagioclase and alkali feldspar. Typically, such lenses are a few centimetres thick. These mixed rocks are called **migmatites** (Figure 6.9).

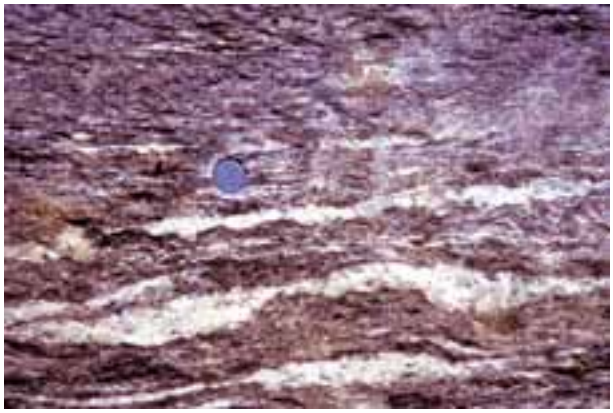


Figure 6.9 An example of migmatite. This is a rock that includes light-coloured lenses of granitic composition within a dark, biotite-rich gneiss (from Langtang Valley, Nepal).

More extensive melting will allow the melts to coalesce and migrate upwards as a granite pluton. With high enough temperatures, the hydrous phases in all crustal compositions will be dehydrated. Initially (above ~700 °C at normal crustal depths), muscovite will melt, leaving an aluminosilicate-rich residue. Biotite will melt at higher temperatures (above ~800 °C) to leave a garnet-rich or pyroxene-rich residue. If igneous rocks of basic and/or intermediate composition are metamorphosed, amphibole will break down to leave a pyroxene-rich residue. The residual dehydrated rocks are known as granulites, and they characterise much of the lower crust, particularly in collision zones, where conditions are too hot for micas or amphiboles to be stable.

Before leaving this introduction to metamorphism, it should be emphasised that metamorphism can occur in any tectonic setting and in both the mantle and the oceanic crust, as well as in the continental crust, in response to changes in temperature or pressure.

- Where have you already encountered metamorphism and metamorphic reactions?
- During the subduction of oceanic crust. Changes in mantle mineralogy in response to increasing pressure are also metamorphic reactions.

Andalusite, the low-pressure aluminosilicate, is commonly formed where aluminous sediments are intruded by magma at shallow depths irrespective of the tectonic setting in which the magma forms (e.g. the andalusite shown in Figure 6.3(b) is from a metamorphosed sediment caused by the intrusion of a granite formed at an active continental margin). Indeed, the continental crust can thicken beneath active continental margins, such as the central Andes, causing metamorphism and deformation very similar to that found in collision zones. But it is in regions of thickened continental lithosphere that crustal rocks are subjected to unusually high temperatures and pressures over a wide area and where evidence for crustal melting is most likely to be found.

6.2 The formation of granites from melting of the continental crust

Ultimately, the origin of all rocks can be traced back to the mantle, although in some cases many cycles of remelting, crystallisation, erosion and deposition may be involved. The great diversity of igneous rocks results from two processes: partial melting and fractional crystallisation. Granite magmas can form from either of these processes.

Before proceeding it is important to clarify the use of the term **granite**. In popular usage it covers a wide range of intrusive rock compositions; indeed, in the building trade, polished gabbro is sometimes referred to as 'black granite'. However, geologists use 'granite' as a more restrictive term for a composition of an intrusive rock that:

- is made up of at least 75% of quartz and feldspars
- contains quartz, plagioclase and alkali feldspar in roughly equal proportions (though in some unusual granites alkali feldspar can be significantly more abundant than plagioclase).

For similar rocks, but with a progressively higher plagioclase/alkali feldspar ratio, the terms granodiorite and **tonalite** (or **quartz diorite**) can be used; and if less quartz and more dark minerals, like amphibole and mica, are present, then the term **diorite** is appropriate. We are not concerned here with the precise definition of each rock type, but collectively these three plutonic compositions are intermediate between the compositions of gabbro and granite. Finally, the term 'granitic rock' is a term loosely applied to any light-coloured, plutonic rock containing quartz as an essential component, together with feldspar. This would include not only granites, but also most rocks with intermediate compositions.

At destructive boundaries granite represents the end-products of fractional crystallisation of basaltic magma, although rocks of intermediate compositions are much more common than granites in the batholiths of continental arcs. In the Andean batholith, for example, granite constitutes perhaps 5% of the rocks present. In collision zones, granite magma is the first product of partial melting of crustal rocks. The property of granite that allows it to form, either from extreme fractional crystallisation or from the initial stages of partial melting, is that granite in its magmatic state is the lowest temperature silicate melt that can exist. To understand why this is so, the following section explores the melting relations of the minerals that constitute granite.

6.2.1 The quartz–albite–orthoclase ternary system

Since granites are made up of over 75% of just three minerals, quartz, albite-rich plagioclase and orthoclase, it is possible to make quite accurate comparisons between the compositions of liquids within the quartz–albite–orthoclase (Qz–Ab–Or) system and natural granites. This three-phase system can be investigated by considering the three binary systems that link the three minerals. Mixtures in the binary Qz–Ab and Qz–Or systems exhibit simple eutectic behaviour similar to that shown in Figure 4.22 for fosterite (Fo) and diopside (Di). The relationships between albite and orthoclase, however, are more complex. The system of these

two feldspars shows both partial solid solution and a temperature minimum, the characteristics of which are explained below. Albite and orthoclase are the sodium and potassium end-members respectively of the feldspar system.

Figure 6.10 is a phase diagram of the alkali feldspar system under fixed pressure conditions (100 MPa). For the pressure at which the experiment was undertaken, there is continuous solid solution between albite and orthoclase. The liquidus surface has a minimum value (M′) with a composition of about Or_{30} (70% albite, 30% orthoclase).

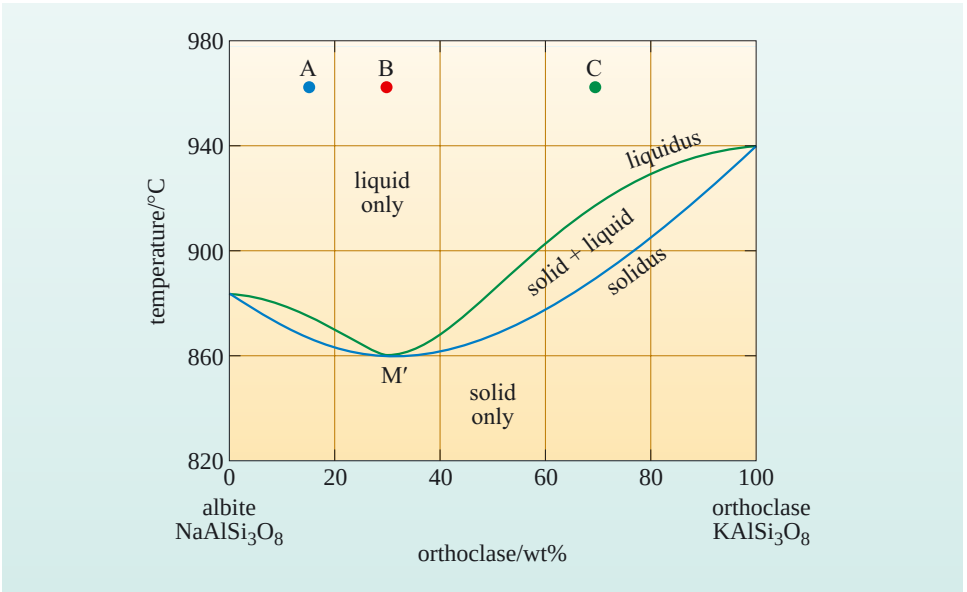


Figure 6.10 The alkali feldspar binary system showing a temperature minimum (M′), determined at a pressure of 100 MPa in the presence of H_2O.

- What is the composition of the first crystals to form on cooling melts of compositions A, B and C in Figure 6.10?
- Melt A has a composition Or_{15}. When this cools to intersect the liquidus, the solid phase that forms is on the solidus at the same temperature and must have the composition Or_7. Melt B has the composition Or_{30}, and when this cools to intersect the liquidus, because the liquidus and solidus coincide, the crystals will have the same composition as the liquid Or_{30}. Melt C has a composition of Or_{70}. The crystals formed when this melt is cooled to the liquidus have a composition of Or_{88}.

- Assuming that bulk equilibrium is maintained between the crystals and the liquid throughout crystallisation, what is the composition of the last crystals to form just prior to complete solidification of samples A, B and C in Figure 6.10?
- If equilibrium between the crystals and the liquid is maintained throughout crystallisation, then the composition of the last crystals to form will be the same as that of the total sample, i.e. Or_{15} for A, Or_{30} for B, and Or_{70} for C.

Contrast your answer with what happens if fractional crystallisation takes place in this system. As more and more crystals are removed, *all* liquids, *regardless of their starting composition*, move towards the minimum point M′. This is arguably the most significant feature of the alkali feldspar system.

The minimum point M′ depicts the composition of the lowest temperature liquid that can exist in the Ab–Or system at this particular pressure. That is a feature it shares with a eutectic point in a binary eutectic system – but can you see what makes a temperature minimum *different* from a eutectic point?

- How many minerals are in equilibrium with the liquid at the eutectic point in a binary eutectic system such as Fo–Di (Figure 4.20)?
- The answer is two. The eutectic point is the only place in a binary eutectic system where *both* minerals (in this case Di and Fo) can coexist with the liquid.

- What is the composition of the mineral(s) in equilibrium with the liquid at the minimum point M′ in Figure 6.10.
- Crystals in equilibrium with a liquid plot on the solidus at the same temperature as that of the liquid. At the minimum point M′ in Figure 6.10, therefore, they have the *same composition as the liquid*.

Therefore, during equilibrium crystallisation, all crystals at a temperature minimum have the same composition, and thus, unlike the situation at the eutectic point, only *one* mineral can be in equilibrium with the liquid at a binary temperature minimum. All crystals at M′ have composition Or_{30}. So, at a temperature minimum in a binary system, one mineral coexists with the liquid and has the same composition as that of the liquid. In contrast, at a eutectic point in a binary system, two minerals (with distinct compositions) coexist with the liquid.

An additional complication of the alkali feldspar system (not shown by Figure 6.10) is that, for temperatures below about 700 °C, alkali feldspar crystals near the middle of the Ab–Or range may separate into two distinct compositions, i.e. one rich in orthoclase and the other rich in albite. This partial solid solution is due primarily to the disparity in ionic radius between K^+ and Na^+. The radii differ by about 35%, and whereas at high temperatures K^+ and Na^+ will substitute for one another and form a complete solid-solution series, if these crystals cool slowly, then element diffusion will occur between two distinct crystal structures, one of which accommodates Na^+ and the other accommodates K^+. This process is called **exsolution** and is the result of one solid crystal separating into two minerals of distinct compositions. In the alkali feldspars, it results in complex intergrowths between albite-rich and orthoclase-rich crystals (Figure 6.11).

Figure 6.11 Photomicrograph of an alkali feldspar exsolution (magnification ×40). Dark areas are orthoclase, light areas are albite. Each phase is homogeneous. (Mackenzie and Guilford, 1980)

The three binary systems (Qz–Ab, Qz–Or and Or–Ab) can now be put together into a three-component Qz–Ab–Or system. The

shape of the liquidus surface in a three-dimensional diagram of composition against temperature is shown in Figure 6.12. However, it is easier to project this surface onto a triangular diagram, as in Figure 6.13, which is the phase diagram (or temperature contour map) of the Qz–Ab–Or system. It consists of two portions, the field in which quartz crystallises first and the field in which alkali feldspar crystallises first.

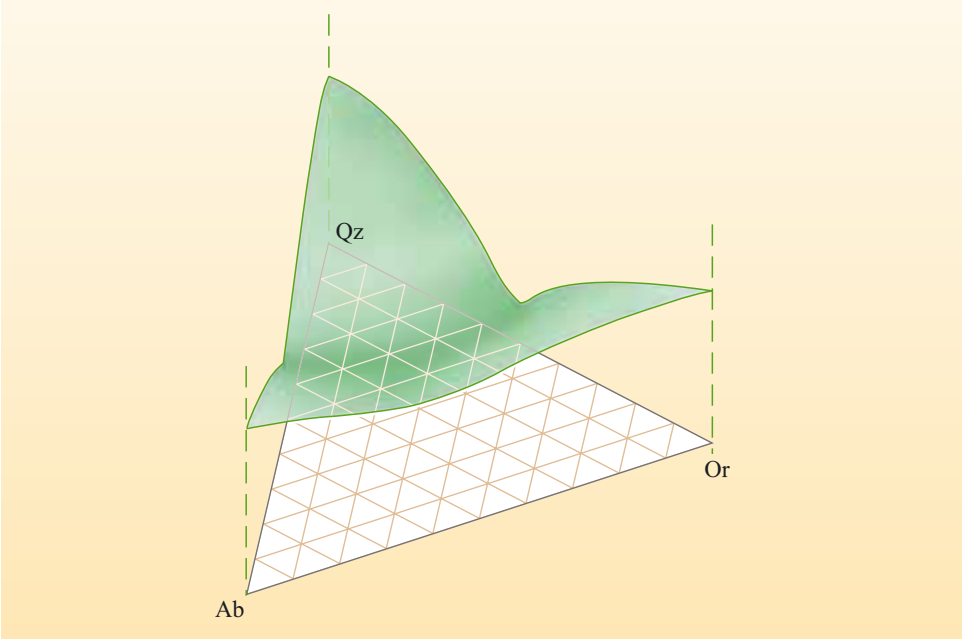


Figure 6.12 Three-dimensional diagram of the liquidus surface of the system Qz–Ab–Or determined at a pressure of 100 MPa in the presence of H_2O.

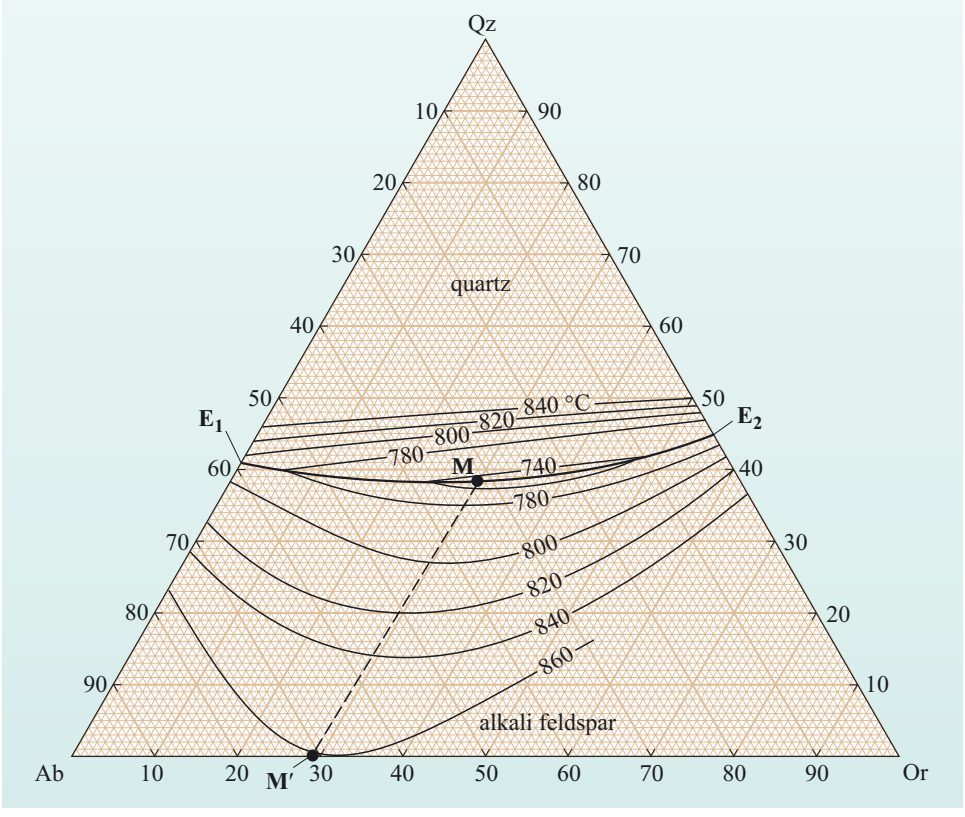


Figure 6.13 Diagram of the Qz–Ab–Or system determined at a pressure of 100 MPa in the presence of H_2O. Temperature contours on the liquidus surface are shown, except for quartz-rich and orthoclase-rich compositions where they are omitted for clarity. M = minimum on curve E_1–E_2; M′ = minimum between albite and orthoclase.

The Qz–Ab and Qz–Or binary systems have *eutectic points* at E_1 and E_2 respectively; these are joined by a cotectic curve E_1–E_2 in the ternary system. This curve slopes 'downhill' towards M from *both* the eutectic points E_1 and E_2; M is the lowest point on the curve and, therefore, the lowest temperature on the entire liquidus surface. This point is known as the **granite minimum**.

The implications of the granite minimum are highly significant. For a melt made up of any combination of quartz, albite or orthoclase, fractional crystallisation will move the composition of the residual liquid firstly towards the cotectic curve E_1–E_2, and then towards the minimum, M. In addition, for a solid made up of any combination of quartz, albite and orthoclase, the composition of the first liquid to form will be that of M.

It follows, therefore, that if naturally occurring granites originated either by fractional crystallisation or by partial melting, their compositions should lie near to that of the granite minimum. To assess this hypothesis, the compositions of a large number of natural granites have been plotted on the phase diagram. The results are shown in Figure 6.14. There is a striking concentration of points within a small area close to the granite minimum, confirming these granites are either the final products of extreme fractional crystallisation or the early magmas formed during partial melting.

There is, however, quite a broad scatter of compositions from the natural rock compositions plotted in Figure 6.14. In addition, there is a displacement between the granite minimum (M) shown in Figure 6.13 and the modal composition of 500 granites in Figure 6.14. This displacement is from $Qz_{35}Ab_{40}Or_{25}$ for the

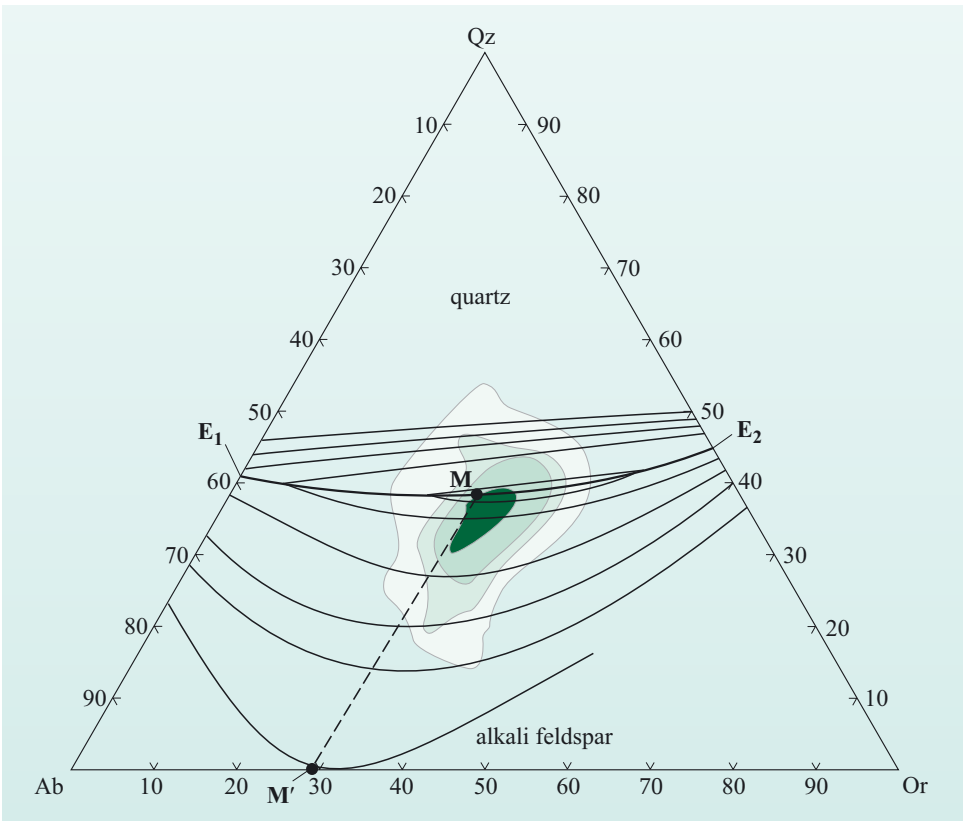


Figure 6.14 Chemical compositions of over 500 granitic rocks in terms of the proportions of quartz, albite and orthoclase. The results have been contoured on the basis of frequency and 90% of the analyses fall in the darkest shaded area in the centre. The curve E_1–E_2 is the cotectic curve for the Qz–Ab–Or system.

experimental plot to $Qz_{37}Ab_{30}Or_{33}$ for natural granites, towards the orthoclase apex. The reason for both the scatter and the displacement is that natural granites result from a wide range of pressures and H_2O conditions, and reflect small chemical variations in elements not represented by the ternary plot (such as Ca). All these affect the precise location of the granite minimum. The most important variable in this case is probably the presence of H_2O; the experimental results that are plotted up in Figure 6.13 are for H_2O-free conditions, whereas most granites form in the presence of some water. A second, perhaps equally important reason for the differences between natural granite compositions and compositions inferred from the topology of the Qz–Ab–Or system is that granites may not always represent the precise compositions of crystallised melts. As granite magma solidifies in the magma chamber, it may include **xenocrysts** of minerals derived from its source region or entrained during ascent or crystallisation. If these are the same types of mineral that are crystallising, e.g. quartz, or feldspar or mica, then it may not be possible to deduce the composition of the pure melt from the composition of the granite.

Whether granites have been derived from fractional crystallisation or by partial melting, their major element composition is much the same. For example, the silica content of melts close to the minimum-melt composition is 70–75%. However, the tectonic implications of the two processes are quite different. In the first case, the granites represent additions to the crust that have formed in a long-lived magma chamber, probably at a destructive boundary. In the second case, the granites are examples of crustal reworking, probably in a collision zone where the crust has thickened and heated sufficiently to melt. In the next section you will look at trace elements as a possible way of distinguishing between granites that result from these different processes.

6.2.2 Trace elements and crustal melting

The concentrations of trace elements in a crystallising magma are controlled by both the initial concentration in the magma and by the fractional crystallisation of minerals from that magma. The relationship between the ratio of the concentration of a trace element in the liquid relative to that present in the source, $\frac{C_L}{C_0}$, the proportion of melt remaining F, and the bulk partition coefficient D is shown graphically in Chapter 5, Figure 5.22 (calculated from Equation 5.7).

It can be seen from Figure 5.22 that the concentration of a trace element in a magma will only be unaffected by fractional crystallisation if the bulk partition coefficient of the crystallising minerals is $D = 1$. For incompatible elements ($D < 1$) the trace-element concentration in the liquid will increase with decreasing F, and for compatible elements ($D > 1$) the trace-element concentration in the liquid will decrease with decreasing F.

A similar series of curves can be drawn that describe the behaviour of trace elements with varying bulk partition coefficients under conditions of partial melting (Figure 5.21), based on the partial melting Equation 5.6. Note that for these curves, F represents the proportion of melt relative to the original rock. You should notice that partial melting will also result in an increase in the concentration of incompatible elements in the melt, particularly for small melt fractions ($F < 0.2$). The melt will be depleted in compatible elements, and for

large bulk partition coefficients ($D > 4$) the melt will contain very low concentrations even for quite high melt fractions. In general, it can be concluded that dramatic increases in the concentrations of incompatible elements can be obtained either by extreme fractional crystallisation or by very small melt fractions during partial melting.

The remainder of this section shows how both sets of curves can be used to see whether trace-element concentrations found in andesites and granites from the same igneous complex (Table 6.2) can distinguish whether granite magmas result from fractional crystallisation of an andesitic magma or partial melting of the continental crust.

The first point to notice about Table 6.2 is that Rb, with a low bulk partition coefficient, is an incompatible element, whereas Sr is a compatible element. In other words, during melting or crystallisation in the crust, Rb will concentrate in the magma whereas Sr will be more strongly partitioned into minerals, such as feldspar, that coexist with the liquid.

Table 6.2 Rb and Sr concentrations for typical andesite and granite compositions from the Andes, with bulk partition coefficients *D*.

	Concentration/ppm		***D***
	Andesite	**Granite**	
Rb	70	140	0.1
Sr	650	100	3.5

■ Using Figure 5.22 and the data in Table 6.2, determine the proportion of melt left if a liquid of andesitic composition undergoes fractional crystallisation to form a liquid of granitic composition based on (i) Rb and (ii) Sr.

■ C_0 and C_L represent trace-element concentrations in the parental andesite and evolved granite respectively.

(i) For Rb: $\frac{C_L}{C_0} = \frac{140}{70} = 2.$

From the curve for $D = 0.1$ on Figure 5.22, $F = 0.45$ for $\frac{C_L}{C_0} = 2$. In other words, 45% of the magma must be left to form a granitic liquid.

(ii) For Sr: $\frac{C_L}{C_0} = \frac{100}{650} = 0.15.$

It is not easy to obtain a precise value of F graphically for $D = 3.5$, but for $\frac{C_L}{C_0} = 0.15$, $F \approx 0.5$ from the curve for $D = 4$, so F will be <0.5 for D = 3.5

The results could be determined accurately from Equation 5.7, which yields values of $F = 0.41$ and $F = 0.48$ respectively.

Depending on the trace element examined, somewhere between 41% and 48% of the original andesitic magma is required to be removed as cumulates by fractional crystallisation to generate the observed Rb and Sr concentrations in the granite. One reason why the value for F differs according to the trace element

used is that the assumed value of D, the bulk partition coefficient, may be inaccurate, and indeed may vary during fractional crystallisation, partly because the mineralogy of the crystallising phases changes and partly because the mineral/liquid partition coefficients change with the composition of the liquid.

Question 6.1

Using Figure 5.21 and the data in Table 6.2, what proportion of the andesite would be required to melt to generate the granite based on (i) Rb and (ii) Sr trace-element concentrations?

The answer to Question 6.1(ii) suggests that partial melting of the andesite could not generate the granite, and the conclusion is that either the values of D are incorrect or that a granite with the trace elements indicated in Table 6.2 could not have formed by partial melting of the associated andesite.

In order to look at the formation of granites by crustal melting more closely, it is necessary to consider what the likely crustal sources might be. The crust is made up of a wide range of compositions, many of which will not melt at all at the temperatures that the continental geotherm can reach in the crust. For example, the temperature at the base of thickened crust is unlikely to exceed about 1000 °C (unless in the immediate proximity of hot mantle melts) and such temperatures will not melt sandstones or indeed a wide range of quartz-rich sediments. You might think that pre-existing granites would be the first lithologies to melt, but this is not the case. This is because, although the three minerals albite, orthoclase and quartz are present, the melting relations shown in Figures 6.12 and 6.13 have all been determined under H_2O-rich conditions. Dry rocks in the lower crust require much hotter conditions to melt. In fact, the most fertile of common crustal rocks are metamorphosed mudstones (often called metapelites). Prior to melting, these contain quartz, plagioclase and muscovite. The breakdown of muscovite (Equation 6.2) releases both alkali feldspar and H_2O which, together with quartz and plagioclase (albite), forms a granite melt (Figure 6.8). The melting reaction can be written as:

$$\text{muscovite} + \text{quartz} + \text{albite} = \text{granite melt} + \text{sillimanite} + \text{alkali feldspar} \quad (6.3)$$

At pressures appropriate to the mid crust (400–700 MPa) crustal melting can occur at temperatures as low as 680–760 °C provided muscovite is present in the source rock (Figure 6.8).

The next example considers a granite intruding high-grade metamorphic rocks where field relations suggest partial melting from the metapelitic migmatite that it intrudes. The Rb and Sr concentrations for both metapelite and granite are given in Table 6.3.

Table 6.3 Rb and Sr concentrations for metapelitic migmatite and intruding granite.

	Concentration/ppm		D_{bulk}
	Metapelite	**Granite**	
Rb	200	360	0.5
Sr	120	65	2.0

Question 6.2

What is the melt fraction F required to generate a granite of this composition from the metapelitic source?

The trace-element data, therefore, suggest that between 11% and 15% melting of the metapelitic source will yield the appropriate granite composition.

If you try to generate the granite in Table 6.3 by fractional crystallisation of the andesite in Table 6.2, then the required melt fraction (from manipulation of Equation 5.7) is 0.16 for Rb and 0.40 for Sr. Such a large difference suggests that fractional crystallisation from an andesite of this composition is not the correct process by which this granite formed.

It is now possible to draw a general conclusion from considering simple quantitative modelling of the Rb and Sr contents of granites and their possible source rocks. *Either* partial melting *or* fractional crystallisation can explain the high concentrations of incompatible elements like Rb, or low concentrations of compatible elements like Sr. From the two examples considered, granites resulting from fractional crystallisation (Table 6.2) have lower Rb/Sr ratios than those resulting from partial melting (Table 6.3). Although this is often true, it cannot be used on its own to distinguish between the two mechanisms of granite formation, because a *higher* melt fraction would reduce the Rb/Sr ratio of a partial melt. Equally, a *lower* melt fraction would increase the Rb/Sr ratio of a melt residual from fractional crystallisation. However, by applying more than one trace-element ratio it is often possible to exclude one or other of the processes by the inconsistencies between the results from the incorrect model. In practice, it requires a combination of field observations with a range of trace-element analyses to determine the origin of a granite. The most useful tool of all is probably the use of isotopic ratios. Neither fractional crystallisation nor partial melting should change the isotopic ratio, so the granite should preserve the isotopic ratio of its source, whether an andesitic magma or a metamorphosed sediment. This approach will be used later in the chapter.

6.3 The India–Asia collision

The important features of continental collision on the scale of plate tectonics are best illustrated by the boundary between the Indo-Australian and Eurasian plates. This boundary is complex and changes its character along its length, as is exemplified by the northwestern margin of the Indo-Australian plate illustrated in Figure 6.15. Also shown are the epicentres of the 14 largest earthquakes recorded in the region since 1900; all but one are located on, or near, the plate boundary. The largest of these in the southeast corner of the figure was responsible for the tsunami of 26 December 2004. This marks the point where the boundary swings northwards from the destructive boundary of the Sunda Arc (Indonesia) to the strike–slip system of Burma. The region of intense seismicity in northeast India marks the point where the boundary twists sharply to form the eastern boundary of the Himalayan arc. The 2000 km stretch of the plate boundary between the mountain of Namcha Barwa, South Tibet and Nanga Parbat, North Pakistan separates the Himalayan mountains to the south from the Tibetan Plateau to the north, and defines the world's most active continental collision zone.

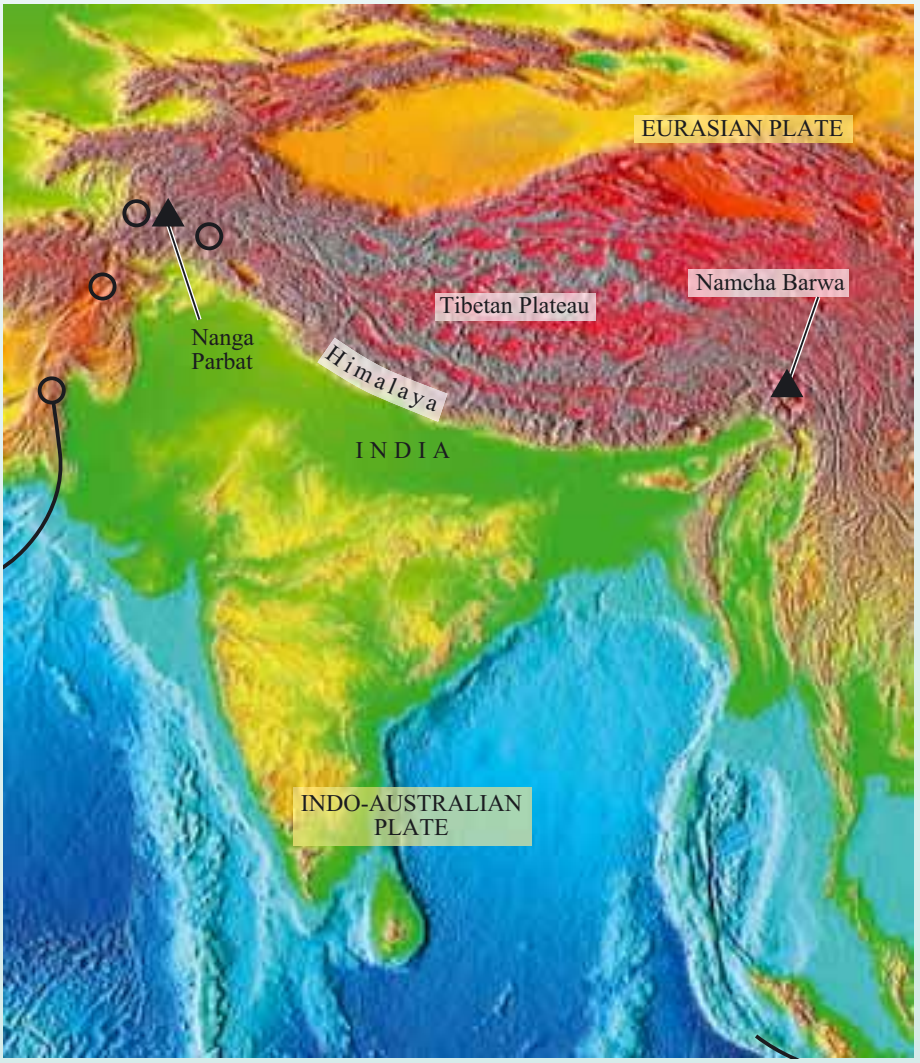


Figure 6.15 Computer-generated map showing the topography of India and Tibet. Grey land areas indicate altitudes >5000 m, red 1000–5000 m (Tibetan Plateau), yellow 500–1000 m, and green <500 m (coastal regions). The solid line indicates the boundary (suture) between the Indian Plate and Eurasian Plate. The Himalaya stretch from Nanga Parbat to Namcha Barwa. Open circles represent the epicentres of the 14 most powerful earthquakes recorded in the region since 1900.

Tibet is a plateau of 2×10^6 km^2, about the size of western Europe. It is on average 5 km above sea level and includes over 80% of the world's land surface higher than 4 km. The Himalayan and Karakoram Mountains, which define its southern and western margins, include the only peaks on Earth reaching more than 8 km above sea level. Why is this part of the southern Asian continent so anomalous?

India lay well south of the Equator 100 Ma ago. The northern part of the Indian Plate was oceanic crust lying beneath the Tethys Ocean. Subduction of this oceanic lithosphere resulted from the northerly migration of the Indian continental landmass on a collision course with Eurasia (Figures 6.16 and 6.17a). As a result of the subduction, magmas were generated beneath the southern margin of Eurasia, intruding and thickening the crust in the same way that the Andes are being thickened today.

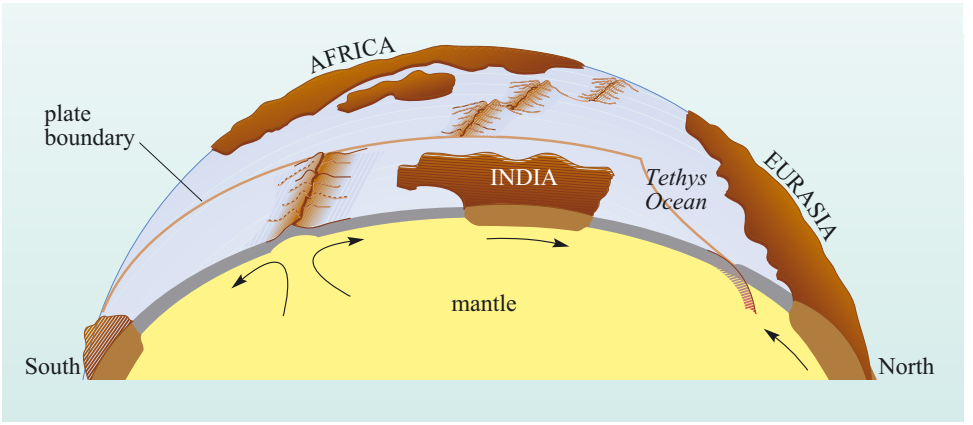


Figure 6.16 Diagram showing the northward movement of India between 100 million and 50 million years ago, causing shrinkage of the Tethys Ocean as India and Eurasia converged. (Adapted from Prentiss, 1970)

By about 50 Ma, i.e. by the mid-Eocene, that part of the Tethys Ocean separating the Indian and Eurasian Plates had completely closed. The two continents then collided, with the result that the continental crust on both sides of the collision zone was thickened and the surface elevation rapidly increased (Figure 6.17b). Today, the join between the Indian and Eurasian Plates can be traced along the northern edge of the Himalayan chain in southern Tibet by a line of outcrops comprising rocks (mainly gabbros and basalts) characteristic of oceanic crust, i.e. ophiolites. These rocks were once part of the oceanic lithosphere underlying the great Tethys Ocean.

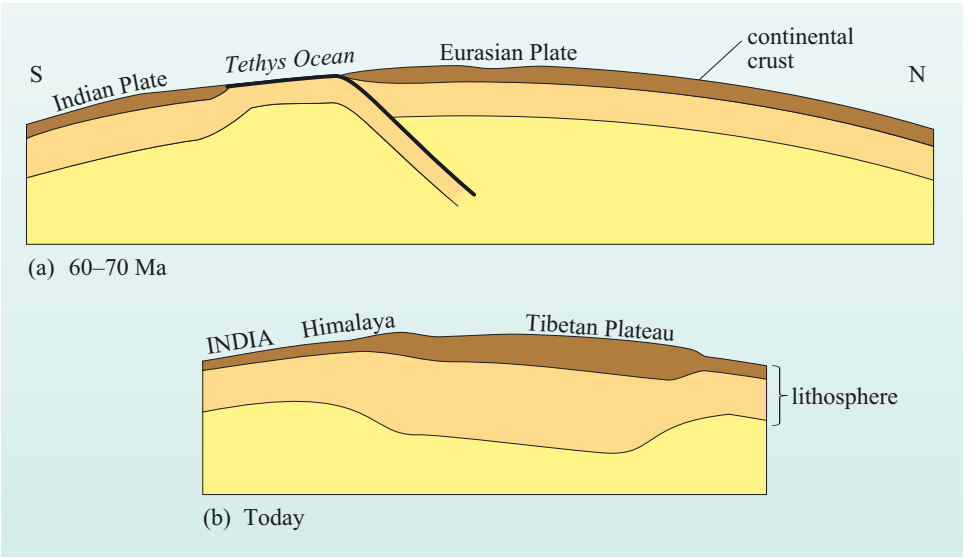


Figure 6.17 The India–Asia collision in cross-section (the scale is realistic, i.e. there is no vertical exaggeration): (a) at about 60–70 Ma; (b) today.

The Himalaya are by no means the only mountains on Earth formed by continental collision. The Alps are the result of a collision between Africa and Eurasia that began 120 Ma ago. During earlier geological times, collisions initiated the building of many of the world's lesser mountain belts, including the Urals and the Scottish Highlands. Why then did the Himalaya, and their hinterland, the Tibetan Plateau, become so high?

The uplift of the Himalaya and of the Tibetan Plateau is a response to considerable crustal thickening. Normal continental crust is 35–40 km thick, whereas the crust beneath the Himalaya and Tibet is 70–100 km thick. Some

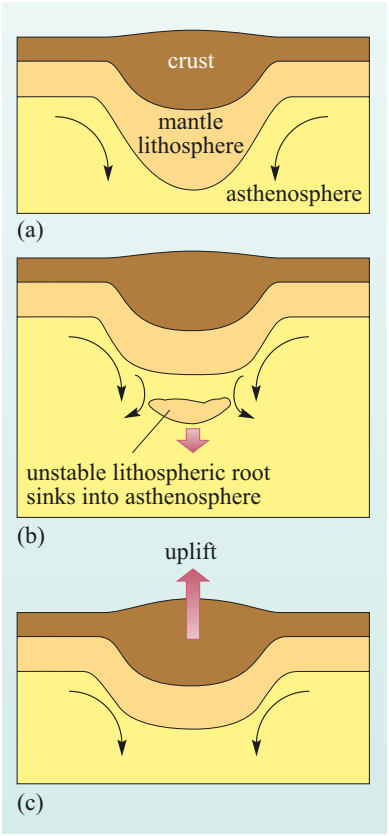


Figure 6.18 Schematic diagram to illustrate how convective thinning of the lithosphere may have contributed to the uplift of the Tibetan Plateau. (a) The lithosphere thickened as a result of collision. (b) Melting and erosion of the dense lithospheric root by convection currents in the asthenosphere. (c) Uplift of the plateau in response to the removal of the base of the lithosphere.

geologists maintain that the continental edge of Eurasia was unusually thick *before* collision, due to magmatic thickening above the subduction zone. In any event, substantial further thickening occurred after collision, by both folding and faulting of the rocks caught near the leading edges of the colliding plates (Figure 6.1). Such dramatic crustal thickening resulted from two unusual characteristics of the collision. First, it was 'head-on' rather than oblique, with a high closure velocity of ~200 mm y^{-1}. Second, the convergence persisted long after the intervening ocean had closed; India has migrated nearly 2000 km northwards since then, and a good deal of the convergence has had to be taken up by squashing and uplifting the rocks of the Himalaya.

The simple model for uplift outlined above works well for rather narrow mountain ranges like the Himalaya, but has Tibet been uplifted in the same way? It is not immediately clear why an area as broad as the Tibetan Plateau should be uplifted so high following collision. However, in recent years, geophysicists have suggested an additional mechanism for rapid uplift.

Although the distribution of major earthquakes is largely restricted to plate boundaries (Figure 6.15), earthquakes of moderate intensity are distributed across the Tibetan Plateau, as indeed is the evidence for deformation resulting from continental collision. This suggests that during continental collision the lithosphere does not behave in a brittle manner, i.e. as a rigid plate deforming only along its edges. Instead, it undergoes internal deformation by ductile thickening because of the higher temperatures generated within the thickened crust by the high concentrations of heat-producing elements. Under compression, the continental lithosphere behaves as a **thin viscous sheet**. Thus, following collision between two continental plates, the entire lithosphere will be thickened creating a lithospheric 'root' that protrudes down into the asthenosphere (Figure 6.18a). Since the early 1980s, geophysicists have been using computer models to study the thermal properties of the lithospheric root and the surrounding asthenosphere. These studies suggest that, following lithospheric thickening, the deep, cold root will eventually become unstable due to convection in the asthenosphere and be recycled back into the convecting mantle. This process, illustrated in Figure 6.18, is known as **convective thinning** of the lithosphere.

- What would be the mechanical consequence of removing the dense base from the lithosphere?
- Isostasy requires that the remaining lithosphere will bob upwards to an elevated position.

Convective thinning, therefore, provides a mechanism for the rapid elevation of the Tibetan Plateau (Figure 6.18(c)).

So, to summarise: the high-impact velocity between India and Eurasia, followed by continuing compression over tens of millions of years, led to an unusual degree of lithospheric thickening beneath Tibet, and the convective thinning that resulted allowed the lithosphere to rebound. This event may be the principal reason why Tibet is so much higher than other plateaux on the Earth today.

6.4 Metamorphism and melting in the Himalaya

The Himalayan orogen provides evidence of two contrasting examples of magmatism that can be found in collision zones. The first of these is the Trans-Himalayan batholith, which is emplaced into the southern margin of the Asian plate, just north of the suture that marks the boundary between pre-collision Asia and India (Figure 6.19). This forms a vast igneous arc, over 3000 km long and 50 km wide (only a segment of it is shown in Figure 6.19), made up of plutons of granites, granodiorites, diorites and a small proportion of gabbros (Figure 6.20). Rb–Sr isochron ages from the plutons cover a range from about 100 Ma to 40 Ma. These magmas intrude volcanic rocks of similar age and also sediments that were deposited along a continental margin between 100 Ma and 200 Ma ago.

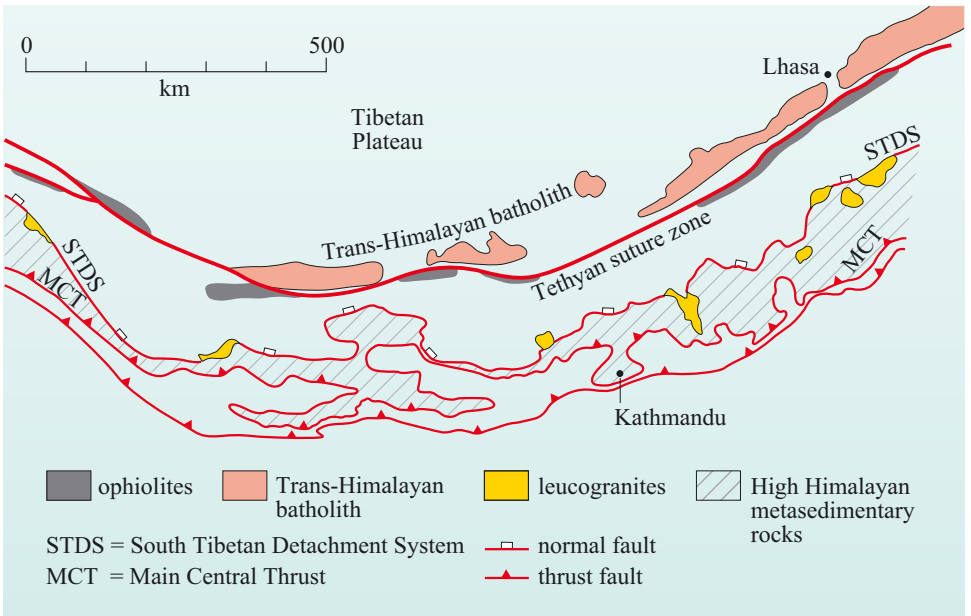


Figure 6.19 Geological sketch map of the central and eastern Himalaya and southern Tibet showing the distribution of plutonic rocks. Hatching indicates the High Himalayan metasediments bound by the South Tibetan Detachment System (STDS) and the Main Central Thrust (MCT). The thick, red line marks the suture zone.

The initial Sr-isotope ratio of a sample $(^{87}Sr/^{86}Sr)_i$ is the isotope ratio at the time of its formation. It reflects the Rb/Sr ratio of the source rock from which it was derived. The initial Sr-isotope ratio of plutons from the Trans-Himalayan batholith, therefore, can shed light on their origin (Figure 6.21).

- Study Figure 6.21; were the granites in the Trans-Himalayan batholith derived ultimately from the mantle or from the continental crust?
- The initial $^{87}Sr/^{86}Sr$ ratios of the trans-Himalayan granites are closer to the mantle evolution line than that of the sediments they intrude, so they were probably derived from the mantle.

In detail, the early granites formed at about 100 Ma have a similar $^{87}Sr/^{86}Sr$ ratio to the upper mantle and so are likely to result from fractional crystallisation of mantle-derived melts.

Figure 6.20 Inclusion of diorite within Eocene granodiorite from near Lhasa. Hammer handle is ~3 cm wide.

However, the initial ratio of the younger samples shows a range of values; some are similar to the upper mantle, but others are distinctly elevated towards the values of the sediments they intrude. This suggests some contamination by crustal material, probably by partial melting and assimilation of crustal rocks during magma ascent.

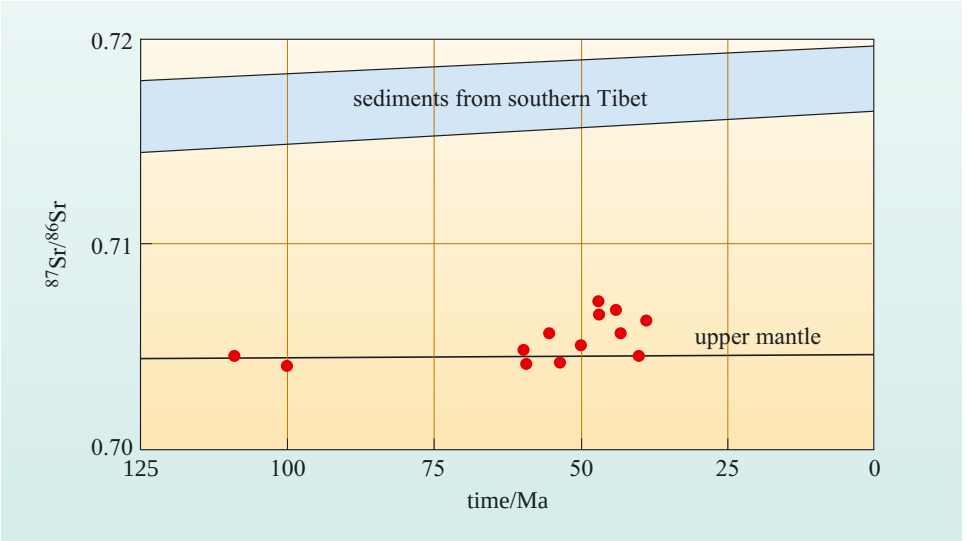


Figure 6.21 Sr-isotope evolution diagram of plutons from the Trans-Himalayan batholith (red circles) and of the sediments that they intrude.

It is now possible to draw together tectonic and geochemical evidence for the formation of the Trans-Himalayan batholith. The magmas were emplaced during subduction of oceanic lithosphere, but magmatism ceased after collision between India and Asia 40–50 Ma ago. Therefore, they were emplaced above a subduction zone at an active continental margin (Figure 6.17(a)). From the range of rock types (granites to gabbros), the suite of magmas appears to be similar to magmas generated in the Andes, and detailed geochemical studies support such an analogy. Therefore, the Trans-Himalayan batholith is an example of pre-collision magmatism at an active continental margin.

- Why do you think mantle-derived granites younger than 40 Ma are largely absent from the Himalaya?
- Once collision occurred, subduction of oceanic lithosphere stopped, cutting off the supply of fluids into the mantle wedge. As a result, further melting in the mantle virtually ceased.

The second example of Himalayan magmatism contrasts strongly with the Trans-Himalayan batholith. A series of light-coloured granites (the High Himalayan leucogranites) are emplaced as lenses and sheets into the metamorphosed sediments of the Himalayan mountains south of the suture (Figure 6.19). They are known as **leucogranites** because they are light in colour (Figure 6.22), being composed almost entirely of quartz, plagioclase and alkali feldspar. All the intrusions of this type are granites containing 70–75% silica. Rb–Sr isotope data from several of these granites indicate an age of 20 ±3 Ma and an initial Sr-isotope ratio varying between 0.738 and 0.785 (Figure 6.23). Therefore, the granites are intruded well after collision and have a much higher initial Sr-isotope ratio than any samples analysed from the Trans-Himalayan batholith (Figure 6.21).

- Suggest a possible source for the High Himalayan leucogranites (based on the isotope data and rock types described so far) and relate this to what you know of Himalayan tectonics.

- The high initial Sr-isotope ratio strongly suggests a crustal source. The restricted, silica-rich compositions of the granites (70–75% SiO_2), equivalent to minimum-melt compositions, support partial melting in the quartz–plagioclase–alkali feldspar system. From Section 6.3, we know that crustal melting can occur in thickened crust due to the high content of heat-producing elements in crustal rocks. Since the intrusion age of 20 Ma is at least 20 million years after initial collision, it seems likely that the granites formed after the thickened sedimentary pile heated up due to the high concentrations of heat-producing elements in crustal rocks.

Figure 6.22 Pinnacle of High Himalayan leucogranite forming Shivling (6543 m) in the Garhwal Himalaya, India.

The High Himalayan leucogranites are not intruded into unmetamorphosed sediments like the Trans-Himalayan batholith, but into migmatites formed from the partial melting of much older aluminous (pelitic) sediments (Figure 6.9). The sediments found to the north of the suture in southern Tibet have $(^{87}Sr/^{86}Sr)_i < 0.720$ (Figure 6.21), but aluminous metasediments further south in the Himalaya have extremely high $(^{87}Sr/^{86}Sr)_i$ of 0.735–0.790. Careful examination of the minerals in the dark parts of the migmatites shows that muscovite is breaking down to form sillimanite by reaction with quartz. This reaction provides alkali feldspar and a hydrous fluid which, when combined with quartz and plagioclase from the sediment, will form a granite melt (Equation 6.3). Therefore, the mineral reaction that is observed in the migmatite provides a likely source for the granite.

Why rocks should melt at all in collision zones can be understood from Figure 6.24. The aluminosilicate stability fields, the muscovite + quartz breakdown reaction and the granite melting curve have been transferred from Figure 6.8, and superimposed on the diagram are two geotherms, A and B. Geotherm A is a typical geotherm through continental crust of normal thickness. Geotherm B is an elevated geotherm that is a consequence of internal heating within thickened continental crust.

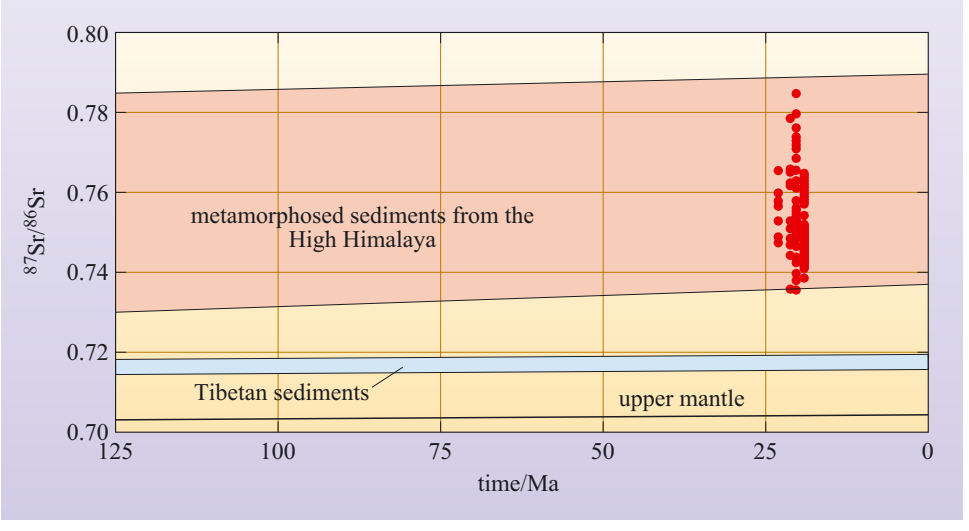


Figure 6.23 Sr-isotope evolution diagram of High Himalayan leucogranites (red circles) and the metamorphosed sediments they intrude.

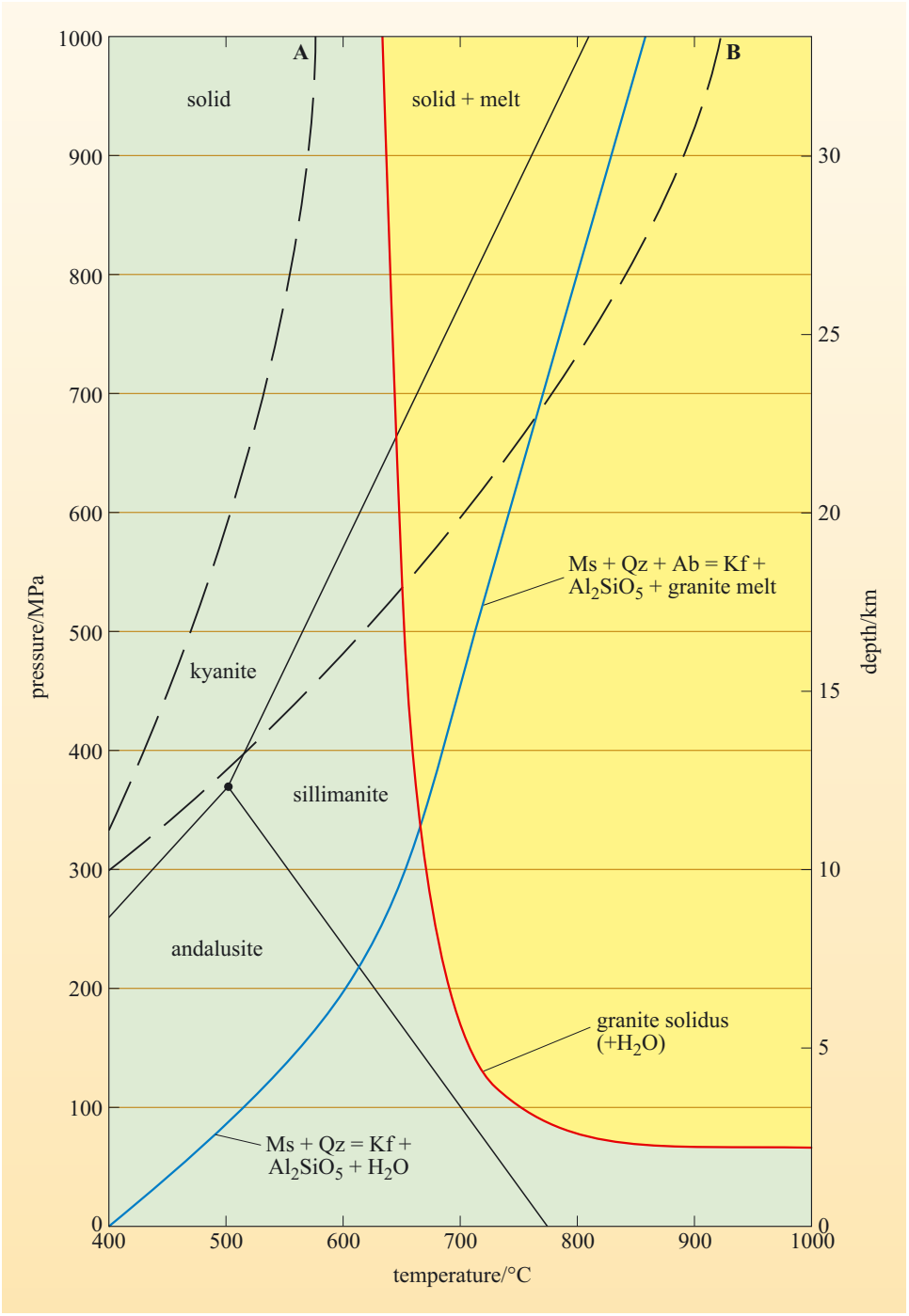


Figure 6.24 Two continental geotherms A and B plotted on phase boundaries taken from Figure 6.8. Ms: muscovite, Qz: quartz, Kf: alkali feldspar and Ab: albite.

Question 6.3

(a) At which depths would you expect melting to occur in response to (i) geotherm A and (ii) geotherm B of Figure 6.24?

(b) Which polymorph of Al_2SiO_5 will be formed by the melting reaction?

At this stage, some general conclusions can be derived from comparing the two groups of Himalayan granites. At active continental margins, such as the Andes or the Trans-Himalayan batholith, granites form a minor component within a suite of less-evolved plutonic rocks of intermediate compositions, like diorites, tonalites or even granodiorites. They result from fractional crystallisation of mantle-derived basaltic magmas combined with assimilation of crustal rocks. Such suites of granitic rocks, therefore, represent net additions to the mass of continental crust. In contrast, collision-zone granites occur in the absence of less-evolved compositions. They require no contribution from the mantle and provide an example of crustal reworking rather than crustal growth. On a global scale, the latter group are relatively small in volume compared with the great batholiths located along continental margins, but they provide insights into the behaviour of the continental crust when heated above its solidus temperature.

6.5 Modelling the Himalayan orogen

In this chapter so far, the Himalayan orogeny has been treated as the simple consequence of two continental margins colliding and exerting extreme compression on the rocks around the collision zone. However, as the structure of the Himalaya has become better known, it has become clear that the detailed picture is more complex. Not only are there thrust faults thickening up the crustal slices, but also there are normal faults suggesting zones of extension within the overall picture of regional compression.

Schists, gneisses and migmatites of an aluminous composition outcrop across a swathe of the High Himalaya bounded by two faults (Figure 6.19): the South Tibetan Detachment System (STDS) (a normal fault system) and the Main Central Thrust (MCT). Melting of these metasedimentary formations may have been critical in shaping the architecture of the Himalayan orogen. Metamorphic studies of the mineral assemblages from these rocks suggest they have been exhumed from depths of about 20 km, and structural studies suggest they have experienced ductile deformation whilst being brought to the surface. The mechanism of their exhumation is well established; both the STDS and the MCT that define the boundaries of these rocks have a shallow northward dip (Figure 6.25). Working together, movement on these two faults has resulted in the upward and southward motion of a sheet or a wedge of High Himalayan metasediments (Figure 6.25).

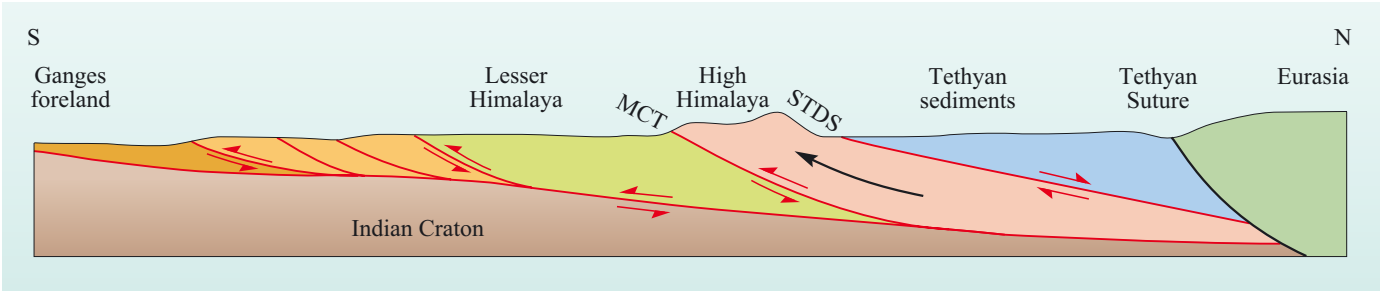


Figure 6.25 Sketch vertical section across the Himalaya. Note that simultaneous movement on the STDS and MCT will exhume the High Himalayan metasedimentary rocks between them (see arrow).

The Himalayan collision is such a clear-cut and recent example of collision tectonics that it has become the focus of intense study by geoscientists. One of their goals is to incorporate the geological characteristics that can be observed at the surface and the deep-level structures imaged by seismic surveys into a thermo-mechanical model of the thickening lithosphere to understand more about the processes that drive lithospheric evolution during continental collision. One such model, the thin viscous sheet model, has been briefly described already. This explains many broad-scale features of the deformation of the Tibetan lithosphere, but it cannot explain why rocks from the High Himalaya are apparently being extruded southwards, as shown in Figure 6.25.

In recent years, the **channel-flow model** has been developed to account for this southern extrusion together with other aspects of the Himalayan orogen. This model takes into account the changing viscosity of crustal rocks with depth. Viscosity is a measure of how freely material is able to flow, and gives an indication of the strength of the material; as rocks are subject to increased temperatures and pressures over time, they may flow like a fluid and so deform in a ductile way, as is implicit in the thin viscous sheet model for deformation of the lithosphere. The viscosity of the lithosphere is determined by the prevailing geothermal gradient and by the nature of the rocks that are present at different depths within it.

A geophysical study of south Tibet has identified a zone of low seismic velocities at a depth of about 15–20 km within the Tibetan crust.

- Given the steep geothermal gradient that results from radioactive decay within thickened crust, and the ability of pelitic rocks to melt at temperatures as low as 700 °C at crustal depths, can you think of a realistic explanation for this low-velocity layer in the mid-crust?
- One explanation is that it represents a zone of partial melting. Other geophysical properties, such as electrical conductivity, are consistent with this interpretation.

Partial melting initially results in a mix of solid and liquid with the liquid distributed along grain boundaries. Even at low melt fractions (F <0.1) a dramatic decrease in viscosity, and therefore strength, of the rock is observed, once the isolated melt pockets establish connectivity along the grain boundaries. Indeed, the viscosity of partially melted granite varies by about 14 orders of magnitude between its liquidus and solidus temperatures. This process is known as **melt weakening**.

The results of incorporating both the temperature and viscosity changes into a model of thickened lithosphere are shown in Figure 6.26. By 21 million years after collision, a low-viscosity channel is created by partial melting of the Indian crust at temperatures above 700 °C (Figure 6.26a). By 33 million years after collision the low-viscosity channel is migrating southwards, driven by the differential pressure between the thickened Tibetan crust to the north and the ‘normal’ Indian crust to the south (Figure 6.26b). At this stage the channel can be likened to toothpaste being squeezed, along a tube.

However, the modelling suggested that this channel would not reach the surface unless the topographic rise of the southern Himalaya was being actively eroded.

If the effects of strong erosion are incorporated into the model then the isotherms are folded into an antiformal structure beneath the erosion front (Figure 6.26b) and by 42 million years after collision the low-viscosity channel would break the surface (breaking the cap off the toothpaste tube) (Figure 6.26c). Of course this model is much simpler than reality. In practice, a partially melted channel would never reach the surface without cooling rapidly, resulting in solidification of the melts followed by brittle deformation as the rocks thrust southwards through the uppermost crust. However, the channel-flow model does explain the southward extrusion of rocks brought up from the mid crust. It also predicts that they will be extruded along the southern topographic front of the Himalaya, where precipitation, from the monsoon, is most intense.

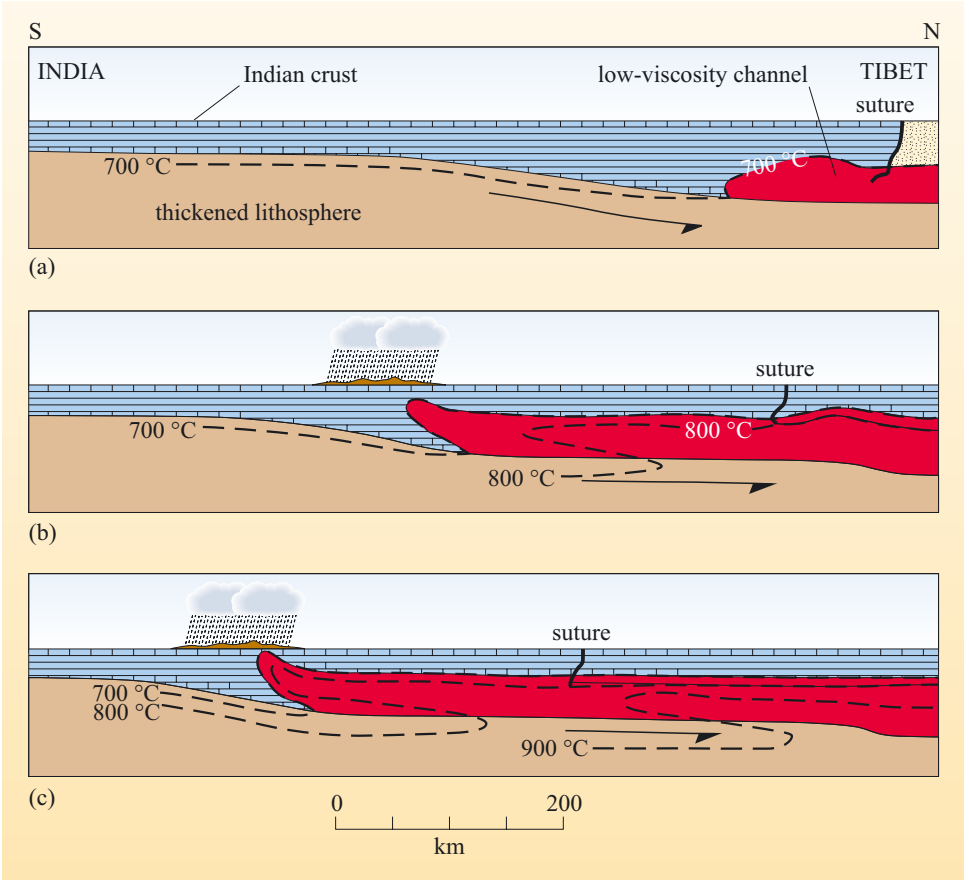


Figure 6.26 Results of modelling the thermal and mechanical behaviour of thickened crust, assuming melting is initiated at 700 °C, and that the southern Himalayan front is undergoing active erosion. A threefold time sequence is illustrated at (a) 21 Ma, (b) 33 Ma and (c) 42 Ma after collision. (Adapted from Jamieson et al., 2004)

The application of the model to the present-day structure of the Himalaya and southern Tibet is shown in Figure 6.27. The low-viscosity channel extruding southwards is represented by the High Himalayan metasediments (in pink), bound by the STDS above and the MCT below (Figure 6.19). In applying the model to actual geological relationships, the model has yielded some further insights. In order for the channel to be extruded at the surface today, melting must have been initiated by about 20 million years after collision (Figure 6.26(a)); in the Himalayan context this is equivalent to an age of around 30 Ma. The large High Himalayan leucogranites are about 10 million years younger than this (Figure 6.23) so cannot represent this early melting event. However, dating of zircons from within migmatites in the High Himalaya has revealed that partial melting of the metasediments occurred as early as 30 Ma. Migmatites can result

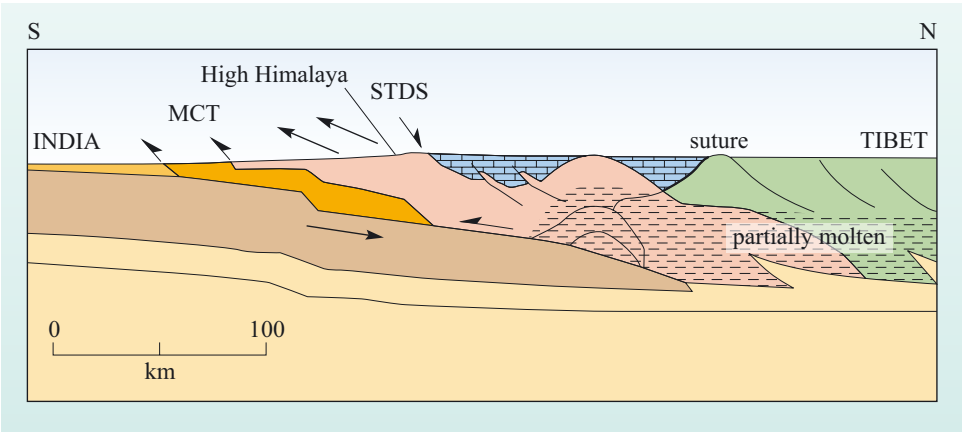


Figure 6.27 Section through the Himalaya and southern Tibet, indicating the southward extrusion of the lower crust from a partial melt zone beneath southern Tibet. (Adapted from Beaumont et al., 2004)

from a very small melt fraction, perhaps as low as 0.05, whereas the larger granites result from higher melt fractions (0.1 to 0.2) being squeezed out of their source rocks into larger magma chambers. This suggests that only small-degree partial melting is needed to initiate the formation of a low-viscosity channel.

Like any scientific model, channel flow offers a simplified view of natural processes. Over the next few years it will be tested and modified by geological and geophysical observations. Other models may eventually supplant it. However, the significance of this particular model for our understanding of continental collision is that it is the first to link deep-seated processes (ductile deformation, metamorphism and magmatism) with surface processes (erosion and precipitation). Until a few years ago, no geoscientist would have considered the possible linkage between atmosphere, hydrosphere and lithosphere, because climate research and the study of tectonics and metamorphism would have taken place in different departments, possibly even at different universities. Now, traditional boundaries between research areas are breaking down and there is an increasing awareness that the Earth behaves as a single system and the components of that system are all interrelated.

Summary of Chapter 6

- Collision between two continental plates results in thickening and metamorphism of the crust.
- Varying pressures and temperatures in the Earth's crust result in reactions between metamorphic minerals.
- The stability field of a mineral, or mineral group, may be represented on a pressure–temperature diagram that then provides information on the pressure or temperature to which a rock containing the mineral(s) have been subjected.
- High pressures favour high-density minerals; therefore, reactions involving large changes in densities provide geobarometers. Geothermometers are provided by reactions that result in dehydration or decarbonation.
- Increasing temperatures and pressures of sediments that contain clay minerals result in the formation of slates, schists, gneisses and migmatites. Gneisses may also form from the metamorphism of igneous bodies.

- The reaction of muscovite and quartz can produce granite melts in sediments of aluminous composition (known as pelites).
- Magmas of granitic composition may be formed either by fractional crystallisation of basic magmas or by partial melting of crustal rocks of pelitic composition.
- The ternary system Qz–Ab–Or is characterised by a liquidus surface that slopes downwards towards a point referred to as the granite minimum. Consequently, fractional crystallisation of all samples in this system drives the composition of the residual liquid towards that of the granite minimum irrespective of their initial composition. Conversely, on heating a rock consisting largely of quartz, plagioclase and alkali feldspar (or a mica that will break down to form an alkali feldspar), the first liquid to form is also close in composition to the granite minimum.
- The trace-element compositions of a granitic rock can be used to distinguish between an origin from fractional crystallisation of an andesitic magma and one from partial melting of a crustal source.
- The collision between India and Eurasia at about 50 Ma was responsible for crustal thickening and uplift of the Tibetan Plateau.
- In the Himalaya, pre-collision magmatism is represented by the Trans-Himalayan batholith that was intruded between 110 Ma and 40 Ma. Its initial Sr-isotope ratio $(^{87}Sr/^{86}Sr)_i$ is indicative of a mantle source with a component of crustal contamination that increases through time. It is typical of magmatism formed at an active continental margin.
- The High Himalayan leucogranites (intruded at about 20 Ma) result from partial melting of mica schists and migmatites in response to thickening and radioactive heating of the continental crust following collision.
- Continuous thickening of the lithosphere beneath Tibet, as proposed by the thin viscous sheet model, may have led to convective thinning followed by rapid uplift.
- The channel-flow model incorporates the effects of reduced viscosity of lower crustal rocks by partial melting, a process called melt weakening. The modelled channel will be extruded laterally by differential pressure exerted by thickened crust and upwards by focused erosion of surface rocks.
- The channel-flow model provides an explanation for the southwards extrusion of High Himalayan metasediments along the MCT and the STDS.

Learning outcomes for Chapter 6

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

6.1 The different mineral assemblages that can form in metamorphosed continental crustal rocks and how they provide information on the pressure–temperature conditions under which the rocks formed.

6.2 How mineral reactions provide information on the conditions of partial melting within the continental crust.

6.3 Why variations in the compositional range of granites occur, by referring to the alkali feldspar binary quartz–albite–orthoclase ternary phase diagrams.

6.4 How the mode of formation (e.g. fractional crystallisation, partial melting) and origin (e.g. source material) of granites can be determined from trace element and isotopic analyses.

6.5 How different approaches to modelling the thermal and mechanical consequences of continental collision provide insight into the timing and interactions between deep-level and surface processes.

Chapter 7

The deep mantle and global cycles

In Chapter 1 you explored the present gross structure of the Earth and in the subsequent chapters you have concentrated on what happens in the upper 100–300 km of the Earth. However, to understand the interactions between the atmosphere, hydrosphere, biosphere, lithosphere and upper mantle it is important to return to a view of the planet as a whole because many of the driving mechanisms that influence these interactions originate deep within the Earth. Perhaps the most obvious process that originates at depth but has a profound influence at the Earth's surface is volcanism. Volcanic eruptions can emit large volumes of volatiles, such as water, carbon dioxide and sulfur dioxide, which can significantly affect the Earth's climate. An obvious question to ask is whether these volatile compounds have always been inside the Earth or whether some proportion of them has been recycled from the Earth's surface back into the mantle by plate tectonics or other processes.

In this chapter you will concentrate on information from geophysical and geochemical studies that provide information on the nature and composition of the deep mantle. Geophysics provides a means of imaging the inside of the whole of the Earth and provides an insight into the fate of surface materials carried by plates as they are subducted back into the mantle, and their possible return to the Earth's surface. Geophysical studies are limited to providing a snapshot of what the planet's interior is like at the present and perhaps some indication of the last few hundred million years. Geochemistry, by contrast, exploits radiogenic isotopes to provide information about timescales of processes in the mantle from days to billions of years, but it is limited to studying materials that reach the Earth's surface as a result of volcanic or tectonic activity.

The origins of ideas about the deep Earth stem back to the earlier part of the twentieth century, when seismic studies allowed the gross structure of the Earth to be defined. At about the same time, Arthur Holmes, who was a great advocate of continental drift, suggested that currents within the mantle might provide a mechanism that could move the continents around the Earth's surface. Unfortunately, his ideas coincided with a rejection of the continental drift hypothesis by the geological community. When plate tectonics became an accepted theory to explain continental drift and sea-floor spreading, researchers started to re-evaluate the idea of solid-state convection within the mantle. Since then, seismic studies have become increasingly sophisticated and with the advent of a global digital seismic network, images of the Earth's interior are improving on a monthly basis. Seismic tomographic images (equivalent to a computer-assisted tomography (CAT) scan of a human body) can now resolve slabs subducting deep into the planet and are starting to resolve increasingly narrow zones of hot material upwelling within the mantle. Experimentalists investigate the mineralogy of the deep Earth by taking mantle materials and subjecting them to the pressures and temperatures prevailing throughout the whole 2900 km depth of the mantle. Geochemists study the isotopic and trace element contents of the products of mantle melting and the materials that are recycled into the mantle to build up geochemical models that are the equivalent to the well-known carbon cycle, but on a mantle length-scale and a geological timescale.

During the 1970s and 1980s, the most common model for the deep mantle was the so-called two-layer model in which the boundary between the upper and lower mantle at the 670 km seismic discontinuity represented a physical barrier to material moving both upwards and downwards. It was also thought that the boundary represented a distinct compositional difference between the depleted upper mantle and a lower mantle that had essentially a primitive (chondritic) composition. Mantle convection took place within each of the two layers and hot-spot volcanism originated at the 670 km discontinuity, although later models allowed for material for some hot-spot volcanism to leak periodically from the lower mantle. The two-layer model seemed consistent with most of the geophysical and geochemical data available at the time. However, the two-layer model restricts any recycling of surface materials to a well-mixed upper mantle, and only this reservoir could influence the atmosphere, hydrosphere and biosphere. More recently, there has been a convergence towards a single-layer model in which the whole mantle convects. This allows a much larger reservoir to influence processes at the Earth's surface, and it even includes possible interaction with the Earth's outer core.

7.1 Discontinuities in the deep Earth

The seismic velocity structure of the Earth has been well known since the 1930s and recent work has only served to refine certain parts of the velocity structure. Figure 7.1 is a profile of P- and S-wave velocities in the Earth illustrating the key changes in the seismic velocity structure with depth.

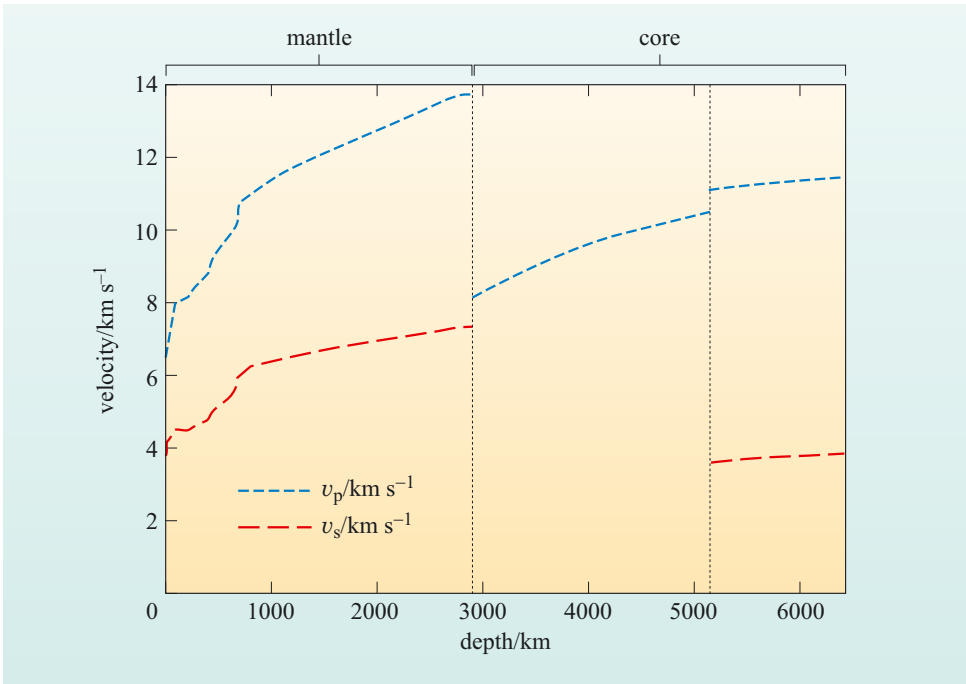


Figure 7.1 P- and S-wave velocity section of the Earth, with a modern Earth reference model. (Fowler, 2005)

- What key velocity changes in the mantle can you pick out from Figure 7.1?
- The major discontinuity occurs at 2900 km with a decrease in P-wave velocity at the core–mantle boundary (CMB) and a corresponding lack of S-waves in the outer core. Within the mantle there is a lower velocity region in the upper mantle at 100–200 km (known rather obviously as the low-velocity zone). There are marked increases in seismic velocity between ~410 km and ~670 km, which is known as the **transition zone**. The lower mantle is rather featureless until just above the CMB, where velocity structure becomes complex, although not apparent on Figure 7.1 because of the scale.

- What are the likely causes of seismic velocity changes in the mantle?
- You should recall from Chapter 1 that seismic velocity is dependent on density and elastic modulus. Also, the presence of any liquid phase will reduce seismic velocities. The discontinuities, therefore, reflect density and possibly phase changes.

The general equation relating seismic velocity, density and elastic modulus is:

$$\text{wave velocity} = \sqrt{\frac{\text{elastic modulus}}{\text{density}}} \qquad (7.1)$$

A critical characteristic of most solids is that when the density increases the elastic modulus also increases, but more rapidly. So, if a phase change in the mantle results in a material with a higher density, then the seismic velocity also increases because the elastic modulus increases at a greater rate than the density. At a given temperature, the density of the mantle depends on both its mineralogy and its bulk composition, and it is obviously important to resolve whether simple changes in mineralogy or compositional layering cause the observed stratification.

7.2 The mineralogy of the deep mantle

Chapters 4 and 5 explored phase diagrams for the upper 100 km of the mantle, in which a number of phase changes occur, particularly the spinel-peridotite to garnet-peridotite transition, that are important when considering the generation of basaltic melt. These phase changes occur in a peridotite mantle of constant composition and serve to show that the physical properties of the mantle can change even if the composition remains the same.

To study the next 2800 km of mantle involves some very special experimental equipment in order to replicate the extreme pressures and temperatures expected in the deep interior of Earth. These conditions can be achieved in the laboratory using **multi-anvil** presses made of very strong materials, such as tungsten carbide and even diamond. To generate pressures equivalent to the depth of the 670 km discontinuity using tungsten carbide anvils required larger and larger presses, the largest of which were the size of a large room. Eventually it was realised that two diamonds cut with perfect facets could be used as an alternative anvil material, and these allow experimentalists to squeeze rock powder to pressures equivalent to those of the outer core, while heating the sample

to very high temperatures with a laser. Fortunately, such **diamond anvil** presses fit on a laboratory bench and allow experimental petrologists to characterise the mineralogy of the mantle right down to the CMB; including a phase change at ~2800 km, which was only discovered in 2004. In these experiments, a tiny sample (a few milligrams) of very fine powder of an appropriate mantle composition is enclosed in a metal container and placed between the anvils. Pressure is then applied with a hydraulic press and the capsule is heated using an electric current or, in the case of diamond anvils, a laser. After a period of time at the required temperature and pressure, the experimental charge is cooled rapidly to quench the high-pressure, high-temperature phases allowing the products of the experiment to be characterised.

Figure 7.2a is a summary of the different phase changes that occur in a fertile peridotite (equivalent to a lherzolite at low pressures) mantle, assuming an adiabatic temperature gradient with a potential temperature of about 1280 °C. The first change sees orthopyroxene dissolve into garnet to produce a new form of garnet called **majorite** at about 5 GPa pressure. This phase change is progressive, and eventually also affects clinopyroxene, such that at pressures of ~13 GPa all pyroxene has been replaced by majorite. Olivine, which makes a significant proportion of a mantle peridotite, undergoes a phase change at ~14 GPa. This involves a reorganisation of the structure of olivine into the so-called **β spinel** structure, producing a mineral called **wadsleyite**, with an associated 10% increase in density. Another phase change occurs at 17–18 GPa, when the β spinel structure changes to a **γ spinel** phase called **ringwoodite** (after the famous experimentalist Professor Ted Ringwood). At ~24 GPa a series of complicated reactions takes place; the γ spinel phase dissociates to produce **magnesium (Mg)-perovskite**, and **magnesiowüstite (Mg-wüstite)**. The most recent experiments on representative mantle compositions demonstrate that this transformation occurs over a very narrow depth range, and the net effect of all the transformations at this pressure is to produce a 10% increase in the density of the mantle. Majorite breaks down between 20 and 25 GPa, forming **calcium (Ca)-perovskite** and more Mg-perovskite. At higher pressures, peridotite composition consists of 79% Mg-perovskite, 16% magnesiowüstite and 5% Ca-perovskite.

- How do the pressures of these phase transitions compare with the depth of seismic transitions in Figure 7.1?

- They correspond remarkably well. The transition of olivine to β and γ spinel structures occur between ~410 km and 520 km, equivalent to the seismic transition zone, and the appearance of Mg-perovskite occurs at ~670 km, coincident with the boundary between the upper and lower mantle. Finally, the lack of seismic discontinuities in the lower mantle is consistent with a lack of phase changes at pressures greater than 25 GPa.

The mineralogy of the peridotitic lower mantle is essentially constant to great depths, but at 100–300 km above the CMB there are changes in the seismic behaviour, including suggestions of dense layers and considerable seismic anisotropy. This region is known as the D′′ (D-double prime) layer, and there has long been speculation over the causes of these seismic anomalies. Recently, experimental studies at extreme pressures (134 GPa), equivalent to 2800 km depth,

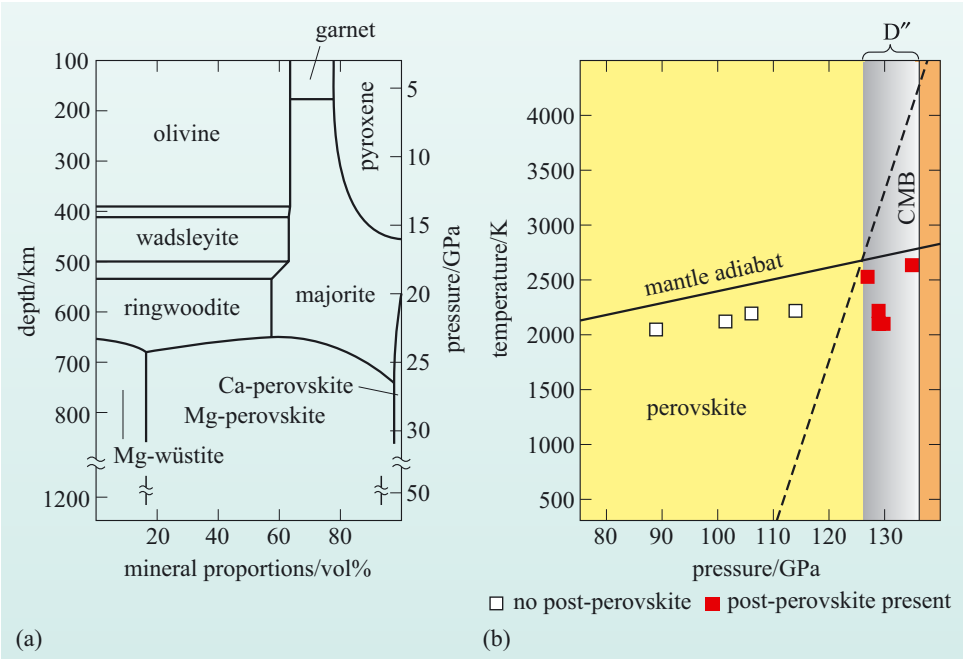


Figure 7.2 (a) Mineral proportions in Earth's mantle for 100–1200 km depth for a fertile peridotite and assuming an adiabatic gradient. (b) Plot of experimental data for the perovskite (open squares)–post-perovskite (filled squares) transition in the lower mantle. Shaded area represents the position of the D″ and the sloping dotted line is the best estimate of the position of the perovskite and post-perovskite transition. ((a) Poli and Schmidt, 2002; (b) Helmberger et al., 2005)

have demonstrated that Mg-perovskite transforms to a **post-perovskite phase**, with a 3% increase in density (Figure 7.2b). This phase has a structure that will generate seismic anisotropy, although layers of melt from the outer core and ancient subducted slabs have also been invoked as a way of producing the same seismic effects.

■ Can you see a relationship between the depths of the phase changes in Figure 7.2a and the changes in seismic velocities in Figure 7.1?

■ Yes, the 410 km and 670 km seismic discontinuities coincide with the transition of olivine to β spinel and of γ spinel to perovskite.

The key to understanding these phase changes is to remember that they happen without any change in the bulk composition of the peridotite. The phase changes occur over a limited depth range and they correspond to the depths of seismic discontinuities. It is reasonable, therefore, to assume that there is no need for major changes in the chemical composition of mantle at the major transition zones. Although this has been debated for the last 30 years, there is now a consensus that there is no major chemical stratification between the upper and lower mantle.

7.3 Seismic tomography

Seismic tomography is often likened to whole-body CAT scans used in medicine to image the internal structure of the human body. Whereas CAT scans use X-rays, seismic tomography utilises seismic energy from earthquakes. Seismic tomography was first developed in the 1970s, and it has become an increasingly powerful tool to investigate the structure of the Earth as computers have become more and more powerful.

The **preliminary Earth reference model (PREM)** (Section 1.2.4) and the velocity profiles illustrated in Figure 7.1 represent an average velocity structure for the Earth.

Seismic tomography produces a three-dimensional (3D) image of the Earth in terms of deviations in seismic velocity away from this average. When an earthquake occurs (E), seismic waves propagate out from the focus like ripples on a pond. If the position of the earthquake is known accurately, then it is possible to calculate the expected travel time for seismic waves to a seismic station (S) based on one of the Earth reference models (Figure 7.3). If there is a structure in the mantle that either slows or speeds up the seismic waves, then the actual arrival time will be either later or earlier than expected. By comparing arrival times from different receiving stations, an image of the structure can be progressively built up. From the example in Figure 7.3 it is obvious that the larger the number of seismic stations the better the resolution of the structure that produced the deviations. Because the Earth is a sphere and seismic stations are located all over its surface, tomography images can resolve not just the two-dimensional structure of an anomalous region, as in the schematic section in Figure 7.3, but also its 3D form.

The simplified explanation of seismic tomography given above is a good first approximation to the methodology; but in practice, modern-day tomography uses the arrival times of many other seismic signals in addition to the travel times of P- and S-waves. A modern study can use records from up to 18 000 different earthquakes and hundreds of seismic stations. A best-fit 3D velocity model of the Earth is then calculated using inversion techniques. Calculations that produce global tomographic images involve millions of calculations and can take up to 3 days to run, even on supercomputers!

The resolution of the Earth's interior by seismic tomography is not uniform in either horizontal or vertical planes. This is because seismic stations are more common in the Northern Hemisphere and are concentrated on the continents. Also, fewer ray paths pass through a given volume in the lower mantle than in the upper mantle, so that the resolution of the lower 1000 km of the mantle is more poorly constrained. The increase in ocean-bottom seismographs and specific studies of areas of interest mean that local tomographic images have improved resolutions down to 'cubes' with 100 km sides, which can be imagined as 3D pixels.

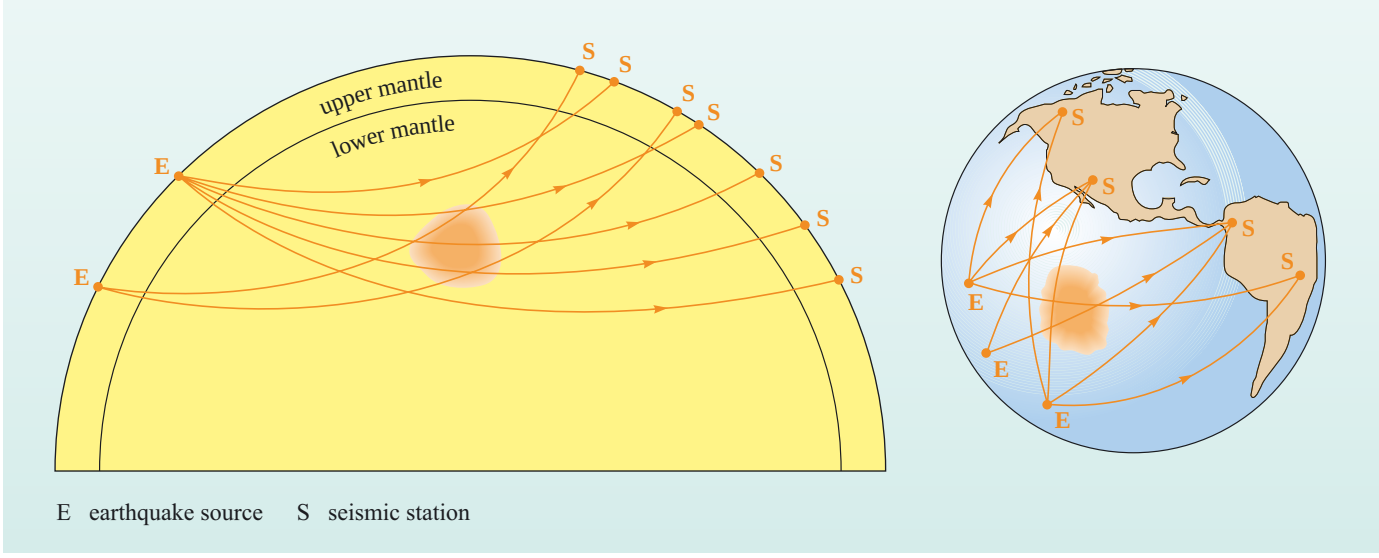


Figure 7.3 The principle of seismic tomography. Information from criss-crossing P-waves (in this case) through the Earth is combined to isolate velocity anomalies such as the shaded area shown here.

Just how good the resolution of the various tomographic images might be and whether they really image structures is a matter of some contention. A formal 'best fit' to millions of calculations is difficult to evaluate, but it is possible to generate 'synthetic' (theoretically generated) seismic data for a given structure, such as a subducting slab. These tests recover the original structure down to a length scale of about 100 km under ideal conditions and suggest that modern tomography is easily capable of resolving features such as subducted slabs and large thermal anomalies. At present discussions are currently focused on the ability of this technique to image narrow (100 km) zones of hot upwelling material in the lower mantle.

Tomography reports the information in terms of percentage difference from the reference models (usually PREM). Seismically slow regions have a negative deviation and seismically fast regions a positive deviation from PREM and are represented on tomographic images as a range of colours. But what do seismically slow and fast regions represent in terms of physical variations in the mantle?

Seismic velocity depends on density, which in turn depends on the temperature and composition of the medium. When a material is heated its density decreases; although density is inversely proportional to the wave velocity (Equation 7.1), the more rapid change in elastic modulus means that the wave velocity also decreases. Likewise, when a material is cooled and its density increases, the wave velocity increases. Therefore, the simplest interpretation of seismic tomographic images is that seismically fast regions represent cool, dense regions and seismically slow regions represent hot, less-dense regions. However, the effects of composition must not be ignored, and the most recent tomographic studies attempt to resolve the effects of temperature and composition on the seismic velocity anomalies. Traditionally, the velocity perturbations are represented using a sliding colour scale, with seismically slow regions being red and seismically fast areas being blue, consistent with their temperature. White regions signify insufficient data to recover useful information. You should also be aware that seismic velocity variations in the upper mantle are generally much greater than those in the lower mantle and often the false colours in tomographic images are 'tuned' to emphasise velocity variations in either one region or the other, but seldom both. Finally, the images are usually sliced in different ways. One method is to slice the image vertically to produce something akin to a geological cross-section of the Earth. Other images peel off constant-depth layers to reveal the global distribution of seismic velocities at different depths in the Earth. Both types of image are used in this chapter.

7.4 Seismic tomographic images

In order to get your eye in at looking at seismic tomographic images you will first consider two regions in which you may have some expectations about the likely seismic structure. Figure 7.4 is a seismic tomographic image across the North American continent, where the Farallon Plate has been subducting beneath the continent for a long time.

- What are the characteristics of the subducting slab?
- The slab is obviously seismically fast and, therefore, colder and denser than the surrounding mantle.

- What major feature can you recognise in the lower mantle?
- There is a major zone of seismically fast material that extends from the point where the subducted slab meets the upper–lower mantle boundary. The fast zone extends at a steeper angle than the upper mantle fast zone to depths of greater than 2000 km.

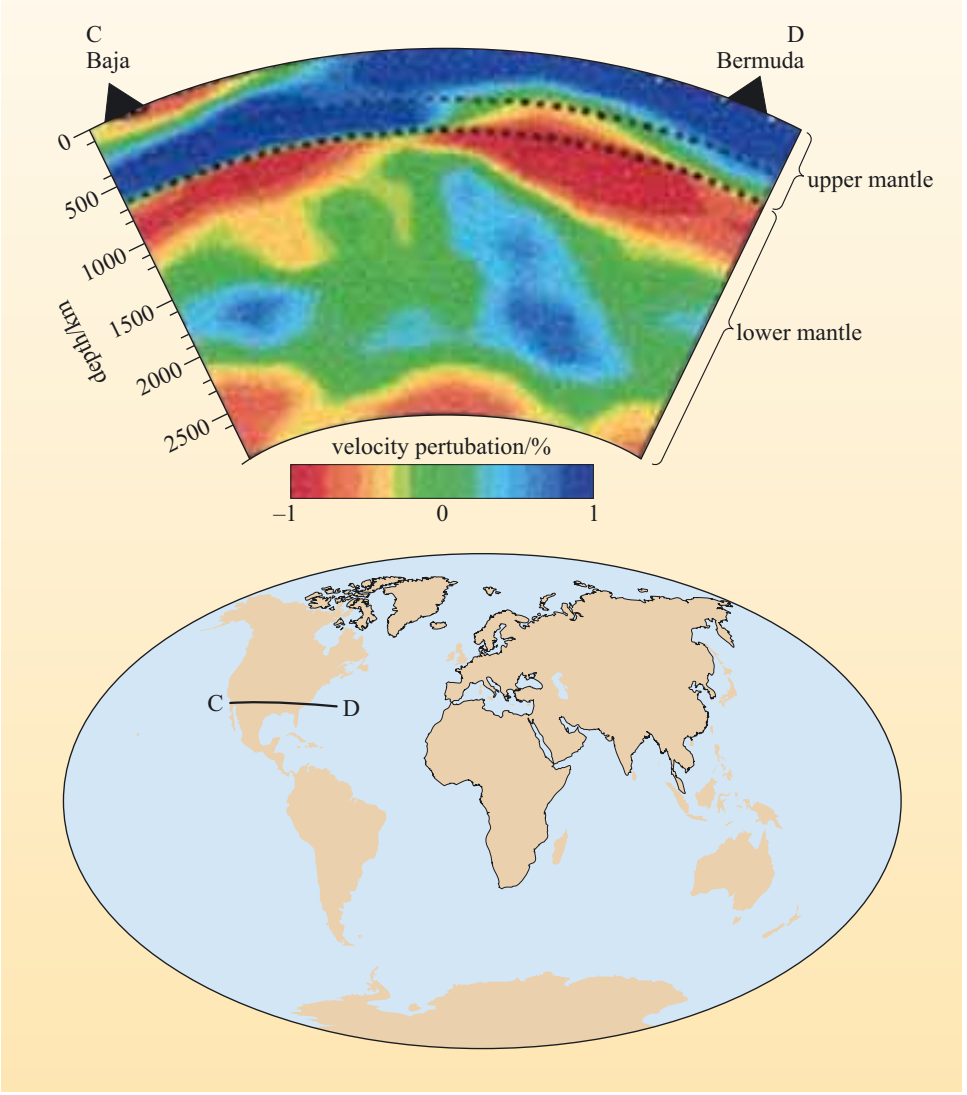


Figure 7.4 Tomographic image along a vertical cross-section across the North American continent. The cross-section runs from Baja to Bermuda and extends to the depth of the CMB. Dotted black lines represent the positions of the 410 km and 670 km discontinuities. Blue (shaded areas to the right of centre in the lower mantle) and red denote the fast- and slow-velocity anomalies respectively. The velocity perturbation scale is shown in per cent. The images are generated from P-waves only. (Lei and Zhao, 2006a)

Figure 7.5 is a N–S seismic tomographic section across the Pacific Ocean from South Pacific to Hawaii.

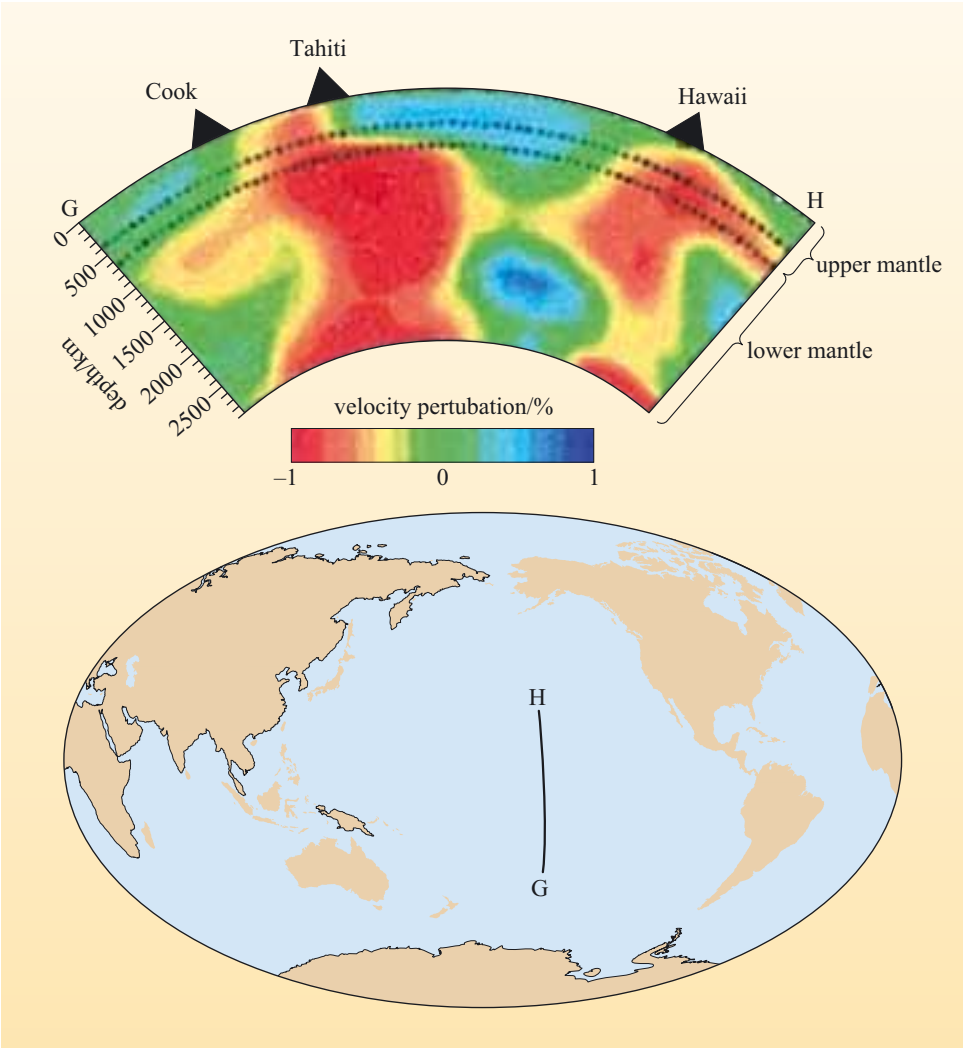


Figure 7.5 Tomographic image along a vertical cross-section across the Pacific Ocean from the South Pacific to Hawaii and extends to the depth of the CMB. Dotted black lines represent the positions of the 410 km and 670 km discontinuities. Red (beneath Tahiti and Hawaii) and blue colours denote the slow and fast velocity anomalies respectively. The velocity perturbation scale is shown in per cent. The images are generated from P-waves only. (Lei and Zhao, 2006a)

- What major features can you recognise on Figure 7.5 in the lower mantle directly below Hawaii and Tahiti?
- There are two major zones of seismically slow material in the lower mantle. One is located beneath the southern Pacific Ocean; the second zone lies beneath Hawaii, but is less laterally extensive and less slow in the middle of the lower mantle.

The large anomaly beneath Tahiti is termed the southwest Pacific super-swell and consists of a broad zone of seismically slow material that extends across the lower mantle. But the question then arises as to what such anomalies might reflect. It is tempting to associate velocity perturbations with just temperature differences, a temptation exacerbated by the use of false colours that have temperature connotations (i.e. blue for cold and red for hot). In the case of seismically fast subduction zones, as in Figure 7.4, the velocity perturbations are consistent with geological and theoretical observations, that slabs should be cold, dense and seismically fast. Tomographic images improve all of the time, and the images presented in this chapter use a variety of P-waves, including those reflected off the outer core. By combining these new P-wave phases with the tomographic inversion, the resolution of the lower mantle is significantly improved. Thus, the imaging of slabs in the lower mantle (Figure 7.4) is generally accepted as being a real temperature contrast.

However, there is considerably more debate about whether geophysicists can separate temperature from compositional effects when it comes to seismically slow zones in the lower mantle. The seismically slow zones in Figure 7.5 are commonly interpreted as being associated with increased temperature, and the sizes of the velocity perturbations are consistent with up to a 250 °C increase in temperature, if temperature was the only variable. It has already been noted in Section 7.3 that calculating uncertainties in tomographic images is difficult, but researchers have recently had success in producing models that calculate the probabilities that determine how much temperature and/or composition play a role in the observed seismic anomalies. Such models can successfully separate the effect of temperature from composition and determine their relative importance to buoyancy forces in the mantle. The major conclusion of these studies is that, in the lower 1000 km of the mantle, compositional variations provide the most important component in the buoyancy forces, although higher temperatures are still present.

In the middle to upper portions of the lower mantle, temperature is the most important component in the buoyancy forces. The difference in the importance of temperature and composition to seismic anomalies relates to the effect of pressure on the thermal expansion of mantle materials. A modern interpretation of the south-west Pacific super-swell is that it represents both a compositional (lower density) and higher temperature anomaly. What causes such large-scale chemical variations is a matter of current discussion; but, as you will explore in Section 7.8, the accumulation of subducted slabs may play an important role. Unfortunately, the ‘snapshot’ aspect of seismic tomography does not indicate whether such large-scale anomalies are stable or long-lived. However, intra-plate volcanism above the current position of the southwest Pacific super-swell has been active for tens of millions of years.

Figure 7.6 shows examples of the most recent (2006) global tomographic images for P-waves, combining a variety of P-wave signals. Each of the 20 images represents a slice of the Earth at a different depth, with the seismic anomalies transposed on to a projection of the Earth’s surface. The centre of the Pacific Ocean is in the middle of the image (located somewhere near Hawaii), with continental Europe at the far right and left edges. The positions of the continental margins and the mid-ocean ridges are marked as dark lines. Black triangles mark the positions of known hot spots.

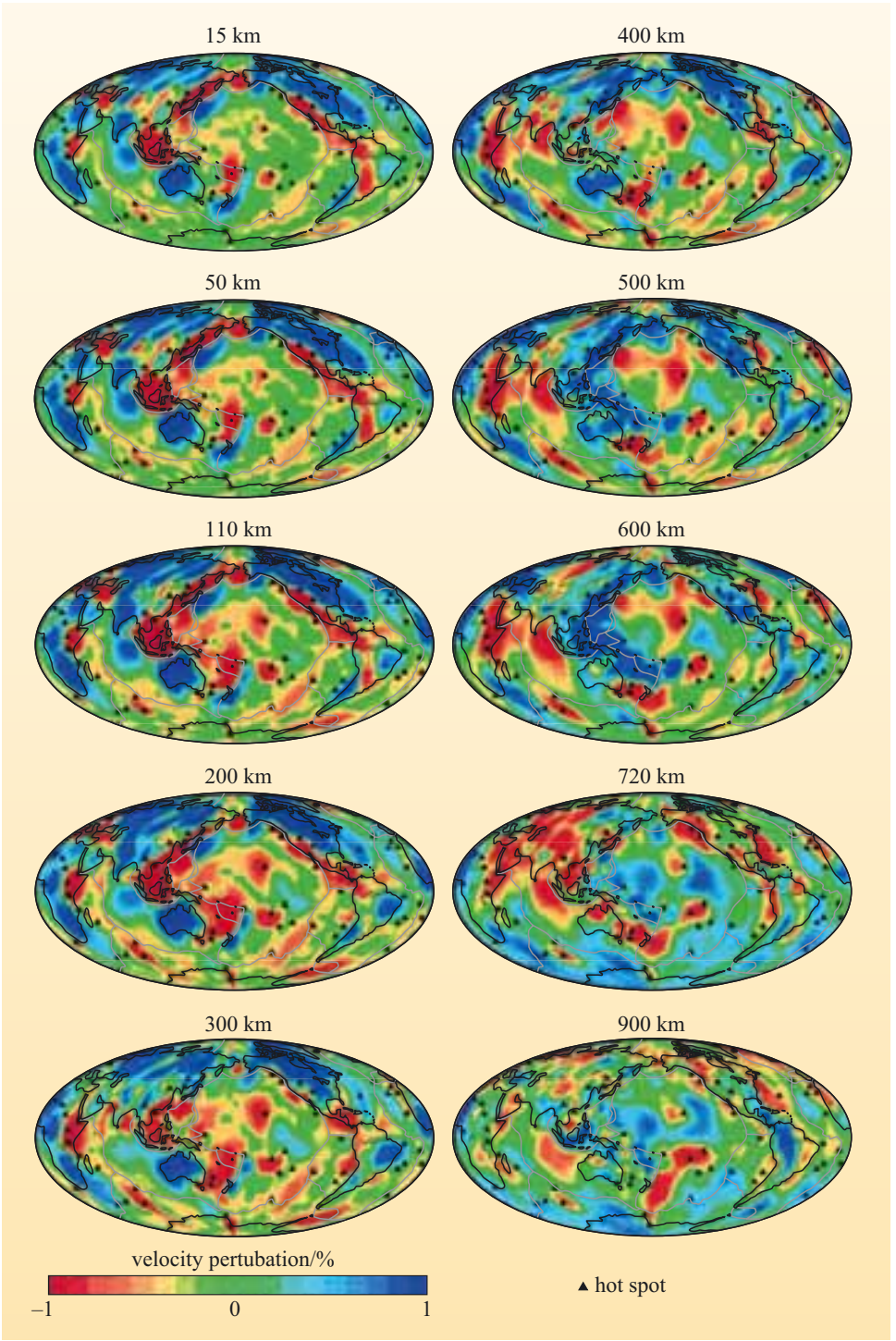


Figure 7.6 Global P-wave tomographic images sliced at different depths throughout the Earth. Red and blue colours denote the slow- and fast-velocity anomalies respectively. Dark lines are continental coastlines and paler lines are tectonic plate boundaries. (Continued overleaf)

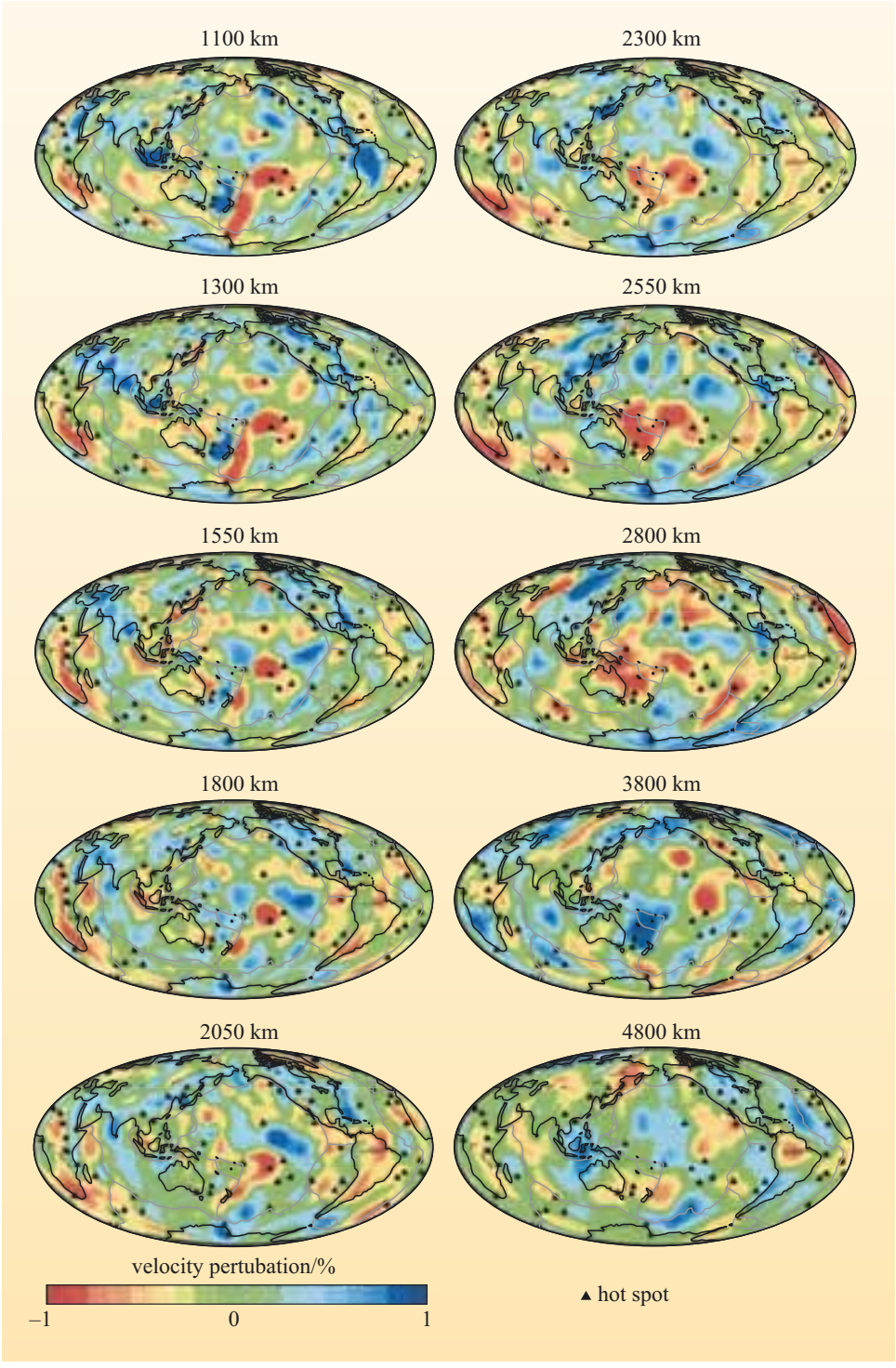


Figure 7.6 Global P-wave tomographic images sliced at different depths throughout the Earth. Red and blue colours denote the slow- and fast-velocity anomalies respectively. Dark lines are continental coastlines and paler lines are tectonic plate boundaries. An enlarged version of Figure 7.6 is in the Appendix. (Lei and Zhao, 2006b)

Question 7.1

With reference to Figure 7.6:

(a) Describe the distribution of seismic anomalies in the continental lithosphere in the top 110 km of the Earth, across North America, Australia and Africa.

(b) Describe the distribution of seismic anomalies along the SE Pacific Rise, which is a mid-ocean ridge, to a depth of 600 km.

(c) Describe the distribution of seismic anomalies along the northern Pacific Rim from North America to Japan and the Mariannas in the upper 600 km of the Earth.

(d) Is there a relationship between surface hot spots (e.g. Hawaii) and seismically slow regions in the upper 2800 km of the Earth?

The answers to the above questions should make you realise that there are some connections between seismic tomographic anomalies and surface features that are clear and unambiguous, and others where the connections are less easy to discern. In particular, the upper mantle structure beneath constructive and destructive plate boundaries and the deeper structure of continents stand out clearly, whereas the association between hot spots and the deep mantle are less apparent. Despite these difficulties, seismic tomography provides an alternative view of a dynamic mantle that contrasts strongly with the rather static image presented by the conventional Earth reference models on which they are based.

7.5 Mantle convection

Several of the chapters in this book describe how solid portions of the Earth move and flow like a fluid, so you should be used to the concept of rocks behaving like a very viscous fluid at high temperatures and over geological timescales.

■ What drives fluid flow?

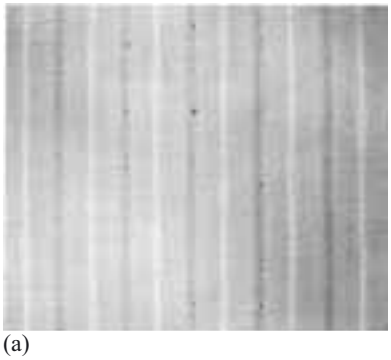
■ Gravitational potential energy and density contrasts.

Dense materials sink and less dense materials float; so, if a part of a fluid body becomes either less dense (or more dense) than its surroundings, then it will flow upwards (or downwards) to restore the system to equilibrium. The forces that drive the fluid to flow are known as **buoyancy forces**.

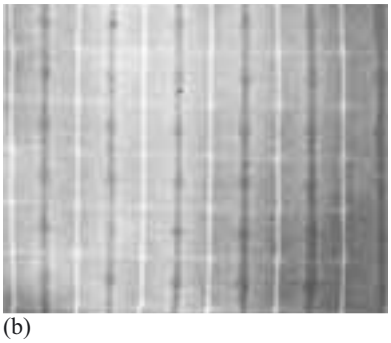
■ Name two causes of density changes in the Earth.

■ Compositional (chemical) and thermal (temperature) variations.

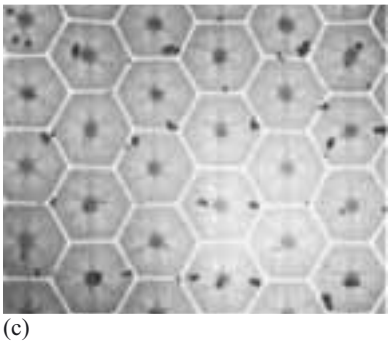
Changes in the composition of any material invariably change its density, and the same is true of the rocks that make up the Earth's mantle. Similarly, when rocks are heated, they expand and their density reduces.



(a)



(b)



(c)

Figure 7.7 (a–c) Evolution of thermal convection within a fluid. Each image is a view of the surface. See text for explanation of the surface patterns. (White, 1988)

Most of you will have come across thermal convection in action if you have ever heated a saucepan of soup on a stove. The soup is a fluid that has two thermal boundary layers: a heated layer at the base and a cooling layer at the surface. Movement within the soup is driven by buoyancy forces that result from changes in density caused by heating the lower boundary layer of the soup.

A slightly more sophisticated experiment entails a tank of viscous fluid (golden syrup is an experimentalist's favourite) heated from below. In this set-up it is possible to observe the onset and evolution of thermal convection within the fluid (Figure 7.7). The fluid at the base of the tank is heated and becomes less dense than the overlying fluid, and this density contrast induces convection: sheets of upwelling hot fluid and downwelling cold fluid will form and transport the heat from the base of the pan (lower boundary layer) to the surface of the syrup (upper boundary layer) (Figure 7.7a). With increased heating the fluid forms a second, weaker, perpendicular set of sheets (Figure 7.7b) and ultimately forms coherent hexagonal convective cells with a central rising plume and six sinking sheets (Figure 7.7c). If heating continues, the cells will break and hot material rises at random. This type of convection in a **Newtonian viscous fluid** is known as **Rayleigh–Bénard convection**, and the convection cells have rather box-like geometries such that their **aspect ratio** (the ratio of the width of the cell to the height of the cell) is approximately one. You may be surprised that a laboratory experiment involving a fish-tank full of golden syrup or a numerical experiment on a computer would be useful in understanding convection in the mantle, which takes places over thousands of kilometres and millions of years. The reason for this comes down to some mathematical expressions that describe the flow properties of convecting fluids (see Box 7.1).

7.5.1 Thermal convection

One of the key properties of a fluid is its viscosity, which is just a measure of the ability of molecules in a fluid to move past each other. Water has a low viscosity of 10^{-3} Pa s (the unit is a pascal second) and golden syrup has a viscosity of 10 Pa s. whereas the mantle is highly viscous and has a value of 10^{21} Pa s. For a Newtonian viscous fluid the stress is proportional to the strain rate:

$$\text{strain rate} \propto \text{stress} \tag{7.2}$$

and the constant of proportionality is the **dynamic viscosity**, which gives

$$\text{strain rate} = \text{dynamic viscosity} \times \text{stress} \tag{7.3}$$

However, the mantle behaves as what is known as a **non-Newtonian fluid**, in which the viscosity varies with the applied strain rate. An everyday example of such a fluid is non-drip paint.

Although the governing equations that describe convection in a heated viscous fluid are well known, they are beyond the scope of this book. However, the flow properties of convective fluids can be described using several **dimensionless numbers** (Box 7.1). Dimensionless numbers are simply mathematical expressions in which the variables (density, times, length, etc.) have been scaled so that the resulting number has no units.

Box 7.1 Thermal convection: the dimensionless numbers

1 The first dimensionless number to consider is called the **Rayleigh number** (*Ra*), which is a measure of the ratio of heat carried by convection relative to heat carried by conduction. (All of the dimensionless numbers described in this box are named after famous physicists.) The *Ra* value essentially determines whether a fluid will convect or not. It is defined as:

$$Ra = \frac{\alpha g d^3 \Delta T}{\kappa \upsilon} \qquad (7.4)$$

where α is the volume coefficient of thermal expansion, g is the acceleration due to gravity, d is the thickness of the layer, ΔT is the temperature difference in excess of the adiabatic gradient across the layer, κ is the thermal diffusivity and υ is the kinematic viscosity, which is given by the following equation:

$$\upsilon = \eta/\rho \qquad (7.5)$$

in which η is the dynamic viscosity and ρ is the density.

The type of flow at a particular *Ra* is always the same irrespective of the size of the system. Therefore, if a laboratory experiment can be carried out that has the same *Ra* as the mantle, then the results from the experiment can be scaled up (with appropriate time, length and viscosity) to those of the mantle and can provide direct information about mantle convection. Similarly, numerical experiments on computers can be used in the same way. So dimensionless numbers are useful!

2 The governing equations for convection can be solved using different *Ra* and different boundary conditions (usually the temperature or heat flux across the top and bottom boundaries). These calculations are not shown here, but one obvious question needs to be asked – at what *Ra* does a fluid begin to convect? This *Ra* is known as the **critical Rayleigh number** (Ra_c) and is defined as

$$Ra_c = \frac{\alpha g d^4 (Q + Ad)}{k \kappa \upsilon} \qquad (7.6)$$

where Q is the heat flow through the lower boundary, A is the internal heat generation and k is the thermal conductivity. Solutions of the governing equations indicate that although the Ra_c depends on several factors, convection will occur when the Ra_c is about 10^3 within a cell with an aspect ratio of 2–3. Vigorous convection occurs when the Ra_c is 10^5 and more irregular convection occurs at $>10^6$.

3 The final dimensionless number you should consider is the thermal **Péclet number** (Pe_t), which is a guide to the relative importance of heat transport by advection relative to that by conduction. It is defined as:

$$Pe_t = \frac{ul}{\kappa} \qquad (7.7)$$

where u is the velocity at which the material is moving, l is the length scale and κ is the thermal diffusion. Equation 7.7 is useful for understanding heat transport where material is being convected upwards, such as under a mid-ocean ridge, where mantle material moves upwards over a discrete distance. Values greater than one indicate that convection rather than conduction is the dominant mechanism of heat transport.

Dimensionless numbers allow mantle convection to be approached somewhat more quantitatively, e.g. by calculating the value of *Ra* for the mantle.

Question 7.2

Calculate the *Ra* value of the mantle from the values below for values of $\Delta T = 1$ and 100 K:

volume coefficient of thermal expansion $\alpha = 2 \times 10^{-5}$ K^{-1}

gravitational acceleration $g = 10$ m s^{-2}

depth of the mantle $d = 2900$ km $= 2.9 \times 10^6$ m

thermal diffusivity of the mantle $\kappa = 10^{-6}$ m^2 s^{-1}

dynamic viscosity $\eta = 10^{21}$ Pa s $= 10^{21}$ kg^{-1} s^{-1}

density $\rho = 3300$ kg m^{-3}

ΔT is difficult to determine for the mantle but is unlikely to be <1 K or much more than 10^2 K. Calculate *Ra* for both cases.

So, does the mantle convect? Calculations for the upper mantle (670 km thick) and the whole mantle (2900 km thick) yield Rayleigh numbers of 10^6 and 6×10^7 respectively, which means that the mantle has an Ra above the Ra_c and will convect. The following question concerns movement of the mantle beneath mid-ocean ridges.

Question 7.3

Mantle decompressing beneath a mid-ocean ridge moves at 0.01 m s^{-1} over 50 km, and the thermal conductivity κ is 10^{-6} m^2 s^{-1}. What is the thermal Péclet number of this system and how is heat transported?

These simple calculations provide some starting points for understanding heat transport in the deeper mantle; they indicate that the mantle convects vigorously and that convection is the dominant form of heat transport within the mantle.

7.5.2 Convection and plate tectonics

The Earth can be seen as a heat engine trying to lose heat, with this convective heat transport in the mantle driving plate tectonics. An alternative view, as developed in Chapter 3, is that plate forces, such as slab pull, drive plate tectonics and that mantle upwelling is a passive response to these forces. Clearly the rigid lithospheric plates are an integral part of the convective system and they dominate the geometry of such a system. Any realistic mantle convection model, therefore, has to incorporate plate-like behaviour.

Experiments with tanks of golden syrup provide important information about some aspects of convection, but are they useful analogues for how the mantle convects? Unfortunately, the answer is no. Convection in these experiments produces cells with aspect ratios of ~1. This is hard to reconcile with plate-scale geometries of 10 000 km (for the Pacific Plate), when the mantle is only 2900 km deep, and more so if the convective system operates only in the upper mantle, which is 670 km thick.

The introduction to the chapter touched on the long-standing debate between whether mantle convection involved two layers of separate convection or a single whole-mantle layer.

- Is there any information from seismic tomography that helps in deciding between the two different models?
- The resolution of slabs subducting through the 670 km boundary reveals a mass flow from the upper to lower mantle. Assuming the lower mantle is not growing and the upper mantle is not shrinking, there must be a return flow of material from the lower mantle back into the upper mantle. Some tomographic images now resolve seismically slow upwelling regions (e.g. Hawaii, East Africa and Iceland), and it is tempting to interpret these as the upwelling of hot material from the lower mantle.

The images indicate that slabs penetrate the lower mantle. In the upper mantle, subducting slabs are very well defined by seismic tomography, whereas in the lower mantle the signal is rather more dispersed and it can be hard to trace slabs all the way to the CMB. Constant-layer slices at depths of 1700–2000 km, however, do record seismically slow regions whose position is consistent with

Cenozoic subduction zones. Many scientists have interpreted the seismically anisotropic behaviour of the D″ layer as being the site of ancient subducted slabs. Therefore, it is clear that slabs do subduct into the lower mantle, but that during the transition through the 670 km boundary and deeper into the deeper mantle their mechanical behaviour changes.

7.5.3 Modelling mantle convection

Computer models are now the preferred method for understanding mantle convection. To illustrate how sensitive models are to changing parameters, Figure 7.8 illustrates three different models for upper mantle convection in which the Rayleigh number is held constant (2.4×10^5) but the source of heat and/or heat flow is varied. In Figure 7.8a the temperature is held constant at the upper boundary and the heat flow is constant at the lower boundary. This produces convection cells with aspect ratios of ~1 and is analogous to the earlier example of convection in a container of golden syrup. In Figure 7.8b the heat flow is held constant across the upper and lower boundaries, which produces much broader convection cells. The final model (Figure 7.8c) has heat supplied from within the mantle and the heat flow is held constant across the upper surface. This produces broad cells with large temperature variations at the boundaries and rapid convection, characteristic of high Rayleigh-number flow. These models illustrate that it is easy to produce high aspect ratio convection cells that have geometries required by plate tectonics.

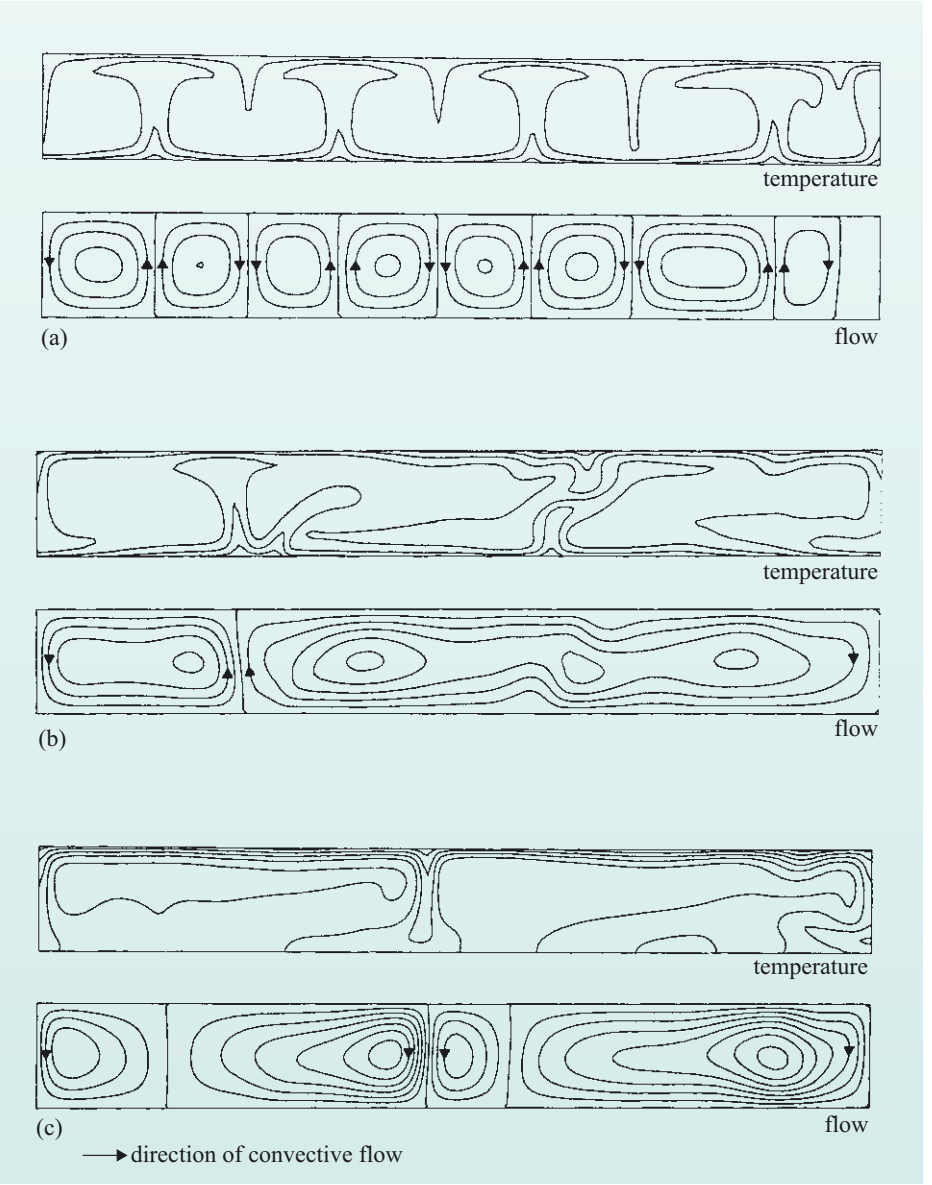


Figure 7.8 (a–c) Temperature distribution and fluid flow for three computer models of convection in the upper mantle. Zones of upward flow relate to upwelling of hot material. The Rayleigh number is (2.4×10^5) and there is no vertical exaggeration. See text for details. (Fowler, 2005)

Successful models of mantle convection must include a range of factors, including the phase changes that you encountered in Section 7.3. Experimental results and thermodynamic calculations have now defined all the likely major phase transitions that occur in the Earth and how these might affect buoyancy forces and convection. An important property of any phase transition is the so-called **Clapeyron slope**, so named after

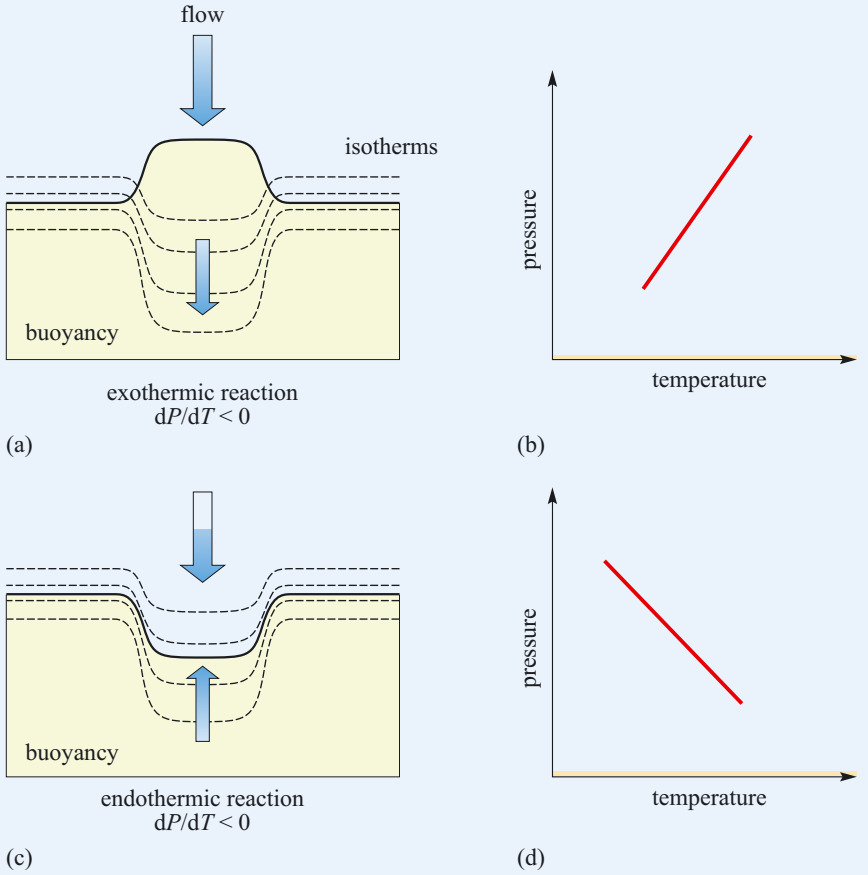
the Clausius–Clapeyron equation in thermodynamics (see Box 7.2) This defines the slope of the phase boundaries on pressure–temperature diagrams and this has a significant effect on the movement of dense material downwards (slabs) and hot material upwards (plumes). If a phase boundary has a positive slope on a pressure–temperature diagram (e.g. the garnet–spinel transition in mantle peridotites discussed in Chapter 4) then it is said to have a positive Clapeyron slope and the reaction is **exothermic**. By contrast, the reaction is **endothermic** for a phase boundary that has a negative Clapeyron slope. The effect on buoyancy is that phase boundaries with negative Clapeyron slopes inhibit slabs penetrating downwards and hotter material moving upwards, whereas a positive Clapeyron slope promotes slab penetration (see Figure 7.9 and Box 7.2). During subduction, as pressure increases, the mineral phases in the subducting slab change progressively to increasingly dense phases. From Figure 7.2, the most important phase changes are γ-olivine (ringwoodite) → Mg-perovskite and the breakdown of majorite → Mg + Ca-perovskite. The first of these has a negative Clapeyron slope whereas the latter has a positive slope. Thus within the cold slab the breakdown of majorite, which may comprise up to 40% of the slab, will occur at a shallower depth than the surrounding mantle and so help to promote deep subduction. By contrast, the negative slope of the breakdown of ringwoodite occurs at a greater depth in the cold slab and may act to prevent a slab penetrating the lower mantle. The net result of these density interactions is that subducting slabs may be stalled at the 670 km boundary. However, the majorite breakdown reaction drives slab penetration if the slab reaches ~800 km depth. The speed of subduction may influence the rapidity of slab penetration with fast-moving old slabs only being stalled at the 670 km boundary for a short period of time (a few million years), but there is no evidence for long-term storage of slabs at the 670 km boundary.

Box 7.2 The Clapeyron law and phase changes

The temperature dependence of a phase transformation is described by the Clapeyron law and is formulated as

$$\frac{dP}{dT} = \frac{\Delta H}{T\Delta V} \qquad (7.8)$$

where dP/dT is simply the change in pressure P as a function of the change in temperature T of a phase boundary (essentially the gradient of the phase boundary on a pressure–temperature diagram), ΔH is the **enthalpy**, ΔV is the volume change of the phase transformation and T is the equilibrium temperature. When a phase transformation has a positive slope ($dP/dT > 1$; Figure 7.9b) the reaction is exothermic ($\Delta H > 0$), whereas the reaction is endothermic ($\Delta H < 0$) for phase transformations with a negative slope (Figure 7.9d). The term ΔV directly relates to the density change that occurs during a phase transformation. The reason the Clapeyron slope is important in both descending subducted slabs and ascending hot mantle is that the phase transformation will occur at a different pressure to the surrounding mantle, causing a buoyancy force. In the case of cold slab that is descending rapidly, a phase transformation with a positive Clapeyron slope will occur at lower pressure than the surrounding mantle. The accompanying density increase and buoyancy decrease will assist in driving the slab downwards (Figure 7.9a). By contrast, a phase transformation with a negative Clapeyron slope will occur at greater depths in the cold slab (Figure 7.9c). This produces a positive buoyancy force that operates against the motion of the slab. To resolve whether such phase transformations have a significant effect of convection, modellers have to balance the various



forces acting on a slab. Integrating such effects into a convection model is complicated, but modellers now agree that transformations with negative Clapeyron slopes only stall convection for geological short periods of time.

Figure 7.9 Effect of Clapeyron slope on the buoyancy forces. (a) The effect of an exothermic reaction on an isotherm in a subducting cool slab. (b) Simplified phase diagram for a phase transformation with a positive Clapeyron slope. (c) The effect of an endothermic reaction on an isotherm in a subducting cool slab. (d) Simplified phase diagram for a phase transformation with a negative Clapeyron slope. (a and c adapted from Poli and Schmidt, 2002)

The other factor that has been added to most models is a depth-dependent viscosity based on theoretical considerations, as shown in Figure 7.10. The viscosity of the lithosphere is high, as it is mechanically strong. Beneath the lithosphere, the viscosity decreases rapidly to a minimum, and then gradually increases with depth to the 670 km discontinuity. Between these two depths the mantle viscosity is lowest, and this corresponds to the asthenosphere. At 670 km there is a viscosity increase of 20–30 times, after which it increases steadily down to the CMB.

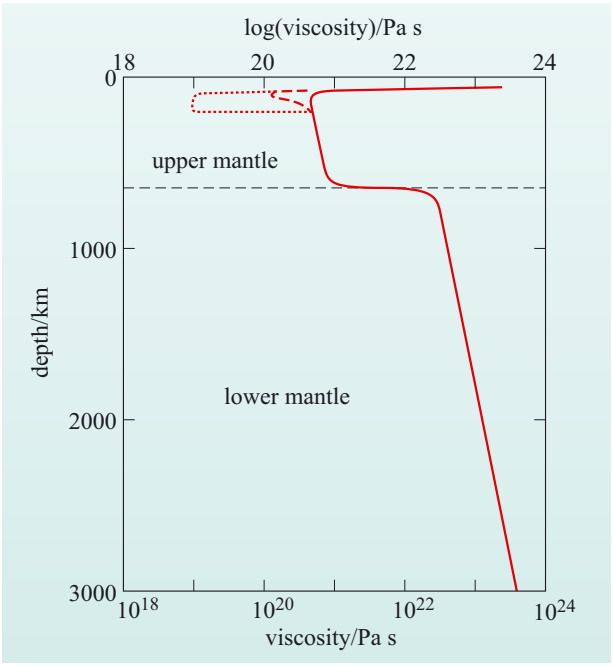


Figure 7.10 A plausible viscosity profile through the mantle. The solid line illustrates the direct, pressure-induced viscosity increase in depth below the lithosphere, and increases due to phase transitions in the transition zone. The dotted lines illustrate the uncertainty in the low-velocity zone in the upper mantle. (Davies and Richards, 1992)

■ How might the viscosity change at the 670 km boundary affect slab subduction?

■ The increase in viscosity will make the transit of slabs into the lower mantle much harder, as it becomes considerably stiffer.

This effect may explain why the tomographic images have the slab more dispersed as it moves below the 670 km boundary, and some images of the Mariana Plate indicate a buckling of the slab at a depth of 670–800 km (Figure 7.11).

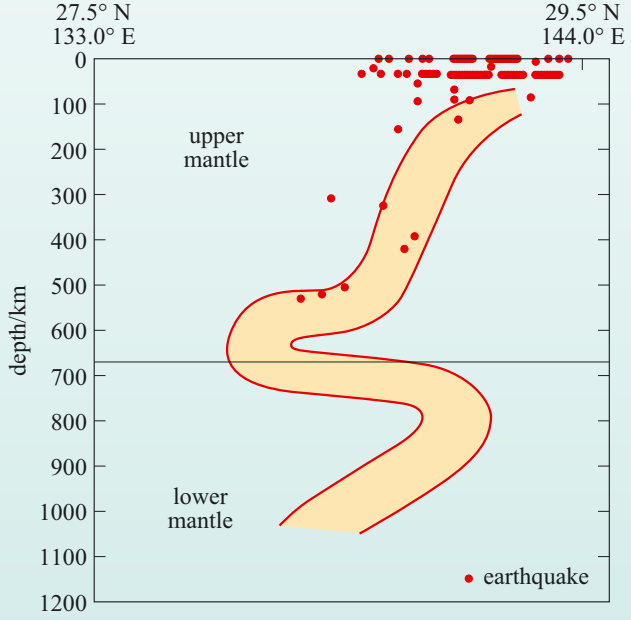


Figure 7.11 Interpretation of the Marianas subduction zone, illustrating the possibility that the slab has buckled upon entering the higher viscosity lower mantle. Shading shows regions of high seismic velocity. (Adapted from Davies and Richards, 1992)

Changes in viscosity and the thermodynamics of mineral reactions are just two of the properties of the mantle that have to be incorporated into realistic models of mantle convection. Other factors include non-Newtonian viscosities, full 3D models and incorporating realistic plate-like features. Linking all these parameters together is mathematically and computationally challenging, but critical results have emerged that impact on our understanding of how the Earth works. These are summarised below.

1 The dominant mode of mantle convection is plate-scale flow with the plates controlling the overall geometry of convection. The plates, therefore, are an integral part of convective flow.

2 Approximately 85% of the Earth's heat loss is through the oceans, and the overall simple topography of the sea floor is consistent with plate-scale flow rather than with many small convection cells in the upper mantle.

3 The 670 km boundary is not a barrier to the flow of slabs into the lower mantle and the upwelling of lower mantle material into the upper mantle. If it were, then the base of the upper mantle would become a thermal boundary layer that would produce a dynamic topography inconsistent with the observed sea-floor topography.

4 An increase in viscosity at the 670 km boundary has several effects. These include: increasing the mixing or stirring time of the lower mantle compared with the upper mantle; producing a time lag between surface plate configurations and deep mantle structure (see tomographic images); complicating the fate of subducted slabs at the bottom of the mantle.

5 Hot upwellings from the CMB can account for 12% of the Earth's heat loss, which is consistent with the heat loss from the core. Such upwellings, known as mantle plumes, are associated with intra-plate volcanism.

Current models suggest that whole mantle convection is the dominant form of mantle flow at the present day and that it explains many large-scale features of the Earth's interior. There are several points in the summary that need some further clarification. For example is there any evidence for hot material moving from the CMB to the Earth's surface? What does the stirring time of the mantle mean and why would anyone want to know? What is the ultimate fate of subducted slabs? Various aspects of these questions will be discussed in the following sections, but it is useful to make some general comments first.

Mantle plumes

The word plume is used in several different ways in the geological literature. It can simply mean an upwelling body of fluid, but because early models of mantle convection were based on simple golden syrup experiments that invoked lots of convection cells in the upper mantle, it is often used to explain all manner of slightly anomalous volcanism that occurs within a plate. Most workers would now regard mantle plumes as expressions of deep-seated mantle upwelling that produce extensive volcanism and have obvious geophysical (gravitational, thermal and seismic anomalies) and topographic signatures. The number of plumes that match these criteria is around 15 and includes Hawaii, Iceland and East Africa. Large igneous provinces from the geological record, such as the Ontong Java Plateau, are also regarded as volcanism associated with a plume. For these types of plume there are well-constrained

dynamic models that draw material from the CMB and move upwards entraining some lower mantle material. This source material, which is hotter than ambient mantle, ultimately melts in the upper mantle at depths of ~100 km.

Stirring times

Stirring times are simply how long it takes to destroy chemical heterogeneities, such as subducted slabs in the convecting mantle. To continue with the cooking analogies for mantle dynamics, if you add some jam to a bowl of thick semolina and start to stir, the effect will be to streak out the jam. However, after some time you will distribute the jam throughout the semolina which will eventually turn pink. By contrast, if you added some red food dye to a bowl of water and stirred the two would mix rapidly. The two systems have different mixing times because their viscosities are different. We know from studies of mid-ocean ridge basalts (MORBs) that their upper mantle source region is compositionally homogeneous and has a relatively uniform temperature (Chapter 4), from which you can infer that it is well mixed. Hence, the mixing time of the upper mantle must be substantially less than the age of the Earth. The mixing time of the lower mantle is thought to be longer because of its higher viscosity; therefore, mantle heterogeneities may be preserved for much longer. The actual stirring time for the lower mantle is unknown, but all models produce reasonably well-mixed lower mantles within about 1–2 billion years and upper mantles that are well mixed in much shorter timescales. An example of a computer simulation of mantle stirring is presented in Figure 7.12.

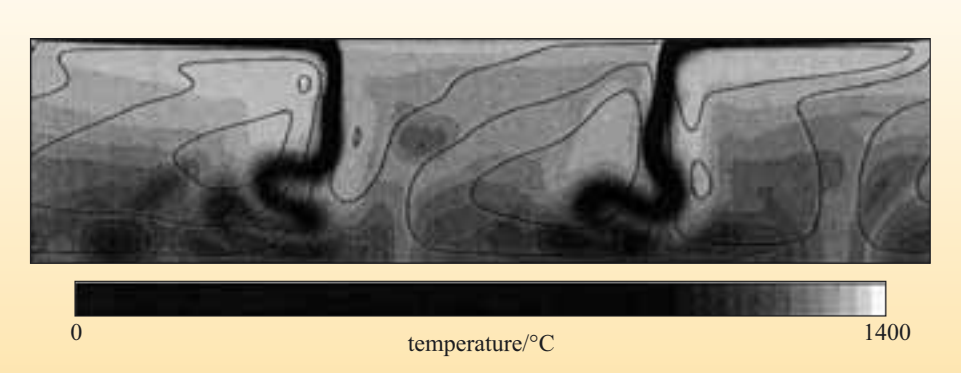


Figure 7.12 Mixing of a passive tracer in a modern, convecting mantle. The tracers are subducted slabs and the mixing is simply shown by temperature variations from 0 °C (black for the slab at the surface) to homogenised higher temperatures (lighter colours). The thin black lines represent streamlines that show the instantaneous direction of fluid flow. (Davies, 2002)

The timescales of stirring have implications for geochemists. The longer a particular compositional heterogeneity is preserved, the more likely it is to generate variations in radiogenic isotope ratios. The observation that some **ocean island basalts** (**OIBs**) have isotope ratios that are different from those in MORB is strong evidence that some regions of the Earth's interior can remain separate from the convecting upper mantle for periods of 1–2 Ga. This geochemical constraint requires some mechanism for preserving portions of the mantle with distinct compositions; therefore, the ultimate fate of subducted slabs is crucial in this debate. The transit of most slabs to 1500–1700 km depth in the Earth can be traced readily by seismic tomography, and there is no dynamic reason why they should not traverse the last 1000 km to the CMB. Impressed by this observation, many scientists have suggested that some portions of the subducted slabs reach the CMB and founder in the D″ layer. This slightly ad hoc arrangement allows material to be stored in the D″ for billions of years, where it slowly warms up before being eventually convected back into the mantle.

7.6 Intra-plate volcanism: the Hawaiian connection

7.6.1 Sources and causes of intra-plate volcanism

Volcanism on most of the planet is associated with plate boundaries such as mid-ocean ridges and subduction zones. However, there is an important class of volcanism that is located away from plate boundaries that is grouped under the general term of intra-plate volcanism. Such volcanism can be found on the continents (e.g. East African Rift and Yellowstone, USA) or in the oceans. There are many examples of oceanic intra-plate volcanism, but the best known and most closely studied is located in the Hawaiian Islands.

The Hawaiian Islands are situated in the centre of the Pacific Ocean on oceanic crust that is approximately 80–100 million years old. In Chapter 3, the age progression along the Hawaiian Chain was used to derive the true motion of the Pacific Plate.

- What does this use of hot-spot volcanism tell you about the location of the source of the heat that produces Hawaiian magmatism?
- It must be located beneath the plate.

You should also recall from Chapter 3 that the oceanic lithosphere cools, subsides and thickens as it ages. As a result, the lithosphere beneath Hawaii is about 60–70 km thick.

- If the lithosphere is 60–70 km thick, what does this suggest about the temperature of the asthenosphere beneath Hawaii?
- For partial melting to occur beneath the lithosphere at 60–70 km, the asthenosphere must have a temperature greater than that of the ambient mantle (see Section 4.5).

The lithosphere behaves as a cool lid at a temperature below the mantle solidus, and so for melting to occur beneath this lid the mantle temperature must exceed the solidus at higher pressures than beneath mid-ocean ridges. With a normal mantle potential temperature the mantle starts to melt at about 50 km depth, and so mantle with a normal potential temperature should not melt beneath the 60–70 km thick lithosphere beneath Hawaii. From Chapter 4, you should recall that the control on the maximum depth of melting in a simple decompression melting system is the potential temperature of the upwelling mantle: the higher the potential temperature, the deeper the onset of melting. Thus, for melting to occur beneath Hawaii the mantle must have a higher potential temperature than the mantle beneath ocean ridges. For Hawaii, it has been estimated that the potential temperature is about 1480 °C, which is about 200 °C higher than the ambient mantle temperature. This inferred higher temperature beneath Hawaii and other intra-plate volcanoes is the reason why they are termed hot-spot volcanoes. However, the cause of these higher than average mantle temperatures away from plate margins has been, and continues to be, the source of much controversy.

The realisation that the heat source for hot-spot volcanism lies beneath the plates led to the development of the ideas behind mantle plumes, i.e. focused columns or jets of hot material rising from deep within the mantle. The active volcanic islands are located on a broad topographic high known as the Hawaiian swell, which is associated with a similarly broad negative free-air gravity anomaly. This association has been interpreted as resulting from uplift caused by upwelling in the mantle beneath the islands. The islands themselves are characterised by a large, narrow positive free-air anomaly that results from the excess mass focused within the volcanic islands on top of and flexing the underlying lithosphere.

The plume model provides an explanation for some of the features of ocean island volcanism, but what is the evidence that Hawaiian basalts do originate from hotter than usual mantle? The answer lies in the compositions of the basalts erupted. Decompression melting beneath Hawaii requires that melting starts at higher pressures and temperatures than beneath mid-ocean ridges and produces hotter melts (Figure 4.24d).

- Can your recall from Chapter 4 which aspect of basalt composition is related to its temperature?
- The MgO content of basaltic magmas increases with temperature.

This observation from experimental studies offers a test of the model that Hawaiian basalts are indeed derived from mantle with higher than normal potential temperatures. If melting beneath Hawaii occurred at higher temperatures, then the primary melts should have higher MgO contents. Submarine glasses from the Kilauea volcano have MgO contents up to 18 wt%, which is higher than the most MgO-rich MORB glasses analysed (10–11 wt%). This information on its own would suggest that the mantle beneath Hawaii is hotter than that below a mid-ocean ridge. However, most MORBs are fractionated, so the comparison is not strictly valid. To assess whether Hawaiian primary melts are generated at higher temperatures than MORB primary melts, the primitive melt compositions have to be calculated, and exactly how this is done is a matter of debate and beyond the scope of this chapter. In summary, most scientists now agree that primitive MORBs have MgO contents seldom greater than 13 wt%, whereas Hawaiian lavas with similarly primitive Mg# have MgO contents of up to 18 wt%, implying significantly higher melting temperatures in the magma source region beneath Hawaii.

A second implication of the plume model for the origin of the Hawaiian basalts is that they originate from greater depths than most MORBs. Again, major element compositions can provide some information in this regard, but it is less easy to relate such compositional variations to pressure than it is to temperature. However, clear pressure indicators can be recognised in the behaviour of certain trace elements in the Hawaiian basalts and their partitioning between the melt and residual mineral phases.

- What is the major difference between the mineralogy of mantle at shallow depths beneath a mid-ocean ridge and that beneath lithosphere up to 75 km thick?
- The shallow mantle has a spinel lherzolite assemblage, whereas at higher pressures the aluminous mineral is garnet (Figure 4.15).

Changes in the nature of the residual minerals during melting have a profound effect on the trace element contents of the melt. Figure 7.13 is a primitive mantle normalised trace element diagram for a primitive Hawaiian (from Kilauea) lava compared with a MORB.

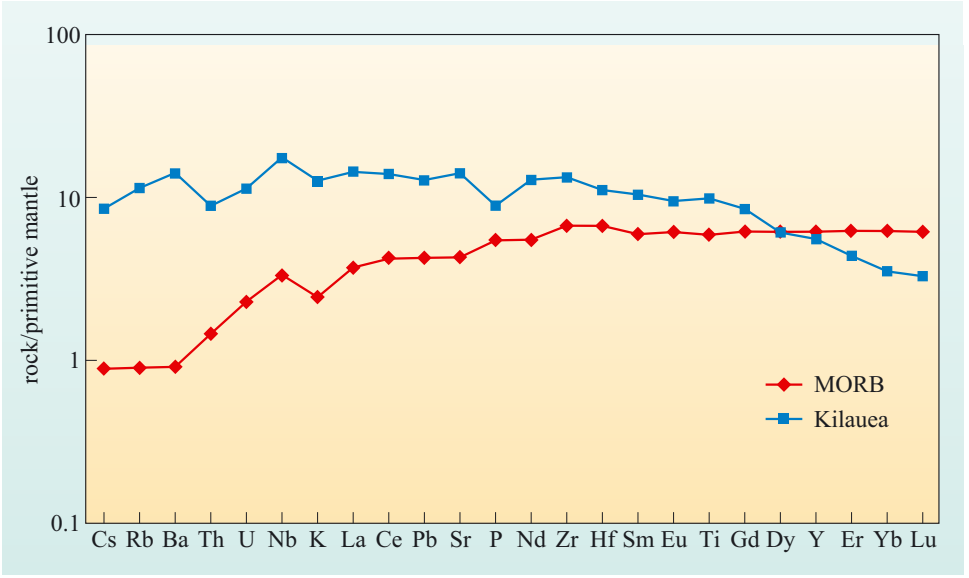


Figure 7.13 Primitive mantle normalised trace element diagram for an average MORB and a primitive lava from Kilauea, Hawaii.

■ How does the Hawaiian trace element pattern differ from that of a MORB?

■ The principal difference with the MORB pattern is that Hawaiian lavas have lower heavy rare earth element (HREE) (Dy–Lu) concentrations and much higher highly incompatible element concentrations (see Section 5.8 for partition coefficients and the Appendix for element names).

The low HREE contents could be explained by higher degrees of partial melting than a MORB, but the higher incompatible element concentrations are more consistent with small degrees of partial melting. So which of these is correct?

If the Hawaiian basalts are generated at depth beneath the lithosphere, then garnet should be an important mineral phase in the source region. Table 5.1 lists the distribution coefficients for a range of elements between garnet and melt.

■ Which element is compatible in garnet?

■ Yttrium (Y).

Yttrium is geochemically similar to the HREEs, and garnet has HREE distribution coefficients that have values between 4 and 7, so the bulk distribution coefficients for the HREEs in a garnet–peridotite are much higher than those for the light rare earth elements (LREEs), so that, at a given melt fraction, the LREEs are more enriched in the melt than the HREEs. By contrast, the bulk partition coefficients for the LREEs for both spinel and garnet lherzolite assemblages are similarly low – the LREEs are highly incompatible during melting at all pressures. The combination of high LREE and low HREE contents relative to MORBs, therefore, implies smaller degrees of melting in the presence of residual garnet. In detail, the trace element patterns are consistent with less than 5% partial melting of the mantle at a depth of 75–120 km beneath Hawaii, which is consistent with melting of hotter mantle beneath the oceanic lithosphere.

Although the amount of melting required to generate the Hawaiian basalts (~5%) is smaller than that required to generate MORBs (15–20%), this does not mean that the flux of magma to the surface is similarly small. At mid-ocean ridges, the relative constancy of the crustal thickness (~7 km) means that the volume of melt extracted from the mantle per unit of time is simply related to the spreading rate. In Hawaii, where there is no sea-floor spreading, the volume of melt generated is related to the amount of mantle that upwells through the melting zone beneath the lithosphere. Observations of eruption rates at Hawaii show that the magma production rate in the source region is about 0.2 $km^3\ y^{-1}$, which compares with a total magma production rate of the entire ocean ridge system of about 20 $km^3\ y^{-1}$. In other words, the single edifice at Hawaii erupts the equivalent of 1% of the volume erupted by the 20 000 km length of the global ocean-ridge system. Generating relatively large volumes of magma in such a focused system implies that upwelling in the mantle plume beneath Hawaii is strong.

Before accepting the hot mantle explanation for Hawaiian magmatism as the only explanation, it is instructive to consider other possibilities. For example, if the mantle solidus is reduced, then it might become possible for melts to be generated by decompression at greater depths without having to invoke higher temperatures.

- From what you have learnt in previous chapters, how might the mantle solidus temperature be lowered?
- By lowering the Mg/Fe ratio (Chapter 4) and by adding H_2O (Chapter 5).

Peridotite minerals with reduced Mg# melt at lower temperatures than those with more magnesian minerals, and so a mantle composition that is richer in iron will melt more easily than one with a typical peridotite Mg# of 88–90. Similarly, if the mantle source region is H_2O rich, then it will melt at a lower temperature, as occurs beneath island arcs. In both cases, such source variations will be reflected in the compositions of the primary melts generated, which you would expect to be more water rich and more iron rich than MORBs. For example, primitive lavas from Hawaii have between 0.3 wt% and 0.5 wt% water, which is slightly higher than most MORBs, but much less than in island-arc basalts.

- Is there a simple explanation for the higher water content of the Hawaiian melts?
- Because water is an incompatible element, it will be more concentrated in small melt fractions; and as Hawaiian primary magmas are the result of smaller degrees of melting than MORBs, this might be a reflection of that aspect of melt generation.

Establishing the iron content of the Hawaiian source region is harder to pin down because the iron content of primary melts increases with pressure, and disentangling the effects of pressure from source composition is not straightforward. Suffice to say that while the high temperature mantle plume model is widely accepted it is not universally so. However, any alternative must also explain the geophysical and topographic effects and the magma production rates. At the time of writing, of all the alternatives only the hot mantle plume model gets anywhere close to explaining the majority of the observations.

7.6.2 The source of the Hawaiian mantle plume

As you have read previously, hot upwellings are common features in laboratory simulations of convection. These experiments demonstrate that convective upwelling originates at a thermal or density boundary. In a saucepan of soup, for example, or in the laboratory container of golden syrup, upwelling plume-like structures originate from the bottom of the saucepan.

- From which boundary layers in the Earth could mantle plumes originate?
- Several boundaries are known about from tomographic and seismic profiles of the Earth. The most important are the 670 km discontinuity and the CMB.

The geophysical evidence for the deeper origins of the Hawaiian plume remains unclear. Hawaii has been the location of extensive seismic tomographic studies primarily because it is the place where the plume hypothesis was first put forward. Seismically slow regions are readily identified beneath Hawaii in the upper mantle demonstrating that the plume has a physical presence at least down to the 670 km discontinuity. Moreover, several studies have located significant seismic anomalies in the D″ layer just above the CMB. But the key question is whether there is a link between the D″ layer and the upper mantle. The most recent tomographic images based on a variety of different waveforms seem to suggest this may be the case (Figures 7.14 and 7.5), but the connection between the Hawaiian plume (as imaged in the upper mantle) and features in the deep mantle (especially the D″ layer) remain speculative. However, given that the Hawaiian plume is the largest that is currently active on Earth, if any plume were to originate from such great depth it will be beneath Hawaii.

Petrological and geochemical information about where melting takes place in the mantle is an important constraint on the conditions in the mantle beneath Hawaii; but if Hawaiian lavas originate from some sort of plume or hot spot that originates deep within the Earth, then there should be some differences in the composition of the source material that are reflected in the compositions of the lavas. The trace element patterns of the Hawaiian lavas presented in Figure 7.13 primarily provide information about partial melting, although there are various trace element ratios that are used to try to understand whether the mantle source is different from MORB. For example, the ratios of some highly incompatible elements, such as barium (Ba)/niobium (Nb), are different in Hawaiian lavas compared with MORBs, and as these two elements are not fractionated during melting this probably reflects a real compositional difference between the two source regions. However, with trace element ratios there is always the possibility that they have been affected in the tortuous passage from the mantle source to the surface.

By contrast, isotope ratios are less likely to be fractionated from each other during these processes. Consequently, isotope ratios, and especially correlations between isotope ratios and other compositional parameters, have been used to speculate on the origins of the material incorporated in mantle plumes. Radiogenic isotopes are especially useful in this respect because they reflect long-term variations in the parent–daughter elemental ratios in the mantle source, as well as placing constraints on the time required for a particular isotopic signature to develop.

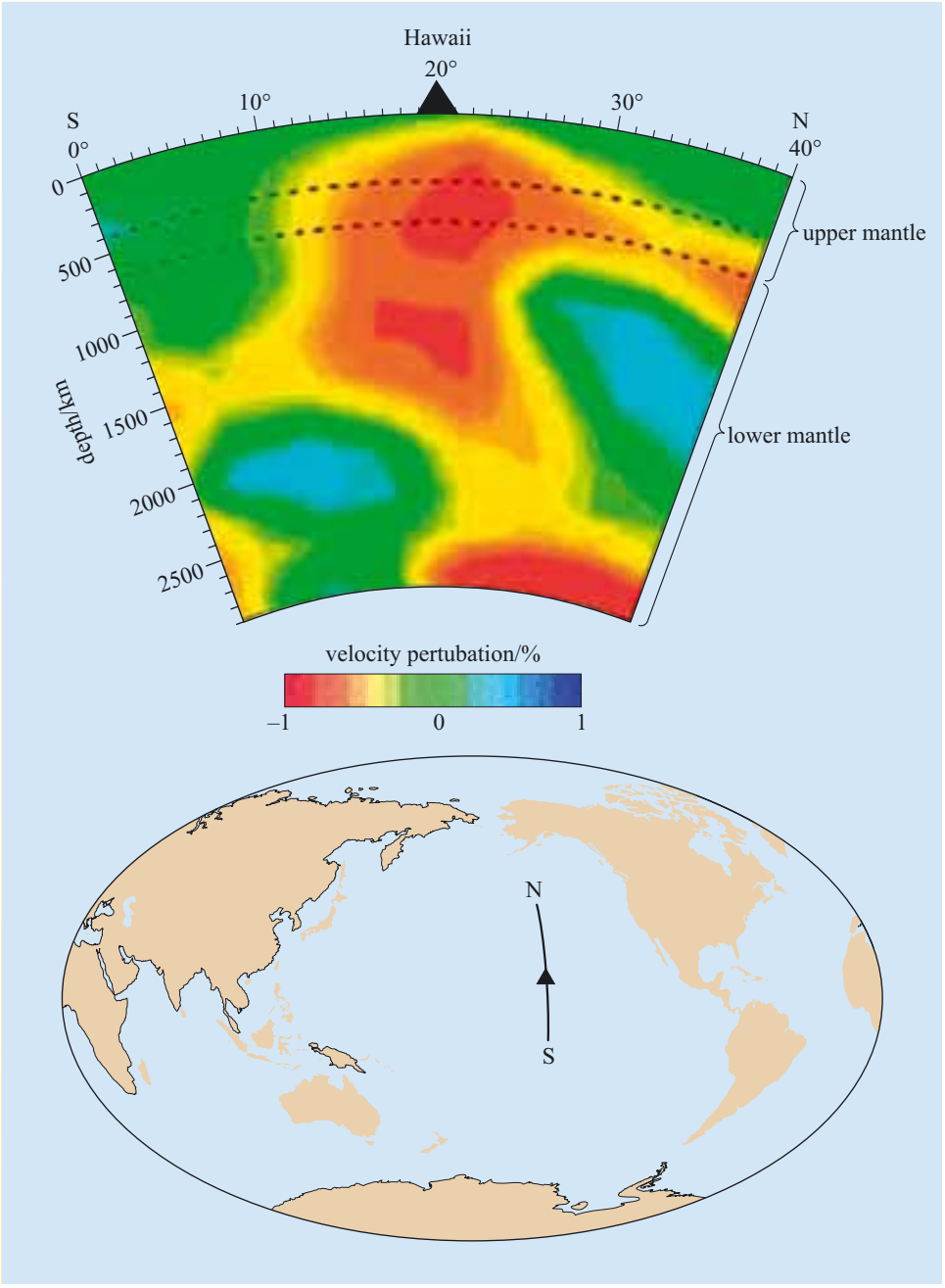


Figure 7.14 P-wave tomographic image along a north–south vertical cross-section passing through the Hawaiian hot spot and extending to the depth of the CMB. Dotted black lines represent the positions of the 410 km and 670 km discontinuities. Red (below Hawaii to bottom right-hand corner) and blue (oval 'blobs' at 40° N in the lower mantle) colours denote the slow- and fast-velocity anomalies respectively. (Lei and Zhao, 2006c)

Hawaiian lavas display a much wider isotope composition than MORBs from all three oceans, suggesting they have a more complicated source region than the well-mixed MORB source. Hawaiian lavas also exhibit good correlations between major elements and radiogenic isotopes. For example, FeO and SiO_2 contents (corrected for olivine fractionation back to an Mg# in equilibrium with mantle olivine) correlate with neodymium (Nd) and osmium (Os) isotope ratios (Figure 7.15). These data confirm the different source characteristics of the Hawaiian basalts, but without further information it is difficult to argue for one particular origin for the Hawaiian plume over another. One popular idea is that the compositional variations in Hawaiian basalts reflect the recycling of ancient subducted oceanic lithosphere, the arguments for which are explored in the following section.

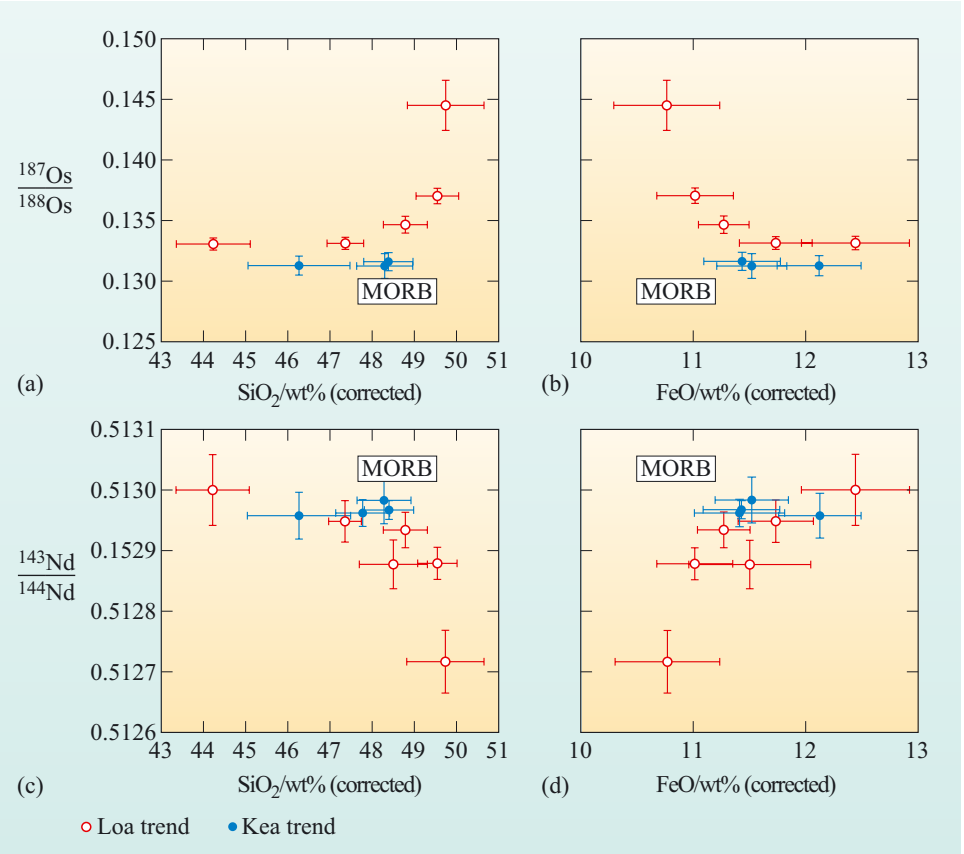


Figure 7.15 Plot of $^{187}Os/^{188}Os$ versus (a) SiO_2/wt% (corrected) and (b) FeO/wt% (corrected) and $^{143}Nd/^{144}Nd$ versus (c) SiO_2/wt% (corrected) and (d) FeO/wt% (corrected). Increasing $^{187}Os/^{188}Os$ and SiO_2, and FeO and decreasing $^{143}Nd/^{144}Nd$ are interpreted as an increased recycled oceanic crust component. (Hauri, 1996)

7.6.3 Recycling of oceanic crust

In Chapter 5 it was demonstrated that a variety of elements are recycled through modern day subduction zones. Elements such as thorium (Th) and beryllium (Be) may be very efficiently recycled from the slab sediment back into the arc crust but other elements may be transferred to the deeper mantle. The composition of the subducted oceanic lithosphere has the same bulk composition as the upper mantle except that it is divided into a basaltic portion and a residual mantle. The basaltic part of the slab has higher SiO_2 and lower MgO and has a lower Mg# than the residual mantle. It is also enriched in incompatible trace elements compared with the residual mantle. In terms of any additional return of elements to the mantle the key aspect is whether hydrothermal alteration of the oceanic crust and subsequent processing during subduction adds or modifies elements or element ratios.

■ Referring back to Chapter 5, which elements are added to the altered oceanic crust during hydrothermal alteration at the ridge crest?

■ The elements uranium (U), caesium (Cs), rubidium (Rb) and potassium (K) are added in significant quantities, and barium (Ba), lead (Pb) and strontium (Sr) to a lesser degree.

In Chapter 5 it was demonstrated that although these elements may be recycled from the basaltic slab by dehydration reaction back into the mantle wedge, there was strong evidence that subducted sediment was the major source of U, Ba and

Rb in arc lavas. This means that the subducted slab may well carry its high U/Pb ratio through the subduction zone and back into the mantle. The generation of high U/Pb ratios in subducted oceanic crust has important implications for Pb isotopes (Box 7.3)

Box 7.3 Lead isotopes

You have already met Pb isotopes in Chapter 2, because they produce the most accurate dates for the age of the Earth. The Pb isotope system is complicated because there are three decay schemes that ultimately produce Pb isotopes: two based on the decay of U isotopes and one based on the decay of a Th isotope. The three decay systems are as follows:

- ^{238}U decays to ^{206}Pb by a combination of alpha (α) and beta (β) decay with a half-life of 4.468 Ga.
- ^{235}U decays to ^{207}Pb by a combination of α and β decay with a half-life of 0.7038 Ga.
- ^{232}Th decays to ^{208}Pb by a combination of α and β decay with a half-life of 14.010 Ga.

^{204}Pb is a stable isotope of Pb, so all Pb isotope compositions are reported as ratios with respect to ^{204}Pb, i.e. $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$.

Geochemists like Pb isotopes because they evolve by decay of two different parent elements that have contrasting geochemical behaviour. Also, the three decay schemes have very different half-lives. This allows processes to be identified that happen over different timescales. ^{238}U decays with a half-life similar to the age of the Earth, whereas ^{235}U has already decayed over six half-lives, so there is now only about 1/64 of the original ^{235}U left in the Earth. Therefore, any process that fractionates U from Pb early in Earth history will potentially produce different $^{206}Pb/^{204}Pb$ ratios relative to $^{207}Pb/^{204}Pb$ compared with a process that happened later in Earth history because of the more rapid decay of ^{235}U relative to ^{238}U.

You are not expected to become experts in the interpretation of Pb isotopes, but consider the following question:

■ What would happen over time if the slab has a higher U/Pb ratio than the surrounding mantle?

■ Over time the slab will evolve to have high $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ ratios relative to the mantle.

The fractionation of U from Pb in the oceanic crust is due in part to alteration that occurs during the hydrothermal interaction close to mid-ocean ridges, and so the presence of high $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ ratios in OIBs is frequently interpreted to reflect the presence of recycled altered oceanic crust in the mantle source region.

Hydrothermal alteration is only one of the processes that control the parent–daughter ratios of subducted oceanic crust. Another occurs during mantle melting

in the generation of MORB magma. This fractionates most radioactive parent-daughter couples by only modest amounts but one is affected more strongly than the rest. During partial melting of the mantle, the elements rhenium (Re) and osmium (Os), which make up a long-lived isotope system (Table 2.1), are strongly fractionated. Re is incompatible and is concentrated in the oceanic crust, whereas Os is compatible and is concentrated in the residual mantle. The ^{187}Re isotope decays to the ^{187}Os isotope by β-decay with a half-life of ~42 billion years, so variations in the ratio of ^{187}Os to another non-radiogenic isotope of Os, usually ^{188}Os, reflect both the time since Re/Os fractionation and the Re/Os ratio. What makes Os isotopes particularly useful is the extreme fractionation of the Re/Os parent/daughter ratio during partial melting. Consequently, MORBs have Re/Os ratios of several hundred, whereas the mantle has an Re/Os ratio of 0.08. Even over tens of millions of years the subducted slab can double its ^{187}Os/^{188}Os ratio, and over a billion years the ratio may increase by a factor of 40. These large shifts in isotope composition are very unusual in most materials that might be recycled into the mantle.

The obvious question to ask is whether there are any lavas that have high ^{206}Pb/^{204}Pb, ^{207}Pb/^{204}Pb and high ^{187}Os/^{188}Os ratios. The answer is yes (Figure 7.16). They are found on the small ocean islands of Tubuai and Mangaia in the Southwest Pacific, associated with the Cook–Austral islands hot spot, and on the island of St Helena in the South Atlantic. The isotope ratios of the lavas from these islands are such that they require up to 2 Ga to evolve from values typical of the source region of MORB to their present-day values and they provide convincing evidence that the mantle is capable of preserving compositional heterogeneities for this period of time.

Compared with basalts from the Hawaiian Islands, the lavas from Tubuai and Mangaia have a much stronger and more uniform isotope signature. In Hawaii, the large range in isotopic data and the correlation of various isotope ratios with major elements is thought to represent mixing of melts from recycled slab material and mantle peridotite. Increasing ^{187}Os/^{188}Os and SiO_2, and decreasing ^{143}Nd/^{144}Nd are all consistent with the compositional properties of (melts derived from) recycled oceanic crust, and the trends illustrated in Figure 7.15 indicate mixing between recycled oceanic crust and surrounding mantle.

It should be noted that although the association of isotope

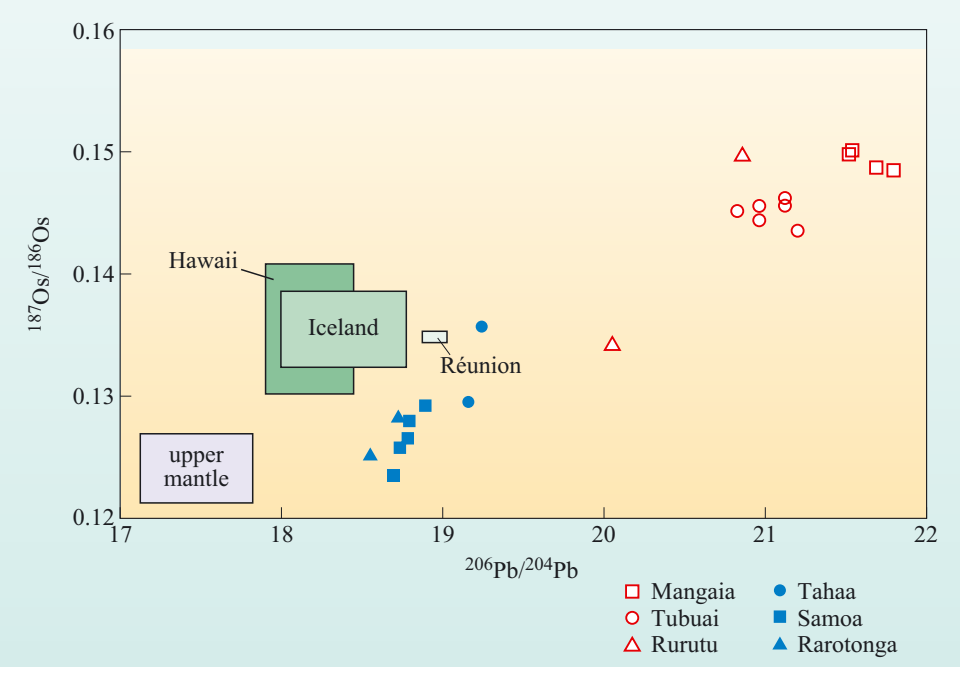


Figure 7.16 Plot of ^{187}Os/^{188}Os versus ^{206}Pb/^{204}Pb for OIBs. MORB is represented by the upper mantle. The three boxes include data from Hawaii, Iceland and Réunion. (Adapted from Hauri and Hart, 1993)

variations in plume-related basalts with recycled oceanic crust and lithosphere is currently very popular, it is not a unique interpretation of the data. For example U/Pb and Re/Os fractionation can be effected by processes other than those occurring at mid-ocean ridges. Any melt generated within the mantle can develop high U/Pb and Re/Os ratios; the separation of sulfide minerals and liquids is known to fractionate the Re/Os ratio and will also fractionate lithophile U from chalcophile Pb, albeit to an unknown extent. But regardless of the process of element fractionation, the degree of isotope variation and the lengths of the parent isotope half-lives require that certain parts of the mantle remain isolated from the well-mixed asthenosphere for periods of up to 2 Ga.

The problem still remains as to where material might be stored for such periods. The great temptation is to link the images from seismic tomography to the conclusions from isotope geochemistry. Perhaps subducted slabs that penetrate the 670 km discontinuity accumulate at the base of the mantle where they constitute the D″ layer. After a period of time, determined by its density and thermal properties, material in the D″ layer becomes buoyant and re-enters the mantle convection cycle. Testing such ideas using models of mantle circulation, isotope geochemistry and mantle tomography is currently at the leading edge of scientific investigation, and even though Hawaii is perhaps the most studied group of islands on the planet, there is still as yet no consensus on the origins of the Earth's most active mantle plume.

7.7 The water cycle in the solid Earth

So far in this chapter, the emphasis has been on the way in which heat and gravity drive the inexorable evolution of the Earth, concentrating low-density, incompatible and volatile components in its outer layers. The dominant expression of that process today is plate tectonics, which cycles material from the Earth's interior to the surface and then back into the interior. Images from seismic tomography graphically illustrate the large scale of the plate tectonic cycle and the way in which oceanic lithosphere is returned to the deep mantle, possibly to the CMB. The implication of such a dynamic view of the Earth is that even the deeper regions of the Earth's interior communicate with the outer layers. This then raises the question of whether conditions at the Earth's surface might influence the way the deeper Earth works. Is it possible for a thin veneer of water only a few kilometres deep to have an effect on the vastly more massive and thermally potent mantle? More intriguingly, is it possible that the presence of an oxidising atmosphere, which is a product of the biosphere, affects the workings of the mantle? The remaining sections of this chapter focus on the role of water within the Earth and reach what you may consider to be some startling conclusions.

In order to understand the behaviour of H_2O in the solid Earth today, and to make inferences of its effects in the past, it is necessary to investigate its behaviour in a number of different processes within the plate tectonic cycle. These processes are summarised in Figure 7.17.

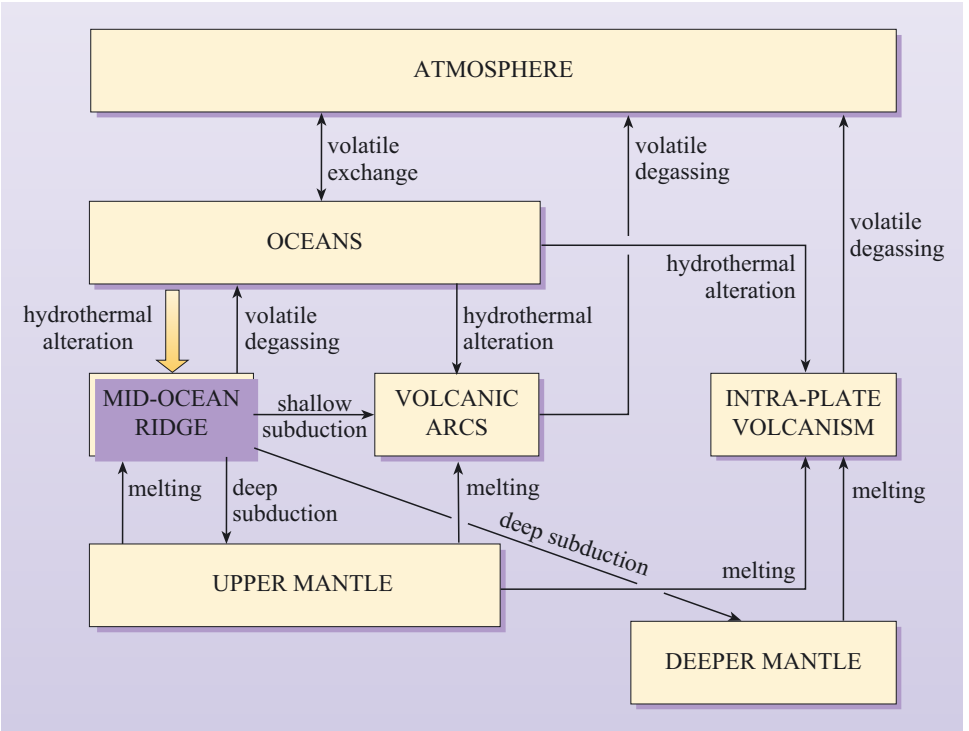


Figure 7.17 Schematic illustration of the H_2O cycle in the Earth.

Note the size of the arrow between the oceans and mid-ocean ridges reflects the importance of this process of returning H_2O to the mantle.

The H_2O cycle can be divided into three key stages: melting, subduction and hydrothermal processes.

Melting

You should recall from Chapter 5 that the presence of H_2O reduces the melting point of peridotite by some 100–200 °C.

- Given this observation, do you think H_2O is compatible or incompatible during melting?
- Incompatible. H_2O reduces the melting point of the mantle because it is highly soluble in silicate melts at mantle pressures.

- What will the effect of melting be on the H_2O content of the residual mantle?
- Melting will effectively strip the H_2O from the mantle; the residue will be virtually anhydrous.

As with any incompatible element, during melting the H_2O in the mantle enters the melt phase and is removed as magma is extracted. Analyses of MORB and OIB suggest that it is about as incompatible as cerium (Ce).

Question 7.4

Using Equation 5.6, calculate the relative amount of H_2O left in the mantle residue after 1, 5 and 10% melting. Assume, equilibrium melting, a bulk partition coefficient of 0.005 and an initial H_2O content of 1 wt%.

These simple calculations reveal that, with a partition coefficient similar to that of Ce (0.005), only a few per cent melting is required to extract >90% of the H_2O into the melt phase.

- What will be the fate of H_2O dissolved in the magma on cooling?
- As the magma rises towards the surface, cools and crystallises, the H_2O is released and is largely exhaled into the ocean (MORB) or the atmosphere (OIB).

Thus, melting effectively strips H_2O from the mantle and transports it to the surface, leaving the residual mantle dry.

Hydrothermal processes

As you saw in Chapter 4, the magmatic heat released as new ocean crust is produced drives hydrothermal systems around the ocean ridge, and water from the oceans penetrates and interacts with the newly generated solid oceanic crust.

- What is the effect of this circulation on the rocks of the ocean crust?
- Individual mineral phases react with the water to form new mineral phases that are stable at lower temperatures and that contain a substantial amount of H_2O.

The end product of hydrothermal activity is a metamorphosed, hydrated basalt that contains up to 5 wt% H_2O. In other localities adjacent to ocean fracture zones and slow-spreading ridges where the uppermost mantle is exposed to the oceans, minerals such as talc and serpentine are also produced from interaction with mantle peridotite. Hydrothermal activity transports many other elements, some of which are fixed in the oceanic crust in addition to H_2O, whereas others are leached and released to the oceans. For the sake of this example, you need only consider H_2O; however, when considering other elements, their mobility or otherwise in hydrothermal systems is a critical part of their geochemical cycle. Thus, hydrothermal circulation replaces some of the water lost to the oceans during the cooling and solidification of basaltic magma.

As an oceanic plate ages and cools, it interacts less and less with the overlying ocean. Consequently, once the ocean crust is altered it remains altered and carries the inventory of H_2O and other elements acquired at the mid-ocean ridge towards the subduction zone and eventually back into the mantle.

Subduction

During subduction, H_2O is progressively released from the various hydrated phases in the altered mafic and ultramafic rocks, as well as any accumulated sediments, as the slab descends into the mantle. Some H_2O release occurs at relatively low pressures and temperatures associated with the breakdown of clay minerals and other phases that are only stable at low temperatures, whereas H_2O can also be transported to greater depths in phases that have higher temperature stability fields.

- Name two H_2O-bearing silicate minerals that are stable at higher temperatures (>400 °C).
- Amphibole and mica are two hydrous minerals that are stable to many hundreds of degrees in the crust and upper mantle.

You have seen in Chapter 5 how subduction-related magmatism is triggered by the release of H_2O in the mantle at 100–150 km depth, and this is thought to be related to the breakdown of hydrous phases in the subducted slab. Clearly, a substantial amount of H_2O trapped in the hydrated oceanic crust penetrates to mantle depths. One of the critical questions in this discussion, however, is whether or not water can be carried beyond the depths of arc magma generation and back into the deep mantle.

During the hydrothermal alteration of mafic and ultramafic rocks at mid-ocean ridges, two of the products are the minerals serpentine and talc, both magnesium silicates with up to 10% H_2O bound in their crystal structures. These two minerals are stable at low pressures and temperatures below about 600 °C. Consequently, during subduction, as the slab gradually heats up, they break down into other mineral phases, releasing their chemically bound H_2O. However, when experimentalists studied these decomposition reactions in some detail they discovered that mineral dehydration is a complex process and that the breakdown products of serpentine and talc may also be hydrous. These mineral phases are known collectively as **dense hydrated magnesium silicate** (DHMS) minerals and are stable at mantle pressures and temperatures of up to 1000 °C.

Subducted slabs follow a variety of temperature–depth paths, depending on the age of the plate when it is subducted (see Figure 5.9 and 5.11).

As DHMS minerals are unstable above 1000 °C, H_2O can be transported to just over 200 km depth by young (hot) slabs, but to depths in excess of 600 km by old (cold) slabs.

- Can H_2O be stored in DHMS in the mantle?
- No. DHMS breaks down above 1000 °C and, as all the mantle is hotter than 1280 °C at depth, they are not stable mantle mineral phases.

Eventually, all subducted slabs heat up above 1000 °C and any H_2O stored within them is released into the mantle. If the mantle is hotter than the H_2O-saturated solidus then it will melt, but if it is cooler then it is possible that H_2O may even exist as a free vapour phase. Thus, DHMS minerals provide a means of transporting H_2O into the mantle, possibly into the lower mantle, but they do not offer a long-term storage reservoir because of their limited temperature stability. Given that is the case, where is the H_2O in the mantle located?

H_2O in the mantle

In the mantle, H_2O can be stabilised in phases that contain potassium. Notable amongst these are magnesium-rich forms of mica and amphibole. Both minerals have potassium as an essential element in their crystal structures, and their stability depends on the availability of both potassium *and* H_2O. You should recall

that many arc-related magmas are enriched in the large ion lithophile elements (Chapter 5) that are mobile in H_2O and that potassium is one of these elements, so the conditions required for the development and stability of mica and amphibole in the mantle are not unreasonable. Indeed, both amphibole and mica have been recognised in mantle xenoliths from kimberlites and basalts, but they are not common.

Away from subduction zones, in regions of the mantle with more normal (i.e. low) potassium contents, more likely hosts for H_2O are, surprisingly, phases such as olivine, pyroxene and garnet and their high-pressure equivalents, such as β-Mg_2SiO_4. As with silicate melts at high pressure, nominally anhydrous silicate phases can incorporate small but significant amounts of H_2O in their crystal structures. Analysis of experimental charges has shown that clinopyroxene can contain between 200 and 500 ppm H_2O, whereas β-Mg_2SiO_4 can contain up to 4000 ppm (0.4 wt%). Indeed, the evidence for H_2O in the nominally anhydrous mantle is probably most convincing in the H_2O content of MORB itself, bringing this discussion of the H_2O cycle neatly back to where it began.

Earlier, you considered the relative proportion of H_2O extracted from the mantle during small degrees of partial melting, but observations on the H_2O content of mantle-derived magmas allow the determination of a more quantitative estimate of the content of H_2O in the mantle. Measurements of MORBs show that they typically have an H_2O/Ce ratio of ~200, and this ratio appears to be characteristic of MORB globally. The similarity of this ratio in MORB worldwide is strong evidence that H_2O and Ce have similarly low partition coefficients during melting, and from this two important aspects of the mantle geochemistry of water can be deduced:

- the H_2O/Ce ratio of the mantle is ~200
- melting effectively strips >95% of Ce and H_2O from the mantle.

Armed with this information, you should now attempt to answer the following question.

Question 7.5

Given that primary MORB has a Ce content of about 7.5 ppm and assuming MORB is derived by 10% melting of the mantle, what is the H_2O content of the mantle?

This figure probably represents a minimum concentration, but it gives a broad indication of the amounts of H_2O that are probably present in the nominally anhydrous mantle source region of MORB. The H_2O contents of basalts from Hawaii and other OIBs appear to be significantly higher than those of MORBs.

■ How can this observation be explained from what you know about the differences between the melting processes that generate MORBs and OIBs?

- Hawaiian basalts and other OIBs are generated from smaller amounts of melting than MORBs. As H_2O behaves like an incompatible element it will be more concentrated in basaltic magmas generated by smaller melt fractions.

H_2O/Ce ratios provide the key to understanding the composition of the source regions of Hawaiian and other OIBs. As the H_2O/Ce ratios of Hawaiian and other OIBs are similar to those of MORB, this suggests that their source regions contain similar amounts of H_2O.

From analyses of oceanic basalts, estimates of mantle H_2O concentrations from OIBs and MORBs vary between 100 and 500 ppm. Neither of these figures sounds like very much, but remember that the mantle represents 66% of the mass of the Earth; so, if this concentration is representative of the mantle as a whole, then even these small concentrations may amount to a considerable mass and volume of water locked away in the deep Earth.

Question 7.6

Using the masses of the mantle and the oceans given below, calculate how many ocean masses there are in a mantle if it contains (a) 100 ppm H_2O and (b) 500 ppm H_2O. Mass of the mantle is 40×10^{23} kg; mass of the ocean is 1.4×10^{21} kg.

Thus, the nominally anhydrous mantle may contain as much H_2O as there is in the present-day oceans. You may find this result surprising, especially considering conclusions about the early extensive degassing of the Earth's atmosphere that resulted from the Earth's hot and violent early history. The continual evolution of the Earth has also entailed extensive melting of the mantle both at plate margins and associated with mantle plumes throughout Earth history (Box 7.4), which must have led to the progressive dehydration of the residual mantle. Yet the current estimates suggest that much of the upper mantle contains substantial amounts of chemically bound H_2O in nominally hydrous mantle minerals. Some authors have even proposed that the upper mantle is saturated in H_2O – it can hold no more without it forming a separate vapour phase, although such statements are continually subject to challenge as experimental evidence improves. Notwithstanding these arguments, the corollary of these rough and ready calculations is that to sustain the present-day concentration of H_2O in the mantle, then H_2O *must* be recycled from the surface back into the interior via subduction.

H_2O released from the mantle via volcanism can find its way back into the mantle via subduction because it reacts with the constituent mineral phases of basalt and peridotite in a complex way. The H_2O cycle in the solid Earth, therefore, is a true cycle, a reversible process and, as it turns out, a critical aspect of what makes the Earth work the way it does.

Box 7.4 How much of the mantle has been melted?

Apart from geochemical arguments about the depleted composition of the upper mantle, you can estimate how much mantle material must have been processed throughout geological time based on present-day plate production rates. Most modern-day melting occurs beneath mid-ocean ridges. If we assume that the present-day rate of spreading, about 2.7 km^2 of new crust per year, is representative of Earth history, then this is equivalent to ~20 km^3 of magma generation per year, assuming an average oceanic crustal thickness of ~7 km (Chapter 4). This in turn implies that about 200 km^3 of mantle is melted each year (assuming 10% melting). Over the course of geological time, 4.5×10^9 years, this means that a total of $4.5 \times 10^9 \times 200$ km^3 of mantle have been processed at mid-ocean ridges, equivalent to 9×10^{11} km^3. Remarkably, this is the same as the volume of the mantle, which is 9×10^{11} km^3.

You might, quite rightly, question the assumptions in this calculation. For example, it is most unlikely that plate and magma production rates have always been the same. However, as you will have seen in Chapter 4, production rates depend largely on mantle temperature; so, as the mantle temperature was greater in the past (Chapter 2), magma production rates were also higher and the volume of mantle processed per year was probably higher. Furthermore, this calculation ignores the volumes of magma generated in mantle plumes; although, today, this volume is <10% of that generated at mid-ocean ridges, on occasions in the past, during the development of large igneous provinces, it may have equalled that of the mid-ocean ridge system. Given these considerations, the estimate that a volume of mantle equivalent to the whole mantle has been processed by partial melting may be conservative!

This is not to say that the whole mantle has at some time experienced melting, but if one part of the mantle has escaped melting then another part must have melted more than once. Notwithstanding all of these possibilities, the rates of melting today and the incompatible behaviour of H_2O during mantle melting require that, in the absence of recycling, the mantle should contain virtually no H_2O at all.

7.7.1 H_2O in the solid Earth

By now you should appreciate that H_2O is a critical component of the whole Earth; not just the hydrosphere, but also the crust and the mantle. It controls many aspects of melting, the dominant process of Earth differentiation, and is cycled from the interior of the Earth to the outer layers and back into the interior. The amount of H_2O hidden in the Earth's interior may well be equivalent to the water in the oceans, despite the outgassing associated with melting and volcanism at the Earth's surface, and the oceans now appear to have been a feature of the Earth from Hadean times onwards.

So what other effects does H_2O have on the way the Earth works? One of the most intriguing is its effect on the strength of minerals. In the previous section you discovered that nominally anhydrous minerals such as olivine and clinopyroxene are capable of dissolving small but significant quantities of H^+ derived from H_2O. Incorporation of H_2O in silicate lattices is accomplished through two substitution mechanisms. The first of these involves the direct replacement of Mg^{2+} with $2H^+$. This is the most common substitution, because all mantle silicates contain Mg. The second mechanism involves the coupled substitution of H^+ and Al^{3+} replacing Si^{4+} in aluminium-bearing minerals, such as pyroxenes, particularly garnet. The introduction of water into the crystal structure of olivine may also replace oxygen atoms with OH^-. All of these substitutions serve to modify the structure of the

minerals and so reduce their strength. Various experiments have investigated the strength of minerals with the amount of dissolved H_2O, and the results of selected experiments are summarised in Figure 7.18.

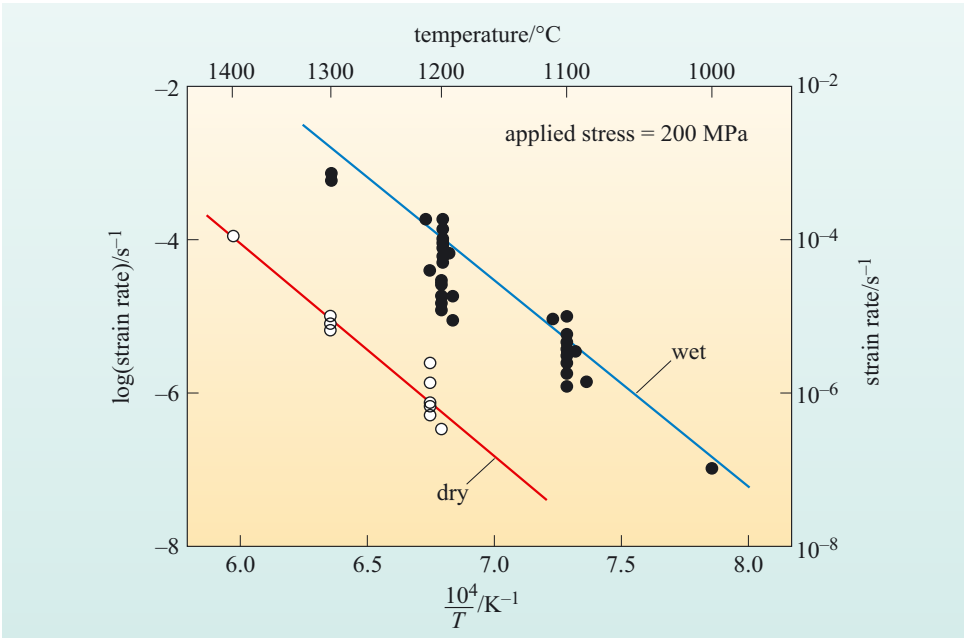


Figure 7.18 A plot of the strain rate of olivine against temperature for dry and H_2O-bearing (wet) systems. The strain rate is a measure of the strength of a mineral under a particular stress. In this case, the applied stress is 200 MPa. The higher the strain rate at a constant stress, the weaker the mineral.(Hirth and Kohlstedt, 1996)

■ Using the information in Figure 7.18, by what factor is the strength of olivine reduced by the addition of water at 1200 °C?

■ At 1200 °C the dry olivine curve has a value of 10^{-6} s^{-1}, whereas the wet olivine curve has a value of 10^{-4} s^{-1}. Thus, dry olivine is 100 times stronger than wet olivine.

■ What temperature change produces the same change in strength in dry olivine?

■ The dry olivine curve has a strain rate of 10^{-4} s^{-1} at 1400 °C, and a value of 10^{-6} at 1200 °C, so a similar change in strength is produced by a 200 °C temperature change.

The amount of H_2O required to reduce the strength of olivine significantly is surprisingly small, of the order of 50 ppm.

■ How does this value compare with your estimate of the H_2O content of the mantle determined in Question 7.5?

■ It is smaller: you calculated previously that the upper mantle has between 100 ppm and 500 ppm water.

Thus, the mantle source region of both MORBs and OIBs has more H_2O than is necessary to reduce the strength of olivine by a factor of ~100. This has clear implications for the way the mantle behaves under stress because olivine comprises 50–60% of the upper mantle. For example, the strength of olivine determines how the mantle responds to isostatic loading by the overlying crust and lithosphere. Clearly, a weaker mantle will respond more rapidly to loading or

unloading than a strong mantle. Weakening olivine allows the mantle to undergo solid-state flow more easily, and so aids mantle flow. To put it another way, water reduces the dynamic viscosity of the mantle. The details of calculating the effects on the dynamic viscosity are beyond the scope of this chapter, but the overall effect of more than 50 ppm water in olivine is to reduce the dynamic viscosity of the mantle by two to three orders of magnitude (a factor of 100–1000).

You should recall from earlier in this chapter that the viscosity of the mantle determines how vigorously the mantle convects. The style of mantle convection is determined by the Rayleigh number *Ra* of the mantle, and mantle viscosity is one of the parameters used to determine *Ra*.

- How will an increase in mantle viscosity effect *Ra*?
- Dynamic viscosity η is a factor in the denominator of the equation for *Ra* (see Equation 7.7 in the answer to Question 7.2). Its effect is to dampen motion rather than drive it – so an increase in viscosity will decrease the Rayleigh number.

For the mantle to convect, the value of *Ra* must be significantly greater than the critical Rayleigh number Ra_c. The current value of *Ra* for the mantle is of the order of 10^5 or 10^6. This is substantially greater than Ra_c, which is around 10^3. Thus, the situation today allows mantle convection and, hence, plate tectonics. Increasing the viscosity by a factor of 10^2–10^3 will decrease the value of *Ra* by a similar amount, bringing it much closer to the critical value.

- What would be the effect of reducing the Rayleigh number of the mantle?
- A lower *Ra* means that mantle convection would become more sluggish.

As a final point, consider the following question:

- Given that the lower parts of oceanic plates are made up of residual harzburgite, how will their strength and viscosity be affected?
- As harzburgites are the product of significant melt depletion, they should have low H_2O contents and so be much stronger than the underlying asthenosphere.

The loss of H_2O from the mantle during the generation of new oceanic lithosphere increases its strength considerably, thus encouraging the thermal development of plate structure.

It should now be apparent that the H_2O content of the mantle has a controlling effect on both the style and vigour of mantle convection, and the development of lithospheric plates, even at the small concentrations present in the mantle today. But the only way in which that H_2O content can be maintained over geological time is as a result of recycling of water from the oceans through hydrothermal modification of the ocean crust and subduction of the resultant hydrated minerals.

7.7.2 Positive and negative feedbacks

On the Earth today, the presence of the oceans allows the mantle to remain hydrated to a sufficient degree to allow it to convect relatively easily and, geologically speaking, rapidly. In addition, the dehydrated residue after melting is significantly stronger than the fertile mantle. As the mantle directly beneath the crust is depleted harzburgite (Chapter 4) this adds to the strength of the oceanic lithosphere, thus aiding the generation and preservation of lithospheric plates. Both effects are essential for maintaining plate tectonics. However, plate tectonics is also the mechanism whereby the mantle is rehydrated via subduction. Thus, it is a system in which the very actions of the system keep that system operating.

- What would happen to the viscosity of the mantle if the rate of subduction increased?
- More H_2O would be returned to the mantle; the viscosity would be reduced.

If mantle viscosity is reduced then subduction would be easier, forcing more H_2O back into the mantle, thereby further decreasing mantle viscosity, and so on. Such a phenomenon is known as a **positive feedback** because the operation of a process has an effect that helps that process to operate more efficiently.

So why isn't there an ever increasing rate of plate motions? The answer lies in the way in which the rate of subduction controls the rate of ocean spreading. You should recall from Chapter 3 that the dominant driving force for plate motion is slab-pull, so if the rate of subduction increases, so does the rate of spreading. An increase in the rate of spreading leads to a concomitant increase in the volume of magma generated at mid-ocean ridges. This, in turn, dehydrates a proportionately larger volume of mantle, increasing the overall viscosity of the mantle and hence slowing convection. Melting, therefore, provides a **negative feedback** on the system, acting to increase mantle viscosity and so slow the system down. Feedbacks, both positive and negative, are important concepts in adopting a systems approach to studying the Earth. In the case of plate tectonics and mantle convection, it is the balance between two processes, i.e. dehydration by melting and rehydration through subduction of 'wet' oceanic plates, that keeps the mantle operating in the way it does today and, as far as we can judge, has done for much of geological history.

7.7.3 The Earth without an ocean

It is instructive now to consider the consequences if the Earth did not have a significant hydrosphere, but had a surface dominated by bare rock and no major oceans. Clearly, the surface of the Earth would be very different. For a start it would be lifeless: the evolution of life as we know it depends on water; and the surface temperature would swing between much greater extremes: the oceans have a controlling and buffering effect on atmospheric temperatures because of their large heat capacity. But in this chapter you are considering the way in which the solid Earth system works, so how would this be affected?

Assume that the initial H_2O content of the mantle and its temperature gradient are both similar to that of today. With this concentration of H_2O the mantle will convect

and hot material will rise towards the surface and melt. As with the plate cycle discussed at length above, melting in a ridge or plume environment will extract H_2O from the mantle and transport it towards the surface. As the magma cools and solidifies, the H_2O will escape to the surface because H_2O is incompatible in most silicate minerals. However, on the waterless planet there is now no ocean to trap it, and so it escapes into the atmosphere.

The next stage in the dry Earth's water cycle should be hydrothermal interaction with the oceans; but, because the oceans are missing, there can be no further interaction. Hence, the plate will cool and subside as now, but it will retain its largely anhydrous igneous composition. Already the cycle has been broken and the path followed by H_2O through the Earth has changed.

As the plate ages and moves towards subduction, it does not accumulate a significant sedimentary layer, again because there is no active surface H_2O cycle to drive erosion and deposition as there is on Earth today. On subduction, therefore, the material being taken down into the mantle is dry: there is no water in sediments and no H_2O in the crust – just dry igneous rock of broadly basaltic composition.

- What is the consequence for destructive plate boundary processes on our dry Earth?
- There will be no magmatic activity associated with subduction.

Subduction-related magmatism is triggered by dehydration reactions at depth in the mantle. Given that the subducting plate is cold and the mantle is not hot enough to melt without decompression, there will be no melting because the subducting plate is not carrying H_2O into the mantle and will simply act to cool the surrounding mantle wedge.

As subduction proceeds further the plate sinks into the deep mantle and gradually equilibrates with its surroundings.

- What is the next consequence for our dry Earth?
- No H_2O is recycled into the mantle because, once again, the plate is dry.

In summary, without an ocean covering the ocean ridge to capture the water released by cooling magma and to rehydrate the plate via hydrothermal and sedimentary processes, there would be no magmatism at destructive plate boundaries and the mantle would become progressively drier. The feedbacks that maintain the current convective movements and plate cycling in the mantle would not be balanced and new continental crust could not be generated because of the lack of subduction-related magmatism. Thus, the hypothetical dry planet will eventually lose any H_2O trapped in the mantle, plate tectonics will cease and there will be no new continental crust formed above subduction zones.

But just how hypothetical is this planet? Can these predictions be tested against another planetary system? Returning to Section 1.1, you should recall that, despite the similarity in radius and mass, observations of Venus have revealed that its surface is too hot for water to exist, so perhaps this is a model for the Earth without water.

- Can you recall from Section 1.1 the differences between the geological structure of the lithosphere of Venus and Earth?
- Venus shows no evidence for plate tectonics and no evidence for a bimodal hypsometric plot.

Although Venus is clearly an active planet, as reflected by the geological youth of many of its surfaces, there is no evidence for active plate tectonics, and the hypsometric plot does not suggest distinctive regions of oceanic and continental crust. Thus, first-order predictions for a dry planet are borne out by primary observation.

Whether Venus has always behaved as it does today or not is a matter of active research and discussion. Remote spectroscopic measurements of the isotopic ratio of traces of water in the atmosphere show it to be much richer in deuterium than water is on Earth, suggesting that Venus may originally have had as much water as the Earth had, but subsequently lost it. Alternatively, the high surface temperature may have allowed its inventory of H_2O to be locked deep in the mantle. Numerous hypotheses abound; but what remains clear is that, without a surface hydrosphere, plate tectonics and the differentiation of the continental crust becomes much less likely.

Summary of Chapter 7

This chapter has considered processes that operate at a planetary scale, and how those processes determine the geochemical and geophysical evolution of the Earth. Particular reference has been made to the deeper (>300 km depth) mantle and how mechanisms that operate at great depth can influence, and be influenced by, surface processes. The mineralogy is known throughout the whole 2900 km of the mantle based on modern experimental methods, along with the physical properties of the minerals. These data feed into seismic tomographic and computer models that strongly support some form of whole mantle convection, in which subducted plates cross the upper/lower mantle boundary and may reach the core–mantle boundary.

The plates are an integral part of the convection system and play a role in determining its geometry. By definition, if slabs enter the lower mantle then there must be a return of material to the upper mantle; this may be in the form of hot upwelling that ultimately generates some intra-plate volcanism. Subducting slabs play a key role in keeping the planet in a dynamic state. They carry chemical elements in to the deeper mantle, which over time (billions of years) develop characteristic isotopic signatures that allow geochemists to quantify the stirring time of the mantle. Most importantly, subducted slabs recycle H_2O back into the mantle, which lowers its viscosity and maintains its convective vigour – allowing the plate cycle to operate.

This chapter gives a taste of how the Earth operates as a system. There are clearly areas where there is much speculation and controversy, but understanding how the planet operates as a dynamic system will be one of the main areas of research during this century.

Learning outcomes for Chapter 7

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

7.1 The discontinuities (e.g. transition zones) within the Earth and how these are related to sequential changes in the mantle mineralogy.

7.2 How understanding of the structure of the Earth and location of melting in the mantle has been enhance through three-dimensional imaging of changing seismic velocities within the Earth by seismic tomography.

7.3 Methods of modelling mantle convection as a means of determining how materials are cycled between different layers within the Earth and over what timescales.

7.4 The potential source and cause of intraplate volcanism, and how this can be identified by changes in the major and trace element geochemistry of resultant magmas, in comparison to normal mantle melts.

7.5 How the recycling of altered oceanic crust into the deep mantle via subduction zones, can be recognised by distinctive major and trace element and isotopic signatures.

7.6 The role of water in the solid Earth and its controlling influence on magmatic and tectonic processes (e.g. melting, subduction, hydrothermal alteration, and plate tectonics) as modelled by positive and negative feedback mechanisms.

Chapter 8

The continental crust

At the start of this book the continental crust was recognised as one of the defining features of what made the Earth unique in the Solar System. So it is fitting that the book should end with a summary of modern ideas of how and when the continental crust formed, and whether or not it is an end-product of planetary differentiation or part of a reversible cycle of material within the Earth.

The continents literally form the bedrock of our existence; they are the source of all our raw materials and provide a platform above the surface of the oceans on which terrestrial life has evolved. Without the continents the Earth would be covered with water because so little of the crust would have sufficient buoyancy to push the solid surface above the globe-encircling ocean. As Bill Bryson puts it in his book *A Short History of Nearly Everything* without the continents 'There might be life in that lonesome ocean, but there certainly wouldn't be football'. More significantly, however, the continents define the inorganic chemical environment within which all life has evolved, even marine life, because critical aspects of ocean chemistry are controlled by continental river run-off.

Elements that have a trace abundance in the crust, because they are chalcophile or siderophile, are more likely to be toxic than those that are lithophile and hence in greater abundance in crustal rocks. For example, selenium is a volatile element during planetary accretion and has strongly chalcophile properties. As a result there is very little Se in the crust, but what there is tends to be soluble in water and finds its way into living organisms. Organisms have subsequently evolved not only to tolerate that small amount of Se but to rely on its presence for particular aspects of their metabolism. Again Bill Bryson puts this much more engagingly:

> Selenium is vital to all of us, but take in just a little too much and it may be the last thing you ever do … We have evolved to expect, and in some cases actually need, the tiny amounts of rare elements that accumulate in the flesh or fibre that we eat. But step up the doses, in some cases by only a tiny amount, and we can soon cross a threshold.

Consequently understanding the evolution of the continental crust is to understand a critical influence on our own evolution.

The continents contain the oldest rocks but so far have been regarded simply as passengers that slowly migrate across the Earth's surface in response to internal processes, currently dominated by plate tectonics and mantle convection. What has not yet been discussed is how the continental crust relates to this dynamic interior and whether the processes operating today lead to the evolution of the continental crust. Is the crust simply the end product of planetary differentiation or is it part of a reversible cycle?

8.1 Reversible cycles and irreversible processes

By now you should appreciate that far from being a solid, static sphere of rock and metal with a thin covering of water and gas, the Earth is a dynamic evolving system that has changed over geological time. After cooling from an initially largely molten state during which period the core, earliest crust, atmosphere and hydrosphere separated from the mantle (Chapter 2), the dominant process in the evolution of the Earth has been plate tectonics (Chapter 3). Material brought to the Earth's surface at mid-ocean ridges carries with it a fraction of the internal heat, which is lost to the ocean by conduction and advection of hydrothermal fluids, especially close to spreading centres (Chapter 4). It is here that the geosphere interacts most directly with the hydrosphere and material exchange takes place. Once the plate cools, interaction becomes more limited and after ~200 Ma, it thickens and subsides into the underlying mantle and is subducted back into the interior (Chapter 5). Sediments are scraped from the plate surface and part of the inventory of volatile and fluid-mobile elements introduced at the mid-ocean ridge are released as the plate warms up during its passage back into the mantle. This in turn induces magmatism in volcanic arcs and the material so generated becomes incorporated into the continental crust during continental collision (Chapter 6). If elements can become fixed in the subducting slab then they can be returned to the deep mantle and evidence of such recycling of surface materials can be found in the magmatic products of mantle plumes (Chapter 7). The most visually compelling evidence for communication between the surface and the deep Earth is apparent in the images generated from seismic tomography that reveal the movement of subducted plates through the upper mantle and into the lower mantle. The overall image is one of a planet in internal turmoil, albeit on a very long timescale, kept mobile by its internal heat but driven by gravity.

The solid Earth system involves a material cycle in which plates generated at mid-ocean ridges are returned to the mantle at subduction zones. Segments of oceanic crust and lithosphere are therefore temporary additions to the outer part of the Earth – oceanic crust is never older than about 180 Ma because after that time it is re-incorporated into the mantle, heats up and becomes indistinguishable from the rest of the mantle. The generation of oceanic crust and lithosphere can therefore be regarded as part of a **reversible cycle** in which the mantle cools by gravity-driven convection.

By contrast, segments of the continental crust and lithosphere are much older, and the record in continental rocks extends back to almost 4.0 Ga as individual rocks in Greenland and Northern Canada, while very rare mineral ages extend back to 4.4 Ga.

- Can you recall why this should be?
- Continental crust is composed of low-density rocks that cannot be subducted back into the mantle (Chapter 6).

Thus the generation of continental crust may be regarded largely as an **irreversible process** that leads to the permanent differentiation of the planet.

- Can you think of other material or energy cycles within the Earth that are irreversible?

- Irreversible processes include the loss of heat (cooling) and the differentiation of the atmosphere from the solid Earth.

The more you consider the workings of the Earth's interior, the more difficult it becomes to define irreversible processes. As you have discovered from previous chapters, simply because a component of the Earth resides at the surface does not mean that it has completely escaped from the interior. The water in the oceans is recycled into the Earth's interior and exerts a controlling influence on the behaviour of the mantle and mantle melting. Similarly in Chapter 5 you explored some of the evidence from the compositional variations of subduction-related magmas that indicate the incorporation of ocean sediments in their source regions, extending the rock cycle to mantle depths. So even material from the continental crust can be recycled into the mantle, although whether that material descends beyond the depths of magma generation is the subject of debate and uncertainty.

8.2 How continental crust is formed

Despite the significance of the continents for our very existence, their evolution and even their bulk composition remain the subjects of scientific debate. From what you have read earlier in this book, you should now appreciate that the oldest rocks are slightly less than 4.0 Ga old and there is evidence in zircons and from miniscule variations in the radiogenic isotopes of Nd that there was a form of continental crust during the Hadean; although the volume and extent of such primordial crust remains unknown.

The age of the Earth's earliest crust depends on how crust is defined. Starting with the primitive Earth enveloped by a magma ocean, the first crust would be analogous to a chilled skin on the surface of the magma ocean, possibly with a composition modified by the floatation of minerals less dense than the liquid magma. An example of ancient primary crust is seen on the Moon in the anorthosites of the lunar highlands. These are characterised by some of the oldest ages yet obtained from lunar material and probably formed by floatation of light plagioclase feldspar on the surface of the lunar magma ocean. Such **primary crust** is no longer represented in the geological record on Earth and it is assumed that it was eradicated by the effects of impacts during the waning stages of accretion while vigorous convection in the underlying hot mantle ensured its rapid recycling back into the interior.

Subsequent to the destruction of the Earth's primary crust, mantle convection resulted in mantle melting to form **secondary crust**.

- Is secondary crust being formed today, and if so what is it?

- Today, partial melting of the mantle generates oceanic crust that is considered to be secondary crust.

The modern-day process of oceanic crustal formation is part of the reversible plate tectonic cycle within the Earth. As stated above there is no oceanic crust surviving beneath the oceans older than 180 Ma, although older fragments do occur as ophiolites in orogenic belts dating back to the late Proterozoic, and may be even older.

The oldest surviving crust is found in the continents and this is known as **tertiary crust**. In considering the evolution of the continental crust it is first important to define where continental crust is created. Today, the destruction of secondary crust by subduction leads to the generation of magmatic rocks in island arc and Andean margins and these are recognised as the current sites of continental crust generation. In the discussion of the evolution of the Himalayan orogen in Chapter 6, distinction was made between two different types of granite intrusion.

- Can you recall what they are?
- The trans-Himalayan batholith evolved during a period of Andean volcanism prior to continental collision while the high-Himalayan granites formed from crustal melting during collision.

- Which of these two represented examples of crustal growth and why?
- The trans-Himalayan batholith has initial $^{87}Sr/^{86}Sr$ ratios close to that of the upper mantle and so has an origin in the upper mantle. It therefore represents an example of recent crustal growth.

The large volume trans-Himalayan batholith comprises a range of igneous rocks related to each other by fractional crystallisation and so represents new material added to the continental crust from the mantle. By contrast the isotopic similarity of the high-Himalayan granites with pre-existing crustal rocks, and their origin as partial melts with no associated mafic counterparts, shows that they are a product of the rock cycle within the crust. While they do not represent a new addition to the crust they reflect an important refining process within the crust whereby silica-rich granitic melts migrate upwards leaving a more mafic residue in the deeper crust. Destructive plate margins are therefore regarded as the most important sites for the growth of new continental crust while collision zones allow the further refinement of the crust into a granitic upper crust and a more mafic mid- and lower crust.

While this appears to be the dominant crust-forming process today, it is not clear whether the same is true for all of geological time. For example, voluminous crustal rocks formed during Proterozoic times do not share the same compositional characteristics as are typical of subduction-related magmas today and there is a growing body of evidence to suggest that at some time in the past crustal growth was dominated by processes occurring within plates rather than at their boundaries, probably controlled by mantle plume activity.

- What is the composition of the magma transferred from the mantle to the crust at both destructive plate boundaries and at locations within plates?
- Basalt.

Notwithstanding the possible different causes of melting, the flux of material from the mantle to the crust throughout geological time is dominated by basalt. The crust itself has formed by a multiplicity of processes related to mantle and crustal melting, fractional crystallisation, and mixing between new material from the mantle and pre-existing continental crust.

8.3 The composition of the continental crust

Estimates of the composition of the continental crust are surprisingly varied. In part this relates to the variety of rock types exposed in the continents, but it is also because the deeper layers of the crust are rarely exposed at the surface. Table 8.1 lists estimated major element compositions of the upper and lower continental crust and an estimate of the composition of the total crust. The composition of the upper crust is the most accessible and the average composition is based on two methods:

- a weighted average of the composition of rocks exposed at the surface
- using the abundances of insoluble elements in fine-grained clastic sediments to infer bulk composition.

The first method involves large-scale sampling of the crust whereas the second exploits the processes of weathering and erosion, transport and sedimentation that serve to average the composition of crustal source regions. The average compositions in Table 8.1 are based on information from both of these methods.

Table 8.1 Average composition of upper, lower and whole crust. (Adapted from Rudnick and Gao, 2005)

	Upper crust/wt%	Lower crust/wt%	Whole crust/wt%
SiO_2	66.6	53.4	60.6
TiO_2	0.64	0.82	0.72
Al_2O_3	15.4	16.9	15.9
FeO	5.04	8.57	6.70
MnO	0.10	0.10	0.10
MgO	2.48	7.24	4.70
CaO	3.59	9.59	6.40
Na_2O	3.27	2.65	3.10
K_2O	2.80	0.61	1.80
P_2O_5	0.15	0.10	0.10
Total	100.07	99.98	100.12

The composition of the lower crust is based on averages of granulites from around the world. You should recall from Chapter 1 that the lower crust is thought to be composed dominantly of relatively anhydrous, high-grade metamorphic rocks known as granulites and these are found both as xenoliths in volcanic pipes and exposed at the surface as a consequence of uplift and erosion. In addition the more uniform physical properties of the lower crust as reflected in seismic velocities can be matched against compositions as these properties vary with both composition and mineralogy.

- What are the major differences between the upper and lower crust?
- The upper crust is richer in silica, potassium and sodium than the lower crust, whereas the lower crust is richer in iron, magnesium and calcium.

These averages clearly show that the lower crust has a more mafic, i.e. less differentiated composition than the upper crust. The lower crust has a composition similar to that of a basaltic andesite whereas the upper crust is closer to dacite.

- Thinking back to earlier chapters, can you suggest how this layering might have developed?
- Mafic residues are produced when magmas evolve and differentiate. Also the extraction of a granitic melt from a crustal precursor will produce a more mafic residue.

Magmatic processes, be they crustal melting or fractional crystallisation, will produce a more evolved upper crust while leaving the lower crust richer in iron, magnesium and calcium.

The final column of Table 8.1 includes an estimate of the bulk major element composition of the whole crust.

- What igneous rock type most closely represents the composition of the whole continental crust?
- The bulk crust has SiO_2 content of 60.6 wt% and an MgO content of 4.7 wt%. Looking back at Figure 5.13, it most closely resembles the composition of andesite (or diorite).

Somewhat unsurprisingly the composition of the bulk crust is intermediate between that of the upper and lower crust and resembles that of an andesite. However, this rather simple observation lies at the root of a fundamental problem in Earth sciences.

- How does the bulk composition of the crust compare with the composition of material supplied from the mantle?
- The bulk crust is andesitic and is therefore more evolved than the basaltic material that is supplied from the mantle.

Despite the association of andesites with the Andes and other regions of modern day and recent crustal growth, andesites are derived from magmas that have fractionated from a basaltic parent (Chapter 5). The flux of material from crust to mantle at subduction zones, as at mid-ocean ridges and above mantle plumes, is dominantly basaltic. Why then, does the crust have an overall andesitic composition?

The short answer to this is that no one really knows, but a number of possibilities have been suggested. One is that the estimated bulk composition of the continental crust is incorrect and that rather than being andesitic it is basaltic. If this is the case then the lower crust is a mafic complement to the dacitic upper crust. You should recall from Chapter 5 that to produce dacite from a parental basalt requires a considerable degree of fractional crystallisation. Calculations based on both major

and trace elements requires up to 85% fractionation before a dacite composition is produced.

- Assuming the upper crust has a thickness of 12 km, how thick would the total crust be if its overall composition was basaltic?
- If 12 km of upper crust is equivalent to 15% (100% – 85%) of the total mass of the crust, then the total crustal thickness should be $\frac{12\text{ km}}{0.15} = 80\text{ km}$.

The calculation is the same if the upper crust developed as a consequence of partial melting because a given amount of basalt can only produce this small amount of melt with a dacite composition. Greater amounts of melting of a basaltic precursor will produce less siliceous compositions. The conclusion from chemical mass balance arguments therefore is a 12 km thick upper crust must be balanced by a lower crust almost 70 km thick.

- How does this total crustal thickness compare with the thickness of the continental crust from seismic refraction (Chapter 1)?
- It is larger.

This simple calculation shows that regardless of the composition of the lower crust, if the flux from mantle to crust is basaltic and the upper crust is dacitic, the total crustal thickness should be 80 km. This is much thicker than the average seismic thickness of the continental crust above the Moho, which is about 40 km.

The fate of the mafic residue from the formation of the upper crust is a matter of current scientific debate but a popular model to explain its absence invokes recycling into the mantle. As continental crust differentiates into a granitic upper crust, the lower crust takes on a more mafic composition.

- What mineralogical transformation occurs in mafic rocks at high pressures?
- The major aluminium-bearing mineral phase changes from low-density plagioclase to high-density garnet, while other chemical components combine to form a dense, sodium-rich clinopyroxene.

You should recall from Chapters 5 and 7 that during subduction, as pressure increases, the ocean crust is transformed into eclogite, which is a metamorphic rock dominated by high-density garnet and clinopyroxene. After repeated melting and dehydration, the altered crustal rocks in the slab become more mafic and, under high-pressure, metamorphic conditions, are transformed into granulites and eclogites with a dense garnet-bearing mineralogy. Similarly, during arc magmatic processes, the fractionation of basaltic magma to an andesitic composition produces large volumes of mafic cumulates dominated by olivine and pyroxenes. Under suitable metamorphic conditions these too can take on a high-density eclogitic mineralogy of garnet and clinopyroxene. The density of eclogite is generally greater than that of peridotite, and so eventually mafic lower crust becomes dense enough to founder into the mantle. This process is frequently referred to as crustal delamination, and is analogous to the convective removal of mantle lithosphere that may occur during continental collision (Chapter 6).

8.4 Crustal evolution through time

The observation that the continental crust contains rocks ranging in age from almost 4 Ga to the present day and that material is still being transferred to the crust from the mantle as a result of melting implies that the volume of crust must have changed through time. But how quickly did the crust grow? Archaean rocks constitute only ~1% of the exposed crust, suggesting that much of the crust has grown since the Archaean.

- Why do you think this conclusion might be incorrect?
- Ancient crust is reworked during continental collision and so old ages can be reset by subsequent metamorphic and igneous events.

As you should appreciate by now, judging rates of major Earth evolution processes from surface observations is fraught with problems and none more so than estimating the rate of crustal growth. Even estimating the rate of crustal growth at the present day is far from simple, largely because much crustal growth happens below the surface with the intrusion of mantle-derived magmas and their evolution towards granitic compositions. The details of how physical growth estimates are made is beyond the scope of this book, but current estimates of crustal growth at destructive plate boundaries lie between 1 $km^3\ y^{-1}$ and 4 $km^3\ y^{-1}$, with a preferred value of 1.6 $km^3\ y^{-1}$.

- The volume of the continental crust is 1×10^{19} m^3. Could this volume have been generated over the age of the Earth at the present rate of crustal growth?
- 1×10^{19} $m^3 = 1 \times 10^{10}$ km^3 (note 10^9 $m^3 = 1$ km^3), so at 1.6 $km^3\ y^{-1}$, it would take:

  $$\frac{1 \times 10^{10}\ km^3}{1.6\ km^3\ y^{-1}} \approx 10^{10}\ y.$$

  to generate the present volume of continental crust, i.e. twice the age of the Earth. The answer is therefore no, the current rate of crustal growth is insufficient to generate the present-day volume of the continental crust over the span of geological time.

Even if the crustal growth rate today is as high as 4 $km^3\ y^{-1}$, it is still barely sufficient to generate the total volume of the continental crust in the full span of geological time. Moreover the problem is exacerbated by the previous discussion of the compositional mass balance that reveals that mafic crustal material must have been recycled into the mantle to balance the dacitic composition of the upper continental crust.

Perhaps this result is not so surprising if you consider the effect of the Earth's internal heat. Heat production was significantly greater early in Earth history leading to greater amounts of mantle melting. As melting is the dominant process for crust generation, the rate of crustal growth must have been greater in the past than it is today.

Although crustal growth rates both today and through geological time are difficult to measure directly, important insights into growth rates can be derived from the

secular evolution of radiogenic isotope ratios in both the mantle and the crust. The details of these models and calculations are complex, but the results can be summarised simply to show the proportion of crust generated through geological time (Figure 8.1).

- Do the three curves in Figure 8.1 show constant or variable rates of crustal growth throughout geological time?
- Constant growth rates would result in a straight line, but all three models show curves that are steeper after 3 Ga and then become shallower towards the present day.

While the three curves shown in Figure 8.1 differ in detail they all show a steeper gradient between 1 Ga and 3 Ga than is apparent today, implying greater crustal growth during the late Archaean and Proterozoic. The lower growth rates prior to 3 Ga may be counter-intuitive – after all the Earth was hottest during the Archaean – but may reflect greater rates of recycling of less-evolved crust during this period.

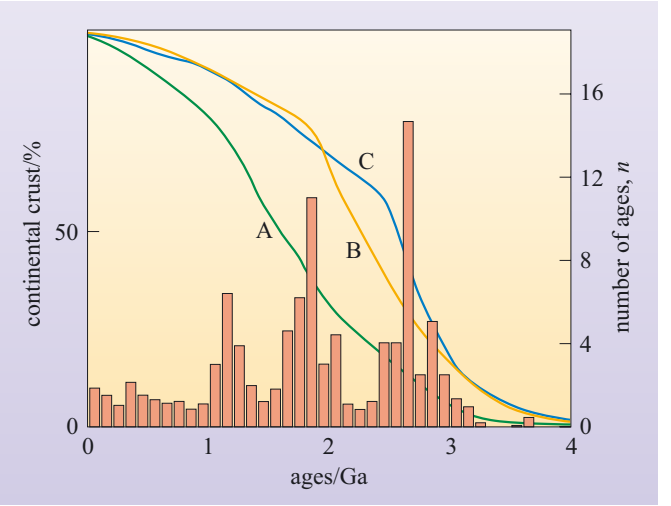


Figure 8.1 The growth of the continental crust through time as estimated from: (A) Nd isotopes in shales; (B) Pb isotope evolution of the mantle; (C) secular changes in the trace element compositions of shales. The histogram shows the ages of individual igneous bodies that contributed to crustal growth. (Hawkesworth and Kemp, 2006)

An alternative method of investigating crustal evolution is to assess the age of new additions to the crust, be they volcanic or plutonic. This is less easy than it sounds because it requires the distinction between new and recycled material although the use of isotope tracers such as initial $^{87}Sr/^{86}Sr$ and other discriminants provide useful indicators. However, a further difficulty arises from the non-random distribution of crust across the surface of the Earth and the variable detail with which different parts of the continents have been explored. First attempts to make such assessments in the 1960s and 1970s revealed distinct peaks in the distribution of ages of new crustal additions as exemplified by large plutons and thick sequences of mantle-derived volcanic rocks. Originally it was thought that these peaks reflected the pattern of preservation of ancient rocks rather than real variations in crust formation ages. However, subsequent additions to this global database, rather than evening out the peaks, have enhanced them, with the current distribution of crust formation ages shown in Figure 8.1. These ages are based largely on the U-Pb (zircon) technique on individual plutons and volcanic units. Each age on Figure 8.1 represents only the most precise age from a unit or pluton to avoid weighting towards well-studied regions of the crust. As is apparent from this diagram the distribution of crust formation ages is far from uniform – crust formation ages show peaks during the Archaean (2.7–3.2 Ga), the mid-Proterozoic (1.8–2.2 Ga) and in the late Proterozoic (1.2–1.5 Ga).

The pulsed pattern in these data clearly contrasts with the smooth curve of crustal growth derived from sediments. Superficially these two results appear at odds with each other, but may be reconciled by considering the formation and development of the sedimentary record. As emphasised above, sediments represent an average of their eroding source regions and so their radiogenic

isotopes are the sum of all of the contributing materials. Thus the very process of formation tends to smooth out any spikes that may be present in the region of crust sampled and the 'age' of any one sample may well be a mixture of the ages of the source terrain. By contrast, the crust formation ages are derived from specific rock units and record the timing of a specific igneous event during which new continental crust was generated.

What could the pulses in crust formation represent? One possibility is that they could reflect plate tectonic Wilson cycles, as discussed in Chapter 3.

- Can you see why Wilson cycles may not be the cause of peaks in crustal growth?
- Wilson cycles have dominated the geological history of the past 1000 Ma of Earth history, yet there are no peaks of crust generation during this time period.

In addition, Wilson cycles occur with a periodicity of about 500 Ma and even though there are no crust formation peaks during the past 1000 Ma, the periodicity before this time is still not 500 Ma.

For evidence for the cause of crustal growth during the peak periods it is necessary to look into the composition of the dominant magmatic rocks in crust formed at those times. For example, large areas of the crust of West Africa were formed during the early and middle Proterozoic, around 2 Ga ago. They developed remote from any pre-existing continental crust and strongly imply the involvement of mantle plumes. Such evidence, although complex, illustrates that crust formation processes have probably not been constant throughout geological time. Since 1000 Ma crust has been generated from the mantle from the quasi-continuous process of subduction at destructive plate margins. Prior to 1000 Ma the evidence for plate tectonics as the dominant process for cooling the Earth becomes less clear and the periodic pulses of crust formation may relate to another process, possibly related to large-scale mantle convection and enhanced mantle plume activity.

Summary of Chapter 8

At present our understanding of the formation evolution and growth of the continental crust remains vague and is defined more by what is not, rather than what is, known. Mantle melting is the dominant process for transporting material from within the planet to its surface. The composition of that material is basaltic, yet the average crust is andesitic and the upper crust is closer to a dacite. Mass balances of current crustal compositions require some crustal material to have been returned to the mantle, but where and how that happens remains controversial. Today destructive plate boundaries are the major sites of crust generation but that may not have been the case for the distant geological past when mantle plume-related activity may have played a significant role.

The continental crust is a product of planetary evolution and has accumulated over the length of geological time. The upper continental crust at least is an end product of an irreversible process in the evolution of the planet. By contrast the

more mafic lower crust is a remnant of the mafic residue that must have been produced during the formation of the upper crust, most of which has been removed and returned into the mantle. So to answer the question posed at the start of this chapter, the continental crust is both the product of an irreversible process and a component in a long-term cycle of material in the solid Earth. Resolving the details of the development of the continental crust remains one of the important challenges facing modern Earth sciences, not least because thick continental crust is a unique terrestrial feature not seen on the other rocky planets of our Solar System.

Learning outcomes for Chapter 8

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

8.1 The relative timing and processes involved in the formation of primary, secondary and tertiary crust on the Earth.

8.2 How the composition of the upper, lower and whole continental crust can be determined using weighted average compositions and/or relative abundances of insoluble elements in clastic sediments.

8.3 The different types of evidence that indicate the lower continental crust is recycled back into the mantle.

8.4 How models of continuous versus episodic crustal formation differ and can be used to explain variations in the rates of crustal formation throughout the evolution of the Earth.

Answers to questions

Question 5.1

$X_{olivine} = 0.6$; $X_{orthopyroxene} = 0.2$; $X_{clinopyroxene} = 0.1$; $X_{garnet} = 0.1$

$D_{olivine} = 0.001$; $D_{orthopyroxene} = 0.003$; $D_{clinopyroxene} = 0.001$; $D_{garnet} = 0.01$ (from Table 5.2)

Therefore, using Equation 5.3:

$$D_{bulk} = (0.6 \times 0.001) + (0.2 \times 0.003) + (0.1 \times 0.001) + (0.1 \times 0.01)$$
$$= 0.0023$$

Question 5.2

$C_0 = 25$, $D = 0.402$ and $F = 0.2$ (remember the amount of crystallisation is $1 - F$). For equilibrium crystallisation (using Equation 5.6):

$$C_L = \frac{25}{0.402 + (1 - 0.402) \times 0.2}$$
$$= 47.9 \text{ ppm}$$
$$= 48 \text{ ppm to two significant figures.}$$

And for fractional crystallisation (using Equation 5.7):

$$C_L = 25 \text{ ppm} \times 0.2^{(0.402-1)}$$

Start by working out the the value in the brackets:

$0.402 - 1 = -0.598$

You can calculate $0.2^{-0.598}$ using the x^y key on a calculator:

0.2 $\boxed{x^y}$ $-0.598 = 2.618$

You can then add this to the original equation:

$C_L = 25 \text{ ppm} \times 2.618 = 65.45 \text{ ppm}$

$= 65$ ppm to two significant figures.

This could also be done in a spreadsheet.

Question 5.3

Using simple geometry, the distance that the slab has to travel from the trench to get to 100 km depth is

depth = distance travelled $\times \sin \theta$

where θ is the angle of subduction. Rearranging this gives

$$\text{distance travelled} = \frac{\text{depth}}{\sin\theta}$$

$$\text{distance travelled} = \frac{100\text{ km}}{\sin 45}$$
$$= 141.42\text{ km}$$

$$\text{time} = \frac{141.42\text{ km}}{60\times10^{-6}\text{ km y}^{-1}}$$
$$= 2.36\times10^{6}\text{ y} = 2.4\text{ Ma or } \sim 1.6\text{ half-lives of } {}^{10}\text{Be}.$$

Therefore, the plate takes 2.4 Ma (to two significant figures) to reach 100 km depth. For a plate travelling at only 10 mm y^{-1} it will take six times as long (~14 Ma).

This simple calculation indicates that subduction zones with relatively rapid convergence rates will have the best chance of preserving some ^{10}Be atoms in the slab.

Question 5.4

The composition of the lavas clearly reflects changes in the composition of the sediment input. Within 2.5 Ma of the change in the sediment U/Th ratio there is a corresponding increase in the U/Th ratio of the arc lavas. This also implies that U is derived from the sediment rather than the basaltic slab in this arc. The data require a total transit time of ~2.5 Ma from sediment subduction to melt eruption. These data are consistent with other pieces of information about this arc. For example the high ^{10}Be ratios in Central America (Figure 5.25e) require rapid (<2 Ma) sediment subduction, consistent with the Cocos Plate being subducted at ~80 mm y^{-1}.

Question 6.1

(i) For Rb: $\frac{C_L}{C_0} = 2.$

From the curve for $D = 0.1$ in Figure 5.21, $F = 0.43$.

(ii) For Sr: $\frac{C_L}{C_0} = 0.15.$

For $D = 3.5$, $\frac{C_L}{C_0}$ does not reach such low values, even at $F = 0$.

These results can be calculated more precisely using Equation 5.6, giving $F =$ 0.44 for Rb. For Sr, $\frac{C_L}{C_0}$ is negative.

Question 6.2

For Rb: $\frac{C_L}{C_0} = \frac{360}{200} = 1.8$. From Figure 5.21, $F = 0.1$ (Equation 5.6 gives 0.11).

For Sr: $\frac{C_L}{C_0} = \frac{65}{120} = 0.54$. From Figure 5.21, F lies between 0.1 and 0.2 (Equation 5.6 gives 0.15), so 10–15% melting would be required.

Question 6.3

(a) (i) Geotherm A does not intersect the melting reaction and so no melting occurs.

(ii) Geotherm B crosses the melting reaction (Equation 6.3) at depths of about 23 km, and this is where melting will occur. No melting is possible where the geotherm crosses the wet granite solidus (at about 18 km) because of the absence of both H_2O and alkali feldspar in the sediments.

(b) At 24 km, geotherm B lies within the sillimanite field, which is the polymorph that is observed in the Himalayan metasediments (Figure 6.3(c)).

Question 7.1

(a) The continental lithosphere beneath North America and Australia is generally seismically fast (blue coloured) in the top 110 km. This is entirely consistent with ancient (Archaean to Proterozoic) lithosphere being present in these continents, as the lithosphere has had a long time to cool. The western edge of North America is clearly seismically slow (red). In continental Africa the region along the East African Rift valley is seismically slow (red), whereas the interior of Africa is seismically fast (blue). This reflects the difference between a region of active volcanism and extension (East African Rift) and cooler Archaean and Proterozoic cratons.

(b) The East Pacific Rise has a small (~0.5%) slow anomaly (yellow) located beneath the position of the ridge to a depth of 600 km. This implies that warmer mantle material is convected upwards from great depth beneath the mid-ocean ridge, even though melting dominantly occurs in the upper 60 km of the mantle.

(c) Down to depths of 200 km there is a strong seismic slow region (red) that defines a narrow strip along the volcanic arc. This seismic slowing relates to arc magmatism. By 300 km depth the slow band is less clear, which may relate to some slab dehydration reactions. Between 400 and 600 km depth a broader zone of seismic slow velocities appears. Because it takes tens of millions of years for the slab to descend through the upper mantle, these images provide evidence for long-term subduction along the Pacific Rim. This is a case where seismic tomography provides more than a simple snapshot. Seismic fast zones located at great depths (>1000 km) have been used to infer plate dynamics back to 60 Ma.

(d) There are clearly zones of seismically slow material (red) that broadly extend down through the mantle to the CMB and which are associated with known

hot spots or at centres of volcanism. These include volcanic centres in the southwest Pacific associated with super-swell, Hawaii, East Africa and some Atlantic Ocean islands that have anomalies that extend through the mantle. This is best seen for the southwest Pacific, whereas the other localities mentioned above have weaker signals between 1300 km and 2300 km depth, suggesting a possible lack of continuity of the buoyant zone, or just poor resolution. Hot spots on continental North America generally have little if any slow signal in the lower mantle, suggesting they are an upper mantle phenomena.

Question 7.2

Incorporating Equation 7.5 ($\upsilon = \eta/\rho$) into Equation 7.4 gives:

$$Ra = \frac{\alpha g d^3 \Delta T \rho}{\kappa \eta} \qquad (7.7)$$

Hence, for $\Delta T = 1$:

$$Ra = \frac{2\times10^{-5}\ \mathrm{K^{-1}} \times 1\ \mathrm{K} \times 10\ \mathrm{m\ s^{-2}} \times 3\times10^{6}\ \mathrm{m^3} \times 3300\ \mathrm{kg\ m^{-3}}}{10^{-6}\ \mathrm{m^2\ s^{-1}} \times 10^{21}\ \mathrm{kg\ m^{-1}\ s^{-1}}}$$

$$= \frac{1.8\times10^{4}\ \mathrm{K\ K^{-1}\ m\ s^{-2}\ m^3\ kg\ m^{-3}}}{\mathrm{m^2\ s^{-1}\ kg\ m^{-1}\ s^{-1}}}$$

The units cancel out, showing that *Ra* is indeed a dimensionless number, therefore $Ra = 1.8 \times 10^4$. So if $\Delta T = 10^2$ K, then $Ra = 1.8 \times 10^6$.

Question 7.3

$$Pe_t = \frac{0.01\ \mathrm{m\ s^{-1}} \times 50\times10^{3}\ \mathrm{m}}{10^{-6}\mathrm{m^2\ s^{-1}}}$$

so $Pe_t = 5 \times 10^8$.

This calculation again illustrates how the units used to calculate a dimensionless number cancel out.

Pe_t is clearly much greater than unity, so the mantle beneath an ocean ridge must advect.

Question 7.4

Using the equation for equilibrium partial melting (Equation 5.6), the concentration of water in the melt phase is given by:

$$C_L = \frac{C_0}{D + F(1 - D)}$$

where C_L is the concentration in the melt, C_0 is the concentration in the source, D is the bulk partition coefficient and F is the melt fraction.

But $C_r = C_L D$, where C_r is the concentration in the residue. Therefore

$$C_r = \frac{C_0 D}{D + F(1 - D)}$$

And so for H_2O with $D = 0.005$, $F = 0.01$ (1% melting) and $C_0 = 1$ wt%

$$C_r = \frac{1 \text{ wt\%} \times 0.005}{0.005 + 0.01\,(1 - 0.005)}$$
$$= 0.33 \text{ wt\%}$$

Therefore 33 wt% of the H_2O remains in the residue.

The values for F = 0.05 (5% melting) and F = 0.1 (10% melting) are calculated in a similar way. So:

for $F = 0.01$, $C_r = 0.33 = 33\%$;

for $F = 0.05$, $C_r = 0.09 = 9\%$;

for $F = 0.1$, $C_r = 0.047 = 4.7\%$.

Question 7.5

The H_2O content of MORB is $7.5 \times 200 = 1500$ ppm.

You have calculated previously that 10% melting extracts 95% of the H_2O from the source (Question 7.4). By proportion, the amount of H_2O in the source can, therefore, be easily calculated to be ~160 ppm using a mass balance equation.

$0.1 \times 1500 \text{ ppm} = 0.95 \times X$; where X is the concentration of H_2O in the mantle.

$$\text{Therefore } X = \frac{150 \text{ ppm}}{0.95} = 157.9 \text{ ppm} = 160 \text{ ppm (2 sig. figs).}$$

Alternatively, the concentration of H_2O can be calculated from $C_0 = C_L \times (D + F(1 - D)) = 157.9$ ppm = 160 ppm (2 sig. figs).

Question 7.6

Mass of H_2O in the mantle = H_2O concentration in mantle × mass of mantle

$$= (100 \times 10^{-6}) \times (40 \times 10^{23}) \text{ kg}$$
$$= 0.4 \times 10^{21} \text{ kg}$$

If the H_2O concentration in the mantle is 500 ppm, then the H_2O mass is 2.0×10^{21} kg.

This compares with 1.4×10^{21} kg water in the ocean.

Appendix A

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	$W\ m^{-1}\ K^{-1}$	$m\ kg\ s^{-3}\ K^{-1}$

Appendix C

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	Y
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

Appendix D

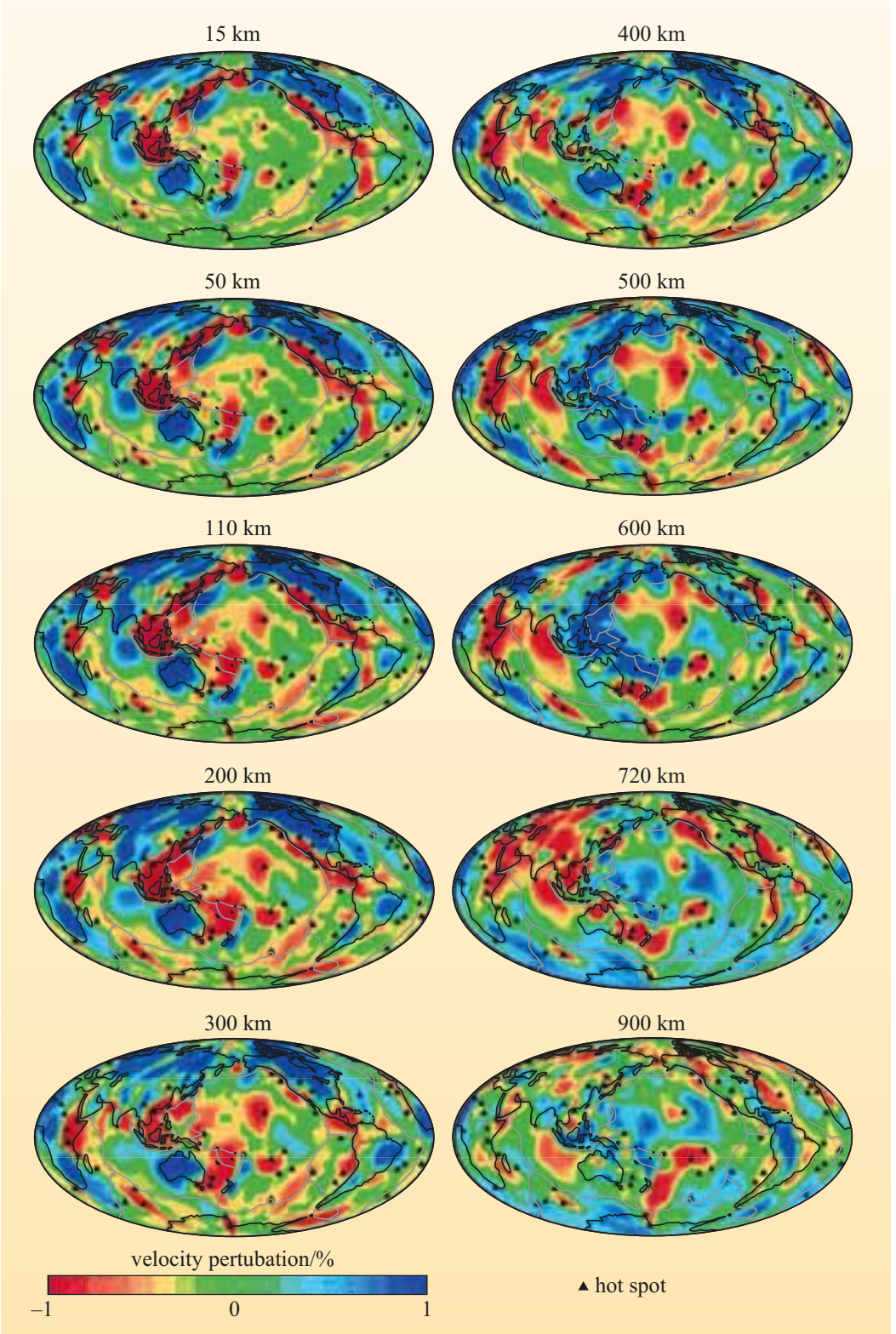
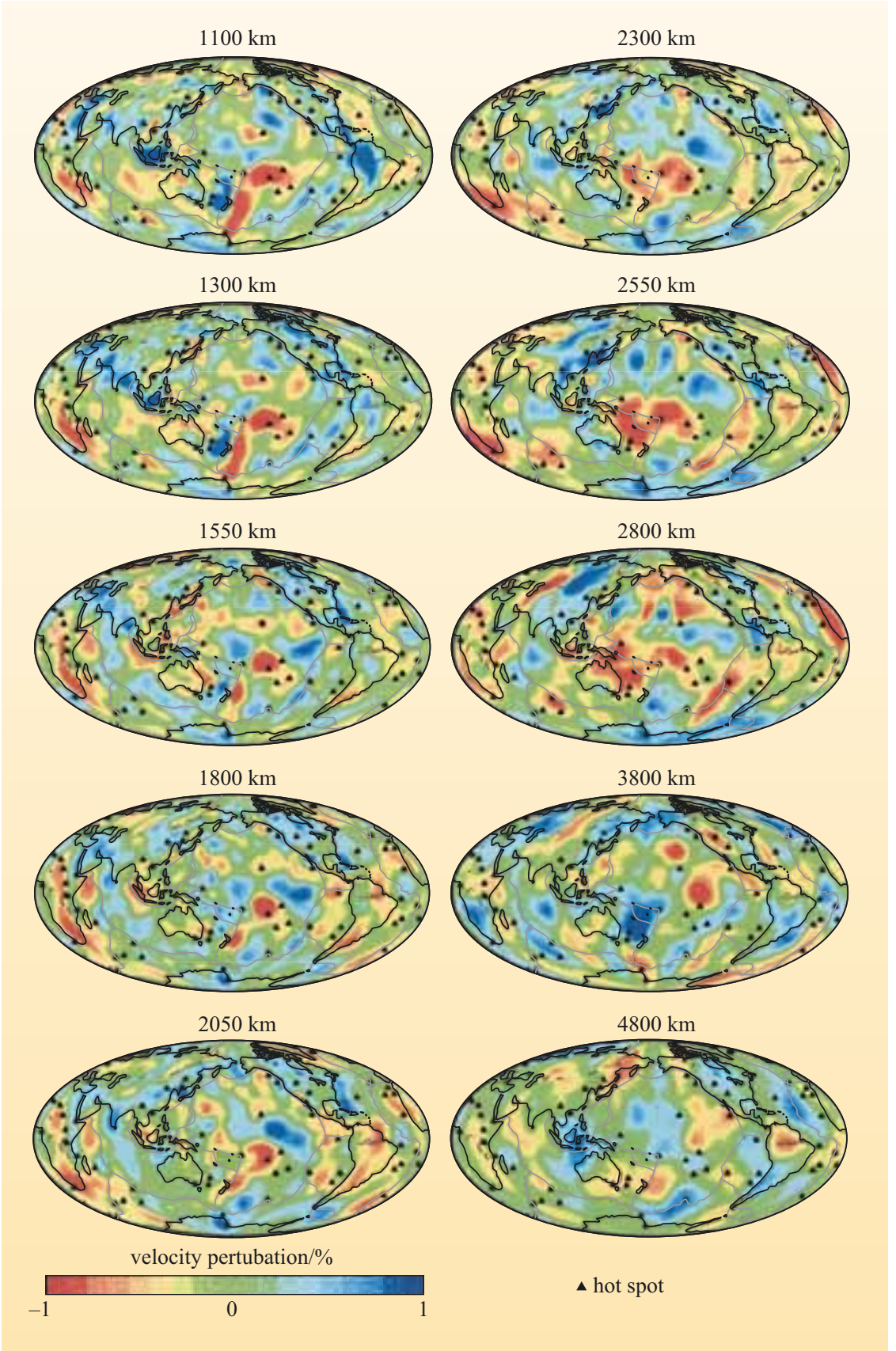


Figure 7.6 Global P-wave tomographic images sliced at different depths throughout the Earth. Red and blue colours denote the slow- and fast-velocity anomalies respectively. Dark lines are continental coastlines and paler lines are tectonic plate boundaries. (Lei and Zhao, 2006b)


1100 km
2300 km
1300 km
2550 km
1550 km
2800 km
1800 km
3800 km
2050 km
4800 km
velocity pertubation/%
–1
0
1
▲ hot spot

Further reading

Chapter 5

Bebout, G.E., Scholl, D.W., Kirby, S.H. and Platt, J.P. (1996) *Subduction: Top to Bottom*, American Geophysical Union Monograph Series, vol. 96, pp. 384. (A collection of papers (high-level reading) on various aspects of subduction zones.)

Chapter 6

Beaumont, C., Jamieson, R.A., Nguyen M.H. and Lee, B. (2001) 'Himalayan tectonics explained by extrusion of a low-viscosity crustal channel coupled to focused surface denudation', *Nature*, vol. 414, pp. 738–742. (A highly technical explanation of the channel flow model.)

Clarke, D.B. (1992) *Granitoid Rocks*, Chapman & Hall, pp. 283. (The complete guide to granites: geochemistry, mineralogy, field relations and origin.)

Harris, N.B.W. (2007) 'Channel flow of the Himalayan–Tibetan orogen – a critical review, *Journal of the Geological Society of London* (in press).

Hodges, K.V. (2000) 'Tectonics of the Himalaya and southern Tibet', *Geological Society of America Bulletin*, vol. 112, pp. 324–350. (A comprehensive review of the Himalayan orogen, but pre-dating the channel-flow model.)

The Open University (2001) S339 Understanding the continents, Block 4 *Mountain building*, Section 7 'The Himalaya and Tibet: a case study of collision', Milton Keynes, The Open University. (The only available comprehensive (but pre-channel flow) account of magma formation in collision zones.)

Yardley, B.W.D. (1989) *An Introduction to Metamorphic Petrology*, Longman Earth Science Series, pp. 248. (The best available introduction to metamorphic petrology.)

Chapter 7

Carlson, R.W. (2005) *The Mantle and Core: 2 (Treatise on Geochemistry)*, Oxford, Elsevier, pp. 586. (High-level, but a very up-to-data view of the geochemistry and evolution of the Earth's mantle and core.)

Fowler, C.M.R. (2005) *The Solid Earth: An Introduction to Global Geophysics*, Cambridge, Cambridge University Press.

Sparks, R.S.J. and Hawkesworth, C.J. (2004) *The State of the Planet: Frontiers and Challenges in Geophysics*, American Geophysical Union Monograph Series, vol. 150, pp. 410. (Some readable reviews of some key geophysical issues.)

van der Hilst, R.D., Bass, J.D., Matas, J. and Trampert, J. (2005) *Earth's Deep Mantle: Structure, Composition, and Evolution*, American Geophysical Union Monograph Series, vol. 160, pp. 334. (A collection of the most recent thoughts on the deeper structure of the planet. High-level, but some readable overviews.)

Chapter 8

Bryson, B. (2003) *A short history of nearly everything*, Doubleday.

Acknowledgements

Grateful acknowledgement is made to the following sources for permission to reproduce material in this book:

Cover photo Dirk Wierma/Science Photo Library.

Figure 5.2 Adapted from Scotese, C. R., Gahagan, L. M. and Larson, R. L. (1998) 'Plate tectonic reconstructions of the Cretaceous and Cenozoic ocean basins', *Technophysics*, vol. 155, Elsevier Science, and Fowler, C.M.R. (1990) *The Solid Earth: An Introduction to Global Geophysics*, Cambridge University Press; *Figure 5.3* Hasagawa, A. et al. (1978) 'Double-planed deep seismic zone', *Geophysical Journal of the Royal Astronomical Society*, vol. 54, Royal Astronomical Society; *Figure 5.4* D. Van der Hilst, Department of Earth, Atmospheric, and Planetary Sciences, MIT; *Figure 5.5* Nakajima, J. et al. (2001) 'Three-dimensional structure of V_p, V_s and V_p/V_s', *Journal of Geophysical Research*, vol. 106, no. B10, American Geophysical Union; *Figure 5.7* Peacock, S. M. (2000) *Thermal structure and metamorphic evolution*, American Geophysical Union; *Figure 5.8* van Keken, P. E., Kiefer, B. and Peacock, S. M. (2002) 'High-resolution models of subduction zones: implications for mineral dehydration reactions and the transport of water into the deep mantle', *G3 – An Electronic Journal of the Earth Sciences*, vol. 3, no 10. The American Geophysical Union; *Figure 5.9, 5.11 and 5.20* Poli, S. and Schmidt, M.W. (2002) 'Petrology of subducted slabs', *Annual Review of Earth and Planetary Sciences*, vol. 30. Copyright © 2002 by Annual Reviews, www.annualreviews.com; *Figure 5.10* Johnson, M. C. and Plank, T. (1999) 'Dehydration and melting experiments constrain the fate of subducted sediments', *G3 – An Electronic Journal of the Earth Sciences*, vol. 1, December 13, The American Geophysical Union; *Figure 5.13* Le Bas, M. J. et al. (1986) 'A Chemical Classification of Volcanic Rocks', *Journal of Petrology*, vol. 27, no. 3, Oxford University Press; *Figure 5.21 and 5.22* Adapted from Cox, K.G., Bell, J.D. and Pankhurst, R.J. (1979) *The Interpretation of Igneous Rocks*, Unwin Hyman; *Figure 5.27 and 5.28* Plank, T. and Langmuir, C. H. (1993) 'Tracing trace elements from sediment input to volcanic output at subduction zones', *Nature*, vol. 362. Copyright © Nature Publishing Group.

Figures 6.3, 6.9, 6.20 and 6.22 photographs courtesy of Nigel Harris; *Figure 6.7* photograph courtesy of Tom Argles; *Figure 6.11* MacKenzie, W. S. and Guilford, C. (1980) *Atlas of rock-forming minerals in thin section*, Longman.

Figure 7.1 and 7.8 Fowler, C.M.R. (2005) *The Solid Earth: An Introduction to Global Geophysics*, Cambridge University Press. Copyright © Cambridge University Press 2005 *Figure 7.2a* Poli, S. and Schmidt, M.W. (2002) 'Petrology of subducted slabs', *Annual Review of Earth and Planetary Sciences*, vol. 30. Copyright © 2002 by Annual Reviews. www.annualreviews.com *Figure 7.2b* Helmberger, D. et al. (2005) 'Deep mantle structure and the perovskite phase transition', *Proceedings of the National Academy of Sciences*, vol. 102, no. 48, 29 November 2005. Copyright © 2005 The National Academy of Sciences of the United States of America, all rights reserved. *Figure 7.3* Nigel Harris; *Figure 7.4 and 7.5* Lei, J. and Zhao, D. (2006) 'Global P-wave tomography: on the

Sources of figures, data and tables

Allègre, C.J. and Rousseau, D. (1984) 'The growth of continents through geological time studied by Nd isotope analysis of shales', *Earth and Planetary Sciences Letters*, vol. 67, pp. 19–34.

Beaumont, C., Jamieson, R.A., Nguyen, M.H. and Medvedev, S. (2004) Crustal channel flows: 1, Numerical models with applications to the tectonics of the Himalayan–Tibetan orogen, *Journal of Geophysical Research*, vol. 109, B06406, doi:10.1029/2003JB002809.

Boher, M., Abouchami, W., Michard, A., Albarède, F. and Arndt, N.T. (1992) 'Crustal growth in West Africa at 2.1 Ga', *Journal of Geophysical Research*, vol. 97, pp. 345–369.

Dietz, R.S. and Holden, J.C. (1970) 'The breakup of Pangaea', *Scientific American*, vol. 223, pp. 30–41, Scientific American Inc., NY.

Kramers, J.D. and Tolstikhin, I.N. (1997) 'Two major terrestrial Pb isotope paradoxes, forward modelling, core formation and the history of the continental crust', *Chemical Geology*, vol. 139, pp. 75–110.

Nelson, K.D., Zhao, W., Brown, L.D., Kuo, J., Che, J., Liu, X., Klemperer, S.L., Makovsky, Y., Meissner, R., Mechie, J., Kind, R., Wenzel, F., Ni, J., Nabelek, J., Leshou, C., Tan, H., Wei, W., Jones, A.G., Booker, J., Unsworth, M., Kidd, W.S.F., Hauck, M., Alsdorf, D., Ross, A., Cogan, M., Wu, C., Sandvol, E. and Edwards, M. (1996) 'Partially molten middle crust beneath southern Tibet: synthesis of project INDEPTH results', *Science*, vol. 274, pp. 1684–1695.

Rudnick, R. L. and Gao, S. (2005) 'Composition of the Continental Crust', *Treatise on Geochemistry*, vol. 3, pp. 1–64.

Taylor, S.R. and McClennan, S.M. (1995) 'The geochemical evolution of the continental crust', *Reviews of Geophysics*, vol. 33, pp. 241–265.