CHAPTER 2

The physical state of the lithosphere

The kindly Earth shall slumber, lapt in universal law
(ALFRED LORD TENNYSON, *IN MEMORIAM* (1850))

SUMMARY

A knowledge of the behavior 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 tectonic processes 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 half of the lithosphere is sufficiently rigid that it is able to store and transmit stresses. Because 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 which 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.

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 nonzero, 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 nonzero 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*.

The lithosphere is able to bend. The shape of the flexure of the lithosphere depends on its rigidity and 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*.

Whereas the upper 15–50 km of the Earth is thought to behave more or less elastically over geological time scales, rocks below this level deform with a creep-like behavior and thus relax stresses.

In order to understand the mechanical behavior of the Earth it is necessary to understand its thermal structure, since this determines its rheology. 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 the convecting fluid 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

equivalent to the "reduced" or basal heat flow from the mantle.

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 reduction of surface gravity data (*Bouguer anomalies*) provides important information on the way in which topography is compensated isostatically. The Airy model assumes that surface topography is compensated by variations in the depth of the underlying crust, whereas the Pratt model assumes variations in the density of the underlying crust. Periodic loads are supported by the flexural strength of the lithosphere, the degree of compensation 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. *Diffusion creep* and *dislocation creep* are both strongly temperature dependent processes, but the latter is more rapid since it is driven by shear stresses acting on crystal imperfections. Crustal rocks may also behave in a ductile manner by pressure-solution creep, or they may deform in a brittle manner by fracturing (*Byerlee's law*). The occurrence of earthquake foci, and laboratory experiments on rock mechanics, suggest that that at depths of about 15 km and temperatures of *c.* 300 °C the continental crust starts to become ductile or plastic. The rheology of the crust is complex because of 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. Continental lithospheric strength is strongly controlled by geothermal gradient and by thickening/thinning of the crust. Strain rate affects whether the lithosphere undergoes a net weakening or net strengthening over time, and therefore has an important role in determining the evolution through time of continental deformation.

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 the next section, 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.

A large number of texts cover topics addressed in this chapter. The reader is particularly directed towards Fowler (1990), Middleton and Wilcock (1994) and Turcotte and Schubert (2002) for a wealth of further details on heat flow, gravity, linear elasticity and flexure, and rheology.

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 normalize rock densities to a zero-pressure value, 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 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 (Table 2.1).

Surface forces act only on the surface area bounding a volume and arise from the interatomic stresses exerted

Table 2.1 Common physical properties of rocks (after Turcotte and Schubert 2002).

	Density $kg\,m^{-3}$	Young's modulus E 10^{11} Pa	Shear modulus G 10^{11} Pa	Poisson's ratio ν	Thermal conductivity $K\,Wm^{-1}K^{-1}$	Coefficient of thermal expansion α, $10^{-5}K^{-1}$
Sedimentary						
Shale	2100–2700	0.1–0.7	0.1–0.3	0.1–0.2	1.2–3	
Sandstone	1900–2500	0.1–0.6	0.04–0.2	0.1–0.3	1.5–4.2	3
Limestone	1600–2700	0.5–0.8	0.2–0.3	0.15–0.3	2–3.4	2.4
Dolomite	2700–2800	0.5–0.9	0.2–6.4	0.1–0.4	3.2–5	
Metamorphic						
Gneiss	2600–2850	0.4–0.6	0.2–0.3	0.15–0.25	2.1–4.2	
Amphibolite	2800–3150		0.5–1.0	0.4	2.1–3.8	
Marble	2670–2750	0.3–0.8	0.2–0.35	0.2–0.3	2.5–3	
Igneous						
Basalt	2950	0.6–0.8	0.25–0.35	0.2–0.25	1.3–2.9	
Granite	2650	0.4–0.7	0.2–0.3	0.2–0.25	2.4–3.8	2.4
Diabase	2900	0.8–1.1	0.3–0.45	0.25	2–4	
Gabbro	2950	0.6–1.0	0.2–0.35	0.15–0.2	1.9–4	1.6
Diorite	2800	0.6–0.8	0.3–0.35	0.25–0.3	2.8–3.6	
Pyroxenite	3250	1.0	0.4		4.1–5	
Anorthosite	2640–2920	0.83	0.35	0.25	1.7–2.1	
Granodiorite	2700	0.7	0.3	0.25	2.0–3.5	
Mantle						
Peridotite	3250				3–4.5	2.4
Dunite	3000–3700	1.4–1.6	0.6–0.7		3.7–4.6	
Miscellaneous						
Halite			0.3	0.15	5.4–7.2	13
Ice	917		0.092	0.31–0.36	2.2	5

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*.

The surface force acting on a unit area at the base of a vertical column of rock is given by

$$\sigma_{yy} = \rho g y \tag{2.1}$$

where ρ is the density of the rock column, g is the acceleration of gravity, and y is the height of the column. For equilibrium to be maintained, we can equate the surface forces due to the differing rock columns under continental and oceanic lithosphere. The isostatic balance at a depth equivalent to the base of the continental block (Fig. 2.1a) is expressed

$$h\rho_c g = b\rho_m g \tag{2.2}$$

and the elevation of the continental block above the mantle is given by

$$h - b = h\left(1 - \frac{\rho_c}{\rho_m}\right) \tag{2.3}$$

Putting $\rho_c = 2700\,kg\,m^{-3}$, $\rho_m = 3300\,kg\,m^{-3}$ and $h = 35\,km$, the elevation of the continental block is 6.4 km.

For the more realistic case of a continental block with its surface at sea level flanked by oceanic crust and an overlying water column (Fig. 2.1b), the isostatic balance at a depth equivalent to the base of the continental block is

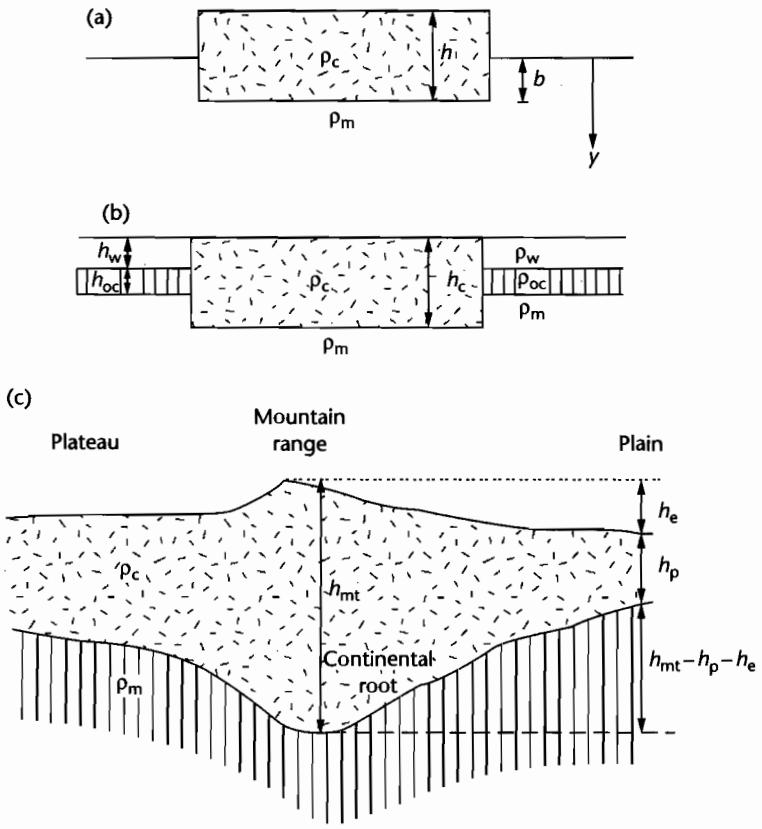

Fig. 2.1 Schematic diagrams illustrating the concepts of isostasy. (a) Continental block "floating" in a fluid mantle; (b) continental block flanked by oceanic crust and overlying water column; (c) continental mountain belt with a root surrounded by a plateau and low-lying plain. Notation explained in text.

$$h_c\rho_c g = h_w\rho_w g + h_{oc}\rho_{oc} g + (h_c - h_w - h_{oc})\rho_m g$$
$$h_{oc}(\rho_m - \rho_{oc}) + h_w(\rho_m - \rho_w) = h_c(\rho_m - \rho_c) \qquad (2.4)$$

Putting $\rho_c = 2700\,\mathrm{kg\,m^{-3}}$, $\rho_{oc} = 2950\,\mathrm{kg\,m^{-3}}$, $\rho_m = 3300\,\mathrm{kg\,m^{-3}}$, $h_c = 35\,\mathrm{km}$, and $h_{oc} = 20\,\mathrm{km}$, the water depth is 6 km. This is roughly the depth of the ocean we should expect above the oceanic abyssal plains.

A more familiar parameterization for isostasy is of a mountain belt with a "root," the excess mass in the elevated continental crust being compensated by the mass deficit at depth of the continental root replacing mantle. The isostatic balance at a depth equivalent to the base of the continental root (Fig. 2.1c) is

$$h_{mt}\rho_c g = h_p\rho_c g + (h_{mt} - h_p - h_e)\rho_m g$$

$$h_{mt} = h_p + h_e\left(\frac{\rho_m}{\rho_m - \rho_c}\right) \qquad (2.5)$$

Putting $h_p = 35\,\mathrm{km}$, $h_e = 5\,\mathrm{km}$, $\rho_m = 3300\,\mathrm{kg\,m^{-3}}$, $\rho_c = 2700\,\mathrm{kg\,m^{-3}}$, the thickness of crust in the mountain belt h_{mt} is 62.5 km. A very deep "granitic" root is therefore required to support a mountain range with an elevation of 5 km.

We will examine a simple application of isostasy in studying the bending of lithosphere in §2.1.4.

Normal surface forces can also be exerted on vertical planes (Fig. 2.2). If the normal surface forces, σ_{xx}, σ_{yy}, and σ_{zz} are all equal and 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

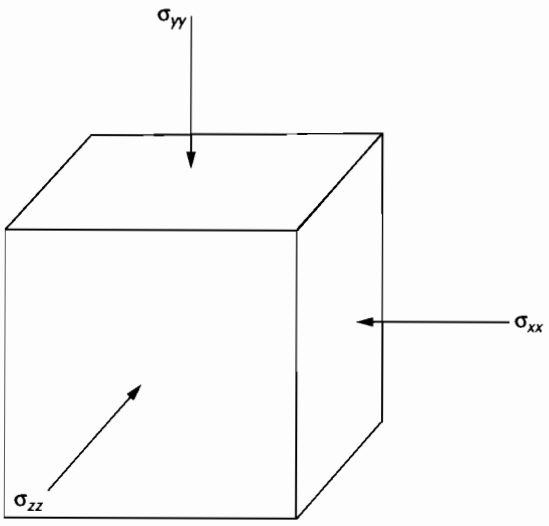

Fig. 2.2 Normal surface forces acting on vertical and horizontal planes.

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 ($\Delta\sigma_{xx}$),

$$\sigma_{xx} = \rho_c g y + \Delta\sigma_{xx} \tag{2.6}$$

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 or may be transmitted from plate boundaries, when they are commonly referred to as *in-plane* or *intraplate stresses* (see Chapters 3 and 8).

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 generalized 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 labeled

σ_1 = maximum principal stress
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

$$p = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3) \tag{2.7}$$

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 nonzero, giving the state of plane stress. Such a system is suggestive of the horizontal stresses in the lithosphere caused by tectonic processes.

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 δ_x, δ_y, and δ_z 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*.

Shear strains, however, may distort the shape of an element of a solid. In the case of a 2-D rectangular element which is distorted to a parallelogram (Fig. 2.4), 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

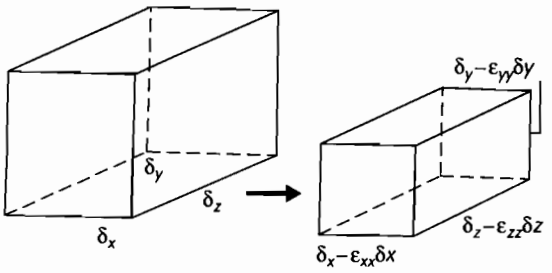

Fig. 2.3 A rectangular block that changes its dimensions but not its shape: this is a deformation involving no shear.

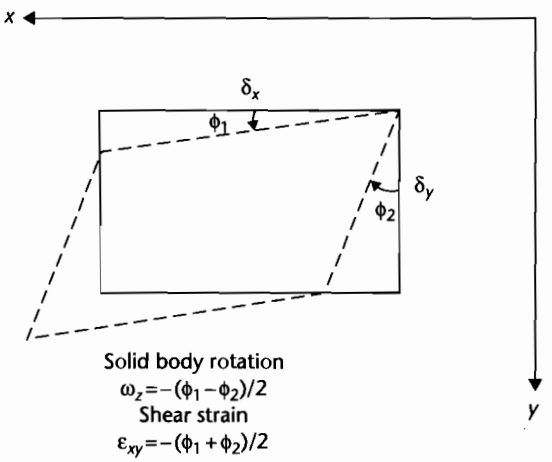

Fig. 2.4 Deformation of a rectangle into a parallelogram by a strain field involving shear.

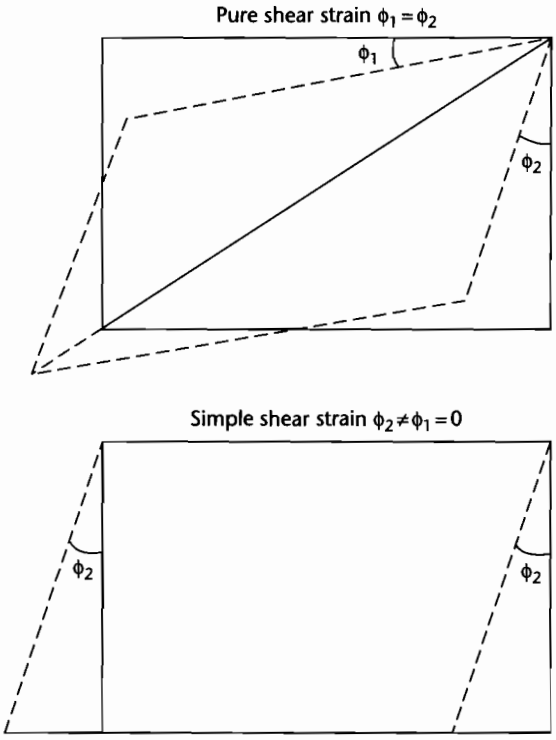

Fig. 2.5 Difference between pure shear strain (no solid body rotation) and simple shear strain (solid body rotation is $\phi_2/2$).

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. Solid body rotations do not involve changes in the distances between neighboring 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*. Figure 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 translithospheric shear zones) have been applied to the formation of extensional basins (Chapter 3).

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 3-D case the condition of isotropic strain is satisfied by

$$e = (\varepsilon_1 + \varepsilon_2 + \varepsilon_3)/3 = \Delta/3 \tag{2.8}$$

where Δ is the dilatation and e 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 nonzero component of principal strain, that is, assuming the nonzero 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 nonzero. 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.

2.1.3 Linear elasticity

It is clearly 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.

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 (§2.4). 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.

Because 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 very great deal of thought on lithospheric mechanics and often constitutes the basic assumption in models of lithospheric behavior.

In a linear, isotropic, elastic solid the stresses are linearly proportional to strains, and mechanical 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.

$$\varepsilon_1 = \frac{\sigma_1}{E} - \frac{\nu\sigma_2}{E} - \frac{\nu\sigma_3}{E} \tag{2.9}$$

$$\varepsilon_2 = -\frac{\nu\sigma_1}{E} + \frac{\sigma_2}{E} - \frac{\nu\sigma_3}{E} \tag{2.10}$$

$$\varepsilon_3 = -\frac{\nu\sigma_1}{E} - \frac{\nu\sigma_2}{E} + \frac{\sigma_3}{E} \tag{2.11}$$

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 nonzero (uniaxial stress), a shortening in the direction of the applied compressive stress will be accompanied by an extension in the two orthogonal directions (Fig. 2.6), 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

$$\sigma_1 = E\varepsilon_1 \tag{2.12}$$

This well-known relationship is called *Hooke's law*.

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 an increase in density. Such a density increase $\delta\rho$ is given simply by

$$\delta\rho = \rho\beta p \tag{2.13}$$

where p is the pressure, ρ is the density of the solid element and β the compressibility. In terms of the

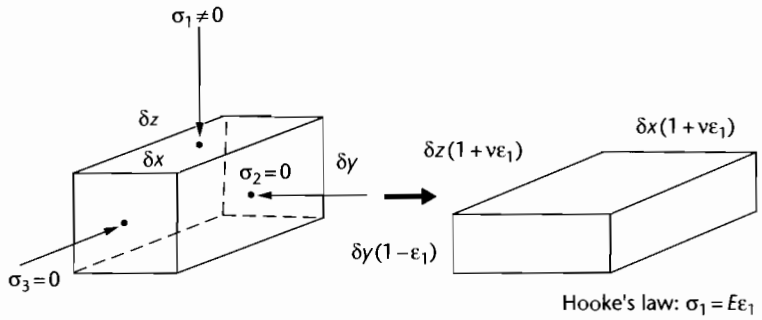

Fig. 2.6 Deformation under a uniaxial stress. Contraction in the direction of the compressive stress σ_1 is compensated by extension in the two orthogonal directions.

previously defined Young's modulus E and Poisson's ratio v,

$$K = \frac{1}{\beta} = \frac{E}{3(1-2v)} \tag{2.14}$$

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

2.1.4 Flexure in two dimensions

Because 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 theoretical background of the way the lithosphere responds by flexure to these applied force systems. A full analysis is provided by Turcotte and Schubert (1982, 2002). The concepts involved in the flexure of an elastic solid may be briefly summarized as follows:

- Flexure results from vertical forces, horizontal forces, and torques (bending moments) in any combination. Horizontal loads are commonly neglected, perhaps unwisely, in geodynamical problems;
- the bending moment is the integration of the fibre (normal) stresses on cross-sections of the plate acting over the distance to the midline of the plate. 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 can be derived which 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.

BOXED TEXT 2.1: Derivation of Flexural Rigidity and the General Flexure Equation

In the simplest terms the flexure of a plate depends on its thickness, elastic properties, and the nature of the applied load. Imagine a plate of thickness h and width L which is fixed at both ends and which is subjected to a line load at its midpoint (Fig. 2.7). In order to attain a force balance, the vertical line load V_a must be counteracted by vertical forces $V_a/2$ at both ends. Assuming that both plate thickness h and deflection w are small compared to the width of the plate L (as required by linear elastic theory), we can study the

Continued

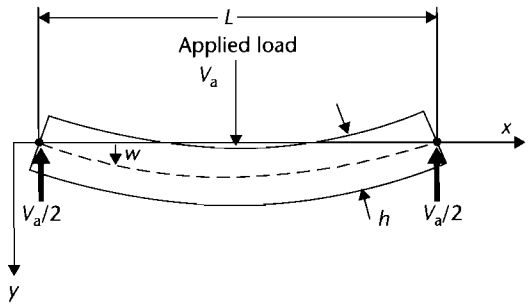

Fig. 2.7 Configuration and notation for a thin plate, pinned at both ends, bending under an applied load.

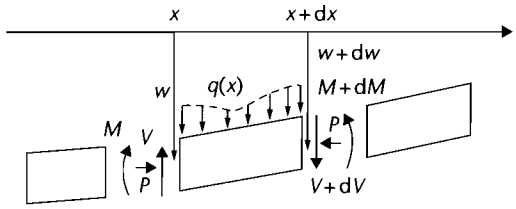

Fig. 2.8 Forces on a small element of a flexed plate. The forces can be balanced vertically and in terms of their tendency to rotate the element (moments or torques).

forces and torques on a small element of the plate. A downward force per unit area $q(x)$ is exerted on the plate by the applied load and on the end sections there is a net shear force per unit length V and horizontal force P per unit length, the latter being independent of x (Fig. 2.8). A bending moment M also acts on the end section related to the effects of normal stresses on the cross-section, σ_{xx}. These normal stresses are known as *fiber stresses* (Fig. 2.9).

The bending moment M is related to the curvature of the plate, since forces on the end section exert a torque about the midpoint of the plate. If the force on an element of thickness dy on the end of the plate is $\sigma_{xx}dy$, then this force will exert a torque about the midpoint ($y = 0$) of $\sigma_{xx}y\,dy$. Integrated over the entire end section we obtain the bending moment

$$M = \int_{-h/2}^{h/2} \sigma_{xx} y dy \tag{2.15}$$

The fiber stresses σ_{xx} result in longitudinal strains in the plate, contractional in the upper half and extensional in the lower half (Fig. 2.9). In the 2-D case

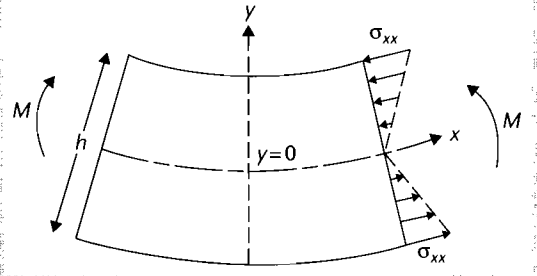

Fig. 2.9 Normal stresses on an end section of a flexed plate. These normal or fiber stresses exert torques about the midpoint of the plate, which when integrated over the end section (from $-h/2$ to $+h/2$) gives the bending moment.

we assume that the plate is infinite in the plane normal to the figure, so there is no strain in the direction perpendicular to the xy plane, that is, $\varepsilon_{zz} = 0$. This is the situation of *plane strain* (§2.1.2). If the plate is thin, it is possible to make the further assumption that stresses normal to the plate's surface are zero ($\sigma_{yy} = 0$). The relations between stress and strain can then be restated as (cf. eqn 2.9)

$$\varepsilon_{xx} = \frac{1}{E}(\sigma_{xx} - \nu\sigma_{zz}) \tag{2.16}$$

and from equation (2.10)

$$\varepsilon_{zz} = \frac{1}{E}(\sigma_{zz} - \nu\sigma_{xx}) \tag{2.17}$$

If the bending is 2-D (there is no strain in the direction perpendicular to the page in Fig. 2.10), $\varepsilon_{zz} = 0$, then (2.16) and (2.17) give

$$\sigma_{xx} = \frac{E}{(1-\nu^2)}\varepsilon_{xx} \tag{2.18}$$

which is the relation between fiber stresses and longitudinal strains. The bending moment can now be related to longitudinal strains by substitution of (2.18) in (2.15)

$$M = \frac{E}{(1-\nu^2)}\int_{-h/2}^{h/2} \varepsilon_{xx} y\, dy \tag{2.19}$$

The bending moment M can also be related to the deflection w. The local radius of curvature R is inversely

proportional to the rate of change in slope of the deflection, or $-d^2w/dx^2$ (if strains are small and $dw/dx \ll 1$, and where the negative sign simply states that w is positive downwards) (Fig. 2.10a). In addition, the longitudinal strain is also related to the radius of curvature. This relation can be derived by geometrical similarity (Fig. 2.10b), since the length of the plate is dependent on the local radius of curvature and ϕ ($l = R\phi$), and the change in length of the plate (Δl) is determined by the distance from the midline of the plate (y) and ϕ ($\Delta l = \phi y$). Using this result, the longitudinal strains ($\Delta l/l$) can be expressed in terms of the deflection,

$$\varepsilon_{xx} = \frac{\Delta l}{l} = \frac{y}{R} = -y\frac{d^2w}{dx^2} \tag{2.20}$$

The bending moment can therefore be rewritten

$$M = \frac{-E}{(1-\nu^2)}\frac{d^2w}{dx^2}\int_{-h/2}^{h/2} y^2\, dy \tag{2.21}$$

which when evaluated between the limits $-h/2$ and $h/2$ gives

$$M = \frac{-Eh^3}{12(1-\nu^2)}\frac{d^2w}{dx^2} \tag{2.22}$$

The coefficient of $-d^2w/dx^2$ in equation (2.22) is defined as the *flexural rigidity, D*. Note that the flexural rigidity D is proportional to the cube of the elastic thickness of the plate h. Flexural rigidity is commonly expressed as this length term. When applied to the lithosphere, the term equivalent elastic thickness Te is preferred, since the value of Te does not reflect the depth of a real physical discontinuity in the lithosphere; it is simply the equivalent thickness of a purely elastic beam.

The bending moment can therefore be written in terms of the curvature,

$$M = -D\frac{d^2w}{dx^2} = \frac{D}{R} \tag{2.23}$$

We can now proceed to establish the general flexural equation by carrying out a force balance on a segment of the plate. The force balance (Fig. 2.8) on a flexed plate with a downward force caused by loading $q(x)$ and horizontal forces (P) on the end sections is as follows.

Equating the forces in a vertical direction we have

$$q(x)dx = -dV$$

$$q(x) = -\frac{dV}{dx} \tag{2.24}$$

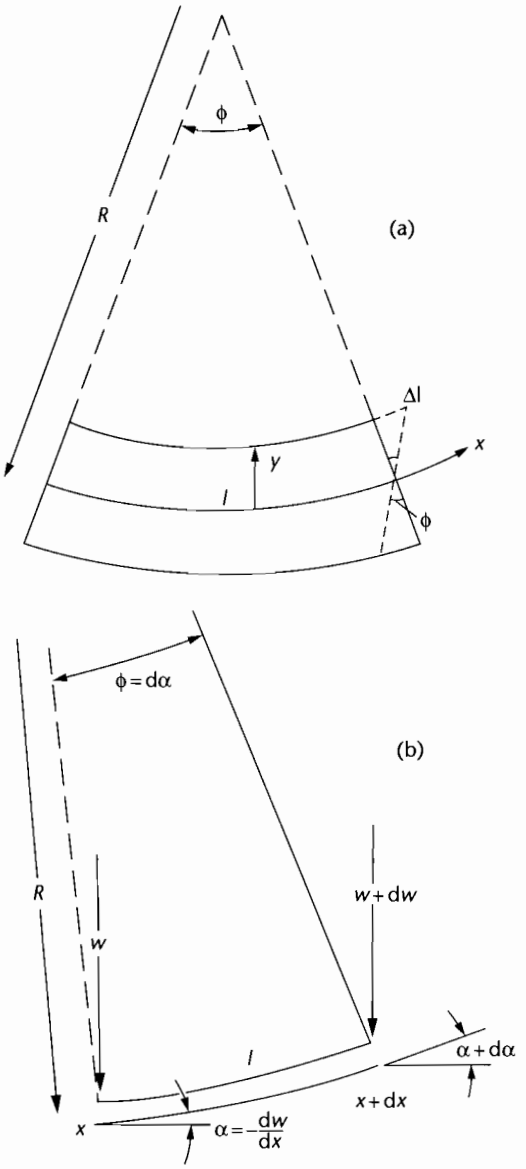

Fig. 2.10 Geometrical aspects of plate bending. (a) Longitudinal strain (extension below the midplane and contraction above the midplane of the plate) is a function of the distance from the midplane of the plate y, and the angle ϕ; (b) Notation to show that the second derivative of the deflection (d^2w/dx^2) gives the rate of change of slope of the plate; this is inversely related to the local radius of curvature R of the plate.

Continued

Balancing torques we have in the counterclockwise direction a force P acting over a moment arm of $-dw$ and the torque dM. In the clockwise sense we have $(V + dV)$ acting over a moment arm of dx. Since the term dV is small we can write

$$-Pdw + dM = Vdx \tag{2.25}$$

Differentiating (2.25) twice, so that (2.24) can be substituted into (2.25) gives

$$-P\frac{d^2w}{dx^2} + \frac{d^2M}{dx^2} = -q(x) \tag{2.26}$$

Since we have already expressed the bending moment in terms of the flexural rigidity and the local curvature, (2.26) can be rewritten as

$$D\frac{d^4w}{dx^4} = q(x) - P\frac{d^2w}{dx^2} \tag{2.27}$$

which is the *general flexure equation* for the deflection of a plate.

In the context of an elastic plate overlying a fluid-like mantle, this equation must be modified to account for a restoring force (buoyancy) acting upwards on the deflected plate (see below).

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.

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 smaller 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 (2.27) therefore becomes

$$D\frac{d^4w}{dx^4} + P\frac{d^2w}{dx^2} + \Delta\rho gw = q_a(x) \tag{2.28}$$

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 Chapter 4.

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, say 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.

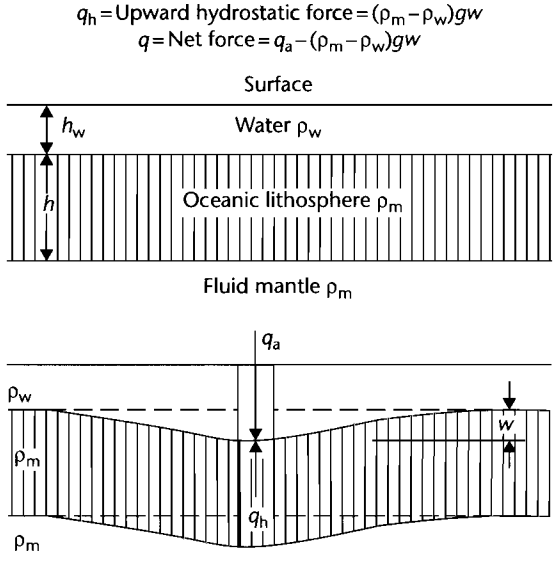

Fig. 2.11 Model for calculating the upward-acting hydrostatic restoring force on an oceanic plate, overlain by water, deflected by an applied force q_a.

However, if the wavelength of a load of the same mass is sufficiently long, there is an effective isostatic response towards hydrostatic equilibrium and the lithosphere appears to have no rigidity. These two situations must

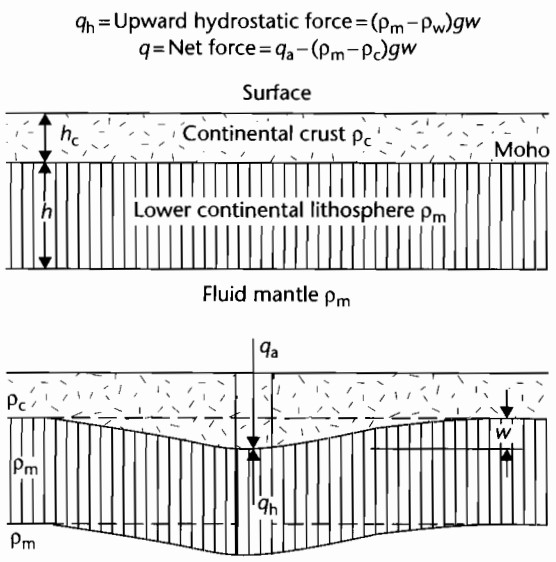

Fig. 2.12 Model for calculating the hydrostatic restoring force on the base of some continental crust where the deflection caused by the applied load q_a is assumed to be filled with material of the same density as the continental crust. This therefore approximates the case of a sediment-filled sedimentary basin on continental lithosphere.

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 §2.3.3 and revisited in §9.3.

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 (§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.

2.2 HEAT FLOW: CONDUCTION AND CONVECTION

2.2.1 Fundamentals

To understand the mechanical behavior of the Earth it is necessary to know something of its thermal structure, since rock rheologies 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 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

$$Q = -K\frac{dT}{dy} \tag{2.29}$$

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.

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 (see § 9.6.1). 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 (Table 2.1). The large range of values for sedimentary rocks is due, to a large extent, to large

variations in porosity (see also §9.3.1). 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 meter per degree Centigrade ($W\,m^{-1}\,°C^{-1}$), and occasionally in $cal\,cm^{-1}\,°C^{-1}$.

Temperature measurements of the ocean floor can also be made by penetrating seafloor sediments with a temperature probe. The same probe contains a heater which enables the *in situ* thermal conductivity to be calculated.

The results of heat flow measurements of the oceans and continents reveal important variations (Table 2.2, Sclater et al. 1980a, 1981). Regions of high heat flow on the continents generally correspond to active volcanic areas, such as the Andes, or to regions of extensional tectonics like the Basin and Range province of western USA. Continental collision zones typically have low to normal surface heat flows (further details in §9.9). In areas devoid of active tectonics and vulcanicity, the heat flow appears to be inversely correlated to the age of the rocks (Jessop and Lewis 1978). This can be explained by the decreasing abundance with age of the radioactive heat-producing isotopes of uranium, thorium and potassium. Surface heat flows are also determined by the underlying rock type. Granites produce large amounts of radiogenic heat, whereas basalts and peridotites produce almost no radiogenic heat (Table 2.3).

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. Mean oceanic surface heat

Table 2.2 Regional variations in surface heat flow (data from Sclater et al. 1980a).

	Mean surface heat flow	
	$mW\,m^{-2}$	HFU
Continents:	56.6	1.35
Africa	49.8	1.19
North America	54.4	1.30
Australia	63.6	1.52
Oceans:	78.2	1.87
North Pacific	95.4	2.28
Indian Ocean	83.3	1.99
South Atlantic	59.0	1.41
Worldwide	69.9	1.67

Table 2.3 Typical concentrations of radioactive elements and heat generation from typical rock types comprising the continental and oceanic crust and undepleted mantle (from Fowler 1990).

	Granite	Tholeiitic basalt	Alkali basalt	Peridotite	Average continental upper crust	Average oceanic crust	Undepleted 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 (%)	3.5	0.2	1.2	0.01	2.0	0.4	0.02
Heat generation ($10^{-10}\,W\,kg^{-1}$)							
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 ($10^3\,kg\,m^{-3}$)	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

Table 2.4 Relative abundance of isotopes and crustal heat generation as a function of age (from Jessop and Lewis 1978).

	Relative abundance					Heat generation	
Age (Ma)	^{238}U	^{235}U	U*	Th	K	Model A†	Model B‡
Present	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	1.08	1.62	1.10	1.03	1.31	1.13	1.17
1000	1.17	2.64	1.23	1.05	1.70	1.28	1.37
1500	1.26	4.30	1.39	1.08	2.34	1.48	1.64
2000	1.36	6.99	1.59	1.10	2.91	1.74	1.98
2500	1.47	11.4	1.88	1.13	3.79	2.08	2.43
3000	1.59	18.5	2.29	1.16	4.90	2.52	3.01
3500	1.71	29.9	2.88	1.19	6.42	3.13	3.81

Notes:

* Assuming a present isotopic composition of 99.2886% ^{238}U and 0.7114% ^{235}U.

† Model A based on Th/U = 4, K/U = 20,000.

‡ Model B based on Th/U = 4, K/U = 40,000.

flows are somewhat larger than their continental counterparts (Table 2.2). Expressed somewhat differently, the Earth's average surface heat flow corresponds roughly to one household light bulb (100 W) over an area of a tennis court. 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. 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), 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.

2.2.2 One-dimensional heat conduction

Heat travels down temperature gradients (eqn 2.29). 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 (Boxed text 2.2). Once such a rule is established, it can be modified to account for 3-D variations in heat flow.

BOXED TEXT 2.2: The 1D Heat Conduction Equation

Consider a volume element of vertical dimension δy with a cross-sectional area a (Fig. 2.13). A heat flow $Q(y)$ enters the element across its top surface. The element has an internal heat generation and is composed of material of density ρ and thermal capacity (specific heat) denoted by c. The specific heat is the amount of heat needed to raise the temperature of 1 kg of material by 1 °C (with units therefore of $Wkg^{-1}°C^{-1}$).

In a unit of time, the heat entering the volume element is $aQ(y)$ and the heat leaving the element is $aQ(y+\delta y)$.

$Q(y+\delta y)$ can be expanded in a Taylor series to give

$$Q(y+\delta y) = Q(y) + \delta y \frac{\partial Q}{\partial y} + \ldots \qquad (2.30)$$

Taking the first two terms only of the Taylor series, the net gain or loss of heat is therefore the difference between the heat fluxes through the top of the element and out of the bottom,

Continued

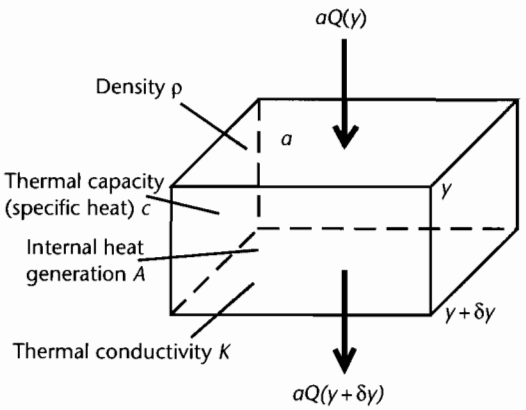

Fig. 2.13 Set-up and notation for conductive heat flow through a volume element of thickness δy, cross-sectional area a, density ρ, thermal capacity (specific heat) c, thermal conductivity K and internal heat generation A. Heat is conducted across the shaded faces only (it is perfectly 1-D).

$$aQ(y) - a\left\{Q(y) + \delta y \frac{\partial Q}{\partial y}\right\} = -a\delta y \frac{\partial Q}{\partial y} \tag{2.31}$$

Let the heat generated internally per unit volume and unit time be A. The heat generated in the element of volume $a\delta y$ is therefore $Aa\delta y$. The total heat gain or loss per unit time is then

$$Aa\delta y - a\delta y \frac{\partial Q}{\partial y} \tag{2.32}$$

If the material making up the volume element has specific heat c and density ρ and undergoes a small temperature change δT in a small time δt, the rate at which heat is gained or lost is

$$ca\delta y\rho \frac{\partial T}{\partial t} \tag{2.33}$$

Equating (2.32) and (2.33),

$$ca\delta y\rho \frac{\partial T}{\partial t} = Aa\delta y - a\delta y \frac{\partial Q}{\partial y} \tag{2.34}$$

which when simplified gives

$$c\rho \frac{\partial T}{\partial t} = A - \frac{\partial Q}{\partial y} \tag{2.35}$$

Substituting Fourier's law (equation 2.29) into (2.35),

$$\frac{\partial T}{\partial t} = \frac{A}{\rho c} + \frac{K}{\rho c}\frac{\partial^2 T}{\partial y^2} \tag{2.36}$$

which is the *1-D heat conduction equation*. We can define the *thermal diffusivity* as $\kappa = \dfrac{K}{\rho c}$. The thermal diffusivity expresses the ability of a material to gain or lose heat by conduction.

Certain boundary conditions can be applied to the 1-D heat conduction equation in order to simplify it for certain physical situations in the lithosphere:

(a) *Steady state solution* where there is no temperature change with time, i.e. $\partial T/\partial t = 0$,

$$\frac{\partial^2 T}{\partial t^2} = -\frac{A}{K} \tag{2.37}$$

(b) *no internal heat generation*, $A = 0$,

$$\frac{\partial T}{\partial t} = \frac{K}{\rho c}\frac{\partial^2 T}{\partial y^2} = \kappa \frac{\partial^2 T}{\partial y^2} \tag{2.38}$$

which is the *diffusion equation*. An equation of this form is used in a wide range of physical problems.

The volume element may be in motion relative to its surroundings and may therefore move through a region where the temperature varies with depth. If the volume element moves at a velocity u_y in the y direction (vertically in the lithosphere), its depth after time t will be $y + u_y t$, and the associated temperature change will be $u_y \partial T/\partial y$. The 1-D heat conduction equation therefore needs to be modified. The temperature change after a small time period δt is given by

$$\frac{\partial T}{\partial t} = \frac{K}{\rho c}\frac{\partial^2 T}{\partial y^2} + \frac{A}{\rho c} - u_y \frac{\partial T}{\partial y} \tag{2.39}$$

The temperature change of a piece of lithosphere is therefore seen to be made of three components: a basal heat flow term, an internal heat generation term; and an *advective term*. The advective term may be the movement towards the surface of a volume element of rock associated with the downcutting action of erosion. Or, with a different sign, it could be the velocity of deposition.

2.2.3 Geotherms

The variation of temperature with depth is called the *geotherm*. In the boxed text, we solve the 1-D conduction equation to derive the geotherm.

Taking reasonable parameter values, it is possible to calculate the temperature as a function of any depth in the crust ($0 \leq y \leq y_c$) using equation (2.43): thermal conductivity $K = 3\,\mathrm{W\,m^{-1}\,°C^{-1}}$, radiogenic heat generation A (in top 30 km) = $1.25\ 10^{-6}\,\mathrm{W\,m^{-3}}$, surface heat flow Q_0 (global average) = $70\,\mathrm{mW\,m^{-2}}$, and surface temperature $T_0 = 10\,°\mathrm{C}$. Extrapolating downwards, the temperatures calculated would cause the mantle to be fully molten below depths of about 100 km, since the geotherm would intersect the solidus for olivine, yet the mantle is able to propagate shear waves. Setting the radiogenic heat generation A to zero leaves the same problem. We conclude that the conduction geotherm in equations (2.43) and (2.46) does not apply to the bulk of the mantle where convection is taking place (§2.2.6).

Nevertheless, a conductive temperature profile appears to match continental geotherms where radiogenic heat production and conductive heat transport safely apply. The radiogenic heat production cannot be assumed to be uniform with depth. The abundances of the heat-producing radiogenic isotopes in a number of common rock types is given in Table 2.3. The heat gen-

BOXED TEXT 2.3: The Geotherm

In the case where there is no advective term and a constant heat flux, we have the equilibrium situation where $\partial T/\partial t = 0$. In this case

$$\frac{\partial^2 T}{\partial y^2} = -\frac{A}{K} \tag{2.40}$$

where A is the internal heat generation and $K/\rho c$ is the diffusivity κ. We can calculate the geotherm as long as we can provide some boundary conditions. Two boundary conditions are that the surface temperature is T_0 ($T = T_0$ at $y = 0$) and that the surface heat flow is $-Q_0$ at $y = 0$. Integrating (2.40) gives

$$\frac{\partial T}{\partial y} = -\frac{A}{K}y + c_1 \tag{2.41}$$

where c_1 is a constant of integration. Since $\partial T/\partial y = Q_0/K$ at $y = 0$, the constant c_1 must be equal to Q_0/K. Equation (2.41) after further integration therefore becomes

$$T = -\frac{A}{2K}y^2 + \frac{Q_0}{K}y + c_2 \tag{2.42}$$

Since $T = T_0$ at $y = 0$, c_2 must be equal to T_0. The temperature as a function of depth is therefore

$$T = T_0 + \frac{Q_0}{K}y - \frac{A}{2K}y^2 \tag{2.43}$$

There are instances where it would be more useful to calculate the geotherm with knowledge of the heat flow from the mantle rather than the surface heat flow (the reason for this is that the basal heat flow may be provided by a forward model of lithospheric stretching – Chapter 3). In this case, we have a pair of new boundary conditions. Firstly, $T = T_0$ at $y = 0$ and secondly, $Q = -Q_m$ at $y = y_c$, where y_c is the depth to the base of the crust, and Q_m is the basal heat flow from the mantle. Since $\partial T/\partial y = Q_m/K$ at $y = y_c$, the constant of integration c_1 in (2.41) is given by

$$c_1 = \frac{Q_m}{K} + \frac{A}{K}y_c \tag{2.44}$$

Substituting into (2.41) and integrating again,

$$T = -\frac{A}{2K}y^2 + \frac{(Q_m + Ay_c)}{K}y + c_2 \tag{2.45}$$

Since $T = T_0$ at $y = 0$, $c_2 = T_0$. Consequently, the geotherm in the region $0 < y < y_c$ is given by

$$T = T_0 + \frac{(Q_m + Ay_c)}{K}y - \frac{A}{2K}y^2 \tag{2.46}$$

Equations (2.43) and (2.46) give the geotherms where we have information of either the surface heat flow or the basal heat flow from the mantle. It would be possible to choose any depth in the lithosphere and apply the appropriate boundary conditions.

eration 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. The question is how we model the radiogenic heat generation in the continental crust. There are essentially two ways in which it can be treated (Boxed text 2.4):

- The lithosphere and sedimentary cover can be dealt with as "slabs" of constant internal heat generation;
- the internal heat generation can be modeled as decreasing exponentially with depth.

BOXED TEXT 2.4: Radiogenic Heat Production

Slab models for radiogenic heat generation

The simplest slab model is to have a single layer of radiogenic crust above a nonradiogenic mantle. The solution is given by equation (2.46) where y_c is the thickness of the crust. In a multilayer model, the same equations would be used but individual layers treated separately. For a two-layer model with a highly radiogenic upper crust, a less radiogenic lower crust and a nonradiogenic mantle (Fig. 2.14), we let

$A = A_1$ over the depth range $0 \le y < y_1$
$A = A_2$ over the depth range $y_1 \le y < y_2$
$Q = -Q_2$ at $y = y_2$
$T = T_0$ at $y = 0$

The solutions for temperature with depth are: for the depth range between $y = 0$ and $y = y_1$,

$$T = T_0 + \left\{\frac{Q_2}{K} + \frac{A_2(y_2 - y_1)}{K} + \frac{A_1 y_1}{K}\right\} y - \frac{A_1}{2K} y^2 \qquad (2.47)$$

and for the depth range between y_1 and y_2,

$$T = T_0 + \left\{\frac{Q_2}{K} + \frac{A_2 y_2}{K}\right\} y + \left\{\frac{A_1 - A_2}{2K}\right\} y_1^2 - \frac{A_2}{2K} y^2 \qquad (2.48)$$

Using this two-layer model, with $A_1 = 2.5 \times 10^{-6}\,\mathrm{W\,m^{-3}}$ in the upper crust ($y_1 = 20\,\mathrm{km}$) and $A_2 = 0.8 \times 10^{-6}\,\mathrm{W\,m^{-3}}$ in the lower crust ($y_2 = 35\,\mathrm{km}$) and the basal heat flow from the mantle $Q_2 = 30 \times 10^{-3}\,\mathrm{W\,m^{-2}}$, the geotherm is as sketched in Figure 2.14. The two-layer solution is compared to a one-layer crustal model with $A = 1.25 \times 10^{-6}\,\mathrm{W\,m^{-2}}$.

Exponential distribution of radiogenic heat generation

In an alternative model, the distribution of radiogenic heat production can be assumed to be exponential with depth (Fig. 2.14). If A_0 is the surface heat generation per unit volume, we can introduce a length scale for the exponential decrease in radiogenic heat production a_r such that the radiogenic heat generation per unit volume A at any depth is given by

$$A = A_0 \exp(-y/a_r) \qquad (2.49)$$

where a_r is between 4.5 km and 16 km for a range of continental heat flow provinces (Table 2.2, Sclater et al. 1980a).

One of the consequences of the exponential distribution of radiogenic heat is a linear relationship between the surface heat flow and the radiogenic heat content of surface rocks. This relationship has been demonstrated from many measurements in continental regions. The reason for this predicted linear relationship is as follows. The steady state solution for 1-D heat conduction (2.37) can be modified to account for an exponential radiogenic heat production to give

$$0 = K\frac{\partial^2 T}{\partial y^2} + A_0 \exp(-y/a_r) \qquad (2.50)$$

Integrating this once gives

$$c_1 = K\frac{\partial T}{\partial y} - A_0 a_r \exp(-y/a_r) \qquad (2.51)$$

Using Fourier's law (equation 2.29), (2.51) can be modified slightly,

$$c_1 = -Q - A_0 a_r \exp(-y/a_r) \qquad (2.52)$$

The constant of integration can be found by applying an appropriate boundary condition. Let us assume that as the depth becomes large ($y \to$ infinity), the heat flow is the basal heat flow from the mantle Q_m. Consequently, (2.51) shows that $c_1 = Q_m$. The heat flux at any depth is therefore

$$Q = -Q_m - A_0 a_r \exp(-y/a_r) \qquad (2.53)$$

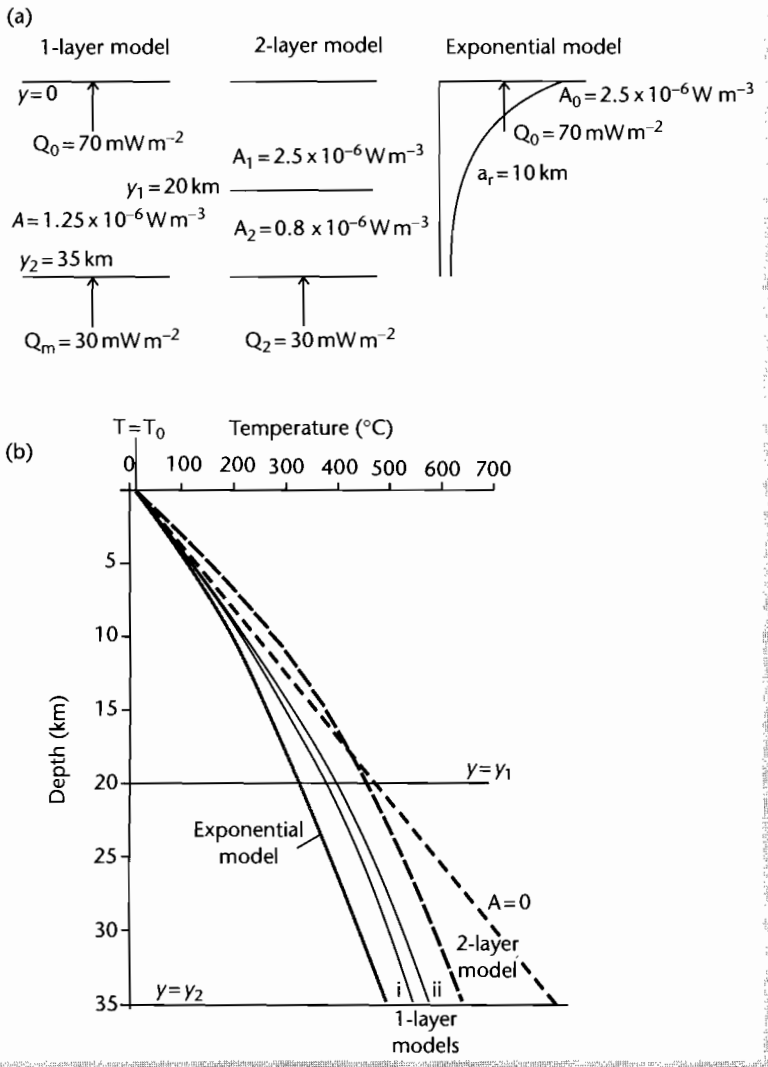

Fig. 2.14 Geotherms for the continental crust. (a) Model set-up; one-layer model with a constant internal heat generation for a 35 km-thick crust, with (i) a surface heat flow of 70 mW m^{-2}, and (ii) a basal heat flow of 30 mW m^{-2}; a two-layer model with a highly radiogenic 20 km-thick upper crust 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.

or in terms of the surface heat flow, where $Q_0 = -Q$ at $y = 0$, is

$$Q_0 = Q_m + A_0 a_r \qquad (2.54)$$

showing that there is a linear relationship between the surface heat flow and the radiogenic heat production of surface rocks, with an intercept equal to the mantle heat flow and a slope of a_r. The example shown in

Continued

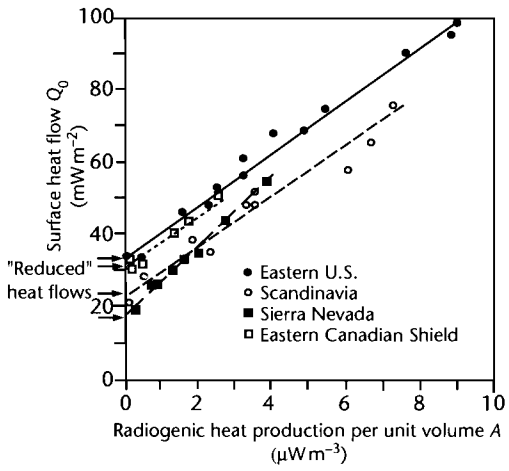

Fig. 2.15 The dependence of the surface heat flow Q_0 on the internal radiogenic heat production rate A in a number of geological provinces. The intercept on the y-axis gives the so-called "reduced" heat flow (20–30 mW m^{-2}). Derived from Roy et al. (1968), Turcotte and Schubert (2002). Reproduced courtesy of Cambridge University Press.

Figure 2.15 from granite plutons in eastern North America (Roy et al. 1968) gives a mantle heat flow, or *reduced heat flow*, Q_m of 30 mW m^{-2} and a length scale a_r of 7.5 km.

A further integration of (2.52) gives the temperature profiles at any depth. Using the boundary condition $T = T_0$ at $y = 0$,

$$T = T_0 + \frac{Q_m}{K} y + \frac{A_0 a_r^2}{K} \{1 - \exp(-y/a_r)\} \qquad (2.55)$$

or in terms of the surface heat flow, by integrating equation (2.53)

$$T = T_0 + \frac{Q_m}{K} y + \frac{(Q_0 - Q_m) a_r}{K} \{1 - \exp(-y/a_r)\} \qquad (2.56)$$

A typical geotherm with $A_0 = 2.5 \times 10^{-6}$ W m^{-3}, $Q_0 = 70$ mW m^{-3}, $K = 3$ W m^{-1} °C^{-1}, $T_0 = 10$ °C, and $a_r = 10$ km, is shown in Figure 2.14.

In some cases, it is possible that heat is conducted in more than one direction, an application common in regions of large lateral variations in surface temperatures. Such lateral variations might be due to topographic effects or to sea–land boundaries. The problem also arises 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.

2.2.4 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 rifting (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 temperature and then continues to lose heat to the cold sea water as the seafloor spreads away from the ridge. The initial cooling can be treated as instantaneous. We know from the 1-D conduction equation (equation 2.38) where $A = 0$ that

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial T^2}{\partial y^2} \qquad (2.57)$$

where $\kappa = K/\rho c$, which is the 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 length^2 time^{-1}):

$$\tau = \frac{l^2}{\kappa} \qquad (2.58)$$

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 with a velocity u (Fig. 2.16). The cooling rocks form the oceanic lithosphere and the boundary between this relatively rigid upper layer and the easily deformed mantle is an isotherm

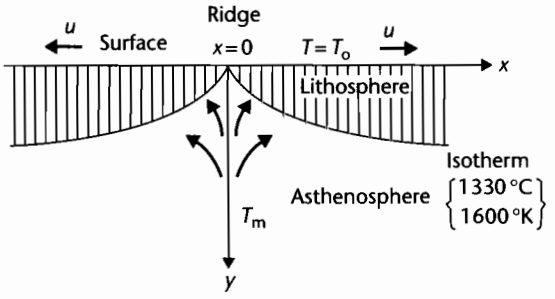

Fig. 2.16 Schematic diagram of the cooling oceanic lithosphere at a mid-ocean ridge. The oceanic plate moves away from the ridge at a velocity u. Its age is therefore determined by x/u, where x is the horizontal distance from the ridge crest.

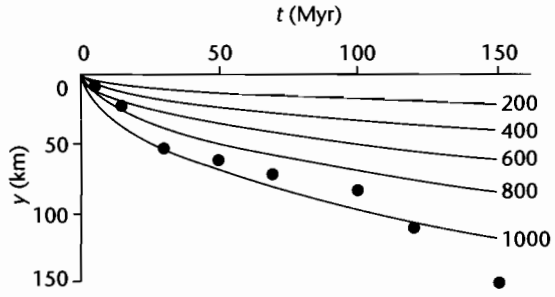

Fig. 2.17 Calculated isotherms for an oceanic lithosphere that is instantaneously cooled. The values of the isotherms are T–T$_s$ °K. The dots are the estimated thicknesses of the oceanic lithosphere in the Pacific, from Leeds et al. (1974).

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.

Using a time-dependent instantaneous cooling model, the isotherms below the seafloor as a function of age are parabolic (Fig. 2.17). The surface heat flow as a function of age is based on the 1-D conduction equation. It is compared with actual heat flow measurements of the ocean floor in Fig. 2.18 (Sclater et al. 1980a). The regions of poor correspondence between the predicted and observed surface heat flow measurements is probably due to temperature changes associated with hydrothermal circulation of sea water through the oceanic crust. These effects become less important with age as impermeable sediments blanket the ocean floor. For large ages (> 80 Myr) however, an additional heat source appears to

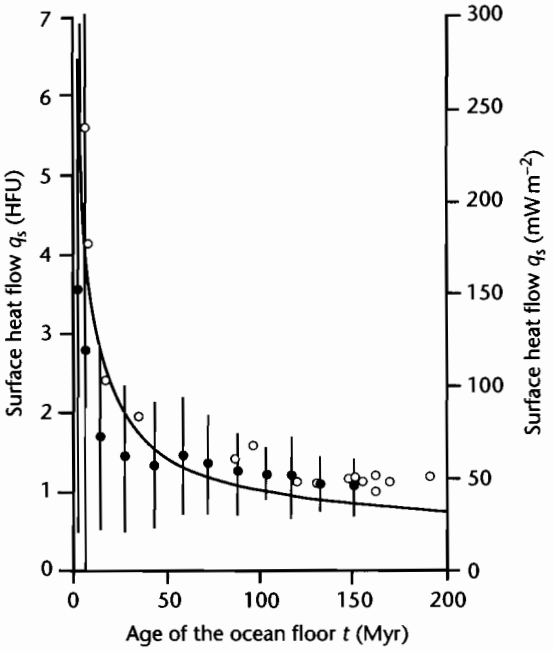

Fig. 2.18 Comparison of measured ocean heat flows (mean and standard deviation) and those predicted using the instantaneous cooling model, as a function of age. Black circles, data from Sclater et al. (1980a). Open circles, data from sediment covered regions of the Atlantic and Pacific Oceans, from Lister et al. (1990).

be recognizable, which may be mantle convection beneath the lithospheric plates.

The solution for the temperature as a function of age is given by

$$\theta = erfc\left(\frac{x}{2\sqrt{\kappa t}}\right) \tag{2.59}$$

where it can be seen that the denominator is in the form of a 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

$$\theta = \frac{T - T_a}{T_s - T_a} \tag{2.60}$$

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 sea water 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. It is a largely arbitrary choice of the definition of the boundary layer, 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 (§2.1.1) 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. 2.19). The mass per unit area in a column of any age is made up of two components: the contribution of the rock column and that of the water column. At the ridge crest, however, there is just the effect of the overlying column of mantle, $w + y_L$ thick.

The observed heat flows in the oceans, such as the Atlantic and Pacific Oceans (Parsons and 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 (Fig. 2.20). 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

$$h = h_{ridgecrest} + Ct^{1/2} \tag{2.61}$$

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 Myr old follows a similar but *inverse root age* relationship:

$$Q = 473t^{-1/2} \tag{2.62}$$

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 away from the site of spreading 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.

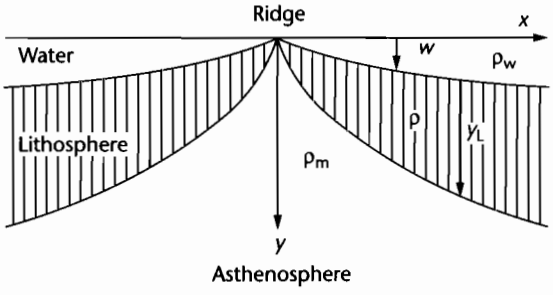

Fig. 2.19 The principle of isostasy requires the ocean to deepen with age to offset the effects of thermal contraction of the oceanic lithosphere. The water depth below the level of the ridge crest is w, the thickness of the oceanic lithosphere is y_L and ρ_m, ρ_w, and ρ are the mantle, water, and lithospheric densities respectively.

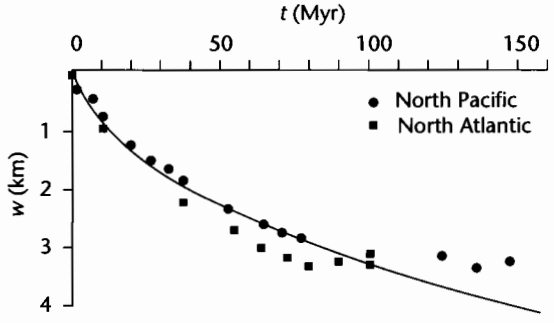

Fig. 2.20 Depth of the ocean floor below the level of the ridge crest as a function of age of the seafloor (Parsons and Sclater 1977). The solid line shows the theoretical result for an instantaneous cooling model. It is in close agreement with observations from the North Pacific and North Atlantic. The oceanic bathymetry follows a root–age relationship.

2.2.5 Thermal expansion

The laws of thermodynamics say that the equilibrium state of a material is governed by any two state variables, examples being temperature T, pressure p, and density ρ.

If a material is subjected to a change in pressure with temperature held constant, its volume will change. The change in volume for a certain pressure change is determined by the *isothermal compressibility* β. The fractional change in specific volume ($v = 1/\rho$) (volume per unit mass) with pressure at constant temperature can be written

$$\beta = -\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T = \rho\left(\frac{\partial v}{\partial p}\right)_T \tag{2.63}$$

If pressure is held constant and temperature is varied, there will also be a fractional change in specific volume. This time, the factor determining the volume change is known as the *volumetric coefficient of thermal expansion* α_v. The fractional change in specific volume with temperature at constant pressure can be written

$$\alpha_v = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_P = \rho\left(\frac{\partial v}{\partial T}\right)_P \tag{2.64}$$

β and α_v are material properties that can be found from laboratory experiments, typical values for rock being $\beta = 10^{-11}\,\mathrm{Pa}^{-1}$ and $\alpha_v = 3 \times 10^{-5}\,^{\circ}\mathrm{K}^{-1}$. β and α_v can be related since any changes in specific volume are due to changes in both temperature and pressure. The total volume change is therefore the net change resulting from the pressure change ($-V\beta \mathrm{d}p$, where $\mathrm{d}p$ is the pressure change) and from the temperature change ($+V\alpha_v\,\mathrm{d}T$, where $\mathrm{d}T$ is the temperature change). Note the opposing signs. With the material properties β and α_v given above, a 100 K change in temperature will result in a pressure change of 3 kbar (300 MPa), implying large changes in pressure and therefore stress for moderately small temperature changes. This is an important result since sedimentary basins are commonly associated with thermal disturbances (see especially Chapter 3).

Let a material have a reference density ρ^* corresponding to the temperature T^*. The difference in density between the material at its reference temperature and the same material at its new temperature T is related to the temperature difference $\mathrm{d}T = T - T^*$ through the volumetric coefficient of thermal expansion:

$$\rho - \rho^* = -\rho^*\alpha_v(T - T^*) \tag{2.65}$$

which by slight rearrangement gives

$$\rho = \rho^*(1 - \alpha_v \mathrm{d}T) \tag{2.66}$$

We could for example set the reference temperature to be 0 °C, or it could be the oceanic crust at the reference temperature of the asthenosphere, as in the mid-ocean ridge spreading problem. Since α_v for rock is typically $3 \times 10^{-5}\,\mathrm{K}^{-1}$, a 100 K temperature change produces a 0.3% change in density. Halite on the other hand has a α_v of $13 \times 10^{-5}\,\mathrm{K}^{-1}$, so a 100 K temperature change causes a 1.3% change in density. Temperature changes therefore cause large changes in stress but relatively small changes in density.

The thermal-associated strains for an isotropic body are the same along each of the principal axes of strain, and it is known from the discussion above that the magnitude of the volume change is proportional to the temperature change by α_v (i.e., $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = -1/3\ \alpha_v \mathrm{d}T$). The minus sign implies an extensional strain by convention. However, it is convenient to consider strains as linear quantities, so it is necessary to introduce a new coefficient, the *linear coefficient of thermal expansion* α_l. This coefficient relates the thermally induced strains to the temperature change, where $\alpha_l = 1/3\ \alpha_v$ (i.e., $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = -\alpha_l \mathrm{d}T$ for an isotropic material). The linear coefficient of thermal expansion is therefore the change in strain per degree of change in temperature.

2.2.6 Thermal structure of the upper mantle: effects of convection

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 is given in Chapter 5.

In the interior of a vigorously convecting fluid, the mean temperature increases with depth along an *adiabat*, so that the *adiabatic temperature gradient* is the rate of temperature increase with depth caused by compression due to the overlying rock column. 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 equations (2.14) and (2.63). 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 (Fig. 2.21). 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 680 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 680 km.

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} (Fig. 2.22). There is an important feature underlying Fig. 2.22, namely that the continental lithosphere as a thermal entity extends to about 200 km below the surface (§1.2).

It has been suggested that mantle convection takes place across the 400 km temperature and density discontinuity, essentially because subducted lithosphere appears to descend smoothly through this discontinuity. The influence of the 680 km discontinuity is less certain. Deep focus earthquakes provide evidence that convection takes place to 680 km but there is no direct evidence below this depth. Two contrasting models of mantle convection are therefore possible, one involving convection of the whole mantle, and one involving two major convecting systems, one shallow system linked to plate tectonics and a deeper system extending to the outer margin of the core. Smaller convecting cells may be superimposed on these major systems.

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

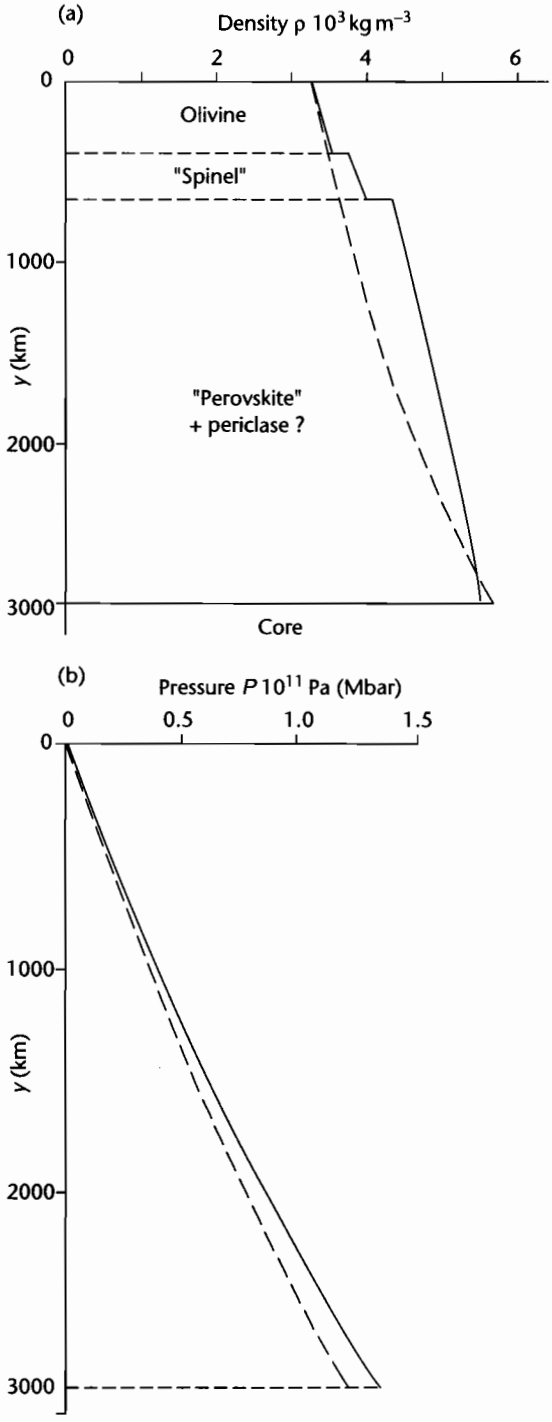

Fig. 2.21 Depth profiles of density (a) and pressure (b) in the mantle. Observed values are shown in solid lines; values calculated for a purely adiabatic behavior are shown in dashed lines.

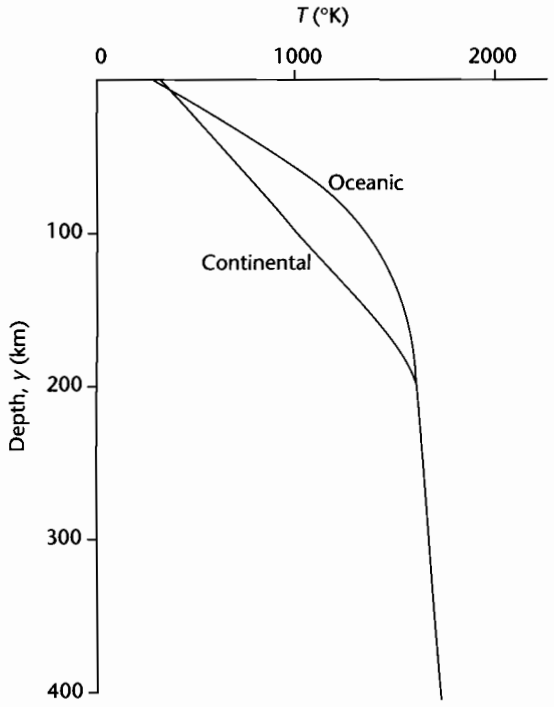

Fig. 2.22 Representative oceanic and continental geotherms in the shallow upper mantle.

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 (more details in §5.1.2).

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 more dense 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.

2.2.7 Mantle viscosity

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.) The problem of direct relevance to sedimentary basin analysis is the viscous flow in the mantle. One way in which the viscosity of the mantle can be estimated is by studying its response to loading and unloading (see also Chapter 4).

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 which 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

$$w = w_m \exp(-t/\tau_r) \tag{2.67}$$

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

$$\tau_r = (4\pi\mu)/(\rho g\lambda) \tag{2.68}$$

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.

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. For example, the northern shore of Lake Superior is rising relative to the south by as much as 0.46 m per century following the Canadian glaciation (Kite 1972). The Baltic region of Scandinavia is also renowned for its raised beaches. After correcting the 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. 2.23). 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. A summary of the possible distribution of viscosity in the mantle (Cathles 1975) is given in Table 2.5.

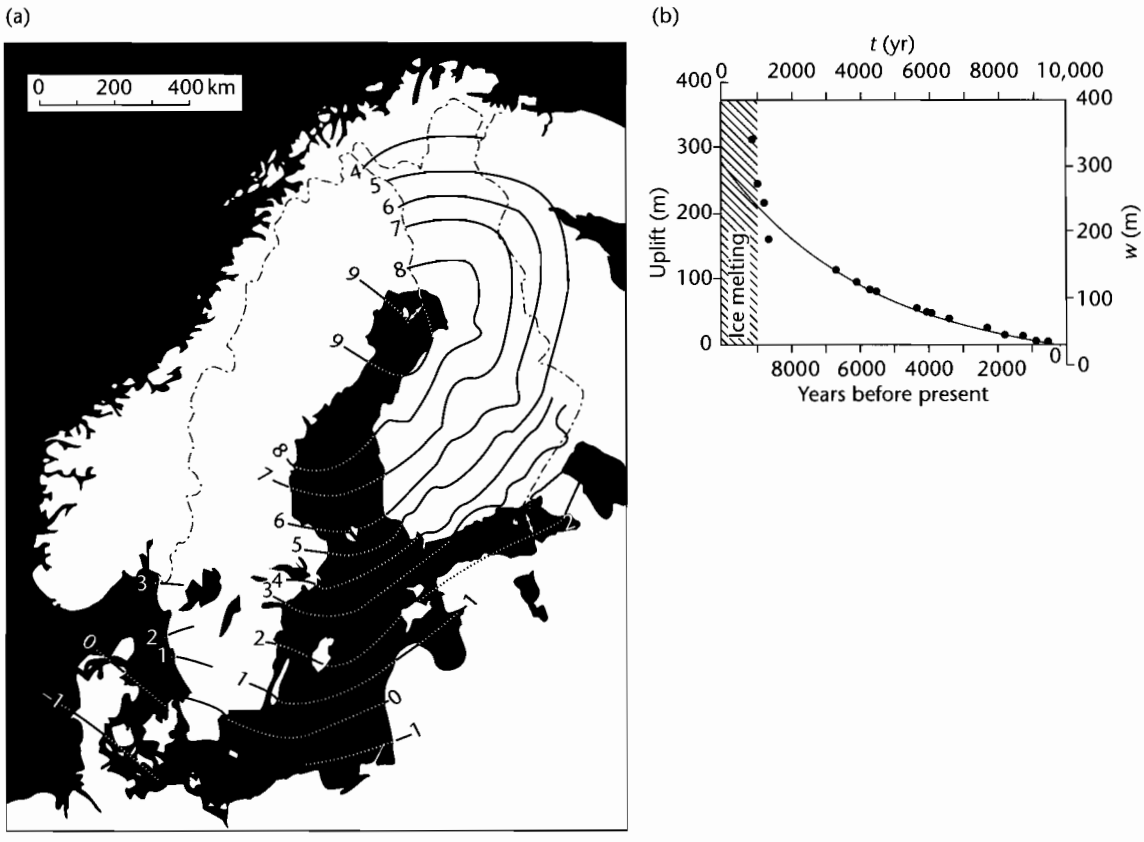

Fig. 2.23 Postglacial rebound of Scandinavia. (a) Present day rates of uplift occurring in Scandinavia (after Flint 1971), reproduced courtesy of John Wiley & Sons; (b) Postglacial uplift of the mouth of the Angerman River, Sweden over the last 10 000 years compared with the exponential relation in equation (2.67) with a constant viscosity of 10^{21} Pa s (after Turcotte and Schubert 2002).

Table 2.5 Viscosity of the mantle from glacial rebound studies (Cathles 1975).

		Dynamic or absolute viscosity	
Region	**Depth (km)**	**(Pa s)**	**(Poise)**
Lithosphere	0–100	Elastic	
Asthenosphere	100–175	4×10^{19}	4×10^{20}
Mesosphere	175–2848	10^{21}	10^{22}

2.3 GRAVITY AND ISOSTASY

2.3.1 Gravity anomalies and the geoid

The reference gravitational field of the Earth is that of an *oblate spheroid*. Topographic effects and density differences within the solid Earth cause local variations in the gravity measured at the Earth's surface. These are known as *gravity anomalies*. Topography causes a positive anomaly because of the mass of rock elevated above the oblate spheroid. However, the fact that the continents do not have positive gravity anomalies suggests that the influence of topography may be compensated by other effects. High topography, for example, is commonly associated with a deep crustal root. Crustal roots cause negative gravity anomalies because of the presence of a density deficit at depth. At large wavelengths, if the topography and root are perfectly balanced in hydrostatic equilibrium, there is a zero surface gravity anomaly.

From Newton's law of gravitation we know that the gravitational acceleration of a point outside a spherical mass such as the Earth is directed radially inwards and is given by

$$g_m = \frac{GM}{r^2} \tag{2.69}$$

where G is the universal gravitational constant (equal to $6.673 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$), M is the mass of the Earth (73.483×10^{21} kg) and r is the radial distance from the center of mass to the point of interest.

A mass within or outside the Earth has a *gravitational potential energy*, which is the work done in bringing the mass from infinity to its position in the gravitational field. The gravitational potential energy per unit mass is the *gravitational potential*. Gravitational equipotential surfaces are where the gravitational potential is everywhere equal. In the absence of barometric and dynamic effects in the ocean–atmosphere system, the sea surface is a gravitational equipotential surface. This reference equipotential surface is known as the *geoid*. Over the ocean the geoid is the mean sea level. Over the continents the geoid is more difficult to recognize, but would be the water elevation in imaginary canals cut through the continents (Fowler 1990, p. 165). The geoid is close to being a spherical surface, with a certain eccentricity (flattening). Small departures from the reference geoid are called *geoid anomalies*. The largest geoid anomaly is about 100 m – small compared with the 21 km difference between the polar and equatorial radii of the reference geoid. A map of geoid anomalies (Fig. 2.24) (Lemoine et al. 1998) shows that some of them can be attributed to density variations related to plate tectonics. For example, geoid highs occur over some subduction zones where dense, cold lithosphere penetrates hot upper mantle (Chile–Peru).

Surface gravity anomalies are caused by mass anomalies on the surface of the Earth or within the solid Earth. From equation (2.69) it can be seen that the gravitational attraction scales on the mass and the square of the distance. Consequently, a mass anomaly of a given magnitude causes a much greater surface gravity anomaly if it is buried at a shallow depth in the upper crust than if it occurs in the upper mantle. Surface gravity anomalies are commonly produced by the gravitational effects of the mass represented by topography. The surface gravity anomaly at a point is given by the *Bouguer gravity formula*

$$g_y = 2\pi G \int_0^h \rho(y)\,\mathrm{d}y \tag{2.70}$$

where g_y is the surface gravity anomaly at a point, G is the universal gravitational constant ($6.673 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$) and the integral gives the mass excess or deficiency below that point, h is the height of the body of anomalous density and y is the vertical coordinate. For the case of topography, h is the height of the topography and $\rho(y)$ is the crustal density ρ_c. The Bouguer gravity formula therefore becomes

$$\Delta g = 2\pi\rho_c G h \tag{2.71}$$

where Δg is the gravity anomaly due to topography. From equation (2.71) it can be seen that if crustal density is $2750\,\mathrm{kg\,m^{-3}}$, a 5 km high topography produces a gravity anomaly of $5.6\,\mathrm{mm\,s^{-2}}$. A more convenient unit of measurement is the *milligal*, equal to $10^{-5}\,\mathrm{m\,s^{-2}}$. Another unit is the *gravity unit*, which is $10^{-6}\,\mathrm{m\,s^{-2}}$. The 5 km high topography therefore produces a gravity anomaly of 560 milligal. Alternatively, if a 5 km high volcanic seamount

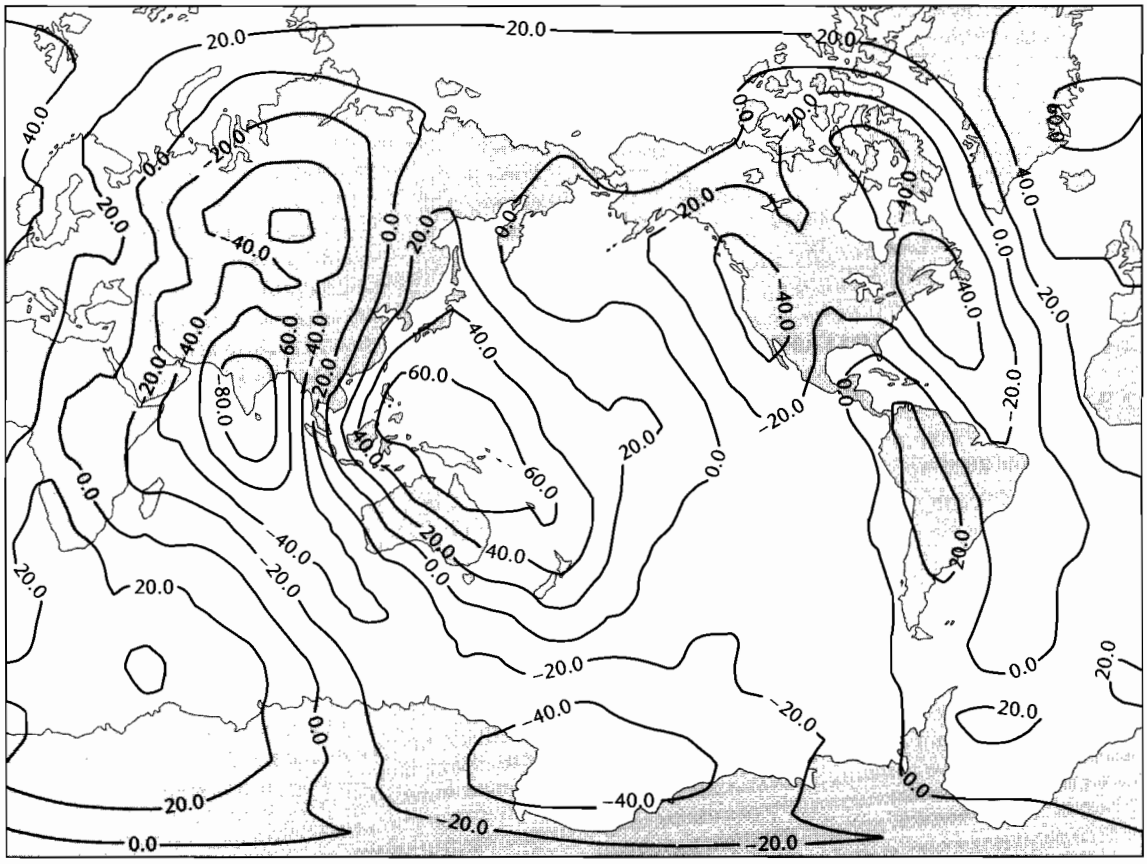

Fig. 2.24 The geoid height above the reference ellipsoid (Lemoine et al. 1998). Reproduced courtesy of NASA.

of density 2900 kg m^{-3} replaced sea water with density 1000 kg m^{-3}, the gravity anomaly would be 4.0 mm s^{-2}, or 400 milligal.

A number of corrections are applied to surface gravity measurements in order to obtain the surface gravity anomaly.

- The reference gravity field is firstly subtracted, which contains a *latitude correction*;
- the variation of gravity with height is then subtracted (*elevation or free-air correction*). This correction is 3.07 mm s^{-2} for an elevation of 1 km. The resulting anomaly after this second step is called a free-air gravity anomaly;
- the gravitational attraction of local (short-wavelength) topography is removed with the Bouguer gravity formula to give the *Bouguer gravity anomaly*. If the topography is particularly steep, a *terrain correction* is also applied.

The Bouguer gravity formula does not account for any compensating changes in mass at depth, such as a light root beneath a mountain chain. Therefore Bouguer anomalies over mountain ranges are strongly negative. Since the mass of the topography is compensated over a long wavelength by the presence of a light root, so that there is no net mass difference beneath the observer, the free-air anomaly over mountain belts should be roughly zero (see also §2.3.3 and Fig. 2.28).

2.3.2 Models of compensation

The simplest way in which the mass excess of topography can be compensated is by the development of a deep, low-density root (Fig. 2.1c). Compensation for topography by changes in the thickness of the underlying crust

(Fig. 2.25a) is termed *Airy compensation* or *Airy isostasy*. From the principle of hydrostatic equilibrium, we can balance the pressure at the base of a lithospheric column through the mountain belt with that at the base of a column through adjacent lowlands at sea level. This gives

$$b = \frac{\rho_c h}{\rho_m - \rho_c} \tag{2.72}$$

where b is the thickness of the root, h is the height of the mountain belt, and ρ_m and ρ_c are the mantle and crustal densities respectively. If ρ_c is 2800 kg m^{-3} and ρ_m is 3300 kg m^{-3}, a 5-km high mountain would require a root thickness of $b = 28$ km. If the topography is negative (h is negative), representing a marine basin, the balancing of lithospheric columns gives

$$b = \left(\frac{\rho_c - \rho_w}{\rho_m - \rho_c}\right) h \tag{2.73}$$

where ρ_w is the density of ocean water. Using the same density terms as above and ρ_w is 1000 kg m^{-3}, the "anti-root" beneath a marine basin of depth 1 km is 3.6 km. The crust must be thinned by 4.6 km under the marine basin compared to the coastal plain at sea level.

An alternative model for isostatic compensation invokes lateral density variations over a depth range to a depth of compensation W (Fig. 2.25b). This model is known as *Pratt compensation*. The variable density ρ_p of lithospheric columns elevated at h is given by

$$\rho_p = \rho_0 \left(\frac{W}{W + h}\right) \tag{2.74}$$

where ρ_0 is the reference density for a lithospheric column with its surface at sea level. For a mountain elevated at 5 km, and taking the depth of compensation W as 125 km, requires a reduction of density beneath the mountain to 96% of ρ_0. Although this is a relatively small density change, there seems little evidence to support its existence. For negative topography (h is negative), the variable density of lithospheric columns below sea level is given by

$$\rho_p = \frac{\rho_0 W + \rho_w h}{W + h} \tag{2.75}$$

The Pratt model of compensation should be suitable where there are lateral density variations at depth, as for instance caused by the presence of plume-related swells. The Hawaiian swell (Crough 1978) reaches water depths of 4 km at its crest (seamounts of course reach sea level). The swell is situated in an abyssal plain at −6 km. What density change is required to elevate the Hawaiian swell 2 km above the adjacent abyssal plain? Putting $\rho_0 =$ 3300 kg m^{-3} under the abyssal plain and $W = 150$ km, the density of the column under the swell crest is 3281 kg m^{-3}, representing a density reduction of 0.6% of the reference density for oceanic rock under the abyssal plain. Since the volumetric coefficient of thermal expansion of oceanic lithosphere is *c.* 3×10^{-5} K^{-1}, this is equivalent to an average heating of the section down to 120 km by *c.* 190 °C. This is a reasonable result since plumes

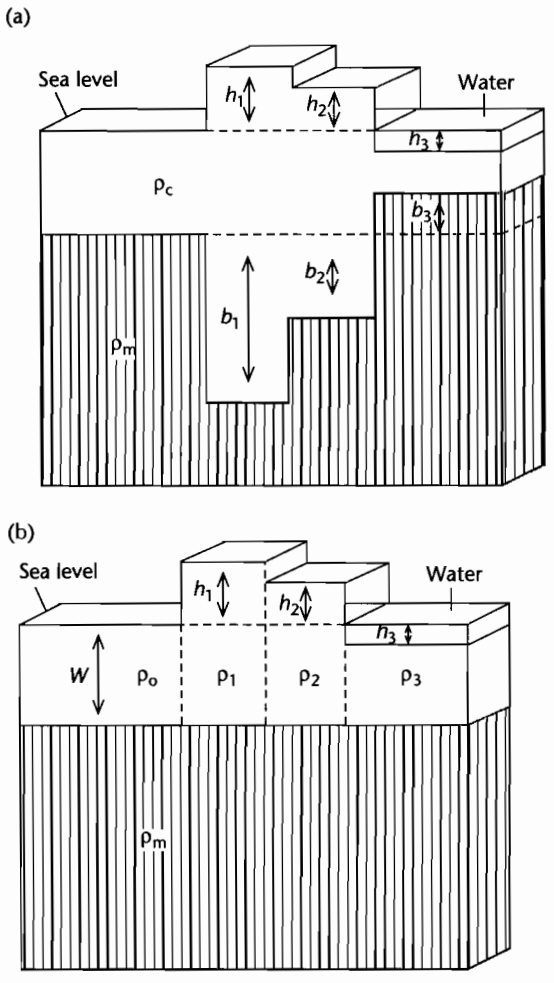

Fig. 2.25 Airy (a) and Pratt (b) models of isostatic compensation. Notation explained in text.

are known to cause excess temperatures of 100–200 °C above the normal asthenospheric temperature.

Whereas the Airy model is a good approximation of isostatic compensation in areas of active tectonics, such as mountain building and sedimentary basin formation, the Pratt model is a good predictor where lateral density variations are caused by temperature changes. Both models make use of the idea of hydrostatic equilibrium, envisaging notional lithospheric blocks able to float freely on a fluid asthenosphere. This in turn implies that the lithosphere is very weak and responds locally to any loads. An alternative model involving a stronger lithosphere able to store elastic stresses for considerable periods of time makes use of the concept of flexural rigidity developed in §2.1.4. In *flexural isostasy*, loads are compensated regionally by a lithosphere approximating an elastic sheet overlying a fluid substratum.

2.3.3 Flexural isostasy

The problem of flexural isostasy can be tackled by considering the way the lithosphere supports periodic topography. This simplified analysis can then be extended to the case of a sedimentary basin of a certain wavelength and amplitude.

BOXED TEXT 2.5: Flexural Isostasy

We start by assuming that topography is represented by a sinusoidal periodic load (Fig. 2.26) given by:

$$h = h_0 \sin(2\pi x/\lambda) \qquad (2.76)$$

where h_0 is the maximum height of the topography and λ is its wavelength. The load actually exerted on the lithosphere has the form of pressure or stress of the form ρgh, so the distribution of pressure under this periodic topography is

$$q_a(x) = \rho_s g h_0 \sin(2\pi x/\lambda) \qquad (2.77)$$

where ρ_s is the density of sedimentary rocks constituting the load. Assuming no horizontal applied forces (i.e., $P = 0$), and returning to the general flexural equation (eqn 2.28), we have

$$D\frac{d^4w}{dx^4} + (\rho_m - \rho_s)gw = \rho_s g h_0 \sin(2\pi x/\lambda) \qquad (2.78)$$

Since the load is sinusoidal and periodic, the deflection of the lithosphere will also be sinusoidal and periodic. The solution for the deflection w can therefore be assumed to be of the type

$$w = w_0 \sin(2\pi x/\lambda) \qquad (2.79)$$

where w_0 is the maximum deflection. Substituting (2.79) into (2.78) and rearranging, the amplitude of the deflection of the lithosphere is

$$w_0 = \frac{h_0}{\left\{D\dfrac{d^4w}{dx^4}\dfrac{1}{\rho_s g w} + \dfrac{\rho_m}{\rho_s} - 1\right\}} \qquad (2.80)$$

The fourth differential of w from (2.79) is $(2\pi/\lambda)^4 w$, so (2.80) simplifies to

$$w_0 = \frac{h_0}{\dfrac{\rho_m}{\rho_s} - 1 + \dfrac{D}{\rho_s g}\left(\dfrac{2\pi}{\lambda}\right)^4} \qquad (2.81)$$

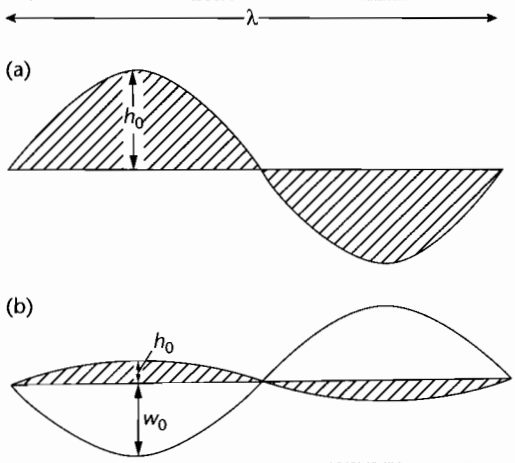

Fig. 2.26 Deflection of the lithosphere under a periodic (sinusoidal) load. In (a) the wavelength of the load is short and there is no deflection of the lithosphere. in (b) the wavelength of the load is long, leading to isostatic compensation of the load by a deflection of the lithosphere. w_0 is the maximum deflection, h_0 is the maximum elevation of the load (after Turcotte and Schubert 1982).

If the wavelength of the load is short, the deflection is very small compared to the maximum height of the load ($w_0 < h_0$). The lithosphere therefore appears to behave very rigidly to loads of this scale. If, however, the wavelength of the load is long, the deflection can be written

$$w = w_{0\infty} = \frac{\rho_s h_0}{(\rho_m - \rho_s)} \qquad (2.82)$$

which is the result obtained for a purely vertical isostatic balance (Airy isostasy) (§2.1). This means that for sufficiently long wavelengths, the lithosphere appears to have no rigidity and periodic loads should be in hydrostatic equilibrium. It is clearly desirable to be able to predict where we are in this range of complete to no compensation. The degree of compensation C of the load (Fig. 2.27) is the ratio of the actual deflection compared with the maximum (Airy) or hydrostatic deflection

$$C = \frac{w_0}{w_{0\infty}} \qquad (2.83)$$

Substituting (2.81) and (2.82) into (2.83)

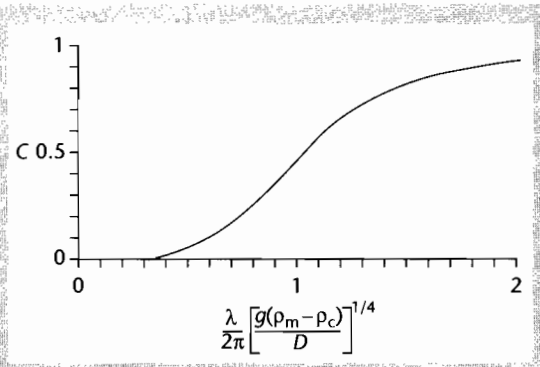

Fig. 2.27 Dependence of the degree of compensation C, on the nondimensional wavelength of periodic (sinusoidal) topography (after Turcotte and Schubert 1982). D is the flexural rigidity, λ is the wavelength of the load, ρ_m and ρ_c are the mantle and crustal densities respectively and g is the gravitational acceleration.

$$C = \frac{(\rho_m - \rho_s)}{\rho_m - \rho_s + \frac{D}{g}\left(\frac{2\pi}{\lambda}\right)^4} \qquad (2.84)$$

$(2\pi/\lambda)$ is sometimes termed the *wave number* (e.g., Watts 1988).

Gravity measurements can be used to test the state of isostasy in a region. Imagine a mountain elevated with respect to a coastal plain at sea level (Bott 1982) (Fig. 2.28). In the case of Airy isostasy, the Bouguer anomaly will be strongly negative over the mountain belt because there is a mass deficiency below sea level beneath the mountain belt. The Bouguer correction has removed the gravitational effect of the mountainous topography. However, the free-air anomaly should be weakly positive, but much smaller in magnitude than the Bouguer anomaly. This is because the gravity effect, measured at the surface, of the crustal rocks elevated above sea level in the mountain belt is stronger than the gravity effect of the more distant compensating root. When the structure is very wide compared to the compensation depth, the free-air anomaly reduces to zero.

If the imaginary mountain belt is only partially compensated, the Bouguer anomaly is reduced but remains negative and the free-air anomaly becomes more strongly positive. Where the mountains are completely uncompensated, that is, they rest on an infinitely rigid lithosphere, the Bouguer anomaly is zero, but the free-air anomaly is strongly positive.

The very common negative Bouguer anomalies associated with mountain belts suggests that they are compensated to some degree, depending on the flexural rigidity of the lithosphere. 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 (Fig. 2.29) (Dorman and 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. The relationship between Bouguer gravity anomaly and topography as a function of its wavelength is called *admittance*. From Fig. 2.29, 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

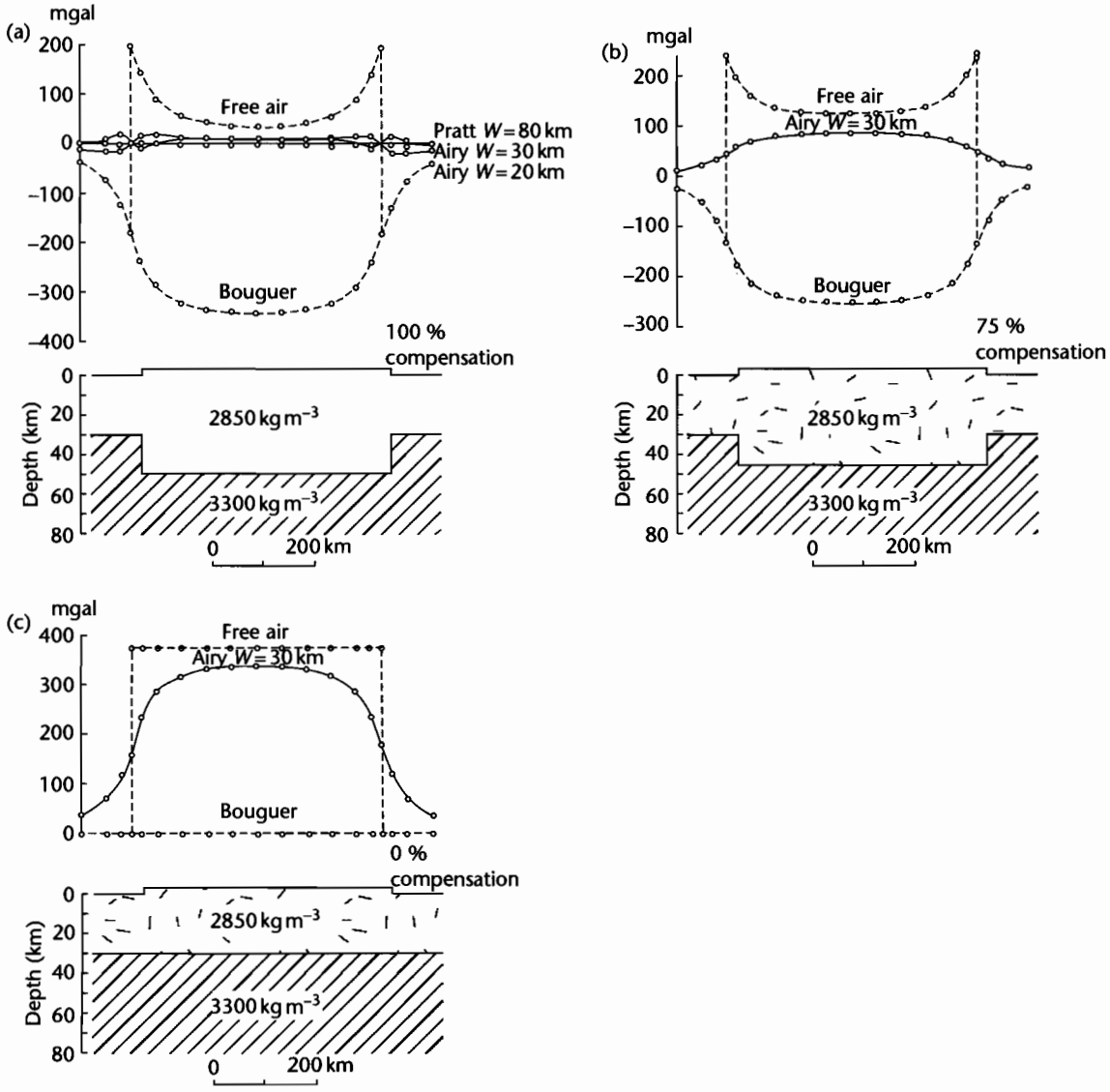

Fig. 2.28 Gravity anomalies over a schematic mountain range (Bott 1982 in Fowler 1990). In (a) the mountain range is 100% compensated (Airy isostasy); in (b) it is 75% compensated (Airy,) and in (c) the mountain range is uncompensated. Solid lines: isostatic anomalies calculated for the different density models shown using Pratt compensation depth $W = 80$ km, Airy compensation depths $W = 20$ km and 30 km. Dashed lines, free-air and Bouguer anomalies that would be measured over the mountain range. Reproduced courtesy of Elsevier.

Fourier transform (Forsyth 1985). The *coherence* function is the square of the correlation coefficient between the topographic and gravity signals. The coherence indicates the degree of compensation for any topographic loads or density contrasts. It can be thought of as the fraction of the gravity field that can be predicted from the topography using a given isostatic response function. A coherence of 1 indicates complete compensation on a plate

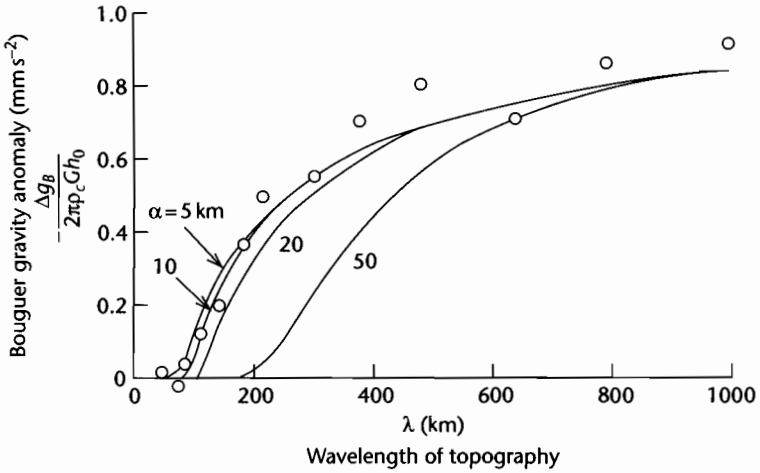

Fig. 2.29 Correlation of Bouguer gravity anomalies with topography (admittance) for the United States (Dorman and Lewis 1972), compared with the gravity formula for loading by periodic topography for different values of the flexural parameter $\alpha = [4D/(\rho_m - \rho_c)g]^{1/4}$ in km. The best agreement is with $\alpha \cong 20$ km, or flexural rigidity $D \cong 10^{21}$ N m, which with $E = 60$ GPa and $\nu = 0.25$, gives an equivalent elastic thickness Te of 6 km. Reproduced courtesy of Cambridge University Press.

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 roll-over 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), so there is a close correspondence between the seismogenic crust, the thickness of the strong elastic layer and the effective elastic thickness (§4.3).

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 (§9.3.4). 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 kind of 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.

2.4 ROCK RHEOLOGY

2.4.1 Fundamentals

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. On the other hand, the widespread occurrence of folds suggests that rocks can also behave in a *ductile* manner. Three parameters are important in determining the brittle to ductile transition (Paterson 1958; Heard 1960): pressure, temperature, and strain rate.

Within the crust, rocks commonly deform by brittle fracture, but may also exhibit ductile behavior under stress by *pressure solution creep* (Rutter 1976, 1983) (§2.4.3). Dissolution of minerals in zones of high pressure and their precipitation in areas of low pressure causes creep even at low temperatures and pressures. Otherwise at low temperatures and pressures rocks may deform in a brittle manner by fracturing.

Mantle convection is thought to be attributed to two processes that allow rocks to deform by thermally activated creep: (a) *diffusion* processes operate at very low stress levels, in which the crystalline solid behaves as a Newtonian fluid with a viscosity that depends exponentially on pressure and the inverse absolute temperature; (b) motion of *dislocations* is more effective at higher stress levels. It is a non-Newtonian fluid behavior with the same exponential pressure and inverse temperature dependence. Creep also takes place in the lower lithosphere where it can relax elastic stresses. In this case the rheology is a combination of an elastic and viscous behavior – a *viscoelastic* rheology (§2.4.4).

The viscosities of both diffusion creep and dislocation creep are temperature dependent. If dislocation creep is the dominant mechanism of flow in the mantle, then the effective viscosity of the mantle will be *stress dependent* as well as temperature dependent. This dependence of effective viscosity on stress is well illustrated by the flow of a power-law rheology fluid in a channel (Fig. 2.30).

For a power-law fluid the strain rate or velocity gradient is proportional to the power n of the stress. 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 center of the flow. This must be true because

$$\mu_{eff} = \tau \Big/ \frac{du}{dy} \tag{2.85}$$

i.e., the viscosity is the coefficient linking the shear 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 cooler boundary at the base of the rigid lithosphere. Shear in the asthenopheric 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.

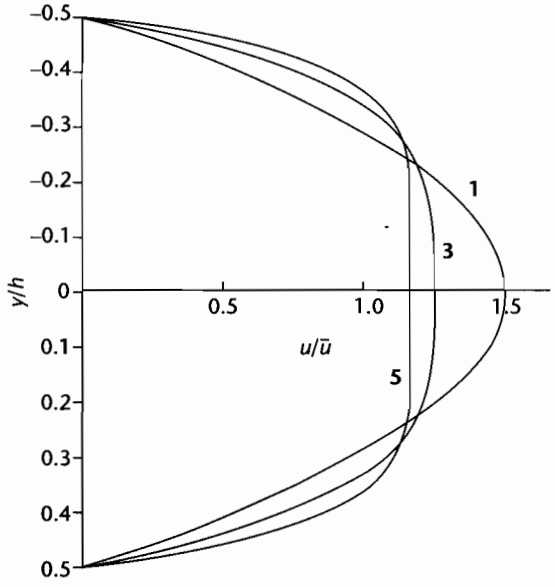

Fig. 2.30 Velocity profiles in a channel of thickness h for power-law fluid rheologies with $n = 1$ (Newtonian), $n = 3$, and $n = 5$. Distance from the channel wall is expressed by y/h. The velocity u is scaled by the average velocity $\bar{u}$.

2.4.2 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 (e.g., Orowan 1967). Alternatively, the mantle may act as a power-law fluid, in which case the mantle's viscosity would be the stress-dependent effective viscosity associated with dislocation creep.

Most views of mantle rheology, however, 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

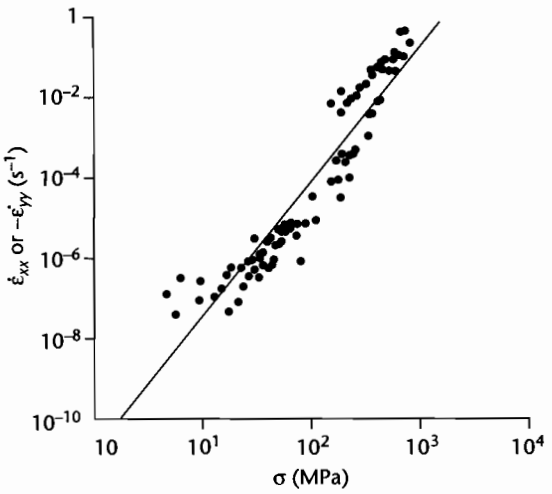

Fig. 2.31 Observed dependence of strain rate on stress for olivine at a temperature of 1400 °C (after Ashby and Verall 1977). Dry olivine obeys an approximate cubic power law rheology in these laboratory experiments.

power-law rheology reasonably well (Ashby and Verall 1977) (Fig. 2.31). Other rocks deform in a nonlinear 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). The theoretical formula connecting strain rate and stress in dislocation creep closely matches the empirical relation for dry olivine in the laboratory, suggesting that dislocation creep is the dominant deformation mechanism. Although the laboratory experiments were at strain rates several orders of magnitude greater than mantle strain rates, the dislocation creep process is also thought to be dominant in the mantle. Only at exceptionally low strain rates, considerably smaller than those expected in the mantle, would diffusion creep become dominant.

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.

2.4.3 Rheology of the 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

$$\frac{F}{N} = f = \tan\phi \tag{2.86}$$

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

$$F = aN + b \tag{2.87}$$

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

$$\tau_c = C + \sigma\tan\phi \tag{2.88}$$

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 0.7–1.7 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 kbar), C is 50 MPa (0.5 kbar) and f is *c.* 0.6–0.85 for a wide variety of rock types (Fig. 2.32). 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 7–8 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 15–20 km

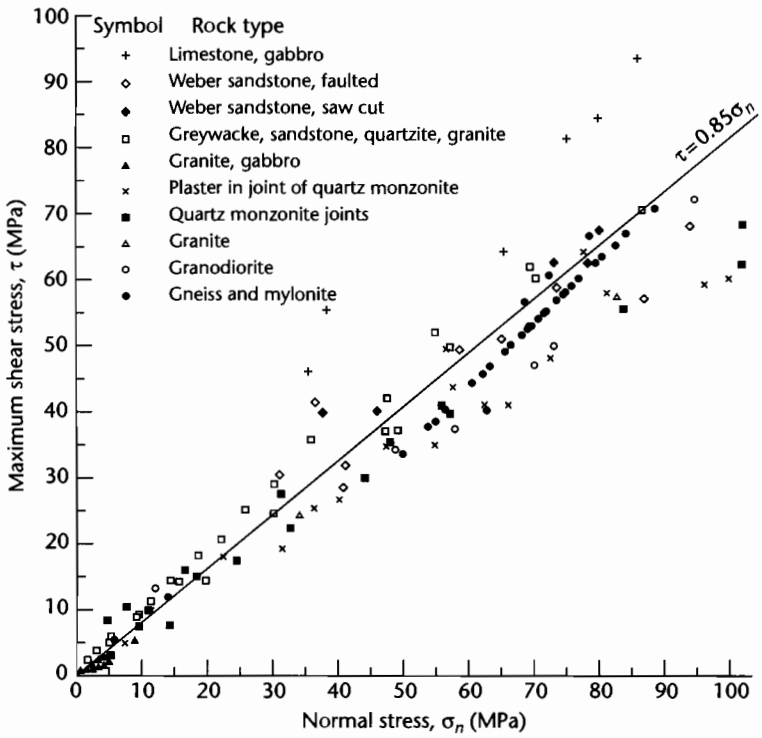

Fig. 2.32 Byerlee's law for the maximum shear stress τ to initiate sliding as a function of the normal stress σ_n, for a variety of rock types. The linear fit defines a maximum coefficient of static friction of 0.85 (Byerlee 1977). Reproduced courtesy of Cambridge University Press.

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 discussed in §2.3.1. 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 sand 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 (§9.2.1).

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. Seismic refraction velocities have been used, for example, to interpret an amphibolite facies upper crust, granulite facies middle crust, and ultramafic lower crust in NW Britain (Hall and Al-Haddad 1976; Bamford 1979). Other authors have suggested much higher quartz contents, with significant proportions of acid and intermediate rocks metamorphosed to granulite and amphibolite facies in the lower crust (Fountain and Salisbury 1981). The layering of the continental crust

affects its strength as a function of depth (§2.4.6). 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 characterized by discontinuous frictional (brittle) faulting passes below *c.* 15 km (*c.* 300 °C) depth into an aseismic continuous quasi-plastic region where dislocation 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 and Watterson 1980). The vertical zonation of deformation mechanism envisaged by Sibson (1983, p. 744) is shown in Fig. 2.33.

2.4.4 Viscoelasticity

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 $\dot{\varepsilon}_v$. The elastic strain of a material under a uniaxial stress is

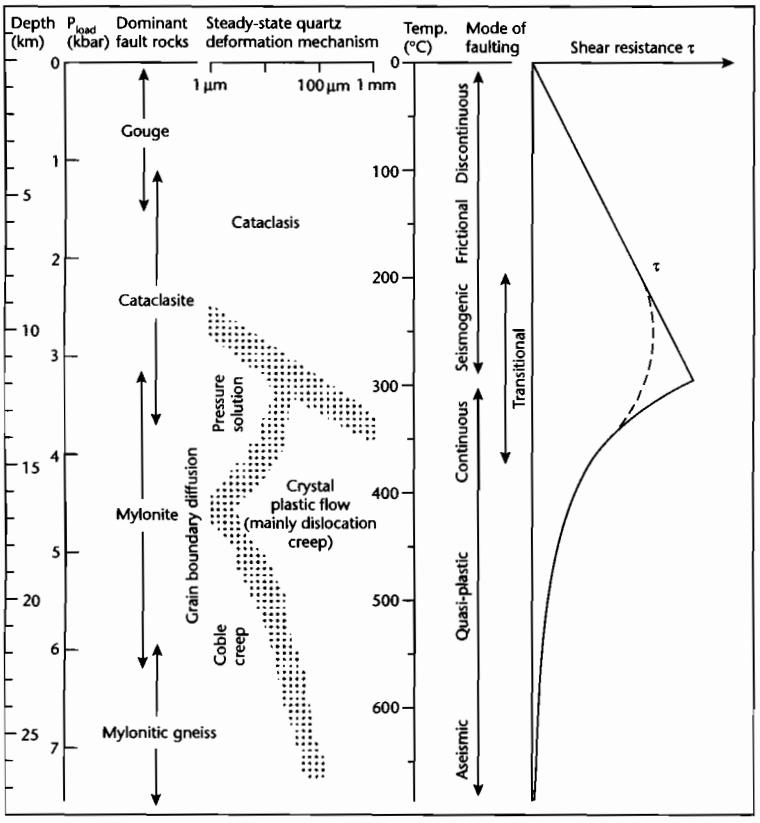

Fig. 2.33 Sibson's (1983) conceptual model for a major fault zone in the continental crust showing the dominant deformation mechanisms.

$$\varepsilon_e = \sigma/E \tag{2.89}$$

as will be recalled from Hooke's law (equation (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 $\partial u/\partial x$. The effective viscosity μ is taken as the ratio between stress and twice the strain rate. Therefore

$$\frac{d\varepsilon_v}{dt} = -\frac{\partial u}{\partial x} = \frac{\sigma}{2\mu} \tag{2.90}$$

(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

$$\frac{d\varepsilon}{dt} = \frac{1}{2\mu}\sigma + \frac{1}{E}\frac{d\sigma}{dt} \tag{2.91}$$

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 (2.91) 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 and Beaumont 1984).

2.4.5 Elastic–perfectly plastic rheology

When confining pressures approach a rock's brittle strength a transition takes place from brittle (elastic) behavior to plastic behavior (Fig. 2.34a). 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. 2.34b). If the deformation continues indefinitely without any addition of stress above σ_0, the material is said to exhibit an *elastic–perfectly plastic* behavior (Fig. 2.34c). Laboratory studies of the mantle rock dunite show that it conforms closely to the elastic–perfectly plastic rheology (Griggs 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 above discussion can be applied to the bending of a plate. If the plate is flexed, a yield stress in the plate may be attained, marking the onset of plasticity. This onset corresponds to a characteristic plate curvature and has associated with it a characteristic bending moment. Consider the situation where the bending moment exceeds the critical value for the onset of plasticity. We have previously seen that the longitudinal strains in a flexed plate are large in its outer parts but small in the interior. As a result, the outer regions may deform plastically while the interior core deforms elastically. After the bending moment for the onset of plasticity is attained there is a rapid increase in plate curvature due to a process known as *plastic hinging*. The plate then bends further, up to a maximum bending moment that is 1.5 times the critical moment for the onset of plasticity (Turcotte and Schubert 1982, p. 344). The observed profile across the Tonga Trench, which shows a curvature far in excess of that predicted of a purely elastic plate, may be due to this process of plastic hinging. A similar process is discussed in §4.3 in relation to the bending of continental plates in mountain belts.

2.4.6 Strength profiles of the lithosphere

Lithospheric strength is controlled by lithospheric rheology, which itself is dependent on the heat flow or geothermal gradient, and the composition of lithospheric rocks. 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 and Morgan 1987).

The lithosphere can deform by brittle (faulting) and 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

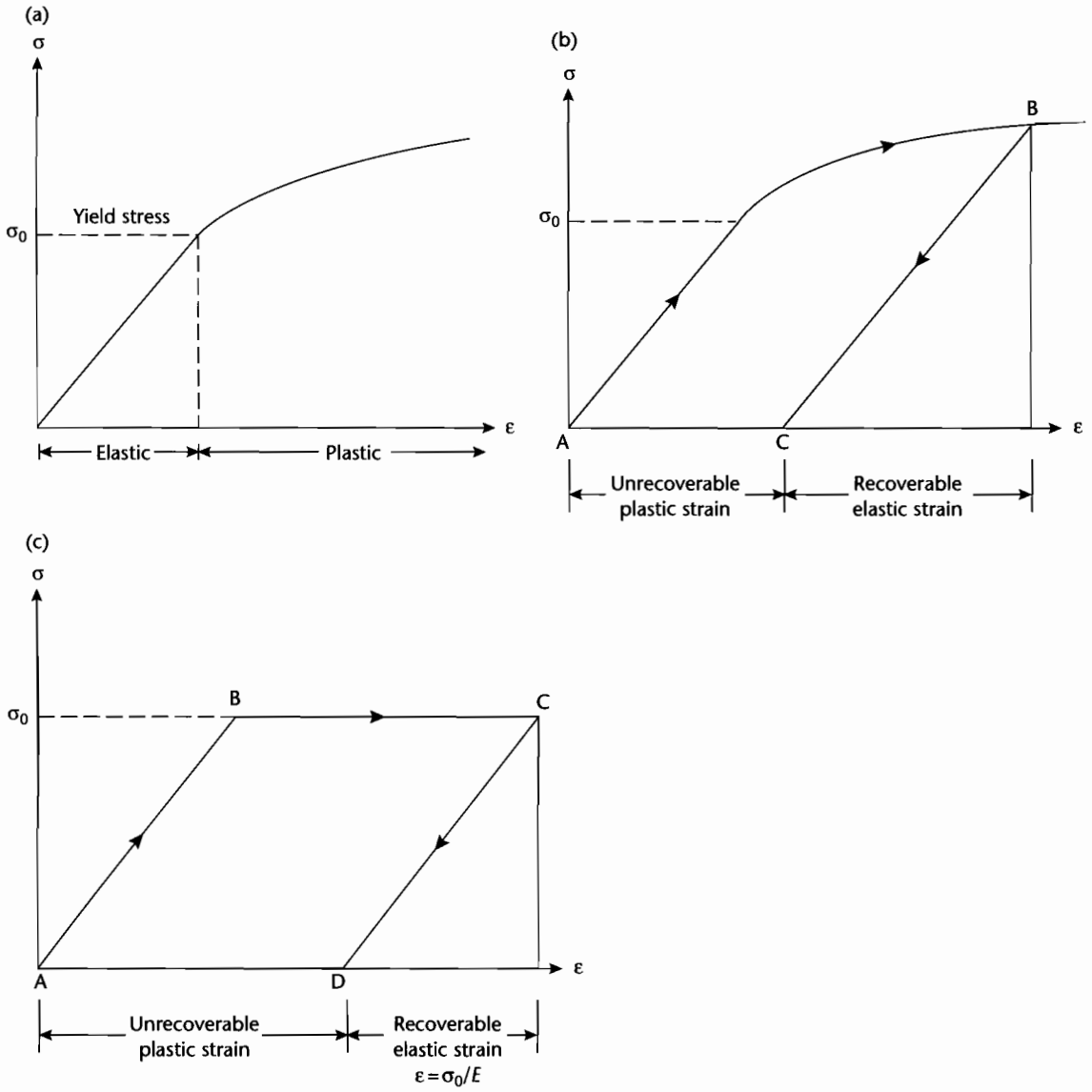

Fig. 2.34 Stress-strain trajectories for (a) a solid showing a transformation from elastic to plastic behavior; (b) loading and unloading of an elastic–plastic material. Unloading of the material once it has passed into the plastic field results in an unrecoverable deformation, or plastic strain; (c) a material with an elastic–perfectly plastic rheology. Plastic strain continues indefinitely without any addition of stress above the yield stress.

strengths for brittle failure and ductile creep (Brace and 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 nonelastic 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 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.

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

$$\Delta\sigma_{xx} = \frac{-2f_s(\rho g y - p_w)}{(1+f_s^2)+f_s} \tag{2.92}$$

and under compression is

$$\Delta\sigma_{xx} = \frac{2f_s(\rho g y - p_w)}{(1+f_s^2)-f_s} \tag{2.93}$$

where $\Delta\sigma_{xx}$ is the deviatoric stress, f_s is that static coefficient of friction, $\rho g y$ 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 $p_w = \rho_w g y$, $\rho = 3300\,\mathrm{kg\,m^{-3}}$, and $\rho_w = 1000\,\mathrm{kg\,m^{-3}}$, and $f_s = 0.6$. The results for tension and compression are sketched in Fig. 2.35.

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 (see also §3.6.4)

$$\dot{\varepsilon} = A\sigma^n \exp\left(\frac{-E_a}{RT}\right) \tag{2.94}$$

For olivine, $A = 4.2 \times 10^5\,\mathrm{MPa^{-3}\,s^{-1}}$, E_a is the activation energy $= 523\,\mathrm{kJ\,mol^{-1}}$, and $n = 3$. Taking strain rates of $10^{-12}\,\mathrm{s^{-1}}$ and $10^{-14}\,\mathrm{s^{-1}}$, the stress as a function of depth for a geothermal gradient of $25\,\mathrm{K\,km^{-1}}$ can be found (Fig. 2.35). 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.

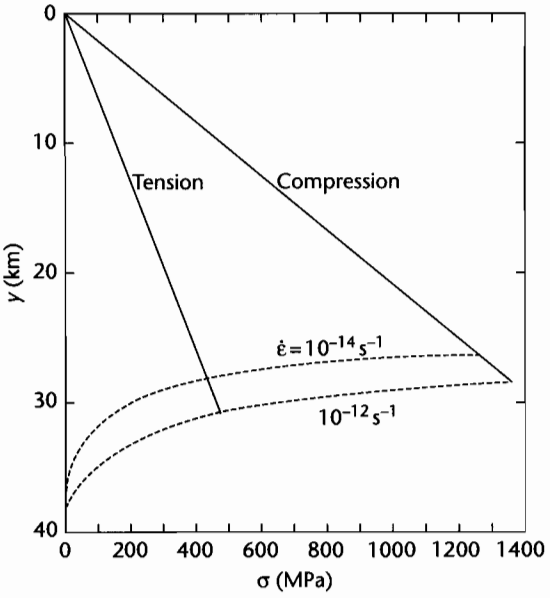

Fig. 2.35 Strength profile for the oceanic lithosphere calculated using equation (2.92) for tension and equation (2.93) for compression. The dashed lines are the stresses associated with creep in the lithosphere at the strain rates indicated. Reproduced courtesy of Cambridge University Press.

The effects of variations in geothermal gradient are illustrated in Figure 2.36 using reduced heat flows of $25\,\mathrm{mW\,m^{-2}}$ representing a cold continental shield, and $59\,\mathrm{mW\,m^{-2}}$ representing the high geothermal gradients typical of the Basin and Range province, USA (Lynch and Morgan 1987) ($K = 2.5\,\mathrm{W\,m^{-1}\,K^{-1}}$, $A_0 = 2.1\,\mu\mathrm{W\,m^{-3}}$, $a_r = 10\,\mathrm{km}$). 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 con-

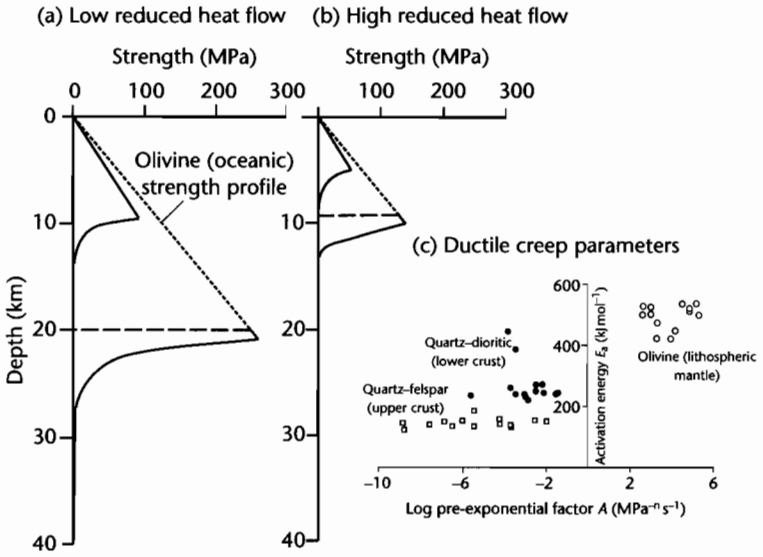

Fig. 2.36 Strength profiles of the continental lithosphere for different geothermal gradients, using (a) a low reduced heat flow of 25 mW m^{-2} for a cold continental shield area, and (b) a higher reduced heat flow of 59 mW m^{-2} for an area undergoing extension. $K = 2.5$ W m^{-1} K^{-1}, $A_0 = 2.1$ mW 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, $A = 10^4$ 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) shows 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, Fernandez and Ranalli 1997).

siderably 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 Myr under both extension and compression are considerably higher than any likely imposed tectonic forces. Continental lithospheric strength is strongly dependent on the basal heat flow or geothermal gradient. For a low heat flow (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 heat flows 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 reduced heat flow of 70 mW m^{-2} (Lachenbruch and Sass 1978), whereas in the eastern USA, which is tectonically inactive, crustal thickness is 40 km and reduced heat flow is 33 mW m^{-2}. A global comparison of tectonic activity and reduced 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 and Park 1987, Fig. 8, p. 42).

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 localized extension and high heat flows, whereas slow strain rates generate broad regions of extension with low surface heat flows.