CHAPTER TWO

The physical state of the lithosphere

Summary

Knowledge of the behaviour of the lithosphere is essential if we are to understand the initiation and development of sedimentary basins. Lithospheric processes are responsible for the highly dynamic nature of the tectonics taking place near the surface of the Earth. The Earth's outer layer or lithosphere can be regarded as a thermal boundary layer between the cool atmosphere or oceans and the hot interior. The lithosphere is therefore a thermal entity, but it also has a physical significance. The upper portion of the lithosphere is sufficiently rigid that it is able to store and transmit stresses. Since the deformations caused by applied forces are generally recoverable, this outer zone is known as the elastic lithosphere.

Applied forces of whatever origin cause stresses that result in deformation or strain. The simplest view of applied forces is of those due simply to the weight of an overlying rock column, known as *lithostatic stress*. The difference between the actual stress and the lithostatic stress is a tectonic contribution known as *deviatoric stress*. Deviatoric stresses can be either tensile or compressive and are commonly transmitted from plate boundaries or originate from the gravitational potential energy due to elevation.

In an elastic solid there is a clear relationship between the stresses and resultant strains. The exact relationship depends on material properties known as Young's modulus and Poisson's ratio. Where only one of the principal axes is non-zero, a *uniaxial* state of stress is said to occur, and the relation between stress and strain is called *Hooke's law*. If there are two non-zero components of principal stress, we have the condition termed *plane stress*. In an analogous fashion, *uniaxial strain* and *plane strain* refer to the coordinate system of principal strains. In a state of stress where all the principal stresses are equal, an *isotropic* state of stress, the fractional volume change caused by isotropic compression is given by the *bulk modulus* or its reciprocal, the *compressibility*. Since rocks have a finite compressibility, they increase in density with depth in the Earth, but they also expand due to heating. The net effect on rock density is critical to understanding the long-term gravitational stability of the continental lithosphere.

The lithosphere is able to bend or flex. The shape of the flexure of the lithosphere depends on its rigidity and on the nature of the applied force or load causing the bending. Bending is accompanied by longitudinal stresses and bending moments in the plate. The bending moment is related to the local radius of curvature by a coefficient called *flexural rigidity*. The bending stresses around the outer arc of flexed lithosphere may be sufficient to cause faulting or fault reactivation.

In order to understand the mechanical behaviour of the Earth it is necessary to understand its thermal structure, since this determines its rheology (Gk. *rheos*, stream, flow). In the lithosphere, heat transfer is predominantly by conduction, whereas in the mantle, convection is extremely important. Convection is the natural result of differential heating above a critical Rayleigh number, and estimates for Earth's mantle strongly suggest that convection must be taking place. A thermal boundary layer develops along the upper surface of a convecting material because heat is lost by conduction to the surface. This cool boundary layer, the lithosphere, detaches and sinks due to gravitational instability along subduction zones. For conduction, the heat flux is related to the temperature gradient by a coefficient, the *thermal conductivity*. Heat fluxes in the continents are determined primarily by conduction from radioisotopic heat sources, whereas in the oceans they reflect cooling of newly created oceanic lithosphere. The variation of temperature with depth is known as the *geotherm*. Measurements of continental heat flows indicate a linear relationship between radiogenic heat generation and surface heat flow, with the intercept indicative of the 'reduced' or basal heat flow from the mantle.

Since radiogenic heat-producing elements (HPEs) are concentrated in the upper crust, they can easily be mobilised by surface processes of erosion and sedimentation, which potentially has an important effect on the continental geotherm. The crust beneath a self-heating basin is hotter because of the effect of the sedimentary blanket. This heating may be sufficient to reactivate old faults or cause new tectonic deformation. Erosion and sediment deposition also have a potentially important transient effect on the underlying geotherm. The time scale for the recovery of the geotherm following an instantaneous erosion or deposition event is of the order of 10^6–10^7 years. Rapid exhumation, where the upward advection of hot crustal rocks outweighs the effect of conductive cooling, causes a marked curvature in the geotherm, with important implications for the interpretation of thermochronometers such as zircon and apatite fission track.

The reference equipotential surface for the Earth is termed the *geoid*. Some geoid anomalies probably represent density differences in the Earth set up by plate tectonic processes. The analysis of surface gravity data (*Bouguer anomalies*) provides important information on the way in which topography is compensated isostatically. Loads are supported by the flexural strength of the lithosphere, the degree of compensation

Basin Analysis: Principles and Application to Petroleum Play Assessment, Third Edition. Philip A. Allen and John R. Allen.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

depending on the flexural rigidity of the plate and the wavelength of the load. Techniques such as admittance and coherence allow the flexural rigidity to be estimated from a correlation of the Bouguer gravity anomaly with the spectral wavelength of the topography.

Mantle convection is thought to take place by means of a thermally activated creep. Crustal rocks may also behave in a ductile manner by pressure-solution creep, or they may deform in a brittle way by fracturing (*Byerlee's law*). The occurrence of earthquake foci and the results of laboratory experiments on rock mechanics suggest that at depths of about 15–25 km and temperatures of c.300 °C the continental crust starts to become ductile or plastic, but the rheology of the crust is complex owing to its compositional heterogeneity. The onset of ductility in the middle crust may serve mechanically to decouple the upper crust from the lower crust and mantle lithosphere. The lower lithosphere may deform as an elastic solid on short time scales but viscously on longer time scales. The time scale of the viscous relaxation of stresses is not fully understood but may be estimated from studies of postglacial rebound. In the case of a flexed plate, plastic deformation may take place if a critical elastic curvature is exceeded.

The key to understanding the deformation of the lithosphere is the profile of lithospheric strength with depth. Oceanic and continental lithospheres have different strength profiles. Strength depends on composition, temperature, volatile content such as water, and presence of heterogeneities acting as weaknesses. The strong part of the lithosphere not only supports loads elastically, but also determines the depth of occurrence of earthquakes. There are two competing models for the strength profile of the continental lithosphere. The 'jelly sandwich' model envisages a strong, brittle, seismogenic upper crust and mantle lithosphere separated by a weak, ductile lower crust, whereas the 'crème brûlée' model proposes that there is one strong, seismogenic layer restricted to the crust.

Sedimentary basins represent a physical deformation of the lithosphere. In order to understand how basins are initiated and evolve through time, it is necessary to have some appreciation of the physical state of the lithosphere. This physical state can be thought of in terms of its thickness, mineralogy, thermal structure, and consequently its strength when acted upon by tangential forces or its rigidity when flexed by orthogonal forces. This chapter introduces some fundamental ideas on the physical state of the lithosphere and how this determines its deformation or flow. In the succeeding chapters in Part 2, the particular cases of the stretching (extension) of the lithosphere (Chapter 3) and the bending (flexure) of the lithosphere (Chapter 4) are considered before investigating the large-scale interaction of the lithosphere and mantle in Chapter 5 and processes in zones of strike-slip deformation in Chapter 6.

2.1 Stress and strain

2.1.1 Stresses in the lithosphere

Body forces on an element of a solid act throughout the volume of the solid and are directly proportional to its volume or mass. For example, the force of gravity per unit volume is the product of ρ, the density, and g, the acceleration of gravity. The body forces on rocks within the Earth's interior depend on their densities, but density is itself a function of pressure. If we normalise rock densities to a zero-pressure value (Appendix 1), typical mantle rocks would have densities of about 3250 kg m^{-3}, ocean crust (basalt and gabbro) would have densities of about 2950 kg m^{-3}, whereas continental granites and diorites would be in the range of 2650 to 2800 kg m^{-3}. Sedimentary rocks are highly variable, ranging from 2100 kg m^{-3} for some shales to 2800 kg m^{-3} for very compact marbles (Tables 2.1, 2.2).

Surface forces act only on the surface area bounding a volume and arise from the interatomic stresses exerted from one side of the surface to the other. The magnitude of the force depends on the surface area over which the force acts and the orientation of the surface. The normal force per unit area on horizontal planes increases linearly with depth. That due to the weight of the rock overburden is known as *lithostatic stress* or *lithostatic pressure*. This concept of surface forces in the Earth's interior is made use of in considering the way in which hydrostatic equilibrium (Archimedes principle) influences the support of the oceanic and continental plates by the mantle, an important application known as *isostasy*. Appendix 2 deals with the isostatic balance for a range of geological settings. We will examine a simple application of isostasy in studying the bending of lithosphere in §2.1.5.

Normal surface forces can also be exerted on vertical planes (Fig. 2.1). If the normal surface forces, σ_{xx}, σ_{yy} and σ_{zz}, are all equal and

Table 2.1 Physical (thermal) properties of common minerals (after Goto & Matsubayashi 2008)

Mineral	Density ρ (kg m^{-3})	Thermal conductivity K (W m^{-1} K^{-1})	Specific heat c (J kg^{-1} K^{-1})	Thermal diffusivity κ (10^{-6} m^2 s^{-1})
Quartz	2648	7.69	741	3.92
Albite	2620	2.2	776	1.08
Anorthite	2760	1.68	745	0.82
Orthoclase	2570	2.32	707	1.28
Muscovite	2831	2.32	796	1.03
Illite	2660	1.85	808	0.86
Smectite	2608	1.88	795	0.91
Chlorite	2800	5.15	818	2.25
Calcite	2710	3.59	820	1.62
Seawater	1024	0.59	3993	0.15
Mud (grain)	2731	3.4	758	1.64

Table 2.2 Physical and thermal properties of common rocks (in part from Turcotte & Schubert 2002; Blackwell & Steele 1989; Jaupart & Mareschal 2011, p. 416, p. 422, thermal conductivities and diffusivities at room temperature)

	Zero-pressure density kg m^{-3}	Young's modulus E 10^{11} Pa	Poisson's ratio ν	Thermal conductivity K W m^{-1} K^{-1}	Thermal diffusivity κ m^2 s^{-1} ($x10^{-6}$)	Coefficient of thermal expansion α 10^{-5} K^{-1}
Sedimentary						
Shale	2100–2700	0.1–0.7	0.1–0.2	1.2–3.0	0.8	
Clay/siltstone				0.80–1.25		
Sand				1.7–2.5		
Sandstone	1900–2500	0.1–0.6	0.1–0.3	1.5–4.2	1.3	3.0
Limestone	1600–2700	0.5–0.8	0.15–0.30	2.0–3.4	1.2	2.4
Dolomite	2700–2800	0.5–0.9	0.1–0.4	3.2–5.0	2.6	
Halite			0.15	5.4–7.2		13
Metamorphic						
Gneiss	2600–2850	0.4–0.6	0.15–0.25	2.1–4.2		
Amphibolite	2800–3150		0.4	2.1–3.8		
Marble	2670–2750	0.3–0.8	0.2–0.3	2.5–3.0	1.0	
Quartzite					2.6	
Igneous						
Basalt	2950	0.6–0.8	0.2–0.25	1.3–2.9	0.9	
Granite	2650	0.4–0.7	0.2–0.25	2.4–3.8	1.0–1.6	2.4
Diabase	2900	0.8–1.1	0.25	2.0–4.0		
Gabbro	2950	0.6–1.0	0.15–0.20	1.9–4.0		1.6
Diorite	2800	0.6–0.8	0.25–0.30	2.8–3.6	1.2	
Granodiorite	2700	0.7	0.25	2.0–3.5		
Mantle						
Peridotite	3250			3.0–4.5	1.7	2.4
Dunite	3000–3700	1.4–1.6		3.7–4.6		
Miscellaneous						
Water				0.6	0.15	
Ice	917		0.31–0.36	2.2	1.2	5.0
Soil				0.17–1.10	0.2–0.6	

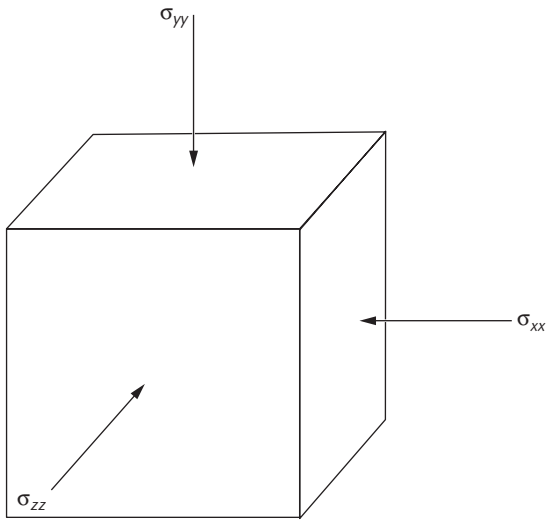

Fig. 2.1 Normal surface forces acting on vertical and horizontal planes. After Turcotte & Schubert (2002).

they are also equal to the weight of overburden, the rock is said to be in a *lithostatic state of stress*. The normal surface forces, σ_{xx}, σ_{yy} and σ_{zz}, are rarely equal when a rock mass is being subjected to tectonic forces. In such a case, the total horizontal surface force (normal stress) acting on a continent, for example, would be made up of two components, a lithostatic term and a tectonic contribution known as a deviatoric stress ($\Delta\sigma_{xx}$),

$$\sigma_{xx} = \rho_c g y + \Delta\sigma_{xx} \qquad [2.1]$$

where ρ_c is the density of the continent. Normal stresses can be either *tensile* when they tend to pull on planes or *compressive* when they push on planes. Horizontal deviatoric stresses may result from uplift producing excess potential energy (Appendices 3, 4) or may be transmitted from plate boundaries, when they are commonly referred to as *in-plane* or *intraplate stresses*.

Surface forces acting parallel to a surface are known as *shear stresses*. Examples are provided by a thrust sheet with a lower fault plane that experiences a frictional resistance, or the gravitational sliding of a rock mass down an inclined plane.

Stress components can be generalised at any point in a material by using the x, y, z coordinate system. At any point we can envisage three mutually perpendicular planes on which there are no shear stresses. Perpendiculars to these planes are known as *principal axes of stress* and can be labelled

σ_1 = maximum principal axis
σ_2 = intermediate principal axis
σ_3 = minimum principal axis,

where convention is that σ is positive for compressional stress and negative for extensional stress. There are certain states of stress that can be described by use of the principal axes notation:

uniaxial stress has a finite σ_1, and $\sigma_2 = \sigma_3 = 0$
biaxial or *plane stress* has $\sigma_1 > \sigma_2$ and $\sigma_3 = 0$
triaxial stress is the general state $\sigma_1 > \sigma_2 > \sigma_3$.

If the principal stresses, σ_1, σ_2 and σ_3, are identical, the state of stress is isotropic and any principal stress is equal to the pressure. In such a case, any set of orthogonal axes qualifies as a principal axis coordinate system. This is known as a *hydrostatic state of stress*. Where the state of stress is not isotropic, the pressure is equal to the mean of the normal stresses

$$p = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3) \qquad [2.2]$$

Subtraction of the mean stress (that is, pressure) from the normal stress component reveals the *deviatoric normal stresses*. It is common for two of the principal stresses to be non-zero, giving the state of *plane stress*. Such a system is suggestive of the horizontal stresses in the lithosphere caused by tectonic processes.

The state of stress of the lithosphere is determined principally by lateral changes of gravitational potential caused by variations in density and elevation of the surface, and by transmission from plate boundaries, basal traction and underlying mantle flow structures. The stress field caused by lateral variations in gravitational potential is called an *ambient lithospheric stress state* (Dahlen 1981), and a plate-scale mean gravitational potential energy can be defined (Coblentz & Sandiford 1994) (Fig. 2.2). Departures of gravitational potential in a lithospheric column from the mean plate-scale value reveals the state of stress at that column, which has important implications for continental tectonics and the development of sedimentary basins. Slow-moving plates surrounded by mid-ocean ridges, such as Africa and Antarctica, are most likely to display an ambient lithospheric stress state since the magnitudes of plate boundary and basal tractional forces are likely to be small.

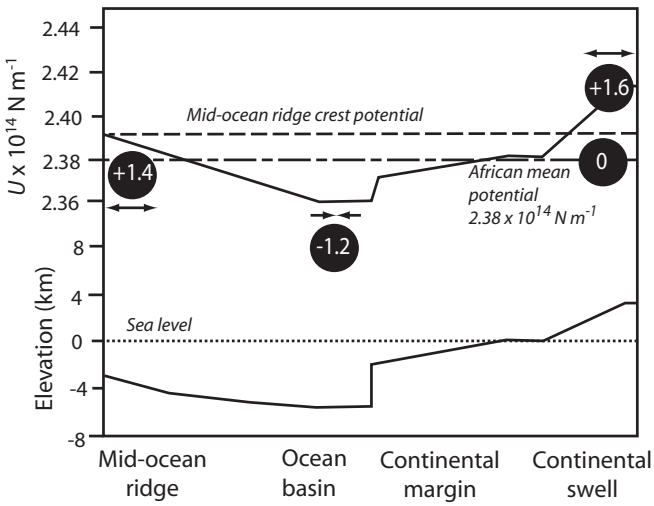

Fig. 2.2 Calculated potential energy U along a transect from an elevated continental swell, an Atlantic-type continental margin, deep ocean basin and mid-ocean ridge, with values of potential difference between calculated values and plate-scale mean potential for Africa in black circles. Profile at base is bathymetry and topography along transect. Modified from Coblentz & Sandiford (1994, p. 832, fig. 1). Calculations of U use a continental crustal density of 2750 kg m^{-3}.

The gravitational potential energy per unit area of a column of rock above a depth y is given by the integral of the vertical stress σ_{yy} from y to the surface h (Molnar & Lyon-Caen 1988). For Africa, the plate-scale mean gravitational potential energy per unit area is 2.4×10^{14} N m^{-1}. This is similar to the potential energy of cooling oceanic lithosphere at a water depth of nearly 4.5 km and continental lithosphere with an elevation of just 70 m (Fig. 2.2). Most of the continental area of Africa (mean elevation 875 m) therefore has a gravitational potential energy greater than the plate-scale mean, implying that it is in deviatoric tension. Highly elevated regions such as the Ethiopian swell, East African Rift and southern Africa have large horizontal extensional stresses of 15 MPa, 9 MPa and 8 MPa respectively – enough to cause continental stretching (see Chapter 3). In other plates, where there is a convergent margin, such as the northern boundary of the Indo-Australian plate, the state of stress is strongly influenced by plate boundary effects (Hillis & Reynolds 2000) (Fig. 1.13).

The global *in situ* stress field was compiled in the World Stress Map (Zoback 1992), and is continually updated and re-released. Stress orientations in the crust are based on focal mechanism solutions of earthquakes from a range of seismogenic depths, borehole break-outs mostly from sedimentary basins <4 km deep, fault slip data from surface outcrops, and fracture orientations from engineering activities generally restricted to the top 1 km. These various data sources are of different reliability and originate from different depth ranges.

Some continental plates (North and South America, western Europe) show horizontal stress orientations parallel to the direction of plate motion, indicating that the forces driving and/or resisting plate motion are also responsible for the regional stress orientations. However, in the Indo-Australian plate (Hillis & Reynolds 2000) stress orientations do not parallel the N to NNE absolute motion direction and appear to be controlled by its complex convergent boundary stretching from the Himalayas to New Guinea in the north and its largely transcurrent boundary in New Zealand in the east, combined with the effects of ridge-push along its long southern boundary (Fig. 1.13).

The magnitude and orientation of intraplate stress of lithospheric plates is therefore most satisfactorily explained by a combination of topographic (gravitational potential energy) and plate boundary forces (Coblentz *et al.* 1998). This intraplate stress field potentially controls the reactivation of old tectonic structures (Hand & Sandiford 1999), the compartmentalisation of sedimentary basins into smaller depocentres, and may be the source of deviatoric stresses driving lithospheric extension (Chapter 3).

2.1.2 Strain in the lithosphere

Strain is the deformation of a solid caused by the application of stress. We can define the components of strain by considering a rock volume with sides δx, δy and δz, which changes in dimensions but not in shape, so that the new lengths of the sides after deformation are $\delta x – \varepsilon_{xx}\delta x$, $\delta y – \varepsilon_{yy}\delta y$ and $\delta z – \varepsilon_{zz}\delta z$, where ε_{xx}, ε_{yy} and ε_{zz} are the strains in the x, y and z directions (Fig. 2.3). As long as the deformation of the volume element is relatively small, the volume change, or *dilatation*, is simply the sum of the strain components ($\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}$). Volume elements may also change their position without changing their shape, in which case the strain components are due to *displacement*.

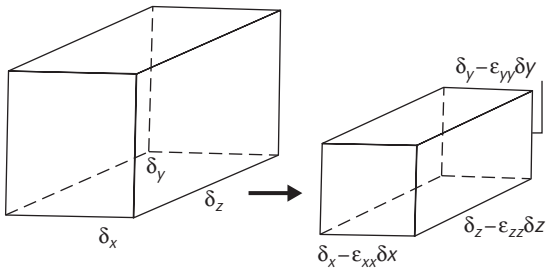

Fig. 2.3 A rectangular block that changes its dimensions but not its shape: this is a deformation involving no shear. After Turcotte & Schubert (2002).

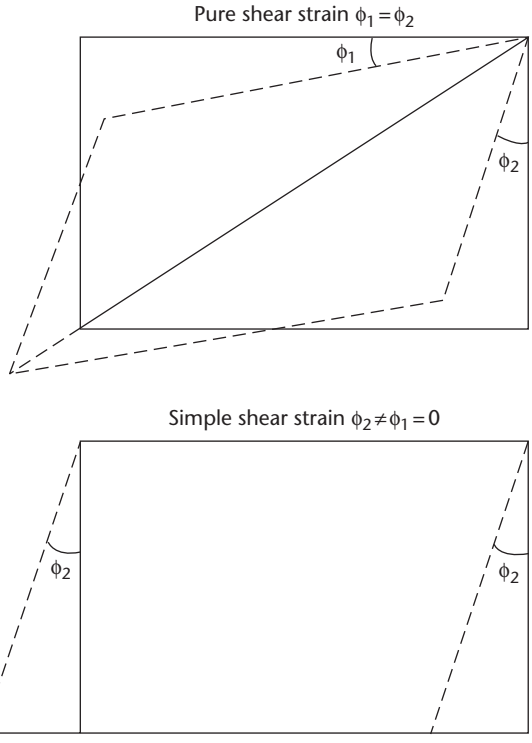

Fig. 2.4 Difference between pure shear strain (no solid body rotation) and simple shear strain (solid body rotation is $\phi_2/2$). After Turcotte & Schubert (2002).

Shear strains, however, may distort the shape of an element of a solid. In the case of a two-dimensional rectangular element that is distorted to a parallelogram, the shear strain is dependent on the amount of rotation of the sides of the rectangular element. Thus, shear strain in the two dimensions, x and y, is determined by the angles through which the sides of the rectangle are rotated (Fig. 2.4).

If the angles through which the sides of the rectangle are rotated (ϕ_1 and ϕ_2 in Fig. 2.4) are not equal, *solid body rotation* is said to have occurred (Fig. 2.5). Solid body rotations do not involve changes in the distances between neighbouring elements of a solid and therefore do not reflect strain.

The deformation of any element can now be described according to the shear strain and the solid body rotation. If no solid body rotation occurs, $\phi_1 = \phi_2$, and the deformation is a result only of shear strains, it is known as *pure shear*. If there is solid body rotation, but $\phi_1 = 0$, the element has undergone *simple shear*. Fig. 2.5 illustrates the two circumstances. We shall see how models of pure shear (e.g. uniform extension with depth) and simple shear (e.g. asymmetrical extension associated with trans-lithospheric shear zones) have been applied to the formation of extensional basins (Chapter 3).

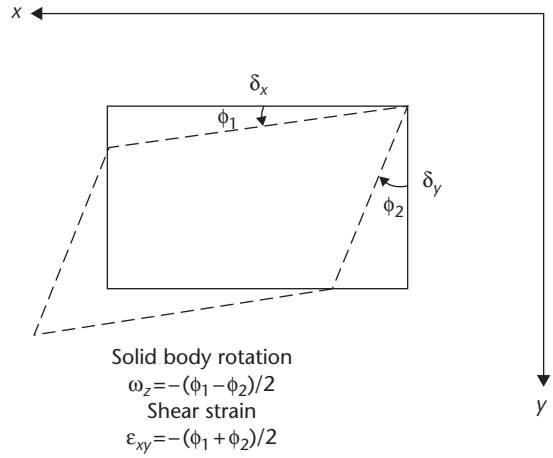

Fig. 2.5 Deformation of a rectangle into a parallelogram by a strain field involving shear. After Turcotte & Schubert (2002).

As in the case of shear stresses, shear strains can be described with reference to a coordinate system that is orientated such that shear strain components are zero. Such a system contains the *principal axes of strain*. The fractional changes in length along the directions of the principal strain axes are the *principal strains*. In the three-dimensional case the condition of isotropic strain is satisfied by

$$\varepsilon = (\varepsilon_1 + \varepsilon_2 + \varepsilon_3)/3 = \Delta/3 \qquad [2.3]$$

where Δ is the dilatation and ε is the mean normal strain.

It is very rare, however, for strain to be homogeneous. *Deviatoric strain* components are strains that are the difference between the actual strain and the mean normal strain. Deviatoric strains invariably result from the operation of tectonic processes. Their analysis therefore greatly aids the interpretation of lithospheric deformation.

In an analogous fashion to the treatment of states of stress, we can define different states of strain:

- *Uniaxial strain* is where there is only one non-zero component of principal strain, that is, assuming the non-zero axis to be ε_1, $\varepsilon_2 = \varepsilon_3 = 0$.
- *Plane strain* is where only one of the principal strain components is zero, for example $\varepsilon_3 = 0$ and ε_1 and ε_2 are non-zero. It is a common starting point for studying lithospheric deformation, since it can be assumed that the strain in the direction of an infinite plate will be zero.

Strain in the lithosphere commonly results from tectonic activity. This tectonic activity, such as the slip on faults, is manifested at the surface as a co-seismic surface strain. Although these surface strains may be very small, creep-like displacements, the strains resulting

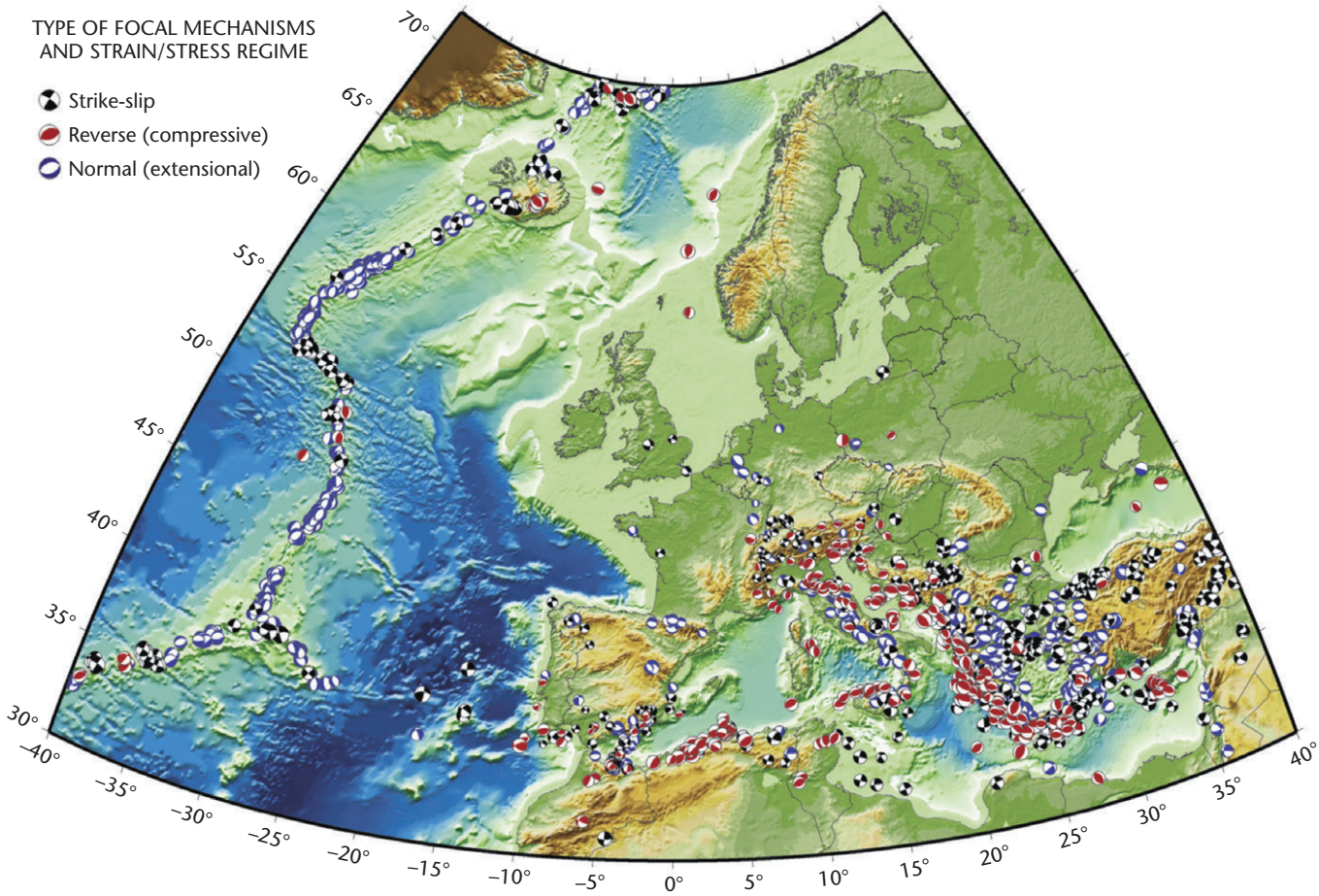

Fig. 2.6 Focal mechanism solutions for earthquakes in the Eurasia-Mediterranean area, overlain by topography, comprising 1608 earthquakes with Mw values between 2.6 and 7.5. Reprinted from Olaiz *et al.* (2009), with permission from Elsevier.

from single large earthquakes may be measured in several metres of displacement at the surface, causing railway tracks, fences and roads to be visibly offset. Repeated co-seismic displacements allow the strain rate to be calculated, with units of $[T]^{-1}$. The first ground motion associated with earthquakes (known as first motion studies or focal mechanism studies) allows the sense of motion to be identified as extensional, contractional or strike-slip (Fig. 2.6).

Classical geodesy involves the construction of benchmarks in a geodetic network and the use of triangulation to measure surface strains with a theodolite. However, tectonic displacements can now be measured from satellites in space, especially using the global positioning system (GPS) (Figs 2.6, 2.7). Synthetic aperture radar interferometry is able to make radar backscatter images of the Earth's surface before and after a tectonic movement, thereby allowing the displacement to be accurately quantified (Fig. 2.8). Radar interferometry allows the distribution of surface strain to be imaged at scales from long plate boundaries to kilometre-scale segments of single faults.

2.1.3 Linear elasticity

It is important to know the relationship between the stress and the strain in a piece of lithosphere. These relationships reflect the basic flow laws of Earth materials (§2.3). *Elastic* materials deform when they are subjected to a force and regain their original shape and volume when the force is removed. For relatively low temperatures, pressures and applied forces, almost all solid materials behave elastically. The relation between stress and elastic strain is linear.

However, at high temperatures and pressures or high levels of stress, rocks do not behave elastically. In near-surface regions where temperature and pressures are low, rocks deform by brittle fracture at high levels of stress. Deeper in the Earth, high temperatures and pressures cause the rock to deform plastically under an applied force, with no fracturing. Brittle materials that have exceeded their yield strength, and plastic materials, do not regain their original shape when the force is removed.

Since much of the lithosphere behaves as a strong material over geological (i.e. $>10^6$ yr) periods of time, it is able to bend under surface loads, to store the elastic stresses responsible for earthquakes, and to transmit stresses over large horizontal distances. This fundamental property of the lithosphere is crucial to an understanding of the formation of sedimentary basins.

The theory of linear elasticity underpins a great deal of thought on lithospheric mechanics and often constitutes the basic assumption in models of lithospheric behaviour. In a linear, isotropic, elastic solid the stresses are linearly proportional to strains, and mechanical

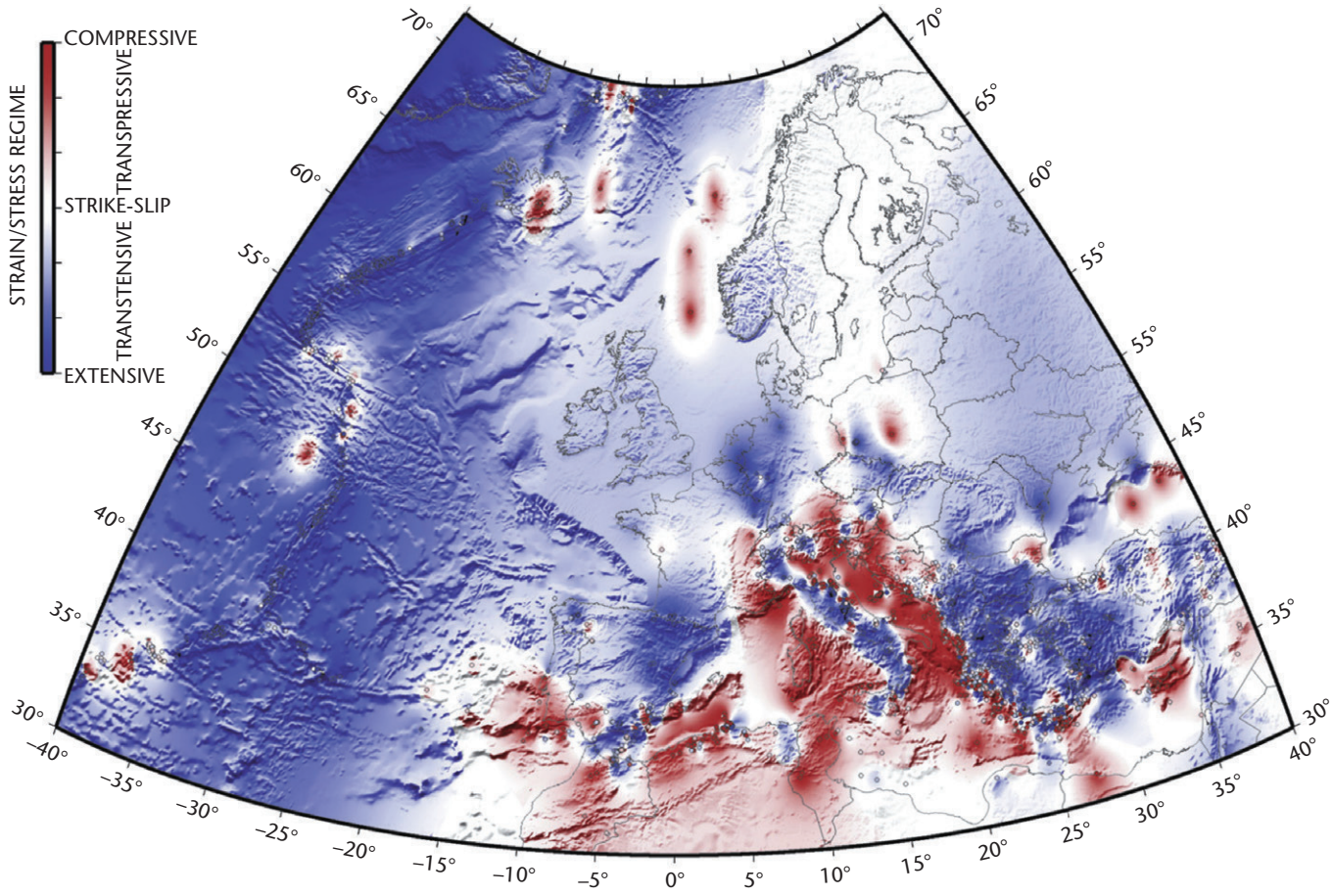

Fig. 2.7 Map of continuous stress-strain for the western Eurasian plate. Colour coding is based on calculations of the ratio of the maximum horizontal strain divided by the maximum vertical strain. Note the high amount of extensional strain in the Aegean Sea. Note also the change from the compressional Adriatic domain to the extensional Tyrrhenian domain in Italy. Reprinted from Olaiz *et al.* (2009), with permission from Elsevier.

properties have no preferred orientation. The principal axes of stress and the principal axes of strain coincide. The relation between the principal strain and the components of principal stress can be stated as follows.

$$\varepsilon_1 = \frac{\sigma_1}{E} - \frac{\nu\sigma_2}{E} - \frac{\nu\sigma_3}{E} \qquad [2.4]$$

$$\varepsilon_2 = -\frac{\nu\sigma_1}{E} + \frac{\sigma_2}{E} - \frac{\nu\sigma_3}{E} \qquad [2.5]$$

$$\varepsilon_3 = -\frac{\nu\sigma_1}{E} - \frac{\nu\sigma_2}{E} + \frac{\sigma_3}{E} \qquad [2.6]$$

The exact partitioning of stresses to give a resultant strain is clearly strongly influenced by E and ν, which are material properties known as *Young's modulus* and *Poisson's ratio* respectively. In general terms, a principal stress produces a strain component σ/E along the same axis and strain components $-\nu\sigma/E$ along the two other orthogonal axes.

Where only one of the principal stresses is non-zero (uniaxial stress), a shortening in the direction of the applied compressive stress will be accompanied by an extension in the two orthogonal directions (Fig. 2.9), and vice versa. Under these conditions where, let us say $\sigma_2 = \sigma_3 = 0$ and $\sigma_1 \neq 0$, there is a simple relation along the axis of uniaxial stress

$$\sigma_1 = E\varepsilon_1 \qquad [2.7]$$

This well-known relationship is called *Hooke's law* (Fig. 2.9).

There is a fractional volume change or dilatation due to uniaxial stress, but the contraction in the direction of uniaxial stress is compensated by expansion by half as much in the two other orthogonal directions.

In an isotropic state of stress, all the principal stresses and principal strains are equal. The pressure under these conditions is related to the dilatation by K, the *bulk modulus*, or its reciprocal β, the *compressibility*. These parameters therefore give the fractional volume change during isotropic compression under a given pressure. If there is a volume change, in order that matter is conserved, there must be

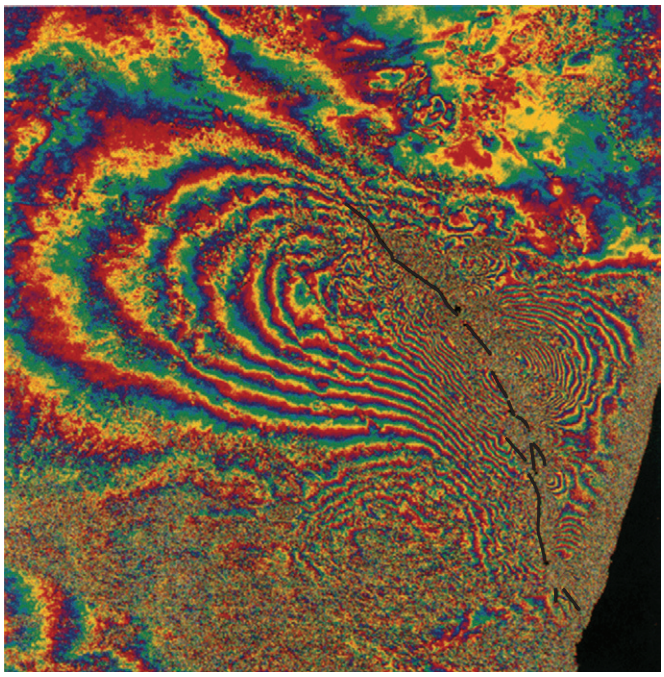

Fig. 2.8 Radar interferometric image of the surface displacement resulting from the Landers earthquake, California. Each colour cycle represents 2.8 cm of displacement. The surface rupture of the fault is shown in black. From Massonnet *et al.* 1993, reproduced in colour at http://southport.jpl.nasa.gov/scienceapps/dixon/report3.html. Reproduced with permission of Nature Publishing Group.

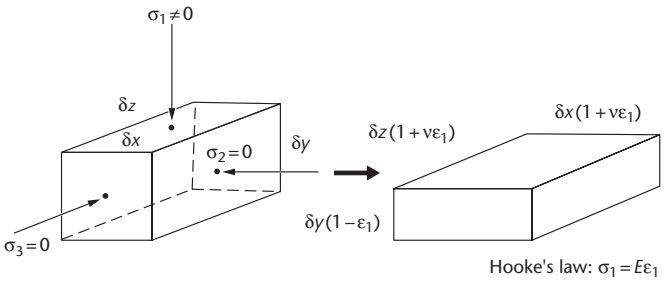

Fig. 2.9 Deformation under a uniaxial stress. Contraction in the direction of the compressive stress σ_1 is compensated by extension in the two orthogonal directions. Hooke's Law is the relation between stress and stress along the principal axis of compressive stress. After Turcotte & Schubert (2002).

an increase in density (see §2.1.6 and Appendix 1). Such a density increase $\delta\rho$ is given simply by

$$\delta\rho = \rho\beta p \qquad [2.8]$$

where p is the pressure, ρ is the density of the solid element, and β the compressibility. In terms of the previously defined Young's modulus E and Poisson's ratio ν,

$$K = \frac{1}{\beta} = \frac{E}{3(1-2\nu)} \qquad [2.9]$$

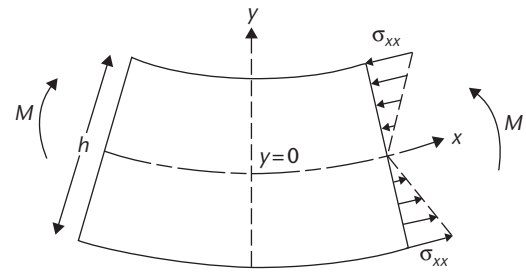

Fig. 2.10 The flexure of a beam of thickness h, showing compressional fibre stresses in the half above the mid-plane, and extensional fibre stresses below the mid-plane. The integration of the fibre stresses over a notional cross section of the flexed beam gives the bending moment M. The bending moment is related to the curvature by a coefficient known as flexural rigidity D. After Turcotte & Schubert (2002). See Appendix 5. © Cambridge University Press, 2002.

showing that as ν approaches 1/2 the bulk modulus tends to infinity, that is, the material becomes essentially incompressible.

2.1.4 Flexure in two dimensions

Since the lithosphere behaves elastically, it is able to bend when force systems or loads are applied to it. We will return to this topic in considerable detail when considering the initiation and maintenance of foreland basins (Chapter 4). The aim here is to provide a brief background of the way the lithosphere responds by flexure to these applied force systems. A full analysis is provided by Turcotte and Schubert (2002), and the derivation of the general flexure equation is given in Appendix 5. The concepts involved in the flexure of an elastic solid may be briefly summarised as follows:

- Flexure results from vertical forces, horizontal forces and torques (bending moments) in any combination. Horizontal loads are commonly neglected, perhaps unwisely, in many geodynamical problems.
- The bending moment is the integration of the fibre (normal) stresses on notional cross sections of the plate acting over the distance to the mid-plane of the plate (Fig. 2.10). The bending moment is related to the local radius of curvature of the plate by a coefficient called the *flexural rigidity*. Flexural rigidity is proportional to the cube of the equivalent *elastic thickness*. When applied to the lithosphere, the equivalent elastic thickness does not represent a real physical discontinuity.
- A general flexural equation (see eqn. [2.10]) can be derived that expresses the deflection of the plate in terms of the vertical and horizontal loads, bending moment and flexural rigidity. This equation can readily be adapted for use in the study of geological problems.

There will be different bending stresses under situations where the plate is pinned at one or both ends, and according to whether it is point loaded (for example, at its free end) or uniformly loaded along its length, or loaded in some other fashion. However, the general flexural equation provides the basic starting point for more specific analyses.

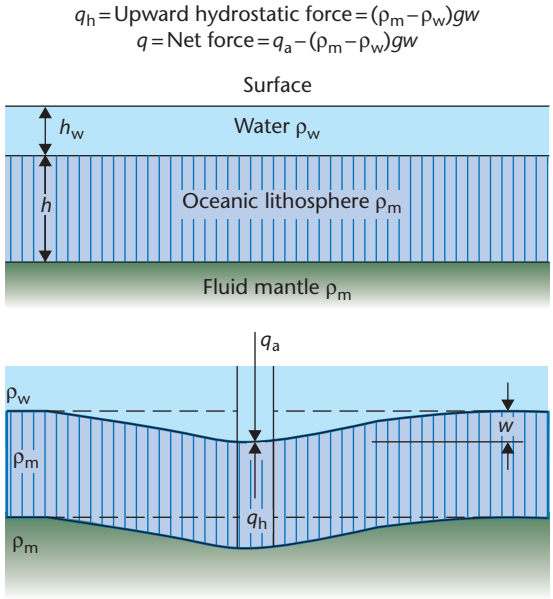

Fig. 2.11 Model for calculating the upward-acting hydrostatic restoring force q_h on an oceanic plate overlain by water, deflected by an applied load q_a. After Turcotte & Schubert (2002).

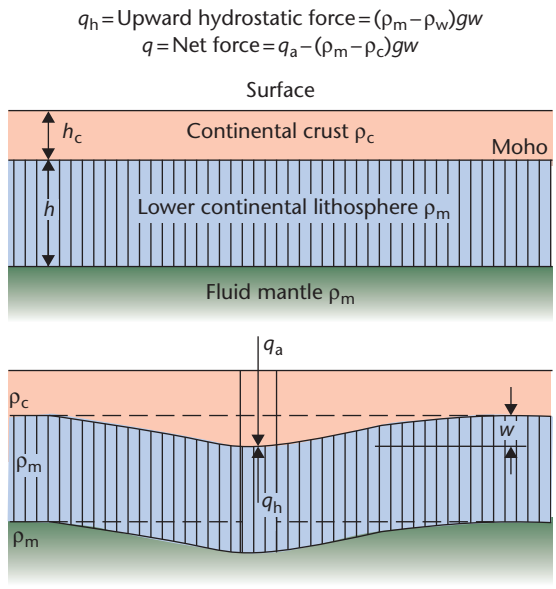

Fig. 2.12 Model for calculating the upward-acting hydrostatic restoring force q_h on the base of some continental crust where the deflection caused by the applied load q_a is assumed to be filled with material of the same density as the continental crust. This approximates the case of a fully compacted sediment-filled basin on continental lithosphere. After Turcotte & Schubert (2002).

When an applied load flexes a plate, the deflected region is filled either with water, as in the case of oceanic lithosphere or a starved continental basin, or with sediment, as in the case of most basins adjacent to hinterlands undergoing erosion. This infilling material has a lower density than the mantle that is being replaced (Figs 2.11, 2.12). The density difference can be denoted by $\Delta\rho$. The magnitude of the restoring force on the base of the deflected plate can be estimated by considering a balance of pressure (ρgh) under the region of maximum deflection and under the unaffected region. This upward restoring force is $\Delta\rho gw$, and the net vertical force acting on the plate is the applied load less this restoring hydrostatic force. The general flexural equation (Appendix 5) therefore becomes

$$D\frac{d^4w}{dx^4} + P\frac{d^2w}{dx^2} + \Delta\rho gw = q_a(x) \qquad [2.10]$$

where $\Delta\rho$ is $(\rho_m - \rho_w)$ for a purely water-filled basin (Fig. 2.11) and $\Delta\rho$ is $(\rho_m - \rho_s)$ for a fully sediment-filled basin (Fig. 2.12). The analytical solution of the general flexure equation in the context of a foreland basin is given in §4.2.

The lithosphere has a different flexural response according to the spatial distribution of the load $q_a(x)$. If the wavelength of a load of a certain mass, for example excess topography, is sufficiently short, the vertical deflection of the lithosphere is small, and the lithosphere can be regarded as infinitely rigid for loads of this scale. However, if the wavelength of a load of the same mass is sufficiently long, there is an effective isostatic response approaching hydrostatic equilibrium, and the lithosphere appears to have no rigidity. These two situations must be regarded as end members. The degree of compensation of the topographic load is the ratio of the deflection of the lithosphere to its maximum or hydrostatic deflection. The response of the lithosphere to an applied load, such as the excess mass in a mountain belt, is identical to its response to sediment loads in a basin. The degree of compensation for sediment loads of varying wavelength is dealt with in §2.1.5 and revisited in §9.3.2.

For reasonable values of plate thickness and flexural rigidity it is found that horizontal forces applied at the end of a plate are generally inadequate to cause buckling. Horizontal forces as buckling agents may, however, be much more important where the lithosphere has been strongly thinned in regions of high heat flow or where it is rheologically layered and decoupled mechanically (§4.4).

The general features of basins such as oceanic trenches and foreland basins adjacent to mountain belts can be explained by flexural models. Applications of the foregoing discussion of elasticity and flexure to sedimentary basin analysis are given in Chapter 4.

2.1.5 Flexural isostasy

The operation of flexure can be recognised in a number of settings in nature. The clearest examples are isostatic rebound during deglaciation, and the downflexing of moats adjacent to oceanic seamounts and island chains. Other important examples are the formation of foreland basins associated with mountain building, and the flexing of ocean plates ahead of subduction zones, whereas less obvious cases are the subsidence beneath sediment loads such as the Amazon fan. In all these cases the lithosphere supporting tectonic, volcanic and sedimentary loads is strong enough to store elastic stresses. Instead of the Airy isostatic model, in *flexural isostasy*, loads are compensated regionally by a lithosphere approximating an elastic sheet overlying a fluid substratum.

The problem of flexural isostasy can be tackled by considering the way the lithosphere supports periodic topography, as given in

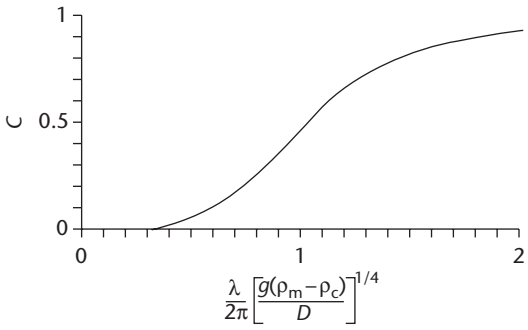

Fig. 2.13 Dependence of the degree of compensation C on the non-dimensional wavelength of periodic (sinusoidal) topography. D is the flexural rigidity, λ is the wavelength of the load, ρ_m and ρ_c are the mantle and crustal or infill densities, and g is the acceleration due to gravity. From Turcotte & Schubert (2002, p. 123, fig. 3.26), © Cambridge University Press, 2002.

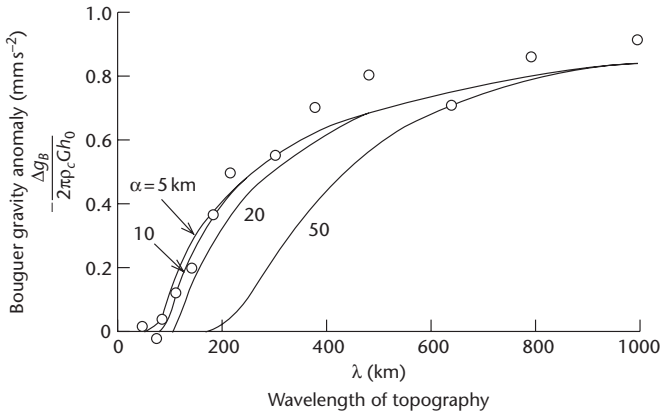

Fig. 2.14 Spectral technique for the estimation of flexural rigidity from the wavelength of the Bouguer gravity anomaly. Correlation of Bouguer gravity with topography (admittance) for the United States (Dorman & Lewis 1972) compared with the gravity formula for loading by periodic topography, for different values of the flexural parameter α. The best agreement is with $\alpha = 20$ km, or flexural rigidity $D = 10^{21}$ Nm, which with $E = 70$ GPa and $\nu = 0.25$, gives an equivalent elastic thickness Te of 6 km. After Turcotte & Schubert (2002).

Appendix 6. This simplified analysis can then be extended to the case of a sedimentary basin of a certain wavelength and amplitude.

The way in which the lithosphere supports loads flexurally, or the degree of compensation C, is related to the rigidity of the plate undergoing flexure and also the wavelength of the load. With a sinusoidal, periodic load, the degree of compensation is the magnitude of the maximum deflection compared to the maximum deflection expected from a plate with zero flexural rigidity (Airy isostasy), or expressed differently

$$C = \frac{(\rho_m - \rho_s)}{\rho_m - \rho_s + \frac{D}{g}\left(\frac{2\pi}{\lambda}\right)^4} \qquad [2.11]$$

where $(2\pi/\lambda)$ is termed the *wave number* (e.g. Watts 1988) (Fig. 2.13).

The same idea of compensation as a function of wavelength can be used to investigate the flexural support for the sediment loads in sedimentary basins (Chapter 9). The sediment load is assumed to have a characteristic wavelength. Taking the example of a sedimentary basin 200 km wide ($\lambda/2 = 200$ km) with a sinusoidal sediment load and an underlying lithosphere of flexural rigidity 10^{24} Nm, $(\rho_m - \rho_s) = 800$ kg m^{-3}, the degree of compensation C is about 0.12. This suggests that the lithosphere behaves very rigidly to this wavelength of load. Changing the wavelength of the load such that $\lambda/2$, the width of the basin, is now 400 km, $C = 0.68$, indicating that the sediment load is only weakly supported. In this case of large compensation, Airy-type isostasy is approached. It must be stressed, however, that rarely can C be estimated accurately in geological situations.

The very common negative Bouguer anomalies extending well beyond the topography of mountain belts suggest that they are flexurally compensated. In present-day settings, the extent of this compensation can be estimated from the correlation of Bouguer gravity anomalies with topography as a function of its wavelength, known as *admittance* (Fig. 2.14) (Dorman & Lewis 1972). Topography with a wavelength of less than 100 km is not compensated, whereas topography with a wavelength of greater than 1000 km is fully compensated. From Fig. 2.14, it can be seen that different curves are predicted for different flexural rigidities. The best fit for data from western United States is for a flexural rigidity of just 10^{21} Nm (equivalent elastic thickness of c.6 km).

A related spectral technique is the conversion of both the topography and Bouguer gravity field by a fast Fourier transform (Forsyth 1985). The *coherence* function is the square of the correlation coefficient between the topographic and gravity signals. A coherence of 1 indicates complete compensation on a plate with no flexural strength, and a coherence of 0 indicates that the load is completely supported by the strength of the plate. On a plot of coherence versus wavenumber ($k = 2\pi/\lambda$), the rollover from a coherence of 1 to 0 gives a measure of the flexural rigidity or equivalent elastic thickness. Lowry and Smith (1994) mapped flexural rigidity variations across the Basin and Range/Colorado Plateau/Rocky Mountain area using coherence analysis. They found that areas of low flexural rigidity (average of 9×10^{21} Nm, Te = 10 km) correlated with areas of high surface heat flows in the Basin and Range. The highest flexural rigidities were in the Rocky Mountains (average of 3×10^{23} Nm, Te = 33 km), with areas of Archaean cratons reaching elastic thicknesses of 77 km.

The good correspondence of flexural rigidity with heat flow indicates that the flexural rigidity of the continental lithosphere is related to the thickness of the elastic layer overlying a temperature-dependent ductile layer. This elastic layer thickness also contains almost all earthquakes on the continents (Maggi *et al.* 2000; Jackson *et al.* 2008) (Fig. 1.10), so there is a close correspondence between the seismogenic crust, the thickness of the strong elastic layer and the effective elastic thickness (§4.3). A global map of elastic thickness based on coherence is found in Audet and Bürgmann (2011) (Chapter 4).

2.1.6 Effects of temperature and pressure on rock density

Since rocks are compressible (they can change in volume), they change in density due to changes in both temperature and pressure.

Owing to their dependence on pressure, rock densities are commonly expressed as a zero-pressure value. For example, the rock type comprising the mantle is peridotite, which has a zero-pressure density of 3250 kg m^{-3}, whereas the gabbros and basalts that comprise the oceanic crust have zero-pressure values of 2950 kg m^{-3}, and the granites making up most of the continental crust have zero-pressure values of about 2650 kg m^{-3} (Table 2.2).

To work out the increase in density with increasing depth within the Earth (Appendix 1), we assume an isotropic state of stress, in which case the pressure is related to the dilatation Δ and the reciprocal of the compressibility β (known as the bulk modulus K), as given in eqn. [2.9]. Since the density change $\delta\rho$ of a solid element with volume V is related to the fractional change of volume δV, which is equal to the dilatation Δ, and since there is no change in mass, the isothermal change in density, in terms of the pressure, the initial density and the compressibility, is given by eqn. [2.8].

The density of rock increases with depth below the surface of the Earth due to isothermal compression. For example, at 1 GPa, equivalent to a depth of 38.5 km in granite, the density of granite increases from its zero-pressure value of 2650 to 2730 kg m^{-3}. Assuming the mantle lithosphere to be peridotitic in composition, peridotite increases in density from 3282 kg m^{-3} at 1 GPa (under 38.5 km of granite) to 3477 kg m^{-3} at 7 GPa, equivalent to a depth of 205 km, close to the base of the thermal lithosphere in continental interiors.

We can incorporate the effects of thermal expansion on rock density. When there is no change in pressure (isobaric), the density of a rock depends on its density at a reference temperature such as 0 °C, denoted ρ^*, the volumetric coefficient of thermal expansion α_v – the fractional change in volume with temperature *at a constant pressure* – and the temperature relative to the reference temperature ΔT

$$\rho = \rho^* (1 - \alpha_v \Delta T) \qquad [2.12]$$

Taking the volumetric coefficient of thermal expansion as 2.4×10^{-5} K^{-1} for granite, a geotherm (§2.2) using a basal heat flow of 30 mW m^{-2}, a radiogenic heat production of 2.5×10^{-6} W m^{-3}, an exponential depth constant of 10 km, a thermal conductivity of 3.0 W m^{-1} °C^{-1}, a surface temperature of 0 °C, and a density at a reference temperature of 0 °C of 2650 kg m^{-3}, the density at a new temperature of 430 °C, which is approximately the Moho temperature (depth of 35 km), is 2620 kg m^{-3}. Comparing the density decrease caused by isobaric thermal expansion of –30 kg m^{-3} and the density increase caused by the isothermal compression of +80 kg m^{-3}, it can be seen that the compression effect dominates in the granitic crust.

Turning to peridotite, and using the same volumetric thermal expansion coefficient of 2.4×10^{-5} K^{-1} and the same geotherm, and a density at 0 °C of 3250 kg m^{-3}, the density at a depth of 35 km is 3216 kg m^{-3}, and the density at a depth of about 200 km is 3088 kg m^{-3}. The density decrease caused by the isobaric thermal expansion between the Moho and the base of the thermal lithosphere is therefore –128 kg m^{-3}. This can be compared with the density increase caused by the isothermal compression of +195 kg m^{-3}. The density increase caused by the increasing pressure with depth therefore also dominates in the peridotitic mantle lithosphere (Fig. 2.15).

The relative effects on rock density of changes in temperature and pressure can also be assessed from their contribution to the state of stress. For the situation of zero horizontal strain, and where the rock behaves in a linear elastic fashion, the horizontal thermal stress relative to the horizontal lithostatic stress is given by a dimensionless number (Wangen 2010, p. 61) N_T, given by

$$N_T = \frac{E\alpha G}{3\nu\rho_b g} \qquad [2.13]$$

where G is the linear geothermal gradient ($\Delta T = Gy$), E is Young's modulus, ν is Poisson's ratio, and ρ_b is the bulk density of the rock

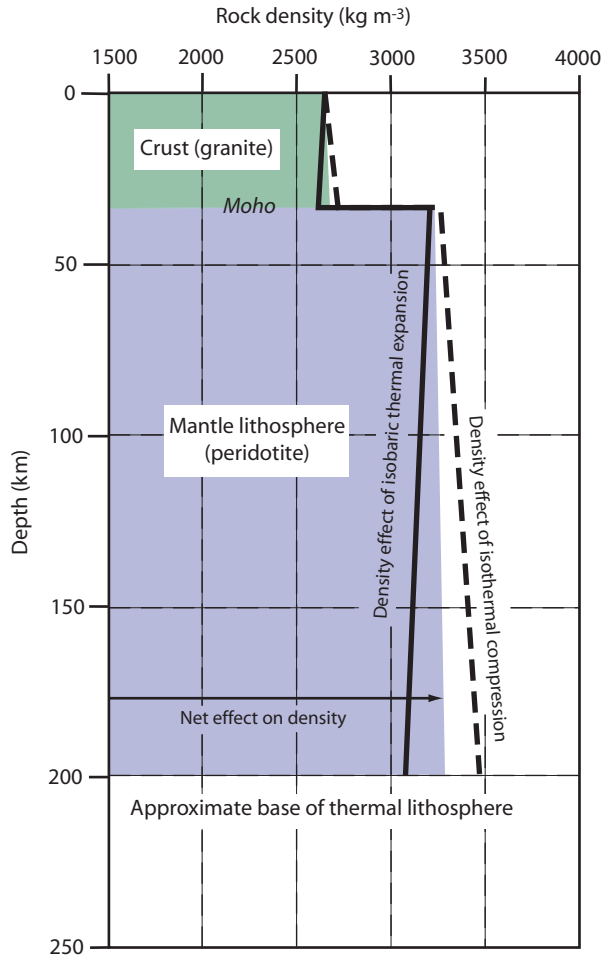

Fig. 2.15 Effects of temperature and pressure on the density of rocks with depth in the Earth. Densities are calculated for a geothermal gradient determined by a basal heat flow of 30 mW m^{-2}, a radiogenic heat production in the crust of 2.5×10^{-6} W m^{-3}, a depth constant for radiogenic heat production a_r of 10 km, thermal conductivity of 3 W m^{-1} K^{-1}, compressibility of 3×10^{-11} and 1×10^{-11} Pa^{-1} for granite and peridotite respectively, volumetric coefficient of thermal expansion of 2.4×10^{-5} K^{-1} for both granite and peridotite, Poisson's ratio of 0.225 and 0.25 for crustal granite and mantle peridotite respectively, and zero-pressure densities of 2650 and 3250 kg m^{-3} for granite and peridotite respectively. The trade-off between the thermal expansion due to increasing temperature with depth versus the increased density due to increased pressure with depth can be gauged by a dimensionless number N_T. For the most plausible range of parameter values, the effect of compression exceeds thermal effects.

column above depth y. For the case of a 30 km-thick granitic crust overlying a peridotitic mantle, a linear geothermal gradient of 20 °C km^{-1}, $E = 60$ GPa, and $\nu = 0.25$, N_T is approximately 0.6, showing that density changes due to thermal expansion and contraction are a significant fraction of the effects of pressure. When $G > 30$ °C km^{-1}, the thermal and pressure effects are approximately equal.

Taking into account the dual effects on density of increasing temperature with depth and increasing pressure with depth, it is clear that where the continental lithosphere is thick, the deep subcrustal part is prone to gravitational collapse into the hot asthenosphere since it is denser. Melt segregation from the thick continental subcrustal lithosphere under continental interiors (cratons) is probably required to reduce the density enough to prevent delamination of continental roots (O'Hara 1975; Jordan 1979). Melting selectively removes garnet, which is heavy, and reduces the iron content of the residual mantle. Changes in density of the residual mantle rock due to melting depends on the percentage of melt removed (Preistley & McKenzie 2006), 20% melt removal, for example, reducing mantle density from 3390 to 3340 kg m^{-3}. This has a significant effect on buoyancy and stability of the mantle lithosphere of continental cores.

2.2 Heat flow

2.2.1 Fundamentals

To understand the mechanical behaviour of the Earth it is necessary to know something of its thermal structure, since rock rheologies commonly depend on temperature, itself a function of depth. The temperature distribution of the Earth must reflect the inputs and outputs of heat to the Earth system. In other words, there is a heat transfer or flow, achieved by processes of conduction, convection and radiation. The essential differences between these processes are as follows. Conduction is a diffusive process whereby kinetic energy is transferred by intermolecular collisions. Convection, on the other hand, requires motion of the medium to transmit heat. Electromagnetic radiation, such as that of the Sun, can also transmit heat, but it is of relatively minor importance in the Earth's *internal* heat budget. It is, however, of paramount importance in determining Earth's *surface* heat budget.

The processes of conduction and convection are of differing importance in different zones. In the lithosphere, heat is transported primarily through conduction, whereas in the mantle, convection of heat from the Earth's deep interior is dominant. Convection is a much faster process of heat transfer than conduction.

The fundamental relation for conductive heat transport is given by *Fourier's law*. It states that the heat flux Q is directly proportional to the temperature gradient, and takes the mathematical form

$$Q = -K\frac{dT}{dy} \qquad [2.14]$$

where K is the coefficient of thermal conductivity, T is the temperature at a given point in the medium, and y is the coordinate in the direction of the temperature variation, generally vertical for heat flow through the lithosphere to the Earth's surface.

The heat flux at the Earth's surface gives a good indication of processes within the interior. Temperature measurements can be made on land in caves, mines, and, better still, in deep boreholes, and allow the heat flux to be calculated as long as the thermal conductivity K is known. K can be measured in the laboratory on rock samples by subjecting them to a known heat flux and measuring the temperature drop across the sample (Beardsmore and Cull 2001) (Table 2.1). Temperature measurements of the ocean floor can also be made by penetrating seafloor sediments with a temperature probe. The same probe contains a heater that enables the *in situ* thermal conductivity to be calculated. The wide range of values of thermal conductivity for sedimentary rocks is due, to a large extent, to large variations in porosity (see also §2.2.6). Heat flow is in units of mW m^{-2} or cal cm^{-2} s^{-1}. Surface heat fluxes are sometimes expressed in heat flow units where 1 HFU is equivalent to 10^{-6} cal cm^{-2} s^{-1} or 41.84 mW m^{-2}. Thermal conductivity is measured in units of watts per metre per degree centigrade (W m^{-1} °C^{-1}), and occasionally in cal cm^{-1} °C^{-1}.

Earth's average surface heat flow (87 mW m^{-2}) corresponds roughly to one household light bulb (100 W) over an area of a tennis court. However, heat flow measurements of the oceans and continents reveal important variations (Table 2.2, Sclater *et al.* 1980a, 1981, §1.4.4). Regions of high heat flow on the continents generally correspond to active volcanic areas, such as the Andes, to those areas underlain by highly radiogenic crust, or to regions of extensional tectonics, such as the Basin and Range province of western USA. Continental collision zones typically have low to normal surface heat flows. In areas devoid of active tectonics and volcanicity, the heat flow appears to be inversely correlated to the age of the rocks (Jessop & Lewis 1978) (Table 2.3). This can be explained by the decreasing abundance with age of the radioactive heat-producing isotopes of uranium, thorium and potassium (§2.2.3). Surface heat flows are strongly influenced by the underlying rock type. Granites produce large amounts of radiogenic heat, whereas basalts and peridotites produce almost no radiogenic heat (§2.2.3) (Table 2.4). In the oceans the surface heat flows are related not to the concentration of radioisotopes but to the age of the seafloor. Newly created oceanic crust cools by conduction as it travels away from the mid-ocean ridge, thereby explaining this relationship (§2.2.7). Mean oceanic surface heat flows (101 mW m^{-2}) are higher than their continental counterparts (65 mW m^{-2}) (Table 2.3). Approximately 60% of the heat loss of the Earth takes place through the ocean floor (Parsons 1982).

Since radioactive isotopes decay to stable daughter products, there must be a steadily decreasing heat production with time from radioactive decay, known as *secular cooling*. The rate at which heat is being transferred to the Earth's surface is therefore also decreasing with time, in turn slowing down the mantle convection system. Analysis of the abundance of the heat-producing radioisotopes and their stable daughter products suggests that heat production was twice the present value 3000 million years ago (Table 2.4, Fig. 2.16), with ^{238}U and ^{232}Th taking over from ^{235}U and ^{40}K as the main heat producers because of the latters' relatively short half-lives.

Heat travels down temperature gradients (eqn. [2.14]). The lithosphere is subjected to changes in temperature with time, particularly associated with tectonic processes. Consequently, it is important to be able to develop a rule for the heat loss or gain of an element of lithosphere (Appendix 7). Once such a rule is established, it can be modified to account for three-dimensional variations in heat flow.

2.2.2 The geotherm

The variation of temperature with depth is called the *geotherm*. In Appendices 7 and 8, we solve the one-dimensional conduction

Table 2.3 Global mean surface heat flow. Data from Pollack *et al*. (1993) and Stein & Stein (1992) in brackets. Continental and oceanic surface heat flows depend on surface geology and age as well as asthenospheric temperatures

	Mean surface heat flow	
	mW m^{-2} and standard error	**Heat flow units (HFU)**
Continents:	65	1.55
Cenozoic (sedimentary/metamorphic) (0–65 Ma)	63.9 ± 0.9	1.53
Cenozoic (igneous) (0–65 Ma)	97.0 ± 5.6	2.32
Mesozoic (sedimentary/metamorphic) (65–251 Ma)	63.7 ± 1.3	1.52
Mesozoic (igneous) (65–251 Ma)	64.2 ± 3.0	1.53
Paleozoic (sedimentary/metamorphic) (251–544 Ma)	61.0 ± 1.2	1.46
Paleozoic (igneous) (251–544 Ma)	57.7 ± 2.6	1.38
Proterozoic (544–2500 Ma)	58.3 ± 1.4	1.39
Archaean (2000–3800 Ma)	51.5 ± 2.4	1.23
Oceans:	101	2.41
Cenozoic (0–65 Ma)	89.3 (125.2) ± 2.8	2.13
Mesozoic (65–251 Ma)	44.6 (51.0) ± 2.8	1.07
Worldwide	87	2.08

Table 2.4 Typical concentrations of radioactive elements from typical rock types comprising continental and oceanic crust and undepleted mantle (from Fowler 1990). ^{238}U and ^{235}U have half-lives of 4.47×10^9 yr and 7.04×10^8 yr, ^{232}Th has a half-life of 1.40×10^{10} yr, and ^{40}K has a half-life of 1.25×10^9 yr. Because of their shorter half-lives, ^{40}K and ^{235}U were the most important heat producers in the distant past, whereas today heat is produced mostly by ^{238}U and ^{232}Th

	Granite	**Tholeiitic basalt**	**Alkali basalt**	**Depleted peridotite (harzburgite)**	**Average continental upper crust**	**Average oceanic crust**	**Undepleted 'fertile' mantle**
Concentration by weight							
U (ppm)	4	0.1	0.8	0.006	1.6	0.9	0.02
Th (ppm)	15	0.4	2.5	0.04	5.8	2.7	0.10
K (%) (10^{-19} W kg^{-1}	3.5	0.2	1.2	0.01	2.0	0.4	0.02
Heat generation (10^{-18} W kg^{-1}							
U	3.9	0.1	0.8	0.006	1.6	0.9	0.02
Th	4.1	0.1	0.7	0.010	1.6	0.7	0.03
K	1.3	0.1	0.4	0.004	0.7	0.1	0.007
Total	9.3	0.3	1.9	0.020	3.9	1.7	0.057
Density (kg m^{-3})	2.7	2.8	2.7	3.2	2.7	2.9	3.2
Heat generation (10^{-6} W m^{-3})	2.5	0.08	0.5	0.006	1.0	0.5	0.02

equation to derive the geotherm. It is commonly said that the mean geothermal gradient in the continental lithosphere is 30 °C km^{-1}, based on the gradient calculated in deep boreholes and coal mines. But extrapolation of this rate downwards from the surface would result in temperatures of 900 °C at a Moho depth of 30 km, and of 3750 °C at the base of a 125 km-thick lithosphere. Comparison with the solidus for peridotite demonstrates that this cannot be the case, since downward extrapolation implies that the mantle lithosphere would undergo wholesale melting, yet it is able to transmit **S**-waves.

If we continue to assume that the heat flow is by conduction in one direction from a hot mantle to a cold Earth surface, with a value of 30 mW m^{-2}, through lithosphere with a thermal conductivity of 3 W m^{-1} K^{-1}, giving a mean geothermal gradient of 10 K km^{-1}, the geotherm would closely approach the solidus for peridotite at the base of the lithosphere (Fig. 2.17). However, this geothermal gradient is considerably lower than observations suggest in the shallow crust and in sedimentary basins.

We conclude that the conduction geotherm given by eqn. [2.14] does not apply throughout the continental lithosphere. The solution to this problem is that the continental lithosphere generates its own heat by radiogenic decay. This self-heating causes the geotherm to be curved, with a higher geothermal gradient at shallow depths and a lower geothermal gradient in the rest of the continental lithosphere. The framework for the understanding of radiogenic heat production is given in §2.2.3.

The radiogenic heat production cannot be assumed to be uniform with depth since the abundances of the heat-producing radiogenic isotopes vary strongly according to rock type (Table 2.4). The heat generation of the crust, being 'granitic' in composition, far exceeds that of the mantle. Despite the fact that the very large volume of the mantle means that it contributes 80% of the Earth's radiogenic heat, it is the crustal contribution that is of importance in basin analysis and that determines the continental geotherm in the lithosphere. The question is how the radiogenic heat generation in the continental crust is modelled (Appendix 9) (§2.2.3).

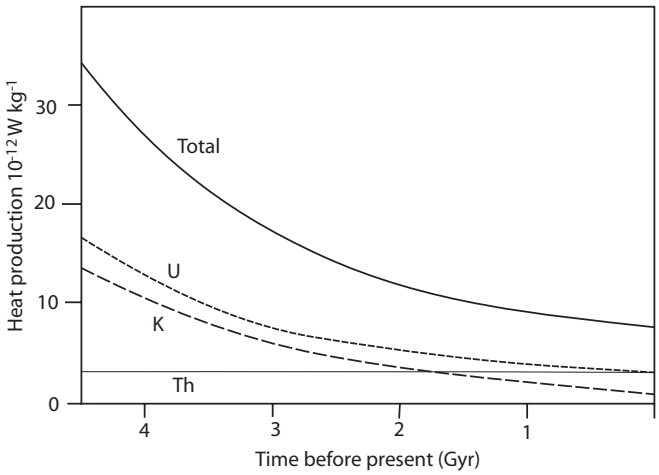

Fig. 2.16 Rate of heat production in W kg^{-1} due to the decay of radiogenic isotopes of U, K and Th, and the total heat production from all isotopes. Uranium concentrations in the mantle are 31×10^{-9} kg kg^{-1}, but heat production is 9.81×10^{-5} W kg^{-1}: Thorium has a higher concentration of 124×10^{-9} kg kg^{-1}, but a lower heat production of 2.64×10^{-5} W kg^{-1}: K is highly concentrated at 31×10^{-5} kg kg^{-1}, but has a low heat production rate of 3.48×10^{-9} W kg^{-1}. The secular reduction of heat production over time is due to the various half-lives of the isotopic systems, which vary from c.10^9 to 10^{10} yrs. From Turcotte & Schubert (2002). © Cambridge University Press, 2002.

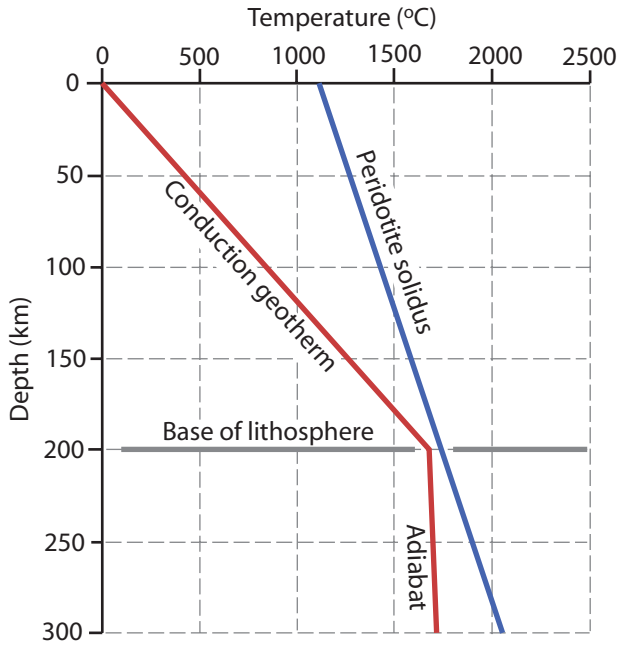

Fig. 2.17 The geotherm for the lithosphere and the solidus for peridotite. The geotherm approaches the solidus at the base of the lithosphere, where partial melting may take place. The geothermal gradient is calculated using 1D conduction with a thermal conductivity of 3.0 W m^{-1} K^{-1}, a basal heat flow of 30 mW m^{-2}, and a surface temperature of 0 °C. Pressure is calculated using a granitic 40 km-thick crust with density 2750 kg m^{-3} and a mantle lithosphere extending to a depth of 200 km with density 3300 kg m^{-3}. The geotherm is adiabatic in the asthenosphere and taken as 0.5 °C km^{-1}. The solidus for peridotite is given by $T = 1500$ (K) + $0.12p$ where pressure (p) is in MPa. Note that the geotherm approaches the solidus near the lithosphere–asthenosphere boundary, leading to partial melting.

In some cases, it is possible that heat is conducted in more than one direction, as may be common in regions of large lateral variations in surface temperatures. Such lateral variations might also arise where the lithosphere is stretched over a relatively narrow zone, as is common in strike-slip basins (Chapter 6), so that there is both upward and lateral loss of heat by conduction.

In the following sections, we make some general observations about radiogenic self-heating of the crust and its effect on the continental geotherm (§2.2.3) before discussing the effect of erosion and sedimentation on the geotherm (§2.2.4 & §2.2.5), and the effect of variable thermal conductivity on the geotherm of the sedimentary basin-fill (§2.2.6). The thermal history of the basin-fill is dealt with in Chapter 10.

2.2.3 Radiogenic heat production

Radiogenic heating is caused by the spontaneous decay of the elements uranium (^{235}U and ^{238}U), thorium (^{232}Th) and potassium (^{40}K). Partial melting of mantle rock, for instance at a mid-ocean ridge, depletes the residuum in incompatible elements such as U, Th and K, which are concentrated in the basaltic melt (Table 2.4). Consequently, ocean crust is concentrated in radiogenic elements relative to 'fertile' (undepleted) mantle, and even more concentrated relative to depleted residual mantle rock. Processes causing growth of continental crust further concentrates the incompatible elements, so concentrations of radiogenic elements are an order of magnitude higher in continental crust than in oceanic tholeiitic basalts. It therefore follows that the radiogenic self-heating of the lithosphere may be highly variable.

In areas devoid of active tectonics and volcanicity, continental heat flow values appear to be strongly correlated with the type of underlying crust. In NE North America, for example, the higher heat flows over the Appalachians (average 58 mW m^{-2}) than over the North American Shield (average 29 mW m^{-2}) may be explained by the different thicknesses of underlying tonalitic crust (Pinet *et al.* 1991) containing the radiogenic HPEs uranium, thorium and potassium (§2.2). In general, granitic terranes have high surface heat flows, whereas basic and ultrabasic igneous rocks and many sedimentary rocks are associated with low surface heat flows.

The contribution of radiogenic heating to the continental geotherm is generally approached in two ways (Fig. 2.18): (i) the assumption of a slab or small number of slabs of uniform radiogenic heat production (in W kg^{-1} or W m^{-3}) (solutions in Appendix 9); and (ii) the assumption that there is an exponential reduction of the radiogenic heat production A from a surface value A_0 at a rate given by the depth constant a_r:

$$A = A_0 \exp(-y/a_r) \qquad [2.15]$$

At a depth of a_r km, A has reduced to 1/e of its surface value A_0.

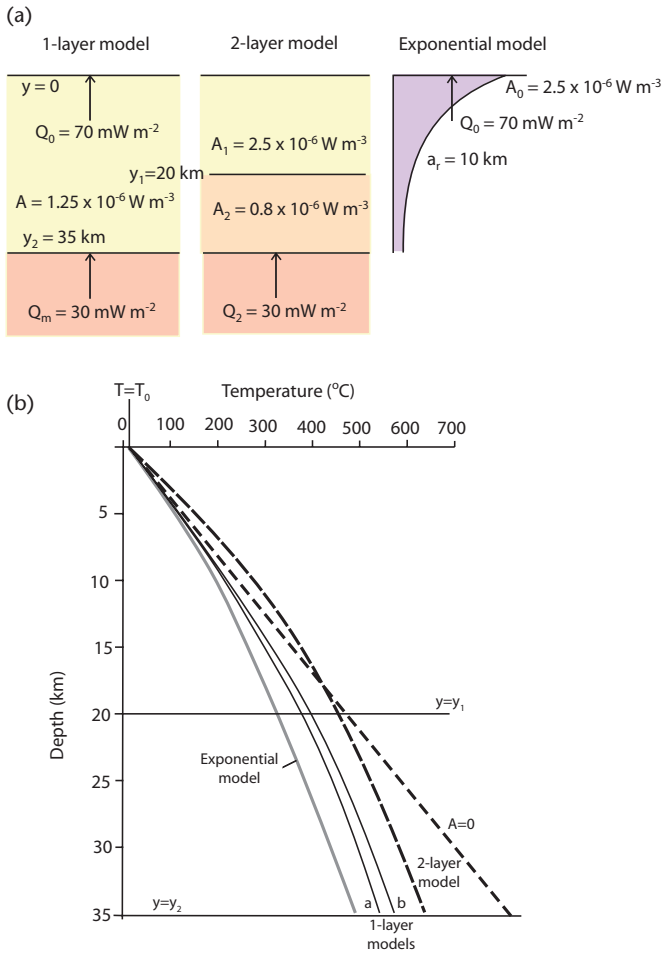

Fig. 2.18 Geotherms for the continental crust. (a) Model set-up: one-layer model with a constant internal heat generation for a 35 km-thick crust with (i) a surface heat flow of 70 mW m^{-2}, and (ii) a basal heat flow of 30 mW m^{-2}; a two-layer model with a highly radiogenic 20 km-thick upper crust and a weakly radiogenic 15 km-thick lower crust; and a model with an exponentially decreasing radiogenic heat production with depth. (b) Resultant geotherms, with a linear geotherm for zero radiogenic heat production ($A = 0$) and surface heat flow of 70 mW m^{-2} shown for comparison.

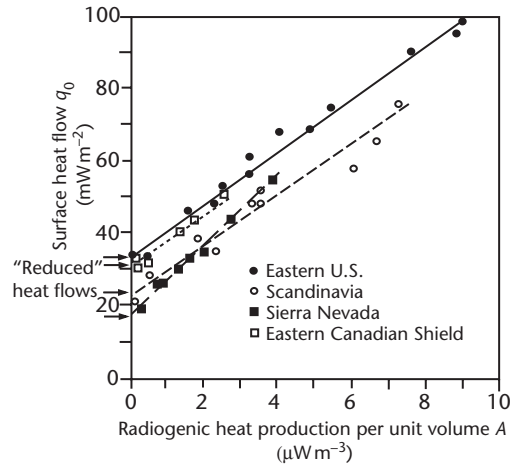

Fig. 2.19 Surface heat flow versus radiogenic heat production for a number of different geological provinces, showing a linear regression. The intercept on the *y*-axis gives a reduced, basal or mantle heat flow. See text for a discussion of the interpretation of these plots. Data from Roy *et al*. (1968).

The heat flow q at any depth y is given by:

$$q = -q_m - A_0 a_r \exp(-y/a_r) \qquad [2.16]$$

where q_m is the basal heat flow from the mantle in W m^{-2}, and A_0 is the radiogenic heat production in W m^{-3}. This can be expressed in terms of the surface heat flow q_0

$$q_0 = q_m + a_r A_0 \qquad [2.17]$$

From eqn. [2.17] it is clear that when the surface heat flow q_0 is plotted against the radiogenic heat production A_0, the linear regression has a slope of a_r and an intercept of q_m. It is therefore possible to obtain an estimate of the basal heat flow and the differentiation (that is, distribution with depth) of heat production from such plots (Fig. 2.19).

However, the interpretation of q_0–A_0 plots is not straightforward (Fig. 2.20). First, the distribution of radiogenic heat production may not be exponential with depth. Second, secular cooling causes a reduction in A_0, which causes a flattening of the regression and makes estimation of q_m problematical. Third, lateral conductive heating of a 'cold' terrane by an adjacent 'hot' terrane reduces the variation in the suite of A_0 values and reduces the slope of the linear regression. For these reasons, some care should be exercised in estimating a_r and q_m. A large number of q_0–A_0 plots, however, indicate that the basal heat flow is commonly less than half of the surface heat flow. Since the global average for the continents is 65 mW m^{-2}, the average basal heat flow from the mantle must be approximately 30 mW m^{-2}.

An alternative method for estimating the vertical distribution of HPEs in the upper part of the lithosphere, which is particularly valuable where the depth distribution of HPEs is variable, is to calculate a depth distribution parameter a that integrates over depth the product of radiogenic heat production $A(y)$ and the depth y (Fig. 2.20) (Sandiford & McLaren 2002). The variable a therefore differentiates between shallow (small a) and deep (large a) sources of radiogenic heat production. The effects of crustal growth, tectonic deformation and surface processes of erosion and sedimentation can then be viewed in a-q_c space, where q_c is the crustal radiogenic heat production (Fig. 2.21). q_c is simply the integration of the radiogenic heat production over the crustal thickness y_c.

The upward emplacement of granitic melts causes no significant change in the total crustal heat production, but a reduction in the depth distribution, since radiogenic material is moved to shallower depths (Fig. 2.21a). Tectonic deformation may cause thinning or thickening of the radiogenic layer. In the former, extension causes a reduction in the total crustal heat production and in most cases an upward movement of the HPEs, causing a reduction in the depth distribution parameter. On the other hand, shortening causes an

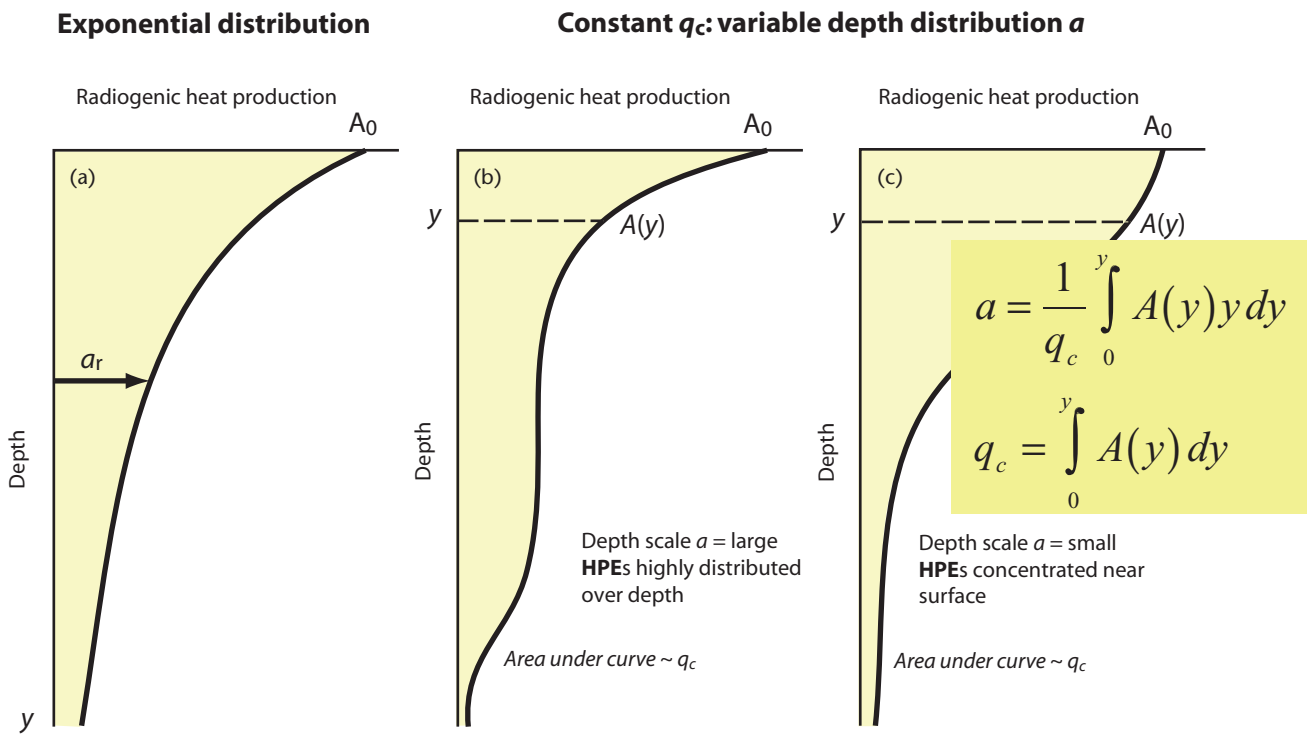

Fig. 2.20 The depth distribution of heat-producing elements (HPEs) can be approximated by an exponential function with a depth constant a_r and a surface heat production A_0, or can be denoted by a variable depth distribution parameter a and crustal heat production q_c. The depth distribution parameter a incorporates the effects of deep versus shallow sources of HPEs, where the value of q_c is identical to the area under the curves in the three profiles shown.

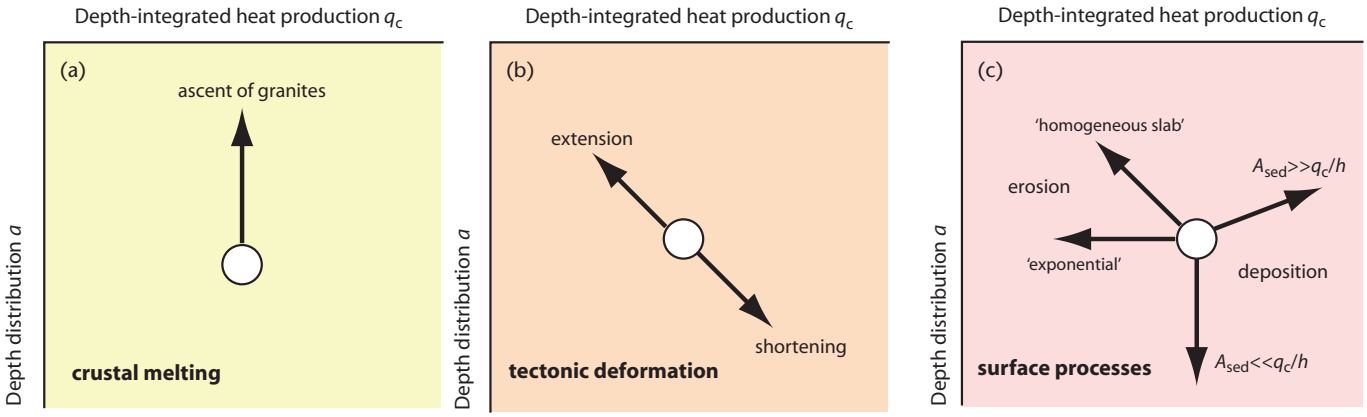

Fig. 2.21 The effects of crustal growth, surface processes and tectonic deformation can be recognised in the q_c-a space (after Sandiford & McLaren 2002). The depth distribution of HPEs a and the depth-integrated heat production q_c are shown in Fig. 2.20. Upward migration of radiogenic melts raises the depth distribution of HPEs while keeping the total heat production the same (a). Tectonic extension and shortening affects both the depth distribution and the total heat production, as shown in (b). Surface processes of erosion and deposition have impacts depending on the magnitude and distribution of HPEs in the crust (c), but in general erosion reduces a and q_c, whereas deposition increases a and q_c. Reprinted with permission from Elsevier.

increase in the total crustal heat production and a deepening of the distribution of HPEs, leading to an increase in the depth distribution parameter (Fig. 2.21b). Surface processes of erosion and deposition may have varied effects (Fig. 2.21c). Erosion reduces the thickness of the radiogenic layer, causing a drop in the value of q_c, accompanied by a reduction in the depth distribution a. Deposition has the opposite effect, but the precise response in q_c-a space depends on whether the heat production of the sediment is large or small in relation to the ratio q_c/a.

2.2.4 Effect of erosion and sediment blanketing on the geotherm

Since radiogenic elements are concentrated in the upper part of the lithosphere, the crustal radiogenic heat production is likely to be strongly affected by the processes of erosion and sedimentation. Deep erosion may potentially strip highly radiogenic crust from a piece of continental lithosphere, thereby reducing its radiogenic heat production, whereas sedimentation may deposit highly radiogenic sediment on top of a normal section of radiogenic crust.

We can initially consider this as a steady-state problem by comparing the geotherm under eroded crust versus under a sediment blanket (see also Appendix 13). Transient effects are considered in §2.2.5 and Appendix 12. The original thickness of the radiogenic crust is y_c. We treat the eroded crust as comprising a slab of radiogenic material that has been thinned erosively by h_e. In this case, the temperature at a depth $0 < y < y_c$ is given by:

$$T = T_0 + \frac{\{q_m + A(y_c - h_e)\}}{K} y - \frac{A}{2K} y^2 \qquad [2.18]$$

where T_0 is the surface temperature, q_m is the basal heat flow, K is the thermal conductivity, and A is the uniform radiogenic heat production in the slab.

The eroded sediment is deposited as a uniform layer of thickness h_b above normal crust comprising an initial radiogenic slab of thickness y_c, giving a total thickness of $(h_b + y_c)$ for the slab. We neglect, for simplicity, the effect of sediment porosity, and assume that the thermal conductivity of the basin-fill is the same as the thermal conductivity of the adjacent and underlying crust (see §2.2.6 for the impact of thermal conductivity variation). The geotherm through and under the basin is

$$T = T_0 + \frac{\{q_m + A(y_c + h_b)\}}{K} y - \frac{A}{2K} y^2 \qquad [2.19]$$

The difference in temperature at any depth between the eroded crust and the basin is linearly dependent on the depth y and the amount of erosion and deposition. It is given by

$$\Delta T = \frac{A}{K} y (h_b + h_e) \qquad [2.20]$$

where h_b is the thickness of the basin-fill and h_e is the depth of erosion. If K is $3\,W\,m^{-1}\,°C^{-1}$, A is $2.5 \times 10^{-6}\,W\,m^{-3}$, and $(h_b + h_e)$ is 10 km, the temperature difference at the base of the sedimentary basin is about 40 °C and at a depth of $y = 10$ km it is about 80 °C (Fig. 2.22).

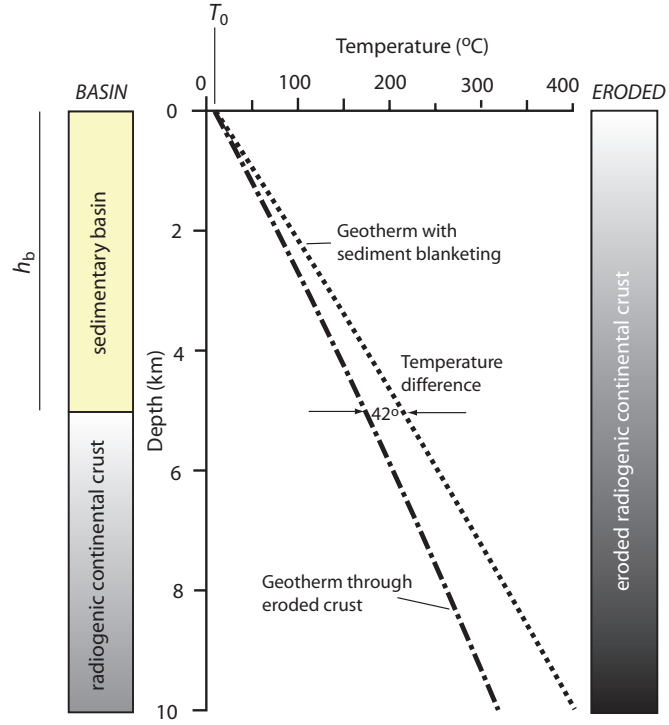

Fig. 2.22 Steady-state crustal geotherms resulting from erosion and sedimentation, using a basal heat flow of $30\,mW\,m^{-2}$ and a radiogenic heat production with $A_0 = 2.5 \times 10^{-6}\,W\,m^{-3}$ and a thermal conductivity of $3\,W\,m^{-1}\,K^{-1}$. The surface temperature T_0 is 10 °C. The steady-state geotherm following erosion by 5 km (long-short dash) shows the effect of stripping off 5 km of highly radiogenic upper crust. The steady-state geotherm resulting from sediment blanketing (short dashes) has elevated temperatures caused by the presence of a homogeneous 5 km-thick basin filled with radiogenic sediment. The temperature difference between the geotherms for the eroded crust and the blanketed crust is 42 °C at the base of the sedimentary basin. Upper crustal rocks under the sedimentary basin are therefore likely to undergo deformation or reactivation of old tectonic structures as a result of the thermal effects of surface processes (see also Sandiford & McLaren 2002).

The effect of erosion and sedimentation can be made more realistic by invoking a two-layer slab model (Appendices 9 and 12) where each slab has its own radiogenic heat production and thermal conductivity, representing the porous basin-fill and the underlying crust. For reasonable parameter values, it can be seen that, at steady state, temperatures are strongly elevated in and under the sedimentary basin, with its radiogenic infill and normal thickness upper crustal radiogenic slab compared to the adjacent eroded crust. The radiogenic structure of the crust and basin-fill are therefore critical to the prediction of the continental geotherm. The deposition of a radiogenic, self-heating sedimentary infill has an important impact on basin-fill and crustal temperatures, and thereby influences the thermal maturation of lithologies within the basin and potentially the deformation of upper crustal rocks (Hand & Sandiford 1999).

2.2.5 Transient effects of erosion and deposition on the continental geotherm

The preceding discussion treats the geotherm as a steady-state problem. However, the geotherm may be affected by time-dependent processes, such as an abrupt change in erosion and sedimentation rate at the surface of the Earth. We assume that the initial geothermal gradient is linear, and that there are no additional radiogenic heat sources. We firstly consider the transient effects on subsurface temperatures following an instantaneous change in erosion rate or sedimentation rate (following Wangen 2010, p. 181–186) (Appendix 12).

The instantaneous removal of rock causes warm crust to be subjected to a cold surface temperature T_0. Over time, the warm crust cools by conduction to the surface of the Earth. The time scale of this cooling depends on the diffusivity of crustal rocks. The depth to which the temperature perturbation is felt is the diffusion distance of the form $\sqrt{\kappa t}$ (see §2.2.7). A similar situation arises with instantaneous deposition, but there is an instantaneous cooling followed by a transient heating with the same time scale.

The temperature following erosion by an amount l (Fig. 2.23) is the sum of the steady-state solution (first two terms on right-hand side) and a transient solution (third term on right-hand side)

$$T(y,t) = T_0 + G(y+l) + Glerfc\left(\frac{y}{2\sqrt{\kappa t}}\right) \qquad [2.21]$$

and, similarly, the temperature following sediment blanketing by a thickness h (Fig. 2.24) is

$$T(y,t) = T_0 + G(y-h) + Gherfc\left(\frac{y}{2\sqrt{\kappa t}}\right) \qquad [2.22]$$

where T_0 is the surface temperature and G is the geothermal gradient, y is zero at the new surface, and *erfc* is the complementary error function. Figs 2.23 and 2.24 show that after ~1 Myr, the geotherm is close to its original linear gradient before the erosion event and before the blanketing event.

Erosion and sedimentation, however, may be rapid, but not instantaneous. Even at fast erosion rates of 1 mm yr^{-1}, it should take 1 million years to erode 1000 m of upper crust. Consequently, it is valuable to consider what effect the speed of erosion and deposition have on the underlying geotherm. To do so requires the Earth's surface to be treated as a *moving boundary*.

When rocks are exhumed rapidly, they move towards the surface of the Earth by advection while at the same time cooling by conduction. During high rates of exhumation due to erosion, hot rocks are brought closer to the surface of the Earth but are unable to cool quickly enough to reach the linear steady-state geotherm due to conduction. That is, the temperature field is determined by two effects: the conductive cooling at a rate given by the thermal conductivity and temperature gradient, and advection of rock towards the surface at the exhumation rate. When the competing rate effects of conduction (or diffusion) and advection are involved in a problem, it is customary to use a *Péclet number*. We assume that the rate of exhumation E is steady over a sufficiently long period of time so that a stable thermal state exists, governed by the balance of advection and conduction (Willett & Brandon 2002).

We firstly non-dimensionalise the vertical depth coordinate and the temperature coordinate, so that $y^* = y/L$ and $T^* = T/L$, where L is the depth to a lower boundary, which might, for instance, be the depth to the base of the crust. Secondly, we make use of the Péclet number, which measures the efficiency of advective versus conductive heat flow:

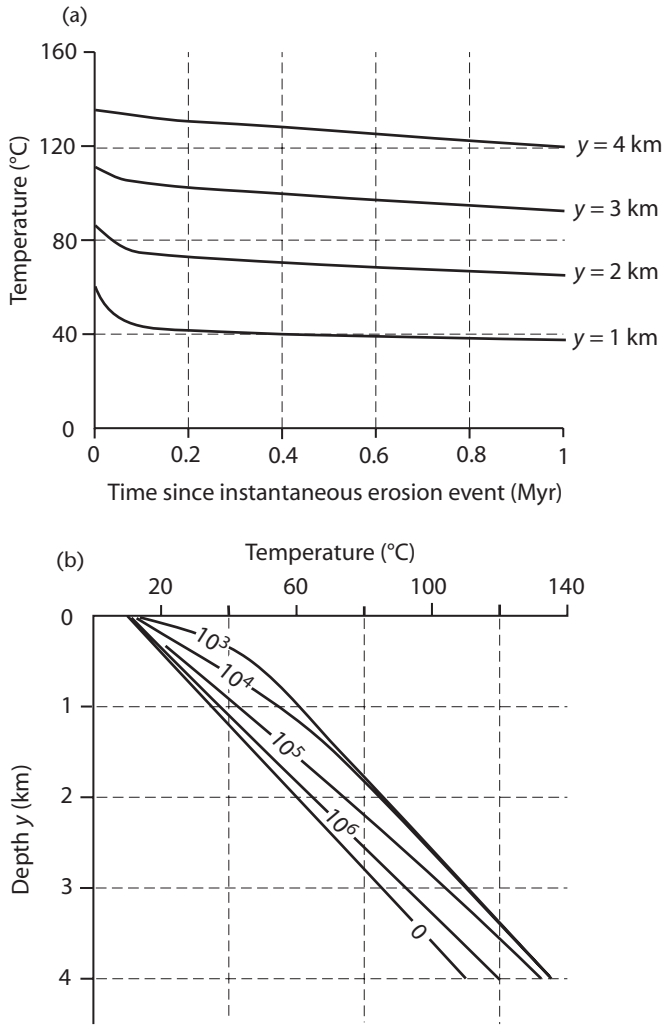

Fig. 2.23 Effect of instantaneous erosion. A thickness of l = 1 km is removed instantaneously by erosion of crust with a surface temperature T_0 of 10 °C, geothermal gradient 25 °C per km, and thermal diffusivity κ of 10^{-6} m^2 s^{-1}. (a) Temperature trajectories for different depths below the surface after erosion, showing a decreasing rate of cooling by conduction following the erosion event. (b) Geotherms as a function of time following erosion, and the linear geotherm prior to erosion (t = 0). After c.10^6 yr, the geotherm has almost returned to its steady-state gradient.

$$Pe = \frac{EL}{\kappa} \qquad [2.23]$$

and the diffusivity $\kappa = K/\rho c$, where K is the thermal conductivity (W m^{-1} K^{-1}), ρ is the rock density (kg m^{-3}), and c is the specific heat (W kg^{-1} K^{-1}). The dimensionless form of the temperature is given by Braun *et al.* (2006, p. 79)

$$T^*(y^*) = \frac{1-\exp(-Pey^*)}{1-\exp(-Pe)} \qquad [2.24]$$

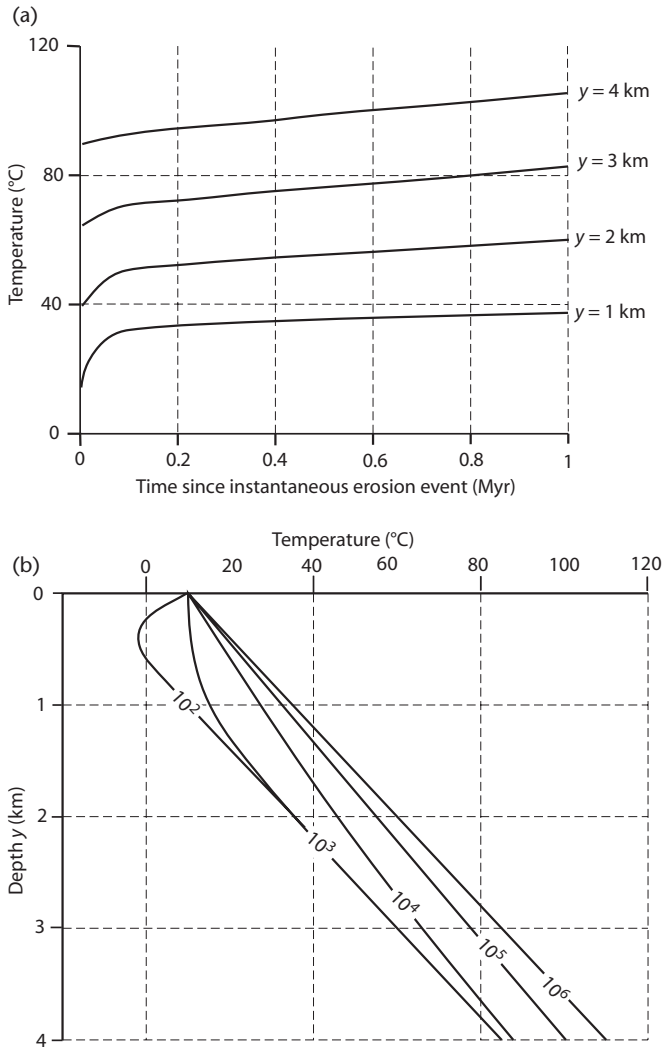

Fig. 2.24 The effect of instantaneous deposition on the crustal geotherm, with a steady-state geotherm G of 25 °C per km, surface temperature T_0 of 10 °C, thermal diffusivity of 10^{-6} m^2 s^{-1}, and depositional thickness h of 1 km. (a) Temperature trajectories for different depths, showing an instantaneous cooling followed by recovery. (b) The geotherm following instantaneous cooling due to sedimentation at times of 10^2 to 10^6 years. It takes in the order of 10^6 years to regain the linear geotherm with gradient G.

and is shown in Fig. 2.25. It can be seen that at Péclet numbers of greater than 1 the geotherm is significantly curved. It is therefore important to know what Péclet numbers are likely to occur in nature. Taking $\rho = 2800\,\mathrm{kg\,m^{-3}}$, $c \sim 1000\,\mathrm{W\,kg^{-1}\,K^{-1}}$, and $K = 2.5\,\mathrm{W\,m^{-1}\,K^{-1}}$, the diffusivity of crustal rocks is approximately $\kappa = 10^{-6}\,\mathrm{m^2\,s^{-1}}$. Using eqn. [2.23], the Péclet number for the crust is >1 for exhumation rates of greater than $1\,\mathrm{mm\,yr^{-1}}$. Rates of this order are common in tectonically active settings. The effect on the geotherm of advection results in a surface heat flow 'anomaly' compared to regions undergoing conduction only. At $E = 2\,\mathrm{mm\,yr^{-1}}$, the surface heat flow is approximately 2.5 times the surface heat flow in an adjacent region where there is no advection ($Pe = 0$). This has profound consequences for the interpretation of surface heat flow data, and for the correct interpretation of thermochronometric data (see §10.4). If these differences in surface heat flow could be confidently attributed to differential rates of exhumation, the surface heat flow 'anomaly' would allow the curvature of the geotherm under the exhuming region to be estimated.

The transient form of the geotherm following a rapid change in the exhumation rate, as may be caused by the initiation of a major tectonic event, or following a rapid change in the sedimentation velocity, as may occur through internally-forced delta switching or externally forced increases in sediment discharge caused by catchment processes, is solved in Appendix 12.

2.2.6 Effect of variable thermal conductivity

Fourier's law (eqn. [2.14]) states that the heat flow is related to the temperature gradient by a coefficient known as the thermal conductivity. Whereas the thermal conductivity is not expected to vary greatly in the crystalline crust, it is highly variable in the sedimentary basin-fill. This variability derives from the wide range of sedimentary lithologies in the basin, their pore-filling fluid, and their diagenetic phases (Giles 1997; Allen & Allen 2005) (see Chapters 9 and 10).

Ignoring, for the moment, lithological variations, thermal conductivities of sediments vary as a function of depth because of their porosity loss with burial (§9.2). Neglecting also radiogenic heat production within the basin-fill, the geotherm in a basin with different thermal conductivities of sedimentary layers is given by

$$T_y = T_0 + (-q)\left\{\frac{l_1}{K_1} + \frac{l_2}{K_2} + \frac{l_3}{K_3} + \ldots\right\} \qquad [2.25]$$

where l_1 to l_n are the thicknesses of the layers with thermal conductivities K_1 to K_n, and $l_1 + l_2 + l_3 \ldots$ must of course be equal to y. The heat flow q is shown negative since heat travels vertically towards the surface in the opposite direction to the depth coordinate y.

Typical thermal conductivities of the minerals comprising framework grains and diagenetic products, and the thermal conductivity of pore-filling water, are given in Table 2.2. Methods of calculating the bulk thermal conductivity of different stratigraphic units are outlined in §10.3.1. The bulk thermal conductivity of most sedimentary rocks ranges between $1.5\,\mathrm{Wm^{-1}K^{-1}}$ (shales) and $4.5\,\mathrm{Wm^{-1}K^{-1}}$ (sandstones). The individual conductivities of framework, matrix and pore fluid are also dependent on temperature, so thermal conductivity values (Table 2.2) are commonly quoted as surface temperature values (Brigaud & Vasseur 1989). The general trend is that non-argillaceous rocks have higher conductivities than argillaceous rocks, and that conductivity increases with increasing porosity.

A fundamental result of maintaining a constant heat flux through a heterogeneous basin-fill with variable thermal conductivity is that the geothermal gradient must vary with depth. A worked example is given in Appendix 13. Taking a basin composed of the stratigraphy given in Table 2.5, the geotherm is highly variable (Fig. 2.26). Comparison with the linear conduction profile using Fourier's law with $q_m = 30\,\mathrm{mW\,m^{-2}}$ and $K = 3.0\,\mathrm{W\,m^{-1}\,K^{-1}}$ emphasises the importance of this approach to the prediction of thermal history. Note the strong insulating effect of the marine shales and the highly conducting effect of quartzites and evaporites (anhydrite).

As we have seen in §2.2.3, the sediments of the basin-fill are also radiogenic. Variations are likely to exist in the internal heat production of the different sedimentary lithologies. We can assume that these different sedimentary layers are the same units that we used

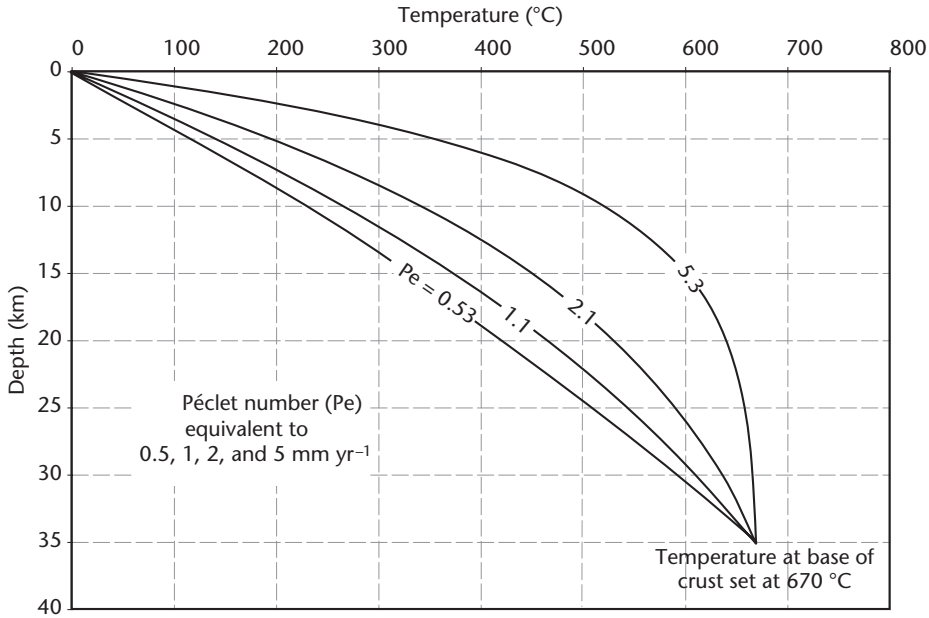

Fig. 2.25 Curves show geotherms for different values of the exhumation rate, shown as a Péclet number. At high exhumation rates, the upward advection of hot rock towards the surface outweighs the conductive cooling, causing highly curved geotherms. Note that in the case of high exhumation rate, the geothermal gradient changes from 40–60 °C km^{-1} in the upper 5 km of the crust to <10 °C km^{-1} in the lower crust. For apatite fission track thermochronometers, the high geothermal gradient at high Péclet numbers expected in rapidly exhuming orogens has the effect of reducing the true depth of erosion compared to estimates using 'normal' or 'standard' geotherms. For U–Th/He thermochronometers, the geothermal gradient used to convert cooling ages to depths of erosion is highly susceptible to the effects of erosion and variable surface topography (Braun *et al*. 2006).

Table 2.5 Bulk thermal conductivities of stratigraphic layers in a model basin-fill (surface temperature values from Brigaud & Vasseur 1989) and radiogenic heat production values

Lithology	Depth range (m)	Bulk thermal conductivity (W m^{-1} K^{-1})	Density (kg m^{-3})	Internal heat production A (× 10^{-6} W m^{-3})
Evaporite	0–500	6.0	2960	2.0
Dolomite	500–1500	3.2–3.5	2870	1.0
Shales	1500–2300	1.2–3.0	2100–2700	4.0
Quartzite	2300–3500	7.0	2650	0.5
Granite	>3500	2.4–3.8	2650	5.0

for calculations of bulk thermal conductivity. For a uniform radiogenic heat production extending to the base of the crust and a uniform thermal conductivity

$$T = T_0 + \left(\frac{q_m + Ay_c}{K}\right)y - \frac{A}{2K}y^2 \qquad [2.26]$$

must be modified to account for n number of layers of thermal conductivity K_1 to K_n, radiogenic heat production A_1 to A_n, and thickness l_1 to l_n

$$T_y = T_0 + \left\{\left(\frac{q_m + A_1y_c}{K_1}\right)l_1 + \left(\frac{q_m + A_2y_c}{K_2}\right)l_2 + \cdots\right\} - \left\{\frac{A_1}{2K_1}l_1^2 + \frac{A_2}{2K_2}l_2^2 + \cdots\right\} \qquad [2.27]$$

where y_c becomes the sum of the sedimentary layers of thickness $l_1 + l_2 + \ldots l_n$ if the geotherm is calculated for the basin-fill only. The solution for a heterogeneous basin-fill using eqn. [2.27] and the data in Table 2.5 gives the geotherm shown in Fig. 2.26, and demonstrates that the linear conduction geotherm may be a very poor approximation of the temperature field in parts of sedimentary basins.

2.2.7 Time-dependent heat conduction: the case of cooling oceanic lithosphere

Many problems involve heat flows that vary in time. An obvious example is the heat flow associated with the intrusion of an igneous body, but the example used here is the cooling of the oceanic lithosphere and its consequent subsidence. This process finds great application in sedimentary basins experiencing a period of cooling following stretching (Chapter 3). Heat sources *within* the medium are unimportant in the time-dependent problem of the cooling of oceanic lithosphere (that is, $A = 0$).

At the crest of an ocean ridge, hot mantle rock injected in dykes and extruded as lava flows is suddenly subjected to a cold surface

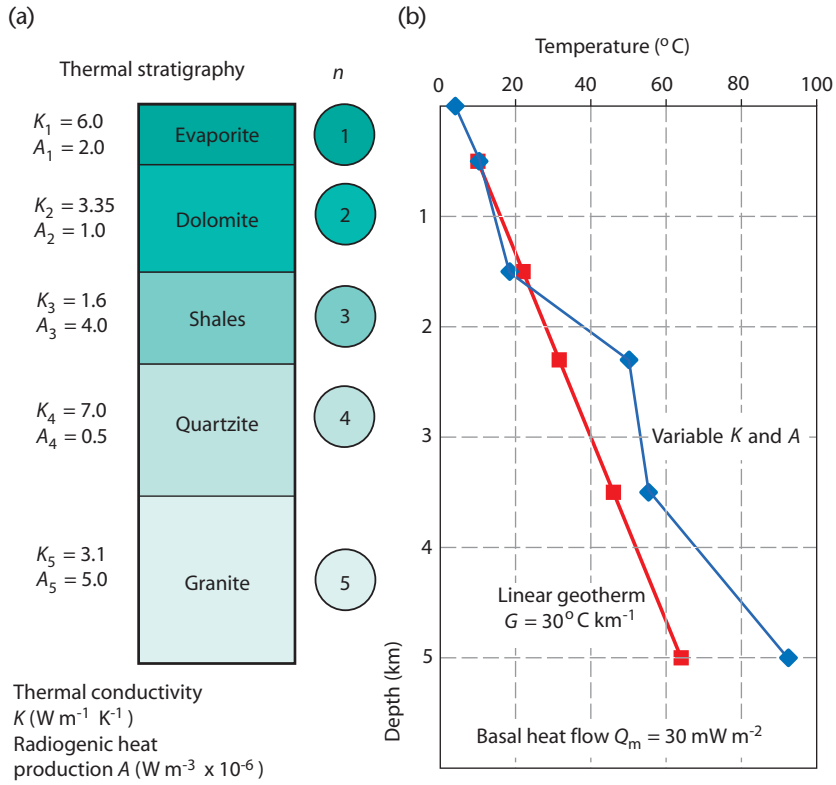

Fig. 2.26 Stratigraphy with variable thermal conductivity (a) used to calculate the geotherm in (b), illustrating the insulating and conducting properties of common rock types.

temperature and then continues to lose heat to the cold seawater as the seafloor spreads away from the ridge. The initial cooling can be treated as instantaneous. We know from the 1D conduction equation (Appendix 7) where $A = 0$ that

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial T^2}{\partial y^2} \qquad [2.28]$$

where the diffusivity $\kappa = K/\rho c$, and K is the thermal conductivity, ρ is the density, and c is the specific heat or thermal capacity. Eqn. [2.28] is a simple diffusion equation. We can define a characteristic time scale τ as the time necessary for a temperature change to propagate a distance l in a material with a thermal diffusivity κ (units of $[L]^2\ [T]^{-1}$):

$$\tau = \frac{l^2}{\kappa} \qquad [2.29]$$

The length scale $l = \sqrt{\kappa\tau}$ gives the distance that a temperature change propagates in time τ. It is known as the *thermal diffusion distance*.

Let the surface plates move away from the ridge along the coordinate x with a velocity u (Fig. 2.27). The cooling rocks form the oceanic lithosphere and the boundary between this relatively rigid upper layer and the easily deformed mantle is an isotherm with a value of about 1600 K (1300 °C). The thickness of the oceanic lithosphere is clearly a function of its age, where age can be expressed in terms of x/u.

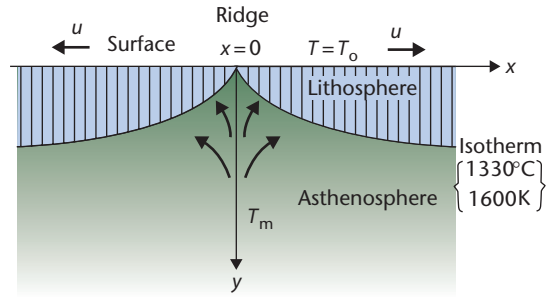

Fig. 2.27 Schematic diagram of the cooling oceanic lithosphere at a mid-ocean ridge. The oceanic plate moves away from the ridge at a velocity u. Its age is therefore determined by x/u, where x is the horizontal distance from the ridge crest. After Turcotte & Schubert (2002), © Cambridge University Press, 2002.

Using a time-dependent instantaneous cooling model, the isotherms below the seafloor as a function of age are parabolic (Fig. 2.28). The surface heat flow as a function of age is based on the 1D conduction equation. It is compared with actual heat flow measurements of the ocean floor in Fig. 2.29 (Sclater *et al.* 1980a). The presence of regions of poor correspondence between the predicted and

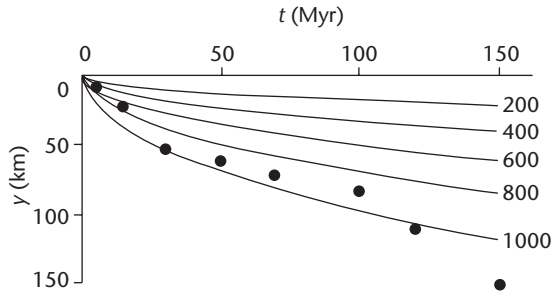

Fig. 2.28 Calculated isotherms for an oceanic lithosphere that is instantaneously cooled. The values of the isotherms are $T–T_s$°K. The dots are the estimated thicknesses of the oceanic lithosphere in the Pacific. From Leeds *et al*. (1974).

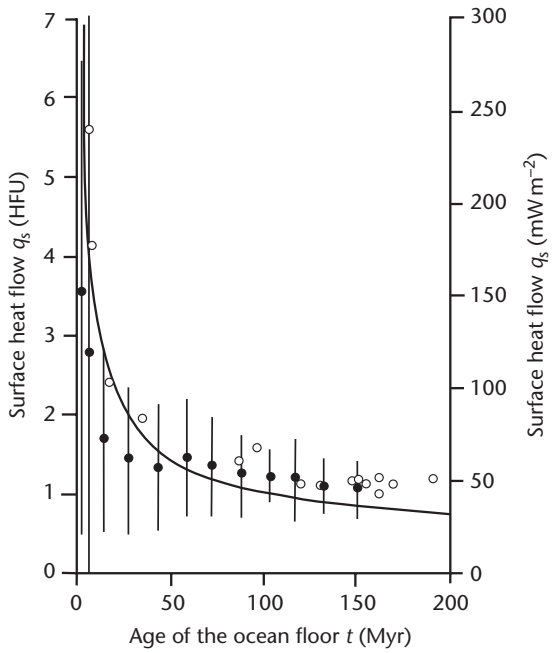

Fig. 2.29 Comparison of measured ocean heat flows (mean and standard deviation) and those predicted using the instantaneous cooling model, as a function of age. Black circles, data from Sclater *et al*. (1980a). Open circles, data from sediment-covered regions of the Atlantic and Pacific oceans, from Lister *et al*. (1990).

observed surface heat flow measurements is probably due to temperature changes associated with hydrothermal circulation of seawater through the oceanic crust. These effects become less important with age as impermeable sediments blanket the ocean floor. For large ages (>80 Ma), however, an additional heat source appears to be recognisable, which may be mantle convection beneath the lithospheric plates.

The solution for the temperature as a function of age has the form

$$\theta = erfc\left(\frac{x}{2\sqrt{\kappa t}}\right) \qquad [2.30]$$

where it can be seen that the denominator in the bracket is twice the thermal diffusion distance, *erfc* is the complementary error function, x is the horizontal distance from the mid-ocean ridge crest, t is the age of the oceanic crust, and θ is a dimensionless temperature ratio given by

$$\theta = \frac{T - T_a}{T_s - T_a} \qquad [2.31]$$

where T_a is the initial temperature (the asthenospheric temperature), T_s is the constant temperature of the space into which the ocean lithosphere is emplaced (the seawater temperature), and T is the temperature at time t.

The cooled oceanic material forms a thermal boundary layer equivalent to the thickness of the new oceanic lithosphere. The choice of the definition of the boundary layer is largely arbitrary, but if we define it as the thickness to where $\theta = 0.1$, it is found that the thermal boundary layer is 2.32 times the thermal diffusion distance $\sqrt{(\kappa t)}$ or $\sqrt{(\kappa x/u)}$. The thickness of the oceanic lithosphere at an age of 50 Myr is therefore approximately 92 km, taking $\kappa = 10^{-6}$ m^2 s^{-1}. The oceanic lithosphere should increase parabolically in thickness with age, and therefore also with distance from the ridge crest.

As it moves away from the ridge crest, the oceanic lithosphere cools and contracts. This increases the density of the oceanic lithosphere and causes a higher lithostatic stress on the underlying mantle. An isostatic balance shows that the seafloor must subside. Any column through the oceanic lithosphere can be balanced isostatically to give the depth of the ocean floor as a function of distance from the ridge crest, or age of the oceanic crust (Fig. 2.27).

The observed heat flows in the oceans, such as the Atlantic and Pacific oceans (Parsons & Sclater 1977), and the observed bathymetry, are in general agreement with a model of instantaneous cooling of new oceanic material and its loss of heat through time by conduction, resulting in subsidence. The general form of the relationship between the bathymetry of the ocean floor and its thermal age for oceanic lithosphere less than 70 Myr old is

$$h = h_{ridgecrest} + Ct^{1/2} \qquad [2.32]$$

where h is the depth of the ocean floor, $h_{ridgecrest}$ is commonly in the region of 2.5 km, t is the age of the oceanic lithosphere, and C is a coefficient approximately equal to 0.35. This is commonly termed a *root-age relationship*. If $h_{ridgecrest} = 2.5$ km, and $C = 0.35$, the ocean depth h at $t = 20$ Myr is 4 km, and at $t = 50$ Myr is 5 km. The heat flow Q (10^{-3} W m^{-2}) for oceanic lithosphere younger than 120 Ma follows a similar but *inverse root-age* relationship:

$$Q = 473t^{-1/2} \qquad [2.33]$$

Since the age of the oceanic lithosphere is directly related to its distance from the ridge crest by the spreading rate, the oceanic bathymetry increases gradually and the heat flow decreases gradually away from the site of spreading. The isostatic balance and the root-age relationship are affected by the presence of marine sediments overlying the ocean crust. However, in the deep sea the cover of pelagic sediments is generally only a relatively thin veneer.

2.2.8 Convection, the adiabat and mantle viscosity

Whereas the thermal structure of the lithosphere is dominated by conduction, that of the mantle is determined primarily by

convection. The lithosphere simply serves as a thermal boundary layer exhibiting high temperature gradients. Extensional basins, other than thin-skinned varieties, involve upwelling of convecting asthenosphere. We therefore present here a very brief account of the thermal structure of the upper mantle. Further information on the impact of mantle dynamics on the Earth's surface and basin development is given in Chapter 5.

In the interior of a vigorously convecting fluid, the mean temperature increases with depth along an *adiabat*. The *adiabatic temperature gradient* is the rate of temperature increase with depth caused by compression due to the overlying rock column. It is roughly linear in the mantle (Appendix 14). The compressional pressure forces cause a decrease in volume and therefore an increase in density. The relationship between the density and pressure changes is given by the *adiabatic compressibility*, β_a. For a solid, it is somewhat smaller than the isothermal compressibility β given in eqn. [2.9].

The purely adiabatic expressions for the variations of density and pressure with depth in the mantle do not perfectly match observed values based on seismic velocities. In particular, at about 400 km there is a density discontinuity thought to be due to the phase change of olivine. This change to the denser spinel structure is thought to occur at a pressure of 14 GPa or 140 kbar and a temperature of 1900 K. The phase change is *exothermic*, causing heating of the rock by c.160 K. The olivine–spinel phase change probably enhances mantle convection rather than blocks it. There is a second discontinuity in density at about 700 km, but its origin is less clear – it is probably also due to a change in mantle composition to perovskite. There is a sharp change in **P**-wave velocity at about 700 km. Close to the outer limit of the core is a seismic discontinuity known as D″ that is related to the change to a denser mineral structure known as post-perovskite (see §1.2.3). This change probably influences the formation of instabilities in the mantle at the core–mantle boundary, some of which may rise through the mantle as vertical conduits or plumes.

If a substance is *incompressible*, its volume is incapable of contracting, so adiabatic heating cannot take place. Rocks, however, are sufficiently compressible that adiabatic temperature changes are extremely important under the large pressure changes in the mantle. The adiabatic geotherms in the upper mantle underlying oceanic and continental lithospheres are different since continental lithosphere has its own near-surface radioactive heat source. Below this near-surface layer the heat flow is assumed to be constant at about 30 mW m^{-2}. It is clear from the shape of the geotherm that the continental lithosphere as a thermal entity extends to about 200 km below the surface (§1.2).

The *onset of convection* is a threshold above which the influence of the temperature difference in the fluid layer exceeds that of the viscous resistance to flow. The temperature difference is conventionally evaluated as the fluid temperature at a point as a departure from the temperature expected from a purely conduction profile. The onset of convection is marked by a critical dimensionless number known as the *Rayleigh number*, *Ra*. A Rayleigh number analysis of Earth's mantle suggests that it must be fully convecting (Appendix 40) (more details in Chapter 5).

There are also large *lateral* temperature heterogeneities in the mantle, such as where cold lithospheric plates are subducted at ocean trenches. These lateral temperature variations are extremely important in providing a driving force for mantle convection. In the case of a descending cold lithospheric slab, the low temperatures of the plate cause it to be denser than surrounding mantle, providing a gravitational body force tending to make the plate sink. A second factor is the distortion of the olivine–spinel phase change since the pressure at which this phase change occurs depends on temperature. The upward displacement of the phase change in the cold descending plate provides an additional downward-acting body force, helping to drive further plate motion. These two processes are often referred to as *trench-pull*. If trench-pull forces are transmitted to the oceanic plate as a tensional stress in an elastic lithosphere with a thickness of 50 km, the resultant tensional stress would be as high as 1 GPa. The forces arising through the elevation of the ridge crest relative to the ocean floor constitute a *ridge-push* force, also helping to drive plate motion. Ridge-push is an order of magnitude smaller than trench-pull, but since the latter is countered by enormous frictional resistances, the two agents of net trench-pull and ridge-push may be comparable.

One of the fundamental differences between fluids and solids is their response to an applied force. Fluids deform continuously under the action of an applied stress, whereas solids acquire a finite strain. Stress can be directly related to strain in a solid, but in fluids applied stresses are related to *rates* of strain, or, alternatively, velocity gradients. In Newtonian fluids there is a direct proportionality between applied stress and velocity gradient, the coefficient of proportionality being known as *viscosity*. (In non-Newtonian fluids there may be complex relationships between applied stresses and resultant deformations, see §2.3.) The problem of direct relevance to sedimentary basin analysis is the viscous flow in the mantle. One way that the viscosity of the mantle can be estimated is by studying its response to loading and unloading (see also Chapter 4 and Appendices 30, 31).

Mountain building is an example where the crust–mantle boundary is depressed through loading, but orogeny is so slow a process that the mantle manages to constantly maintain hydrostatic equilibrium with the changing near-surface events. In contrast, the growth and melting of ice sheets is very rapid, so that the mantle adjusts itself dynamically to the changing surface load; the way in which it does so provides important information on mantle viscosity. The displacement of the surface leads to horizontal pressure gradients in the mantle that in turn cause flow. In the case of positive loading, fluid is driven away from the higher pressures under the load, the reverse being true on unloading. The surface displacement decreases exponentially with time as fluid flows from regions of elevated topography to regions of depressed topography. If w is the displacement at any time and w_m is the initial displacement of the surface, the form of the exponential decrease in surface topography with time is

$$w = w_m \exp(-t/\tau_r) \qquad [2.34]$$

where τ_r is the characteristic time for the exponential relaxation of the initial displacement. It is given by

$$\tau_r = (4\pi\mu)(\rho g\lambda) \qquad [2.35]$$

where μ is the viscosity and λ is the wavelength of the initial displacement. Mantle viscosity can therefore be estimated from postglacial rebound if this relaxation time can be found (Appendix 30).

Elevated beach terraces that have been dated (usually by ^{14}C) provide the basis for quantitative estimates of the rate of postglacial rebound. Large lateral variations in rebound rates are found. The former Lake Agassiz in North America has a number of lake shoreline terraces that have been uplifted by flexural rebound accompanying deglaciation (Watts 2001) (Appendix 31). The Baltic region of Scandinavia is also renowned for its raised beaches. After correcting the

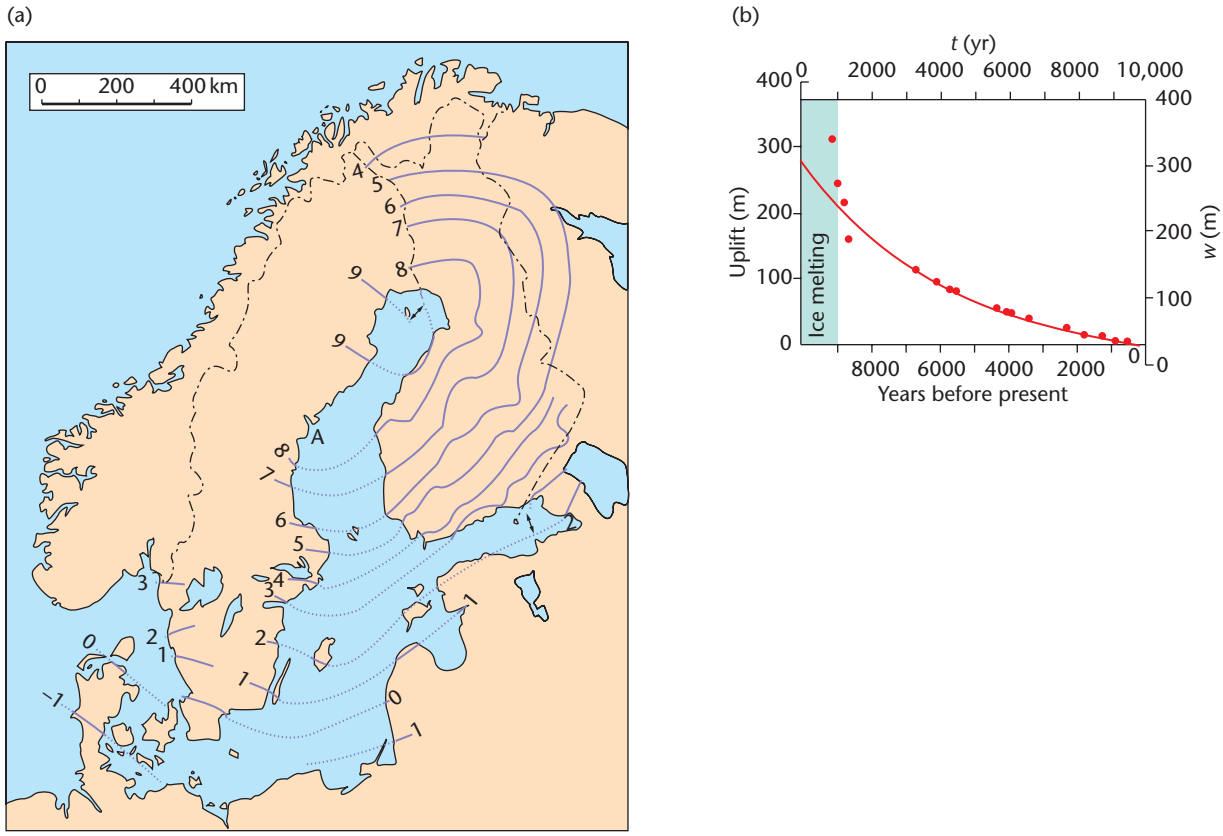

Fig. 2.30 Mantle viscosity can be estimated through studies of postglacial isostatic rebound. (a) Present-day rates of uplift occurring in Scandinavia (after Flint 1971; material reproduced with permission of John Wiley & Sons, Ltd.). Contours are in mm yr^{-1}. (b) Postglacial uplift of the mouth of the Angerman River, Sweden, over the last 10,000 years compared with the exponential relation in eqn. [2.38] with a constant viscosity of 10^{21} Pa s (after Turcotte & Schubert 2002; © Cambridge University Press, 2002). Angerman River mouth in the Gulf of Bothnia is marked by an 'A'.

uplift of Swedish beaches for absolute (eustatic) changes in sea level, a relaxation time of $\tau_r = 4400$ years is found. Assuming a reasonable wavelength of the displacement for the Scandinavian glaciation to be $\lambda = 3000$ km, the viscosity of the mantle is estimated to be $\mu = 1.1 \times 10^{21}$ Pa s (10^{22} poises) (Fig. 2.30). However, this approximate analytical solution does not take into account: (i) the flexural rigidity of the elastic lithosphere; and (ii) the depth-dependency of mantle viscosity.

2.3 Rock rheology and lithospheric strength profiles

2.3.1 Fundamentals on constitutive laws

Rocks in the convecting mantle, the mantle lithosphere and the crust are subjected to different temperatures, pressures, volatile contents and strain rates, which causes them to deform in different ways. For example, in the crust, joints and faults are evidence that rocks can behave as *brittle* materials, that is, they behave elastically up to a limit, beyond which they fail by fracturing. Alternatively, the widespread occurrence of folds suggests that rocks can also behave in a *ductile* manner. Ductile deformation in crustal rocks takes place by pressure-solution creep (Rutter 1976, 1983), whereby dissolution of minerals in zones of high pressure and their precipitation in areas of low pressure causes creep even at low temperatures and pressures. In the mantle, however, convection is thought to take place by a thermally activated creep that depends exponentially on both temperature and on pressure. When the effective viscosity is stress-dependent, it is analogous to the flow of a power-law fluid in a channel. Creep also takes place in the mantle lithosphere where it can relax elastic stresses. In this case, the rheology is a combination of an elastic and viscous behaviour – a *viscoelastic* rheology. Some of these rheologies (flow laws) are discussed in further detail with reference to the deformation of the mantle (§2.3.2) and the continental crust (§2.3.3).

A key concept in rock rheology is that there is a change in deformation behaviour under changing pressures. When confining pressures approach a rock's brittle strength a transition takes place from brittle (elastic) behaviour to plastic behaviour (Fig. 2.31a). This transition is at the rock's yield stress σ_0. The elastic strain can be regarded as recoverable, but the deformation in the plastic field is not recoverable when the stress or load is removed (Fig. 2.31b). If the deformation continues indefinitely without any addition of stress above σ_0, the material is said to exhibit an *elastic–perfectly plastic* behaviour (Fig. 2.31c). Laboratory studies of the mantle rock dunite show that it conforms closely to the elastic–perfectly plastic rheology (Griggs

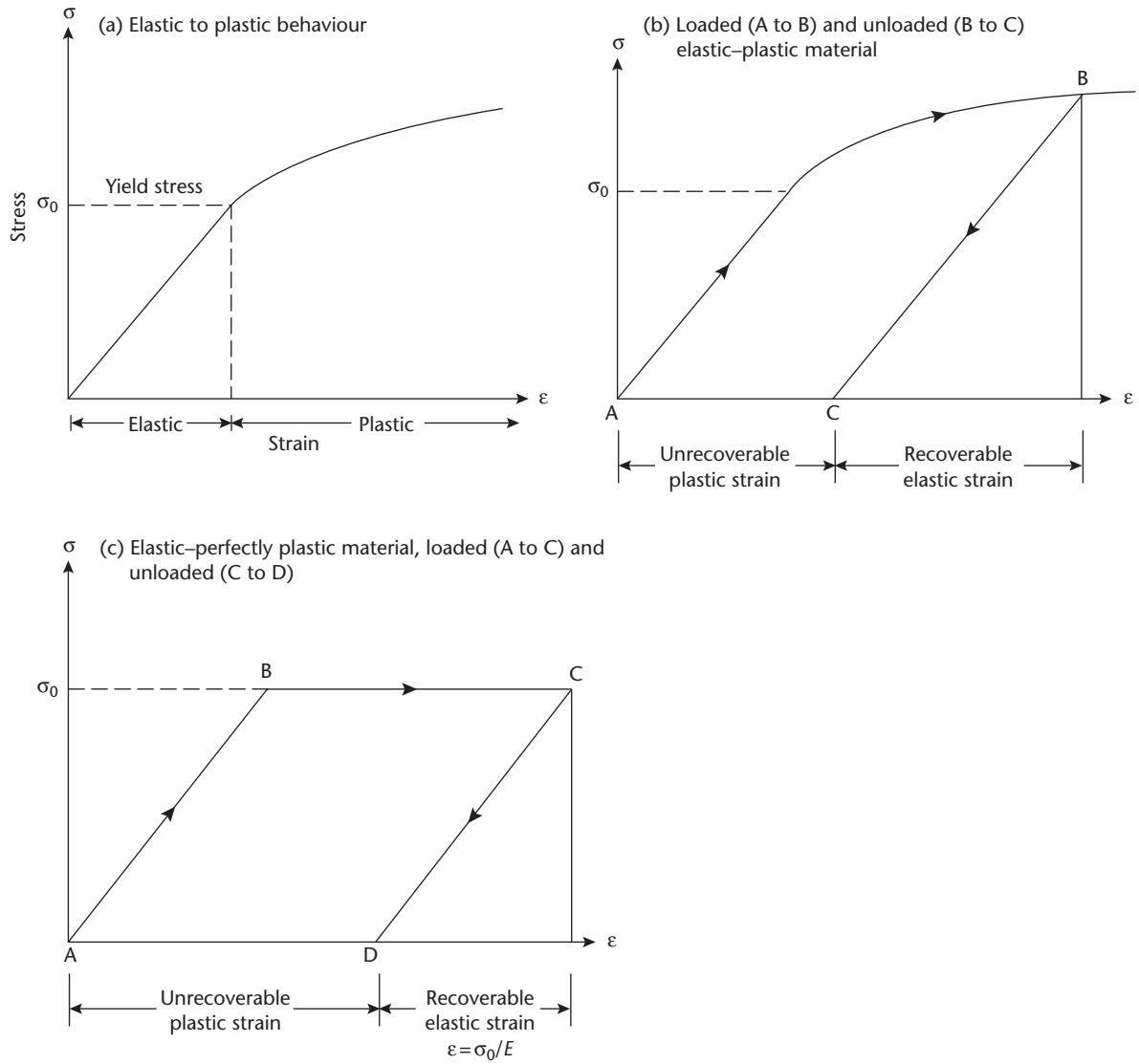

Fig. 2.31 Deformation of solids with different constitutive laws. Stress-strain trajectories for: (a) a solid showing a transformation from elastic to plastic behaviour; (b) loading and unloading of an elastic–plastic material. Unloading of the material once it has passed into the plastic field results in an unrecoverable deformation or plastic strain; and (c) a material with an elastic–perfectly plastic rheology. Plastic strain continues indefinitely without any addition of stress above the yield stress.

et al. 1960). A typical depth at which the brittle–plastic transition would occur for dunite is about 17 km; below this the rock should yield plastically under large deviatoric stresses. The importance of the brittle–ductile transition for the strength profile of the lithosphere is discussed in §2.3.4.

2.3.2 Rheology of the mantle

It was initially believed that at the high temperatures and low strain rates of the upper mantle, Newtonian (linear) flow would take place. Alternatively, the mantle may act as a power-law fluid, in which case the mantle's viscosity would be stress-dependent. For a power-law fluid the strain rate or velocity gradient is proportional to the power n of the stress (Fig. 2.32). Velocity gradients and therefore strain rates are greater near the walls of the channel where shear stress is a maximum, whereas a core region experiences small strain rates. This plug-flow appearance of the velocity profiles for large n is a direct consequence of the stress dependency of the effective viscosity μ_{eff}, which changes from low at the walls to high at the centre of the flow. This must be true because

$$\mu_{eff} = \tau \Big/ \frac{du}{dy} \tag{2.36}$$

that is, the viscosity μ is the coefficient linking the shearing stresses to the resultant velocity gradient or strain rate.

The idea can be applied to shear flow in the asthenosphere, which can be treated as having a heated lower boundary and an upper

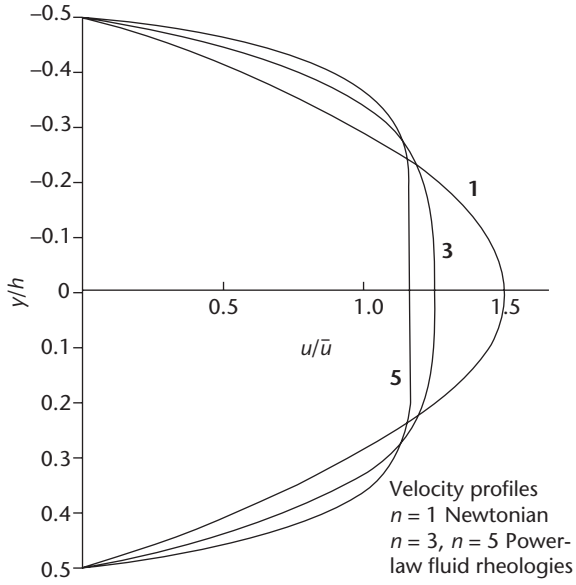

Fig. 2.32 Velocity profiles in a channel of thickness h for power-law fluid rheologies with $n = 1$ (Newtonian), $n = 3$ and $n = 5$. Distance from the channel wall is expressed by y/h. The velocity u is scaled by the average velocity $\bar{u}$. After Turcotte & Schubert (2002).

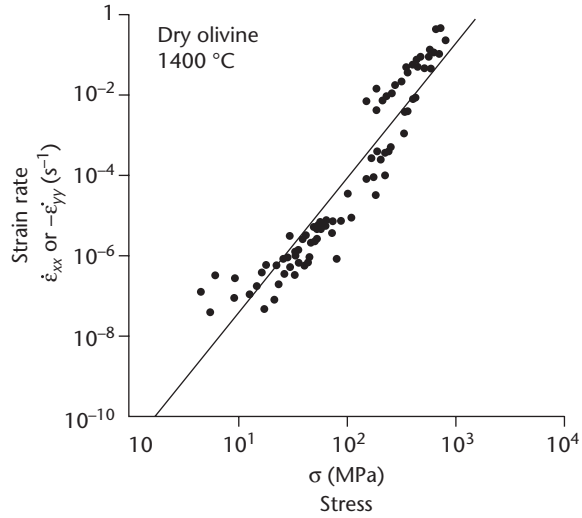

Fig. 2.33 Observed dependence of strain rate on stress for olivine at a temperature of 1400 °C (after Ashby & Verall 1977). Dry olivine obeys an approximate cubic power rheology in these laboratory experiments. After Turcotte & Schubert (2002), © Cambridge University Press, 2002.

cooler boundary at the base of the rigid lithosphere. Shear in the asthenospheric velocity profile is concentrated in zones close to the lower boundary of the asthenosphere where the fluid is hottest and viscosity the smallest. The upper part of the asthenosphere, on the other hand, tends to behave like a rigid extension of the overlying lithosphere as a result of this temperature and stress dependency of viscosity. Frictional heating can also have important consequences for shear flow of a fluid with a strongly temperature-dependent viscosity.

The fact that seismic shear waves can be propagated through the mantle suggests that it is an elastic solid, yet it also appears to flow like a viscous fluid, enabling, for example, postglacial rebound to take place. A material that behaves as an elastic solid on short time scales, but viscously on long time scales, is known as a viscoelastic or *Maxwell* material.

In a Maxwell material, the rate of strain $\dot{\varepsilon}$ is the sum of the linear elastic strain rate $\dot{\varepsilon}_e$ and a linear viscous strain rate $\dot{\varepsilon}_f$. The elastic strain of a material under a uniaxial stress is

$$\varepsilon_e = \sigma / E \qquad [2.37]$$

as will be recalled from Hooke's law (eqn. [2.12]), where E is Young's modulus. The rate of strain is therefore the time derivative of ε_e, $d\varepsilon_e/dt$. The rate of strain in a viscous Newtonian fluid subjected to a deviatoric normal stress σ is a velocity gradient $\delta u/\delta x$. The effective viscosity is taken as the ratio between stress and twice the strain rate. Therefore

$$\frac{d\varepsilon_f}{dt} = -\frac{\partial u}{\partial x} = \frac{\sigma}{2\mu} \qquad [2.38]$$

(where the minus sign simply indicates a tensional strain by convention). Since the total strain is the sum of the elastic and viscous fluid strains, the total strain rate is

$$\frac{d\varepsilon}{dt} = \frac{1}{2\mu}\sigma + \frac{1}{E}\frac{d\sigma}{dt} \qquad [2.39]$$

This is the fundamental rheological equation relating strain rate, stress and rate of change of stress for a *Maxwell viscoelastic* material. When a strain is initially rapidly applied to the viscoelastic medium, the time-derivative terms in eqn. [2.38] dominate and the material behaves elastically. Subsequently, if there is no change in the strain, the stress relaxes to 1/e of its original value in a time $2\mu/E$, which is known as the *viscoelastic relaxation time*. For the asthenosphere, this relaxation time is of the order of 30 to 40 years, that is, longer than the period of seismic waves but shorter than the duration of postglacial rebound. The stress relaxation time is a strong function of temperature, rheological parameters and initial stress, and its range of values for the lithosphere is controversial. Some workers believe that the relaxation time is very short ($<10^6$ yr) while others believe it to be sufficiently long (20–30 Myr) to have a major impact on sedimentary basin geometry and subsidence. A viscoelastic rheology with a long relaxation time was used in early models of foreland basin evolution in particular (Beaumont 1978, 1981; Quinlan & Beaumont 1984).

Most views of mantle rheology come from laboratory studies. Since the mantle is composed primarily of olivine, laboratory studies of the creep of olivine at high temperatures are particularly relevant to mantle rheology. The strain rate in dry olivine $\dot{\varepsilon}_{xx}$ (or $-\dot{\varepsilon}_{yy}$) at 1400 °C as a function of stress appears to follow a cubic power-law rheology reasonably well (Ashby & Verall 1977) (Fig. 2.33). Other rocks deform in a non-linear way at high temperatures in the laboratory, with slightly different coefficients in the power law (ice $n = 3$, halite 5.5, limestone 2.1, dry quartzite 6.5, wet quartzite 2.6, basalt 3). It is reasonable therefore to assume that the mantle has a power-law rheology. This power-law effect is likely to be small,

however, compared with the temperature dependence of mantle rheology.

2.3.3 Rheology of the continental crust

As every field geologist knows, crustal rocks exposed at the Earth's surface are strongly fractured and faulted at a range of scales. The occurrence of brittle failure depends on the stress overcoming the frictional resistance of the rock (Atkinson 1987 and Scholz 1990 for more details). This resistance to brittle failure increases with pressure and therefore depth.

It has been known for a considerable time that there is a roughly constant ratio between the frictional force on a potential failure plane F and the normal stress N. This ratio is equal to the coefficient of friction f, or the tangent of the angle of sliding friction ϕ

$$\frac{F}{N} = f = \tan\phi \qquad [2.40]$$

Byerlee (1978) found a large range of friction coefficients for different rock types at low pressures, but at moderate pressures (5–100 MPa, 50–1000 bar) the correlation was poor, and at high pressures (200–2000 MPa, 2–20 kbar) there was no rock type dependence at all. *Byerlee's law* states

$$F = aN + b \qquad [2.41]$$

where the coefficients a and b are 0.6 and 0.5 kbar at normal pressures of >2 kbar (200 MPa). Byerlee's law has the same form as the Navier-Coulomb failure criterion

$$\tau_c = C + \sigma\tan\phi \qquad [2.42]$$

where the shear stress acts in the direction of and balances the frictional force on a fracture plane, τ_c is the critical shear stress for failure, σ is the normal stress, ϕ is the angle of internal friction (in the range 0.7 to 1.7 for compact, coherent rocks), and C is the strength that exists even at zero normal stress (pressure) called *cohesion*. The shear stress needed to produce failure increases as the confining pressure increases. For initiation of a new fracture σ should be higher than for slip on an existing fracture.

Byerlee (1978) suggested that at normal pressures greater than 200 MPa (2 kbars), C is 50 MPa (0.5 kbars) and μ is 0.6 for a wide variety of rock types (Fig. 2.34). Where in the crust would we expect to find normal pressures of 200 MPa? The weight of a column of rock of density 2800 kg m^{-3}, height 1 km, and g = 10 m s^{-2}, is 28 MPa. We would expect to find 200 MPa of pressure at depths of 7 to 8 km in the crust. Earthquake foci are found at depths up to 15 km in the continental crust, equivalent to about 420 MPa (4.2 kbar). Between the surface and 15 km, therefore, we should expect brittle deformation on unlubricated faults. Below 15 to 20 km, however, earthquakes are much less common, and the rheology is ductile or 'plastic'. Depths of about 10–15 km correspond with a temperature of 300 °C in continental crust.

Although faulting of the brittle upper crust is very familiar, there is also observational evidence that near-surface rocks deform in both a plastic and fluid-like manner. The texture of many folded rocks suggests that the deformation responsible for the folding was the result of diffusive mass transfer. But the lower temperatures involved preclude the thermally activated diffusion processes found in the

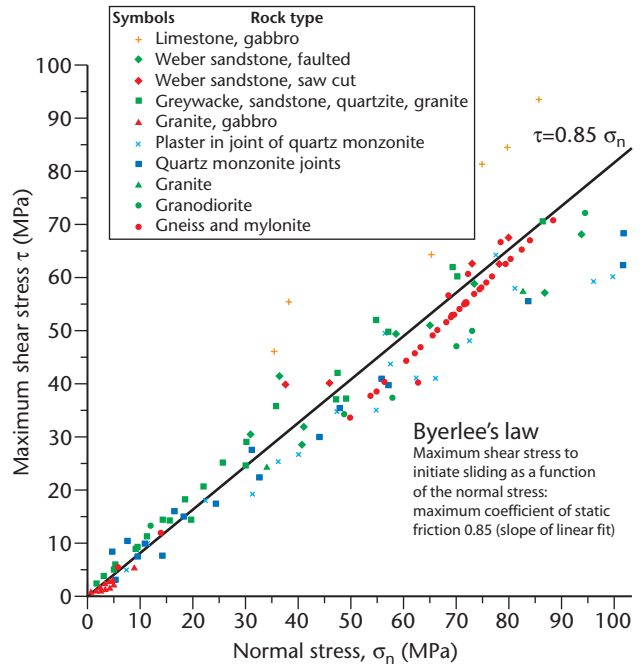

Fig. 2.34 Byerlee's law for the maximum shear stress τ to initiate sliding as a function of the normal stress σ_n for a variety of rock types. The linear fit defines a coefficient of static friction of 0.85 (Byerlee 1977). Reproduced with kind permission from Springer Science+Business Media.

mantle. Instead, the diffusive processes take place through *pressure solution* whereby material is transported in solution from areas of high intergranular pressure and stress and precipitated in regions of low pressure and stress, leading to creep. Point and line contacts between grains in a sandstone would represent such high-pressure/stress zones. The presence of water around the grains acts as the solvent and facilitates the transport of dissolved material. The pressure-solution process also leads to compaction (§9.2). Strain rate is linearly proportional to applied stress in pressure-solution creep, so the deformation is equivalent to that of a Newtonian fluid. It explains viscous folding of rocks at quite low temperatures.

Although pressure solution probably dominates as the main ductile mechanism at low temperatures in the upper crust, the rheology of the crust is likely to be complex as a result of the many compositional changes taking place within it. The layering of the continental crust affects its strength as a function of depth (§2.3.4). In particular, a ductile region in the middle crust is thought to be the location of detachment of major extensional faults (Kusznir *et al.* 1987). This topic of decoupling in mid-crustal regions is dealt with in greater depth in considering extensional basins in Chapter 3.

In summary, the consideration of rock deformation textures, laboratory studies of quartz-bearing rock and continental seismicity suggests that a seismogenic upper crust characterised by discontinuous-frictional (brittle) faulting passes below 10–15 km (c.300 °C) depth into an aseismic continuous quasi-plastic region where creep takes place in mylonitic fault zones (Sibson 1983; Scholz 1990). In the deep crust at temperatures greater than 450 °C, fully ductile continuous deformation is thought to dominate in gneissose shear zones (Grocott & Watterson 1980). The vertical zonation of

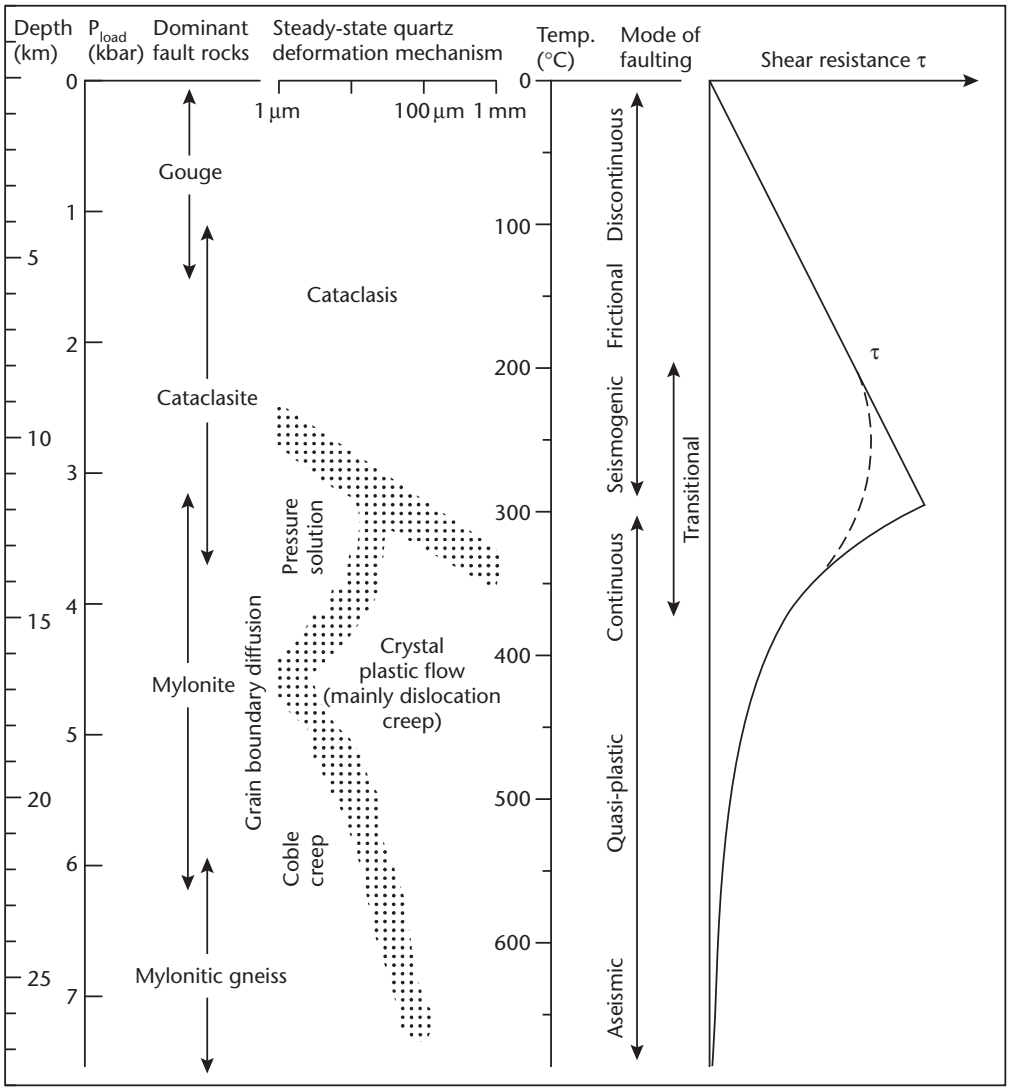

Fig. 2.35 Sibson's (1983) conceptual model for a major fault zone in the continental crust, showing the dominant deformation mechanisms.

deformation mechanism envisaged by Sibson (1983, p. 744) is shown in Fig. 2.35.

2.3.4 Strength profiles of the lithosphere

Lithospheric strength is controlled by lithospheric rheology, which itself is dependent on the heat flow or geothermal gradient, the composition of lithospheric rocks and the presence of volatiles. The thickness of the crust and lithosphere and the geothermal gradient therefore interact to produce distinctive strength profiles. We should expect these profiles to be dramatically different between oceanic and continental lithosphere (Lynch & Morgan 1987).

The lithosphere can deform by brittle (faulting) and by ductile mechanisms, each with its own yield strength. The yield stress (strength) of the lithosphere at any depth is conventionally calculated as the lesser of the two yield strengths for brittle failure and ductile creep (Brace & Kohlstedt 1980). The integral of the yield stress profile over depth is the total yield stress of the lithosphere, equivalent to the horizontal deviatoric force required to cause non-elastic strain. At one or several points in the profile, the brittle and ductile yield strengths may be equal. These points or depths are known as the *brittle–ductile transition*.

The brittle yield strength is assumed to be related to the pressure and therefore to the depth in the crust. It is independent of strain rate, temperature and rock type. The ductile yield stress, however, is strongly dependent on strain rate, temperature and rock composition. Let us firstly take the relatively straightforward case of the oceanic lithosphere (Fig. 2.36).

The Anderson theory of faulting expresses the dip of a fault in terms of the coefficient of static friction. The tectonic stress required to cause faulting under tension is then

$$\Delta\sigma_{xx} = \frac{-2f_s(\rho g y - p_w)}{\left(1+f_s^2\right)+f_s} \qquad [2.43]$$

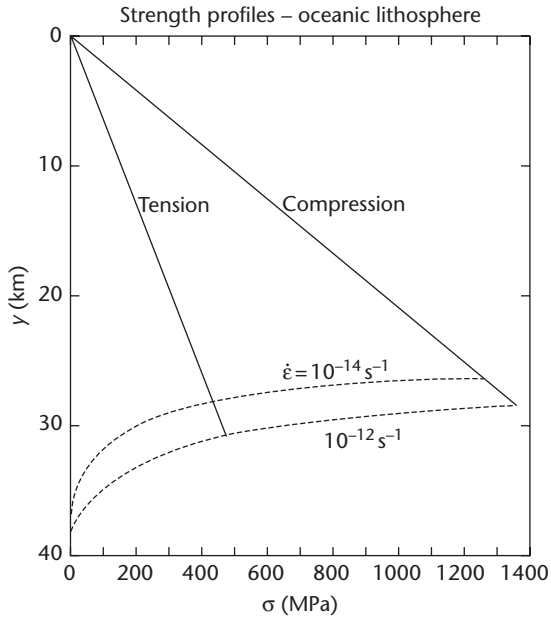

Fig. 2.36 Strength profile for the oceanic lithosphere calculated using eqn. [2.43] for tension and eqn. [2.44] for compression. The dashed lines are the stresses associated with creep in the lithosphere at the strain rates indicated. After Turcotte & Schubert (2002), © Cambridge University Press, 2002.

and under compression is

$$\Delta\sigma_{xx} = \frac{2f_s(\rho g y - p_w)}{\left(1+f_s^2\right) - f_s} \qquad [2.44]$$

where $\Delta\sigma_{xx}$ is the deviatoric stress, f_s is the static coefficient of friction, ρgy is the lithostatic stress, and p_w is the pore pressure. To find the failure stress of the oceanic crust, we assume the pore pressure is hydrostatic, so $p_w = \rho_w gy$, $\rho = 3300\,kg\,m^{-3}$ and $\rho_w = 1000\,kg\,m^{-3}$, and $f_s = 0.6$. The results for tension and compression are sketched in Fig. 2.36.

But how far downwards in the oceanic lithosphere does the Anderson theory extend? Power-law creep as a function of temperature is given by an equation of the form

$$\dot{\varepsilon} = A\sigma^n \exp\left(\frac{-E_a}{RT}\right) \qquad [2.45]$$

For olivine, $A = 4.2 \times 10^5\,MPa^{-3}\,s^{-1}$, E_a is the activation energy = $523\,kJ\,mol^{-1}$, and $n = 3$. Taking strain rates of $10^{-12}\,s^{-1}$ and $10^{-14}\,s^{-1}$, the stress as a function of depth for a geothermal gradient of $25\,K\,km^{-1}$ can be found (Fig. 2.36). We then assume that the lower of the frictional stress and the creep stress determines the strength of the oceanic lithosphere. The change from a frictional stress to a power-law stress takes place at a depth of c.28 km for the conditions given above. This is the brittle–ductile transition. Note that its depth depends on the geothermal gradient and strain rate, as well as on the 'sign' of the deviatoric stress (tension or compression).

The same concepts apply to the continental lithosphere, but geothermal gradients are likely to be more complex because of the internal heat generation of radiogenic minerals.

The effects of variations in geothermal gradient are illustrated in Fig. 2.37 using basal heat flows of $25\,mW\,m^{-2}$ representing a cold

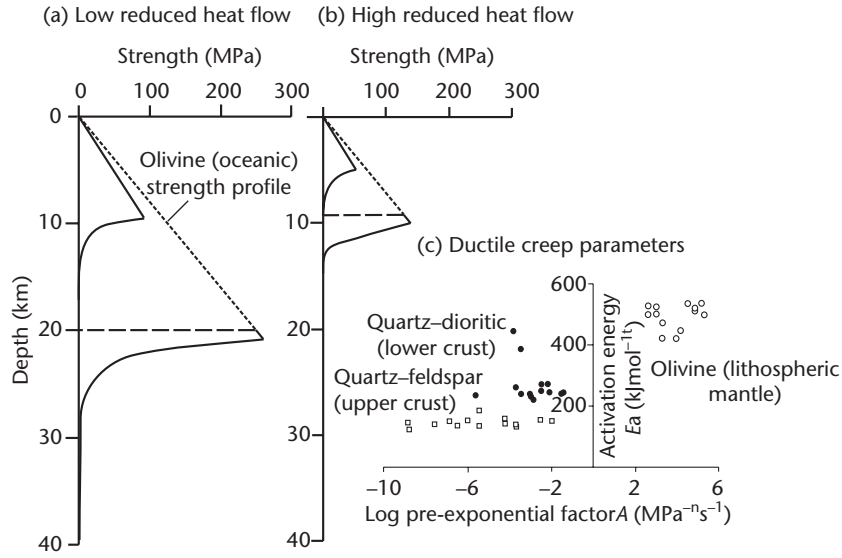

Fig. 2.37 Strength profiles for the continental lithosphere for different geothermal gradients, using (a) a low reduced heat flow of $25\,mW\,m^{-2}$ for a cold continental shield area, and (b) a higher reduced heat flow of $59\,mW\,m^{-2}$ for an area undergoing extension. $K = 2.5\,W\,m^{-1}\,K^{-1}$, $A_0 = 2.1 \times 10^6\,W\,m^{-3}$, $T_0 = 15\,°C$, universal gas constant = $8.31451\,J\,mol^{-1}\,K^{-1}$, strain rate is $10^{-12}\,s^{-1}$, $n = 3$. For the olivine rheology, $A = 104\,MPa^{-3}\,s^{-1}$, $E_a = 500\,kJ\,mol^{-1}$; for brittle failure under tension, $f_s = 0.6$, $\rho = 3300\,kg\,m^{-3}$. For the lower crustal rheology, $A = 100\,Pa^{-3}\,s^{-1}$, $E_a = 300\,kJ\,mol^{-1}$; and for brittle failure under tension, $f_s = 0.6$ and $\rho = 2750\,kg\,m^{-3}$. (c) Plot of activation energy E_a versus the pre-exponential factor A for a range of rock types typical of the lithospheric mantle, lower crust and upper crust. Derived from table 1 in Fernandez & Ranalli (1997).

continental shield, and 59 mW m^{-2} representing the high geothermal gradients typical of the Basin and Range province, USA (Lynch & Morgan 1987) ($K = 2.5\,\mathrm{Wm^{-1}K^{-1}}$, $A_0 = 2.1\,\mu\mathrm{W\,m^{-3}}$, $a_r = 10\,\mathrm{km}$). There are a number of points to note. First, there are two high-strength brittle regions in the lithosphere, one in the upper crust and one in the upper mantle, each underlain by brittle–ductile transitions. Second, the brittle–ductile transition occurs at a much shallower level in the crust for the hot geotherm than for the cold geotherm. Third, the area to the left of the strength profile, representing the total yield strength of the lithosphere, is much greater for the cold lithosphere. We should expect cold lithosphere to be 'stronger' than hot lithosphere when flexed or subjected to horizontal extensional or compressional deviatoric stresses. The total yield strength of the lithosphere is considerably smaller (about half) under tension than under compression.

Oceanic lithosphere has a yield strength that is a strong function of its age. The entire oceanic crust is brittle, and the yield strengths of oceanic lithosphere older than about 10 Ma under both extension and compression are considerably higher than any likely imposed tectonic forces. Continental lithospheric strength, on the other hand, is strongly dependent on the basal heat flow or geothermal gradient. For a low geotherm (basal heat flow of 25 mW m^{-2}) the strength under extension is an order of magnitude greater than any likely imposed forces; consequently, cold continental lithosphere is unlikely to undergo significant deformation. At higher geotherms the strength under extension is much reduced, making the continental lithosphere susceptible to extensional deformation. For example, the actively extending crust in the Basin and Range province, USA, is 30 km thick with a basal heat flow of 70 mW m^{-2} (Lachenbruch & Sass 1978), whereas in the eastern USA, which is tectonically inactive, crustal

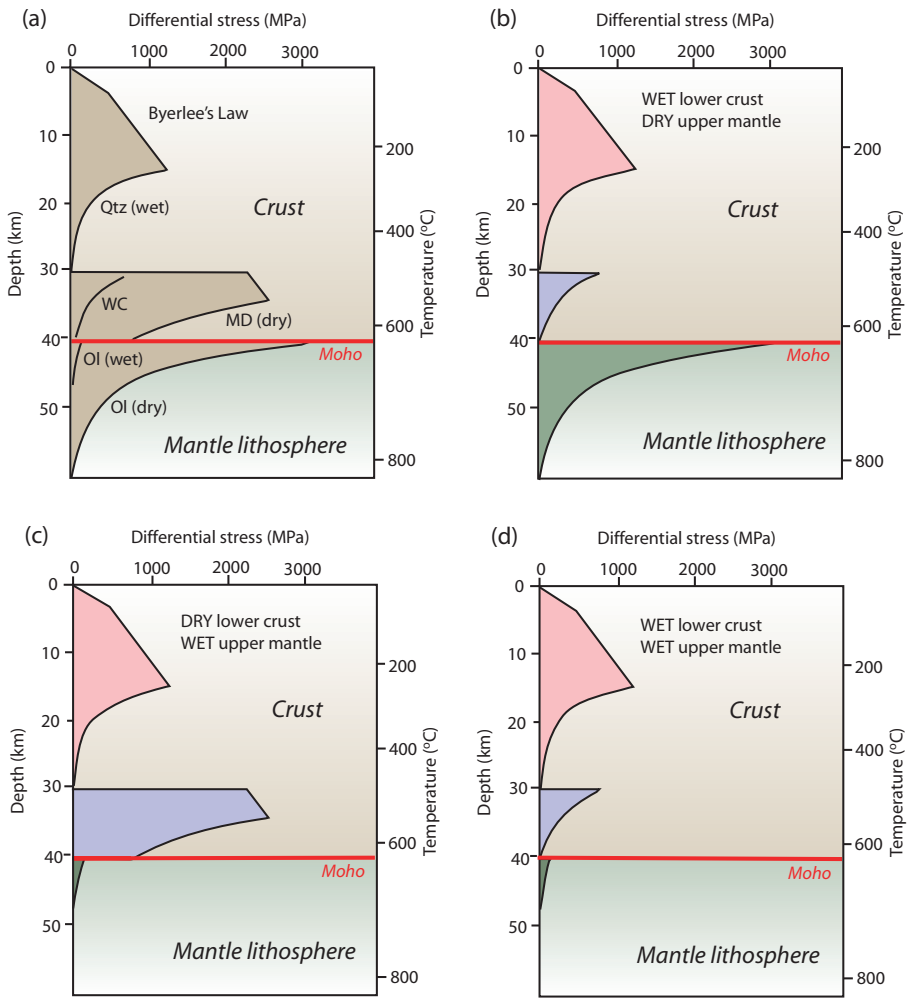

Fig. 2.38 Strength profiles: differential stress (rock strength) of the continental lithosphere (after Mackwell *et al*. 1998; Jackson 2002) based on the presence or absence of water. The Moho is at 40 km, and the geotherm is for a surface heat flow of 70 mW m^{-2}. The strain rate is $10^{-15}\,\mathrm{s^{-1}}$ in (a) to (d). (a) Rheologies are wet quartz (Qtz) in the upper crust, Maryland diabase (MD) or undried granulite (WC) in the lower crust, and olivine (Ol) in the lithospheric mantle. The frictional strength of the upper crust is given by Byerlee's law for quartz. In (b), a wet (weak) lower crust and dry (strong) upper mantle constitutes the jelly sandwich model of lithospheric strength. In (c), a dry (strong) lower crust and wet (no strength) upper mantle may represent conditions in continental cratons. In (d), a wet (weak) lower crust and a wet (weak) upper mantle causes almost all strength to reside in the seismogenic upper crust, known as the crème brûlée model.

thickness is 40 km and basal heat flow is 33 mW m^{-2}. A global comparison of tectonic activity and basal heat flows shows that regions with high heat flows (>60 mW m^{-2}), such as the Rhine Graben (western Europe) and Shansi Graben (China), undergo rifting, whereas regions with lower heat flows (<50 mW m^{-2}) do not show signs of active extensional deformation (Kusznir & Park 1987, p. 42, fig. 8).

Crustal composition is an important factor, particularly at high geothermal gradients. The quartz rheology of the upper crust is weaker than the plagioclase rheology of the lower crust, which is in turn much weaker than the olivine rheology of the mantle. The thicknesses of upper and lower crust are therefore important in determining strength profiles. Crustal thickening has the effect of weakening the lithosphere, since the weaker quartz-plagioclase rheologies may replace the stronger olivine rheology at the depths normally associated with the lithospheric mantle. At crustal levels, a strong plagioclase or ultramafic lower crustal rheology may be replaced by a weaker quartz rheology during thickening. Thickened crust in zones of convergence may therefore be prone to extensional collapse.

During extension of the lithosphere at a finite strain rate, an increase in geothermal gradient weakens the lithosphere, whereas the thinning of the crust simultaneously strengthens it. The balance between net strengthening and net weakening is strongly affected by the strain rate. At high strain rates ($\dot{\varepsilon}$ of 10^{-13} to 10^{-14} s^{-1}) there is less opportunity for heat loss and re-equilibration of the geothermal gradient, leading to net weakening, whereas slow strain rates ($\dot{\varepsilon}$ of 10^{-16} s^{-1}) promote net strengthening. Fast strain rates therefore lead to intense localised extension and high surface heat flows, whereas slow strain rates generate broad regions of extension with low surface heat flows.

In conclusion, the distribution of rock strength in the lithosphere has a profound influence on its mechanical behaviour, including during basin development. Conventional views are that there are two high-strength layers corresponding to the brittle upper crust (quartz) and brittle upper mantle (olivine), separated by a weak lower crust (diabase or granulite), causing it be termed the *jelly sandwich model* (Chen & Molnar 1983) (Fig. 2.38b). An alternative is that there is essentially a strong crust with a dry lower crust overlying a wet, weak mantle, so lithospheric strength resides almost entirely within a single strong crustal layer within which earthquakes would be confined (Fig. 2.38c). In this *crème brûlée model*, the continental lithosphere is more like the oceanic lithosphere in its strength profile. The presence of volatiles such as water, even in very low quantities, is known to dramatically lower the creep strength of rocks (Maggi *et al.* 2000). Consequently, small concentrations of water, perhaps related to migration of metasomatic fluids or derived from subduction, might lower the strength of the continental lithospheric mantle. In contrast, the lower crust under continental cratons (Fig. 2.38d) is most likely of an anhydrous granulitic composition, with a permeability barrier to underlying wet mantle and where volatiles have been driven off. It is clear, therefore, that the lithospheric strength profile, which controls the storage of elastic stresses, the deformation of the continents and the mechanical behaviour during extension (Chapter 3) and flexure (Chapter 4), is potentially complex.