Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration … he should seek an economical description of natural phenomena.

George E. Box, *Science and Statistics*, 1976.

The sciences do not try to explain … they mainly make models … a mathematical construct the justification [of which] is that it is expected to work.

John von Neumann, *Methods in the Physical Sciences*, 1955.

CHAPTER 3

Shallow Water Systems

CONVENTIONALLY, 'THE' SHALLOW WATER EQUATIONS describe a thin layer of constant density fluid in hydrostatic balance, rotating or not, bounded from below by a rigid surface and from above by a free surface, above which we suppose is another fluid of negligible inertia. Such a configuration can be generalized to multiple layers of immiscible fluids of different densities lying one on top of another, forming a stably-stratified 'stacked shallow water' system, which in many ways behaves like a continuously stratified fluid. These types of systems are the main subject of this chapter. We also introduce the notion of available potential energy, which involves thinking about a continuously stratified system as if it were a stacked shallow water system.

The single-layer model is one of the simplest useful models in geophysical fluid dynamics because it allows for a consideration of the effects of rotation in a simple framework without the complicating effects of stratification. A model with just two layers is not only a simple model of a stratified fluid, it is a surprisingly good model of many phenomena in the ocean and atmosphere. Such models are more than just pedagogical tools — we will find that there is a close physical and mathematical analogy between the shallow water equations and a description of the continuously stratified ocean or atmosphere written in isopycnal or isentropic coordinates, with a meaning beyond a coincidental similarity in the equations. Let us begin with the single-layer case.

3.1 DYNAMICS OF A SINGLE SHALLOW LAYER OF FLUID

Shallow water dynamics apply, by definition, to a fluid layer of constant density in which the horizontal scale of the flow is much greater than the layer depth. The fluid motion is fully determined by the momentum and mass continuity equations, and because of the assumed small aspect ratio the hydrostatic approximation is well satisfied, and we invoke this from the outset. Consider, then, fluid in a container above which is another fluid of negligible density (and therefore negligible inertia) relative to the fluid of interest, as illustrated in Fig. 3.1. Our notation is that $\boldsymbol{v} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ is the three-dimensional velocity and $\boldsymbol{u} = u\mathbf{i} + v\mathbf{j}$ is the horizontal velocity. $h(x, y)$ is the thickness of the liquid column, H is its mean height, and η is the height of the free surface. In a flat-bottomed container $\eta = h$, whereas in general $h = \eta - \eta_b$, where η_b is the height of the floor of the container.

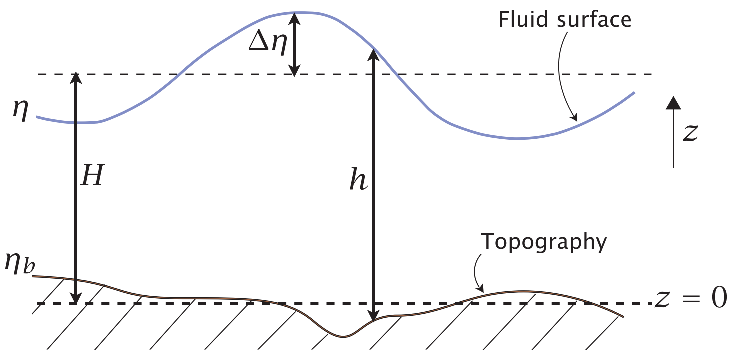

Fig. 3.1 A shallow water system. h is the thickness of a water column, H its mean thickness, η the height of the free surface and η_b is the height of the lower, rigid, surface above some arbitrary origin, typically chosen such that the average of η_b is zero. $\Delta\eta$ is the deviation free surface height, so we have $\eta = \eta_b + h = H + \Delta\eta$.

3.1.1 Momentum Equations

The vertical momentum equation is just the hydrostatic equation,

$$\frac{\partial p}{\partial z} = -\rho_0 g, \tag{3.1}$$

and, because density is assumed constant, we may integrate this to

$$p(x, y, z, t) = -\rho_0 g z + p_o. \tag{3.2}$$

At the top of the fluid, $z = \eta$, the pressure is determined by the weight of the overlying fluid and this is assumed to be negligible. Thus, $p = 0$ at $z = \eta$, giving

$$p(x, y, z, t) = \rho_0 g(\eta(x, y, t) - z). \tag{3.3}$$

The consequence of this is that the horizontal gradient of pressure is independent of height. That is

$$\nabla_z p = \rho_0 g \nabla_z \eta, \tag{3.4}$$

where

$$\nabla_z = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} \tag{3.5}$$

is the gradient operator at constant z. (In the rest of this chapter we will drop the subscript z unless that causes ambiguity. The three-dimensional gradient operator will be denoted by ∇_3. We will also mostly use Cartesian coordinates, but the shallow water equations may certainly be applied over a spherical planet — 'Laplace's tidal equations' are essentially the shallow water equations on a sphere.) The horizontal momentum equations therefore become

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho_0}\nabla p = -g\nabla\eta. \tag{3.6}$$

The right-hand side of this equation is independent of the vertical coordinate z. Thus, if the flow is initially independent of z, it must stay so. (This z-independence is unrelated to that arising from the rapid rotation necessary for the Taylor–Proudman effect.) The velocities u and v are functions of x, y and t only, and the horizontal momentum equation is therefore

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \frac{\partial \boldsymbol{u}}{\partial t} + u\frac{\partial \boldsymbol{u}}{\partial x} + v\frac{\partial \boldsymbol{u}}{\partial y} = -g\nabla\eta. \tag{3.7}$$

That the horizontal velocity is independent of z is a consequence of the hydrostatic equation, which ensures that the horizontal pressure gradient is independent of height. (Another starting point

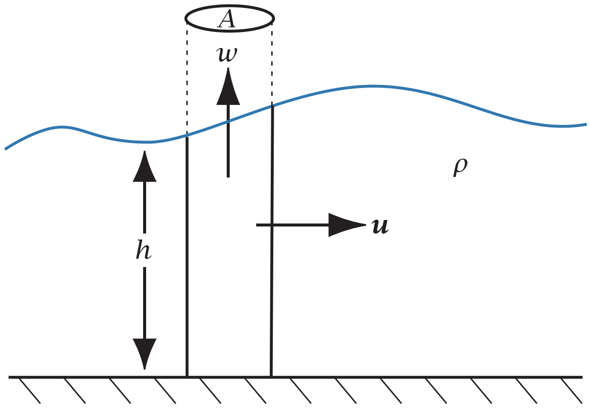

Fig. 3.2 The mass budget for a column of area A in a shallow water system. The fluid leaving the column is $\oint \rho_0 h\boldsymbol{u} \cdot \boldsymbol{n}\, dl$ where $\boldsymbol{n}$ is the unit vector normal to the boundary of the fluid column. There is a non-zero vertical velocity at the top of the column if the mass convergence into the column is non-zero.

would be to take this independence of the horizontal motion with height as the *definition* of shallow water flow. In real physical situations such independence does not hold exactly — for example, friction at the bottom may induce a vertical dependence of the flow in a boundary layer.) In the presence of rotation, (3.7) easily generalizes to

$$\boxed{\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta,} \tag{3.8}$$

where $\boldsymbol{f} = f\mathbf{k}$. Just as with the primitive equations, f may be constant or may vary with latitude, so that on a spherical planet $f = 2\Omega \sin\vartheta$ and on the β-plane $f = f_0 + \beta y$.

3.1.2 Mass Continuity Equation

From first principles

The mass contained in a fluid column of height h and cross-sectional area A is given by $\int_A \rho_0 h \, \mathrm{d}\boldsymbol{A}$ (see Fig. 3.2). If there is a net flux of fluid across the column boundary (by advection) then this must be balanced by a net increase in the mass in A, and therefore a net increase in the height of the water column. The mass convergence into the column is given by

$$F_m = \text{mass flux in} = -\int_S \rho_0 \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{S}, \tag{3.9}$$

where S is the area of the vertical boundary of the column. The surface area of the column is composed of elements of area $h\boldsymbol{n}\,\delta l$, where δl is a line element circumscribing the column and $\boldsymbol{n}$ is a unit vector perpendicular to the boundary, pointing outwards. Thus (3.9) becomes

$$F_m = -\oint \rho_0 h\boldsymbol{u} \cdot \boldsymbol{n}\, \mathrm{d}l. \tag{3.10}$$

Using the divergence theorem in two dimensions, (3.10) simplifies to

$$F_m = -\int_A \nabla \cdot (\rho_0 \boldsymbol{u} h)\, \mathrm{d}A, \tag{3.11}$$

where the integral is over the cross-sectional area of the fluid column (looking down from above). This is balanced by the local increase in height of the water column, given by

$$F_m = \frac{\mathrm{d}}{\mathrm{d}t}\int \rho_0 \,\mathrm{d}V = \frac{\mathrm{d}}{\mathrm{d}t}\int_A \rho_0 h\,\mathrm{d}A = \int_A \rho_0 \frac{\partial h}{\partial t}\,\mathrm{d}A. \tag{3.12}$$

Because ρ_0 is constant, the balance between (3.11) and (3.12) leads to

$$\int_A \left[\frac{\partial h}{\partial t} + \nabla \cdot (\boldsymbol{u}h) \right] \mathrm{d}A = 0, \tag{3.13}$$

and because the area is arbitrary the integrand itself must vanish, whence,

$$\frac{\partial h}{\partial t} + \nabla \cdot (\boldsymbol{u}h) = 0 \qquad \text{or} \qquad \frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla \cdot \boldsymbol{u} = 0. \tag{3.14a,b}$$

This derivation holds whether or not the lower surface is flat. If it is, then $h = \eta$, and if not $h = \eta - \eta_b$. Equations (3.8) and (3.14) form a complete set, summarized in the shaded box on the facing page.

From the 3D mass conservation equation

Since the fluid is incompressible, the three-dimensional mass continuity equation is just $\nabla \cdot \boldsymbol{v} = 0$. Writing this out in component form

$$\frac{\partial w}{\partial z} = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = -\nabla \cdot \boldsymbol{u}. \tag{3.15}$$

Integrate this from the bottom of the fluid ($z = \eta_b$) to the top ($z = \eta$), noting that the right-hand side is independent of z, to give

$$w(\eta) - w(\eta_b) = -h\nabla \cdot \boldsymbol{u}. \tag{3.16}$$

At the top the vertical velocity is the material derivative of the position of a particular fluid element. But the position of the fluid at the top is just η, and therefore (see Fig. 3.2)

$$w(\eta) = \frac{\mathrm{D}\eta}{\mathrm{D}t}. \tag{3.17a}$$

At the bottom of the fluid we have similarly

$$w(\eta_b) = \frac{\mathrm{D}\eta_b}{\mathrm{D}t}, \tag{3.17b}$$

where, apart from earthquakes and the like, $\partial \eta_b / \partial t = 0$. Using (3.17a,b), (3.16) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}(\eta - \eta_b) + h\nabla \cdot \boldsymbol{u} = 0 \tag{3.18}$$

or, as in (3.14b),

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla \cdot \boldsymbol{u} = 0. \tag{3.19}$$

3.1.3 A Rigid Lid

The case where the *upper* surface is held flat by the imposition of a rigid lid is sometimes of interest. The ocean suggests one such example, since the bathymetry at the bottom of the ocean provides much larger variations in fluid thickness than do the small variations in the height of the ocean surface. If we suppose that the upper surface is at a constant height H, then from (3.14a) with $\partial h / \partial t = 0$ the mass conservation equation is

$$\nabla_h \cdot (\boldsymbol{u}h_b) = 0, \tag{3.20}$$

The Shallow Water Equations

For a single-layer fluid, and including the Coriolis term, the inviscid shallow water equations are

momentum: $$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta. \tag{SW.1}$$

mass continuity: $$\frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla\cdot\boldsymbol{u} = 0 \qquad \text{or} \qquad \frac{\partial h}{\partial t} + \nabla\cdot(h\boldsymbol{u}) = 0, \tag{SW.2}$$

where $\boldsymbol{u}$ is the horizontal velocity, h is the total fluid thickness, η is the height of the upper free surface and η_b is the height of the lower surface (the bottom topography). Thus,

$$h(x, y, t) = \eta(x, y, t) - \eta_b(x, y) \tag{SW.3}$$

The material derivative is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}, \tag{SW.4}$$

with the rightmost expression holding in Cartesian coordinates.

where $h_b = H - \eta_b$. Note that (3.20) allows us to define an incompressible *mass-transport velocity*, $\boldsymbol{U} \equiv h_b\boldsymbol{u}$.

Although the upper surface is flat, the pressure there is no longer constant because a force must be provided by the rigid lid to keep the surface flat. The horizontal momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho_0}\nabla p_{lid}, \tag{3.21}$$

where p_{lid} is the pressure at the lid, and the complete equations of motion are then (3.20) and (3.21).[1] If the lower surface is flat, the two-dimensional flow itself is divergence-free, and the equations reduce to the two-dimensional incompressible Euler equations.

3.1.4 Stretching and the Vertical Velocity

Because the horizontal velocity is depth independent, the vertical velocity plays no role in advection. However, w is certainly not zero for then the free surface would be unable to move up or down, but because of the vertical independence of the horizontal flow w does have a simple vertical structure; to determine this we write the mass conservation equation as

$$\frac{\partial w}{\partial z} = -\nabla\cdot\boldsymbol{u}, \tag{3.22}$$

and integrate upwards from the bottom to give

$$w = w_b - (\nabla\cdot\boldsymbol{u})(z - \eta_b). \tag{3.23}$$

Thus, the vertical velocity is a linear function of height. Equation (3.23) can be written as

$$\frac{\mathrm{D}z}{\mathrm{D}t} = \frac{\mathrm{D}\eta_b}{\mathrm{D}t} - (\nabla\cdot\boldsymbol{u})(z - \eta_b), \tag{3.24}$$

and at the upper surface $w = \mathrm{D}\eta/\mathrm{D}t$ so that here we have

$$\frac{\mathrm{D}\eta}{\mathrm{D}t} = \frac{\mathrm{D}\eta_b}{\mathrm{D}t} - (\nabla\cdot\boldsymbol{u})(\eta - \eta_b). \tag{3.25}$$

Eliminating the divergence term from the last two equations gives

$$\frac{\mathrm{D}}{\mathrm{D}t}(z - \eta_b) = \frac{z-\eta_b}{\eta-\eta_b}\frac{\mathrm{D}}{\mathrm{D}t}(\eta - \eta_b), \tag{3.26}$$

which in turn gives

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{z-\eta_b}{\eta-\eta_b}\right) = \frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{z-\eta_b}{h}\right) = 0. \tag{3.27}$$

This means that the ratio of the height of a fluid parcel above the floor to the total depth of the column is fixed; that is, the fluid stretches uniformly in a column, and this is a kinematic property of the shallow water system.

3.1.5 Analogy with Compressible Flow

The shallow water equations (3.8) and (3.14) are analogous to the compressible gas dynamic equations in two dimensions, namely

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho}\nabla p \tag{3.28}$$

and

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\boldsymbol{u}\rho) = 0, \tag{3.29}$$

along with an equation of state which we take to be $p = f(\rho)$. The mass conservation equations (3.14) and (3.29) are identical, with the replacement $\rho \leftrightarrow h$. If $p = C\rho^\gamma$, then (3.28) becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}\rho}\nabla\rho = -C\gamma\rho^{\gamma-2}\nabla\rho. \tag{3.30}$$

If $\gamma = 2$ then the momentum equations (3.8) and (3.30) become equivalent, with $\rho \leftrightarrow h$ and $C\gamma \leftrightarrow g$. In an ideal gas $\gamma = c_p/c_v$ and values typically are in fact less than 2 (in air $\gamma \approx 7/5$); however, if the equations are linearized, then the analogy is exact for all values of γ, for then (3.30) becomes $\partial\boldsymbol{v}'/\partial t = -\rho_0^{-1}c_s^2\nabla\rho'$ where $c_s^2 = \mathrm{d}p/\mathrm{d}\rho$, and the linearized shallow water momentum equation is $\partial\boldsymbol{u}'/\partial t = -H^{-1}(gH)\nabla h'$, so that $\rho_0 \leftrightarrow H$ and $c_s^2 \leftrightarrow gH$. The sound waves of a compressible fluid are then analogous to shallow water waves, which are considered in Section 3.8.

3.2 REDUCED GRAVITY EQUATIONS

Consider now a single shallow moving layer of fluid on top of a deep, quiescent fluid layer (Fig. 3.3), and beneath a fluid of negligible inertia. This configuration is often used as a model of the upper ocean: the upper layer represents flow in perhaps the upper few hundred metres of the ocean, the lower layer being the near-stagnant abyss. If we turn the model upside-down we have a perhaps slightly less realistic model of the atmosphere: the lower layer represents motion in the troposphere above which lies an inactive stratosphere. The equations of motion are virtually the same in both cases.

3.2.1 Pressure Gradient in the Active Layer

We will derive the equations for the oceanic case (active layer on top) in two cases, which differ slightly in the assumption made about the upper surface.

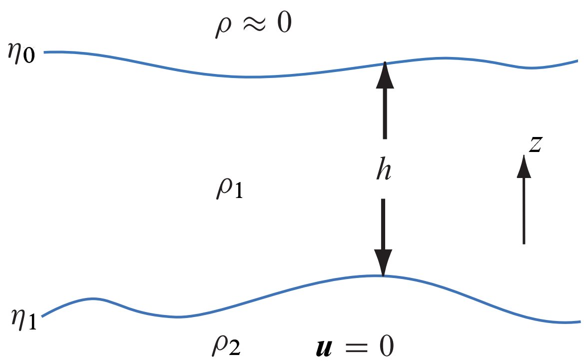

Fig. 3.3 The reduced gravity shallow water system. An active layer lies over a deep, denser, quiescent layer. In a common variation the upper surface is held flat by a rigid lid, and $\eta_0 = 0$.

I Free upper surface

The pressure in the upper layer is given by integrating the hydrostatic equation down from the upper surface. Thus, at a height z in the upper layer

$$p_1(z) = g\rho_1(\eta_0 - z), \tag{3.31}$$

where η_0 is the height of the upper surface. Hence, everywhere in the upper layer,

$$\frac{1}{\rho_1}\nabla p_1 = g\nabla\eta_0, \tag{3.32}$$

and the momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} = -g\nabla\eta_0. \tag{3.33}$$

In the lower layer the pressure is also given by the weight of the fluid above it. Thus, at some level z in the lower layer,

$$p_2(z) = \rho_1 g(\eta_0 - \eta_1) + \rho_2 g(\eta_1 - z). \tag{3.34}$$

But if this layer is motionless the horizontal pressure gradient in it is zero and therefore

$$\rho_1 g\eta_0 = -\rho_1 g'\eta_1 + \text{constant}, \tag{3.35}$$

where $g' = g(\rho_2 - \rho_1)/\rho_1$ is the *reduced gravity*, and normally $(\rho_2 - \rho_1)/\rho \ll 1$ and $g' \ll g$. The momentum equation becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} = g'\nabla\eta_1. \tag{3.36}$$

The equations are completed by the usual mass conservation equation,

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla\cdot\boldsymbol{u} = 0, \tag{3.37}$$

where $h = \eta_0 - \eta_1$. Because $g \gg g'$, (3.35) shows that surface displacements are *much smaller* than the displacements at the interior interface. We see this in the real ocean where the mean interior isopycnal displacements may be several tens of metres but variations in the mean height of ocean surface are of the order of centimetres.

II The rigid lid approximation

The smallness of the upper surface displacement suggests that we will make little error if we impose a *rigid lid* at the top of the fluid. Displacements are no longer allowed, but the lid will in general

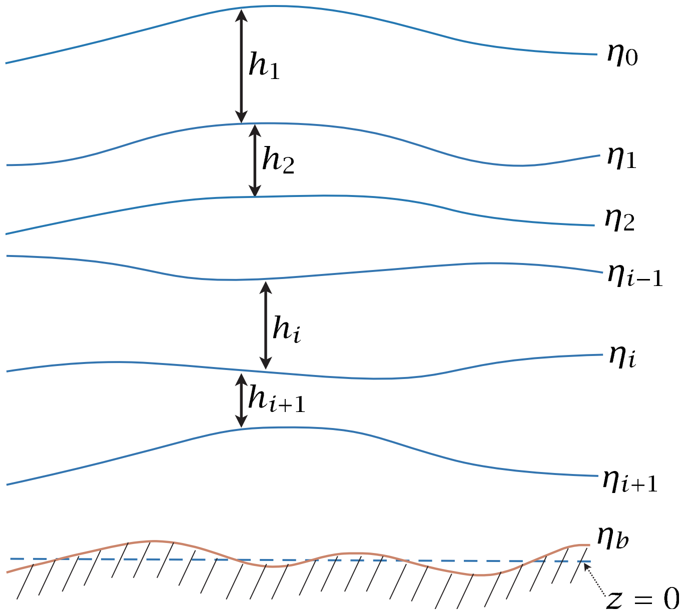

Fig. 3.4 The multi-layer shallow water system. The layers are numbered from the top down. The coordinates of the interfaces are denoted by η, and the layer thicknesses by h, so that $h_i = \eta_{i-1} - \eta_i$.

impart a pressure force to the fluid. Suppose that this is $P(x, y, t)$, then the horizontal pressure gradient in the upper layer is simply

$$\nabla p_1 = \nabla P. \tag{3.38}$$

The pressure in the lower layer is again given by hydrostasy, and is

$$p_2 = -\rho_1 g \eta_1 + \rho_2 g(\eta_1 - z) + P = \rho_1 g h - \rho_2 g(h + z) + P, \tag{3.39}$$

so that

$$\nabla p_2 = -g(\rho_2 - \rho_1)\nabla h + \nabla P. \tag{3.40}$$

Then if $\nabla p_2 = 0$ (because the lower layer is stationary) we have $g(\rho_2 - \rho_1)\nabla h = \nabla P$, and the momentum equation for the upper layer is just

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g' \nabla h, \tag{3.41}$$

where $g' = g(\rho_2 - \rho_1)/\rho_1$. These equations differ from the usual shallow water equations only in the use of a reduced gravity g' in place of g itself. It is the density *difference* between the two layers that is important. Similarly, if we take a shallow water system, with the moving layer on the bottom, and we suppose that overlying it is a stationary fluid of finite density, then we would easily find that the fluid equations for the moving layer are the same as if the fluid on top had zero inertia, except that g would be replaced by an appropriate reduced gravity.

3.3 MULTI-LAYER SHALLOW WATER EQUATIONS

We now consider the dynamics of multiple layers of fluid stacked on top of each other. This is a crude representation of continuous stratification, but it turns out to be a powerful model of many geophysically interesting phenomena as well as being physically realizable in the laboratory. The pressure is continuous across the interface, but the density jumps discontinuously and this allows the horizontal velocity to have a corresponding discontinuity. The set up is illustrated in Fig. 3.4.

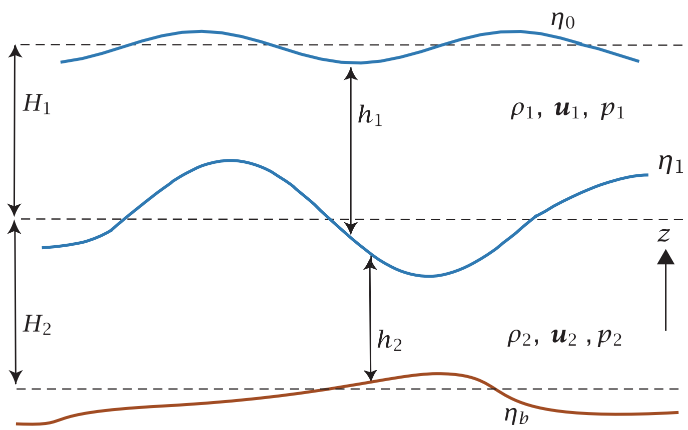

Fig. 3.5 The two-layer shallow water system. A fluid of density ρ_1 lies over a denser fluid of density ρ_2. In the reduced gravity case the lower layer is arbitrarily thick and is assumed stationary and so has no horizontal pressure gradient. In the 'rigid-lid' approximation the top surface displacement is neglected, but there is then a non-zero pressure gradient induced by the lid.

In each layer pressure is given by the hydrostatic approximation, and so anywhere in the interior we can find the pressure by integrating down from the top. Thus, at a height z in the first layer we have

$$p_1 = \rho_1 g(\eta_0 - z), \tag{3.42}$$

and in the second layer,

$$p_2 = \rho_1 g(\eta_0 - \eta_1) + \rho_2 g(\eta_1 - z) = \rho_1 g\eta_0 + \rho_1 g_1'\eta_1 - \rho_2 gz, \tag{3.43}$$

where $g_1' = g(\rho_2 - \rho_1)/\rho_1$, and so on. The term involving z is irrelevant for the dynamics, because only the horizontal derivative enters the equation of motion. Omitting this term, for the nth layer the dynamical pressure is given by the sum from the top down:

$$p_n = \rho_1 \sum_{i=0}^{n-1} g_i'\eta_i, \tag{3.44}$$

where $g_i' = g(\rho_{i+1} - \rho_i)/\rho_1$ (and $g_0 = g$). The interface displacements may be expressed in terms of the layer thicknesses by summing from the bottom up:

$$\eta_n = \eta_b + \sum_{i=n+1}^{i=N} h_i. \tag{3.45}$$

The momentum equation for each layer may then be written, in general,

$$\frac{\mathrm{D}\boldsymbol{u}_n}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_n = -\frac{1}{\rho_n}\nabla p_n, \tag{3.46}$$

where the pressure is given by (3.44) and in terms of the layer depths using (3.46). If we make the Boussinesq approximation then ρ_n on the right-hand side of (3.46) is replaced by ρ_1.

Finally, the mass conservation equation for each layer has the same form as the single-layer case, and is

$$\frac{\mathrm{D}h_n}{\mathrm{D}t} + h_n\nabla\cdot\boldsymbol{u}_n = 0. \tag{3.47}$$

The two- and three-layer cases

The two-layer model (Fig. 3.5) is the simplest model to capture the effects of stratification. Evaluating the pressures using (3.44) and (3.45) we find:

$$p_1 = \rho_1 g\eta_0 = \rho_1 g(h_1 + h_2 + \eta_b), \tag{3.48a}$$

$$p_2 = \rho_1[g\eta_0 + g_1'\eta_1] = \rho_1\left[g(h_1 + h_2 + \eta_b) + g_1'(h_2 + \eta_b)\right]. \tag{3.48b}$$

The momentum equations for the two layers are then

$$\frac{\mathrm{D}\boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_1 = -g\nabla\eta_0 = -g\nabla(h_1 + h_2 + \eta_b), \tag{3.49a}$$

and in the bottom layer

$$\begin{aligned}\frac{\mathrm{D}\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_2 &= -\frac{\rho_1}{\rho_2}\left(g\nabla\eta_0 + g_1'\nabla\eta_1\right) \\ &= -\frac{\rho_1}{\rho_2}\left[g\nabla(\eta_b + h_1 + h_2) + g_1'\nabla(h_2 + \eta_b)\right].\end{aligned} \tag{3.49b}$$

In the Boussinesq approximation ρ_1/ρ_2 is replaced by unity.

In a three-layer model the dynamical pressures are found to be

$$p_1 = \rho_1 g h, \tag{3.50a}$$
$$p_2 = \rho_1\left[gh + g_1'(h_2 + h_3 + \eta_b)\right], \tag{3.50b}$$
$$p_3 = \rho_1\left[gh + g_1'(h_2 + h_3 + \eta_b) + g_2'(h_3 + \eta_b)\right], \tag{3.50c}$$

where $h = \eta_0 = \eta_b + h_1 + h_2 + h_3$ and $g_2' = g(\rho_3 - \rho_2)/\rho_1$. More layers can obviously be added in a systematic fashion.

3.3.1 Reduced-gravity Multi-layer Equation

As with a single active layer, we may envision multiple layers of fluid overlying a deeper stationary layer. This is a useful model of the stratified upper ocean overlying a nearly stationary and nearly unstratified abyss. Indeed we use such a model to study the 'ventilated thermocline' in Chapter 20 and a detailed treatment may be found there. If we suppose there is a lid at the top, then the model is almost the same as that of the previous section. However, now the horizontal pressure gradient in the lowest model layer is zero, and so we may obtain the pressures in all the active layers by integrating the hydrostatic equation upwards from this layer. Suppose we have N moving layers, then the reader may verify that the dynamic pressure in the nth layer is given by

$$p_n = -\sum_{i=n}^{i=N} \rho_1 g_i' \eta_i, \tag{3.51}$$

where as before $g_i' = g(\rho_{i+1} - \rho_i)/\rho_1$. If we have a lid at the top, and take $\eta_0 = 0$, then the interface displacements are related to the layer thicknesses by

$$\eta_n = -\sum_{i=1}^{i=n} h_i. \tag{3.52}$$

From these expressions the momentum equation in each layer is easily constructed.

3.4 ✦ FROM CONTINUOUS STRATIFICATION TO SHALLOW WATER

In this section we show that the *continuously stratified* equations have a close correspondence to the shallow water equations, without breaking the fluid into discrete layers of differing densities. In particular, if the continuous equations are linearized and the flow is stably stratified, then each vertical mode of the continuous equations has the same form as the shallow water equations, with the modes being distinguished by the phase speed of the associated gravity waves.[2]

3.4.1 Vertical Normal Modes of the Linear Equations

We begin with a hydrostatic Boussinesq system, linearized about a state of rest and with fixed stratification, $N(z)$, noting that a similar derivation can be applied to an ideal gas using pressure coordinates. The equations are

$$\frac{\partial u}{\partial t} - fv = -\frac{\partial \phi}{\partial x}, \qquad \frac{\partial v}{\partial t} + fu = -\frac{\partial \phi}{\partial y}, \qquad 0 = -\frac{\partial \phi}{\partial z} + b, \tag{3.53a,b,c}$$

$$\nabla \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0, \qquad \frac{\partial b}{\partial t} + wN^2 = 0. \tag{3.53c,d}$$

The first line above contains the u and v momentum equations and the hydrostatic equation, and the second line contains the mass continuity equation and the buoyancy or thermodynamic equation, with the ∇ operator being purely horizontal (or at constant pressure), and we will take $N^2 > 0$. We assume a lid at the bottom and top of the domain. Including a free surface at the top, as appropriate for an ocean, is a slight extension. Including a 'leaky' tropopause with a stratosphere above is a more major extension.

The difficulty with these equations is that there are five independent variables in three spatial coordinates so that even the linear problems are algebraically complex, especially when f is variable. The equations are more general than is needed, because it is often observed that the vertical structure of solutions is relatively simple, especially in linear problems. A solution is to project the vertical structure onto appropriate eigenfunctions, and then to retain a very small number — often only one — of these eigenfunctions.

To determine what those eigenfunctions should be, we first combine the hydrostatic and buoyancy equations to give

$$\frac{\partial}{\partial t}\left(\frac{\phi_z}{N^2}\right) + w = 0. \tag{3.54}$$

Differentiating with respect to z and using the mass continuity equation gives

$$\frac{\partial}{\partial t}\left(\frac{\phi_z}{N^2}\right)_z - \nabla \cdot \boldsymbol{u} = 0. \tag{3.55}$$

It is this equation that motivates our choice of basis functions: we choose to expand the pressure and horizontal components of velocity in terms of an eigenfunction that satisfies the following Sturm–Liouville problem:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{N^2}\frac{\mathrm{d}C_m}{\mathrm{d}z}\right) + \frac{1}{c_m^2}C_m = 0, \qquad \frac{\mathrm{d}}{\mathrm{d}z}C_m(0) = \frac{\mathrm{d}}{\mathrm{d}z}C_m(-H) = 0. \tag{3.56}$$

The eigenfunctions C_m are orthogonal in the sense that

$$\int_{-H}^{0} C_m C_n \,\mathrm{d}z = \frac{c_m^2}{g}\delta_{mn}\,, \tag{3.57}$$

where $\delta_{mn} = 0$ unless $m = n$, in which case it equals one. The normalization is by convention and the factor of g makes the functions C_m nondimensional. There are an infinite number of eigenvalues, c_m, namely $c_0, c_1, c_2 \ldots$, normally arranged in descending order of size, and for each there is a corresponding eigenfunction C_m. The pressure and horizontal velocity components are then expressed as

$$[u, v, \phi] = \sum_0^\infty \left[u_m(x,y,t), v_m(x,y,t), \phi_m(x,y,t)\right] C_m(z). \tag{3.58}$$

The benefit of this procedure is that the z-derivatives in the equations of motion are replaced by multiplications, and in particular (3.55) becomes

$$\frac{\partial \phi_m}{\partial t} + c_m^2 \nabla \cdot \boldsymbol{u}_m = 0 \quad \text{or} \quad \frac{\partial \eta_m^*}{\partial t} + H_m \nabla \cdot \boldsymbol{u}_m = 0, \tag{3.59a,b}$$

where $\eta^* \equiv \phi/g$. The quantity $H_m = c_m^2/g$ is the *equivalent depth* associated with the eigenmode. Equations (3.59) are evidently of the same form as the familiar linear mass continuity equation in the shallow water equations, namely

$$\frac{\partial \widehat{\eta}}{\partial t} + c^2 \nabla \cdot \boldsymbol{u} = 0 \qquad \text{or} \qquad \frac{\partial \eta}{\partial t} + H \nabla \cdot \boldsymbol{u} = 0, \tag{3.60a,b}$$

where $c = \sqrt{gH}$ and $\widehat{\eta} = g\eta$.

The horizontal momentum equations are simply,

$$\frac{\partial u_m}{\partial t} - f v_m = -\frac{\partial \phi_m}{\partial x}, \qquad \frac{\partial v_m}{\partial t} + f u_m = -\frac{\partial \phi_m}{\partial y}. \tag{3.61a,b}$$

Equations (3.59) and (3.61) are a closed set, once we have calculated the equivalent depth H_m for each mode. If there is a forcing in the momentum equation then the transformed forcing appears on the right-hand sides of (3.61). If there is a source in the buoyancy equation then a corresponding term appears on the right-hand side of (3.59), analogous to a mass source term in the shallow water equations. Note that a thermodynamic source affects $\partial \phi/\partial z$ and not ϕ itself.

Eigenfunctions for buoyancy and vertical velocity

The vertical velocity and the buoyancy do not satisfy the same boundary conditions and so should not be expanded in the same way. Rather, we let

$$\left[w, \frac{b}{N^2}\right] = \sum_0^\infty \left[w_m(x,y,t), \hat{b}_m(x,y,t)\right] S_m(z), \tag{3.62}$$

where the eigenfunctions satisfy

$$\frac{1}{N^2}\frac{\mathrm{d}^2 S_m}{\mathrm{d}z^2} + \frac{1}{c_m^2} S_m = 0, \qquad S_m(0) = S_m(-H) = 0, \tag{3.63a,b}$$

where $S_m = 0$ if $N = 0$, and we may use the orthonormalization,

$$\int_{-H}^{0} N^2 S_m S_n \,\mathrm{d}z = g\delta_{mn}. \tag{3.64}$$

The functions S_m and C_m are related by

$$C_m = \frac{c_m^2}{g}\frac{\mathrm{d}S_m}{\mathrm{d}z}, \qquad N^2 S_m = -g\frac{\mathrm{d}C_m}{\mathrm{d}z}, \tag{3.65}$$

and it is these relationships that motivate the form of (3.63). The vertical velocity may be evaluated from the mass continuity equation, $\partial w/\partial z = -\nabla \cdot \boldsymbol{u}$, which becomes

$$w_m \frac{\mathrm{d}S_m}{\mathrm{d}z} = -C_m \nabla \cdot \boldsymbol{u}_m \qquad \Longrightarrow \qquad w_m = -\frac{c_m^2}{g}\nabla \cdot \boldsymbol{u}_m. \tag{3.66a,b}$$

Buoyancy is obtained from (3.53c) which, using (3.65), gives $\hat{b}_m = -\phi_m/g$.

3.4.2 Examples and Approximations

The values of c_m can be computed by solving the eigenvalue problem for the given stratification, although in general this must be carried out numerically. Consider, though, the simplest case in which N is constant, which is a reasonable approximation for the troposphere, less so for the ocean. The normal modes are sines and cosines, and for $m = 1, 2 \dots$ we have

$$C_m(z) = A_m \cos \frac{m\pi z}{H}, \qquad S_m(z) = B_m \sin \frac{m\pi z}{H}, \qquad c_m = \frac{NH}{m\pi}, \tag{3.67}$$

where, for $m > 0$, $A_m = c_m/\sqrt{gH/2}$ and $B_m = \sqrt{2g/HN^2}$. The equivalent depth is given by

$$H_m = \frac{N^2H^2}{gm^2\pi^2} = \frac{g'H}{gm^2\pi^2}, \tag{3.68}$$

where $g' \equiv HN^2$ and for a Boussinesq fluid $g' = (gH/\rho_0)\partial\rho/\partial z$. Using (3.67) we see that $c_m = \sqrt{gH_m} = \sqrt{g'H}/m\pi$, and note the factors of π are significant in these expressions. The mode with $m = 0$ is a special one and is called the *barotropic mode* with

$$C_0 = A_0/2, \qquad c_0^2 = gH. \tag{3.69}$$

The above expressions allow us to estimate equivalent depths and phase speeds for the atmosphere and ocean, with some caveats. For the atmosphere we should properly take into account its compressibility and a leaky tropopause, but proceeding nevertheless let us take $H = 10\,\mathrm{km}$ and $N = 10^{-2}\,\mathrm{s}^{-1}$ (a typical tropospheric value), whence

$$c_0 \approx 300\,\mathrm{m\,s^{-1}},\ c_1 \approx 30\,\mathrm{m\,s^{-1}},\ c_2 \approx 15\,\mathrm{m\,s^{-1}} \quad \text{and} \quad H_1 \approx 100\,\mathrm{m},\ H_2 \approx 25\,\mathrm{m}. \tag{3.70}$$

These equivalent depths are much smaller than the actual depth of the atmosphere, a fact that transcends our approximations and that greatly affects the properties of atmospheric gravity waves, as we discover in later chapters. The best fits to observations of internal gravity waves in the atmosphere are often in fact made with an equivalent depth of 50 m or less and a speed of about $20\,\mathrm{m\,s^{-1}}$.

The oceanic stratification is in fact not constant, but decreases significantly below the thermocline, which is about 1 km thick. We might proceed by simply using values appropriate for the thermocline in the above, and if we take $N = 10^{-2}\,\mathrm{s}^{-1}$ and $H = 1\,\mathrm{km}$ we find, using (3.67) and (3.68),

$$c_0 \approx 200\,\mathrm{m\,s^{-1}},\ c_1 \approx 3\,\mathrm{m\,s^{-1}},\ c_2 \approx 1.5\,\mathrm{m\,s^{-1}} \quad \text{and} \quad H_1 \approx 1\,\mathrm{m},\ H_2 \approx 0.25\,\mathrm{m}. \tag{3.71}$$

The speed c_0 is (as for the atmosphere) vastly larger than any parcel speed in the ocean. In contrast, the equivalent depths are very small, but this just reflects the smallness of the density variations in the ocean and the fact that H_m is proportional to g'/g.

If the oceanic stratification varies reasonably slowly we can use WKB methods (page 247) to good effect to better evaluate the eigenvalues and eigenfunctions.[3] Roughly speaking NH is replaced by $\int N\,\mathrm{d}z$ in (3.67), and the WKB solution, for $m \geq 1$, is

$$S_m \sim S_0 \sin\left(\frac{1}{c_m}\int_{-H}^{z} N(z)\,\mathrm{d}z\right), \qquad C_m \sim \left(\frac{c_m N S_0}{g}\right)\cos\left(\frac{1}{c_m}\int_{-H}^{z} N(z)\,\mathrm{d}z\right),$$
$$c_m \approx \frac{1}{m\pi}\int_{-H}^{0} N\,\mathrm{d}z, \tag{3.72a,b,c}$$

where $S_0 = (c_m/N)^{1/2}$. Using (3.72c) still gives values of c_1 of around 2–$3\,\mathrm{m\,s^{-1}}$ over the ocean gyres, less in equatorial regions, providing some post facto justification for using $H = 1\,\mathrm{km}$ previously. The eigenfunctions, (3.72a,b), are 'stretched' sines and cosines, with local wavenumbers

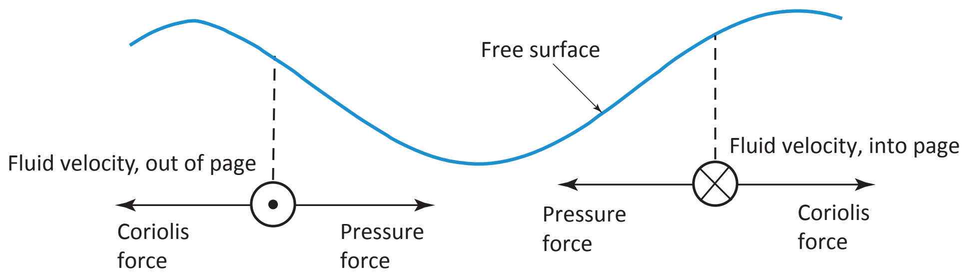

Fig. 3.6 Geostrophic flow in a shallow water system, with a positive value of the Coriolis parameter f, as in the Northern Hemisphere. The pressure force is directed down the gradient of the height field, and this can be balanced by the Coriolis force if the fluid velocity is at right angles to it. If f were negative, the geostrophic flow would be reversed.

proportional to $N(z)$ and so varying more rapidly in the upper ocean than at depth (look ahead to Fig. 12.12). The vertical velocity eigenfunctions, S_m, have a smaller amplitude in the upper ocean but the pressure and horizontal velocity amplitudes are larger.

For the remainder of this chapter we will use the shallow water equations in their conventional form, for if there *is* a region where density changes rapidly in the vertical then the layered equations are quite natural, and allow for the incorporation of nonlinearities more easily.

3.5 GEOSTROPHIC BALANCE AND THERMAL WIND

We now turn our attention to the *dynamics* of shallow water systems, beginning with the effects of rotation. Geostrophic balance occurs in the shallow water equations, just as in the continuously stratified equations, when the Rossby number U/fL is small and the Coriolis term dominates the advective terms in the momentum equation. In the single-layer shallow water equations the geostrophic flow is:

$$\boldsymbol{f} \times \boldsymbol{u}_g = -g\nabla\eta. \tag{3.73}$$

Thus, the geostrophic velocity is proportional to the slope of the surface, as sketched in Fig. 3.6. (For the rest of this section we drop the subscript g, and take all velocities to be geostrophic.)

In both the single-layer and multi-layer cases, the slope of an interfacial surface is directly related to the difference in pressure gradient on either side and so, by geostrophic balance, to the shear of the flow. This is the shallow water analogue of the thermal wind relation. To obtain an expression for this, consider the interface, η, between two layers labelled 1 and 2. The pressure in two layers is given by the hydrostatic relation and so,

$$p_1 = A(x, y) - \rho_1 g z \qquad \text{(at some } z \text{ in layer 1)}, \tag{3.74a}$$

$$\begin{aligned} p_2 &= A(x, y) - \rho_1 g\eta + \rho_2 g(\eta - z) \\ &= A(x, y) + \rho_1 g_1'\eta - \rho_2 g z \qquad \text{(at some } z \text{ in layer 2)}, \end{aligned} \tag{3.74b}$$

where $A(x, y)$ is a function of integration. Thus we find

$$\frac{1}{\rho_1}\nabla(p_1 - p_2) = -g_1'\nabla\eta. \tag{3.75}$$

If the flow is geostrophically balanced and Boussinesq then, in each layer, the velocity obeys

$$f\boldsymbol{u}_i = \frac{1}{\rho_1}\mathbf{k} \times \nabla p_i\,. \tag{3.76}$$

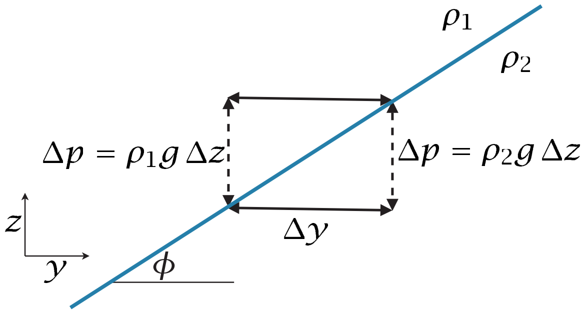

Fig. 3.7 Margules' relation: using hydrostasy, the difference in the horizontal pressure gradient between the upper and the lower layer is given by $-g'\rho_1 s$, where $s = \tan\phi = \Delta z/\Delta y$ is the interface slope and $g' = g(\rho_2 - \rho_1)/\rho_1$. Geostrophic balance then gives $f(u_1 - u_2) = g's$, which is a special case of (3.78).

Using (3.75) then gives

$$f(\boldsymbol{u}_1 - \boldsymbol{u}_2) = -\mathbf{k} \times g_1' \nabla \eta, \tag{3.77}$$

or in general

$$f(\boldsymbol{u}_n - \boldsymbol{u}_{n+1}) = -\mathbf{k} \times g_n' \nabla \eta. \tag{3.78}$$

This is the thermal wind equation for the shallow water system. It applies at any interface, and it implies *the shear is proportional to the interface slope,* a result known as the Margules relation[4] (Fig. 3.7).

Suppose that we represent the atmosphere by two layers of fluid; a meridionally decreasing temperature may then be represented by an interface that slopes upwards toward the pole. Then, in either hemisphere, we have

$$u_1 - u_2 = \frac{g_1'}{f}\frac{\partial \eta}{\partial y} > 0, \tag{3.79}$$

and the temperature gradient is associated with a positive shear.

3.6 FORM STRESS

When the interface between two layers varies with position — that is, when it is wavy — the layers exert a pressure force on each other. Similarly, if the bottom of the fluid is not flat then the topography and the bottom layer will in general exert forces on each other. This kind of force (normally arising as a force per unit area) is known as *form stress,* and it is an important means whereby momentum can be added to or extracted from a flow.[5] Consider a layer confined between two interfaces, $\eta_1(x, y)$ and $\eta_2(x, y)$. Then over some zonal interval L the average zonal pressure force on that fluid layer is given by

$$F_p = -\frac{1}{L}\int_{x_1}^{x_2}\int_{\eta_2}^{\eta_1} \frac{\partial p}{\partial x}\,\mathrm{d}x\,\mathrm{d}z. \tag{3.80}$$

Integrating by parts first in z and then in x, and noting that by hydrostasy $\partial p/\partial z$ does not depend on horizontal position within the layer, we obtain

$$F_p = -\frac{1}{L}\int_{x_1}^{x_2}\left[\frac{\partial p}{\partial x} z\right]_{\eta_2}^{\eta_1} \mathrm{d}x = -\overline{\eta_1 \frac{\partial p_1}{\partial x}} + \overline{\eta_2 \frac{\partial p_2}{\partial x}} = +\overline{p_1 \frac{\partial \eta_1}{\partial x}} - \overline{p_2 \frac{\partial \eta_2}{\partial x}}, \tag{3.81}$$

where p_1 is the pressure at η_1, and similarly for p_2, and to obtain the second line we suppose that the integral is around a closed path, such as a circle of latitude, and the average is denoted with an overbar. These terms represent the transfer of momentum from one layer to the next, and at a particular interface, i, we may define the form stress, τ_i, by

$$\tau_i \equiv \overline{p_i \frac{\partial \eta_i}{\partial x}} = -\overline{\eta_i \frac{\partial p_i}{\partial x}}. \tag{3.82}$$

The form stress is a force per unit area and its vertical derivative, $\partial\tau/\partial z$, is the force (per unit volume) on the fluid. Form stress is a particularly important means for the vertical transfer of momentum and its ultimate removal in an eddying fluid, and is one of the main mechanisms whereby the wind stress at the top of the ocean is communicated to the ocean bottom. At the fluid bottom the form stress is $\overline{p\partial_x\eta_b}$, where η_b is the bottom topography, and this is proportional to the momentum exchange with the solid Earth. This is a significant mechanism for the ultimate removal of momentum in the ocean, especially in the Antarctic Circumpolar Current where it is likely to be much larger than bottom (or Ekman) drag arising from small-scale turbulence and friction. In the two-layer, flat-bottomed case the only form stress occurring is that at the interface, and the momentum transfer between the layers is just $\overline{p_1\partial\eta_1/\partial x}$ or $-\overline{\eta_1\partial p_1/\partial x}$; then, the force on each layer due to the other is equal and opposite, as we would expect from momentum conservation. (Form stress is discussed more in an oceanographic context in Sections 19.6.3 and 21.7.2.)

For flows in geostrophic balance, the form stress is related to the meridional heat flux. The pressure gradient and velocity are related by $\rho f v' = \partial p'/\partial x$ and the interfacial displacement is proportional to the temperature perturbation, b' — in fact one may show that $\eta' \approx -b'/(\partial\overline{b}/\partial z)$. Thus $-\overline{\eta'\partial p'/\partial x} \propto \overline{v'b'}$, a correspondence that will recur when we consider the *Eliassen–Palm flux* in Chapter 10.

3.7 CONSERVATION PROPERTIES OF SHALLOW WATER SYSTEMS

There are two common types of conservation property in fluids: (i) material invariants; and (ii) integral invariants. Material invariance occurs when a property (φ say) is conserved on each fluid element, and so obeys the equation $\mathrm{D}\varphi/\mathrm{D}t = 0$. An integral invariant is one that is conserved after an integration over some, usually closed, volume; energy is an example.

3.7.1 Potential Vorticity: a Material Invariant

The vorticity of a fluid (considered at greater length in chapter 4), denoted $\boldsymbol{\omega}$, is defined to be the curl of the velocity field. Let us also define the shallow water vorticity, $\boldsymbol{\omega}^*$, as the curl of the horizontal velocity. We therefore have:

$$\boldsymbol{\omega} \equiv \nabla \times \boldsymbol{v}, \qquad \boldsymbol{\omega}^* \equiv \nabla \times \boldsymbol{u}. \tag{3.83}$$

Because $\partial u/\partial z = \partial v/\partial z = 0$, only the vertical component of $\boldsymbol{\omega}^*$ is non-zero and

$$\boldsymbol{\omega}^* = \mathbf{k}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = \mathbf{k}\,\zeta. \tag{3.84}$$

Considering first the non-rotating case, we use the vector identity

$$(\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \frac{1}{2}\nabla(\boldsymbol{u}\cdot\boldsymbol{u}) - \boldsymbol{u}\times(\nabla\times\boldsymbol{u}), \tag{3.85}$$

to write the momentum equation, (3.8) with $f = 0$, as

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{\omega}^*\times\boldsymbol{u} = -\nabla\left(g\eta + \frac{1}{2}\boldsymbol{u}^2\right). \tag{3.86}$$

To obtain an evolution equation for the vorticity we take the curl of (3.86), and make use of the vector identity

$$\begin{aligned}\nabla\times(\boldsymbol{\omega}^*\times\boldsymbol{u}) &= (\boldsymbol{u}\cdot\nabla)\boldsymbol{\omega}^* - (\boldsymbol{\omega}^*\cdot\nabla)\boldsymbol{u} + \boldsymbol{\omega}^*\nabla\cdot\boldsymbol{u} - \boldsymbol{u}\nabla\cdot\boldsymbol{\omega}^*\\ &= (\boldsymbol{u}\cdot\nabla)\boldsymbol{\omega}^* + \boldsymbol{\omega}^*\nabla\cdot\boldsymbol{u},\end{aligned} \tag{3.87}$$

using the fact that $\nabla \cdot \boldsymbol{\omega}^*$ is the divergence of a curl and therefore zero, and $(\boldsymbol{\omega}^* \cdot \nabla)\boldsymbol{u} = 0$ because $\boldsymbol{\omega}^*$ is perpendicular to the surface in which $\boldsymbol{u}$ varies. Taking the curl of (3.86) gives

$$\frac{\partial \zeta}{\partial t} + (\boldsymbol{u} \cdot \nabla)\zeta = -\zeta \nabla \cdot \boldsymbol{u}, \tag{3.88}$$

where $\zeta = \mathbf{k} \cdot \boldsymbol{\omega}^*$. Now, the mass conservation equation may be written as

$$-\zeta \nabla \cdot \boldsymbol{u} = \frac{\zeta}{h}\frac{\mathrm{D}h}{\mathrm{D}t}, \tag{3.89}$$

and using this (3.88) becomes

$$\frac{\mathrm{D}\zeta}{\mathrm{D}t} = \frac{\zeta}{h}\frac{\mathrm{D}h}{\mathrm{D}t}, \tag{3.90}$$

which simplifies to

$$\boxed{\frac{\mathrm{D}Q}{\mathrm{D}t} = 0 \qquad \text{where} \qquad Q = \left(\frac{\zeta}{h}\right).} \tag{3.91}$$

The important quantity Q is known as the *potential vorticity*, and (3.91) is the potential vorticity equation. We re-derive this conservation law in a different way in Section 4.6.

Because Q is conserved on parcels, then so is any function of Q; that is, $F(Q)$ is a material invariant, where F is any function. To see this algebraically, multiply (3.91) by $F'(Q)$, the derivative of F with respect to Q, giving

$$F'(Q)\frac{\mathrm{D}Q}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t}F(Q) = 0. \tag{3.92}$$

Since F is arbitrary there are an infinite number of material invariants corresponding to different choices of F.

Effects of rotation

In a rotating frame of reference, the shallow water momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta, \tag{3.93}$$

where (as before) $\boldsymbol{f} = f\mathbf{k}$. This may be written in vector invariant form as

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{\omega}^* + \boldsymbol{f}) \times \boldsymbol{u} = -\nabla\left(g\eta + \frac{1}{2}\boldsymbol{u}^2\right), \tag{3.94}$$

and taking the curl of this gives the vorticity equation

$$\frac{\partial \zeta}{\partial t} + (\boldsymbol{u} \cdot \nabla)(\zeta + f) = -(f + \zeta)\nabla \cdot \boldsymbol{u}. \tag{3.95}$$

This is the same as the shallow water vorticity equation in a non-rotating frame, save that ζ is replaced by $\zeta + f$, the reason for this being that f is the vorticity that the fluid has by virtue of the background rotation. Thus, (3.95) is simply the equation of motion for the total or absolute vorticity, $\boldsymbol{\omega}_a = \boldsymbol{\omega}^* + \boldsymbol{f} = (\zeta + f)\mathbf{k}$.

The potential vorticity equation in the rotating case follows, much as in the non-rotating case, by combining (3.95) with the mass conservation equation, giving

$$\boxed{\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\zeta + f}{h}\right) = 0.} \tag{3.96}$$

That is, the potential vorticity in a rotating shallow system is given by $Q = (\zeta + f)/h$ and is a material invariant. (The same symbol, Q, is commonly used for many of the manifestations of potential vorticity.)

Vorticity and circulation

Although vorticity itself is not a material invariant, its integral over a horizontal material area is invariant. To demonstrate this in the non-rotating case, consider the integral

$$C = \int_A \zeta \,\mathrm{d}A = \int_A Qh \,\mathrm{d}A, \tag{3.97}$$

over a surface A, the cross-sectional area of a column of height h (as in Fig. 3.2). Taking the material derivative of this gives

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \int_A \frac{\mathrm{D}Q}{\mathrm{D}t} h \,\mathrm{d}A + \int_A Q \frac{\mathrm{D}}{\mathrm{D}t}(h \,\mathrm{d}A). \tag{3.98}$$

On the right-hand side the first term is zero, by (3.91), and the second term is just the derivative of the volume of a column of fluid of constant density and so it too is zero. Thus,

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t}\int_A \zeta \,\mathrm{d}A = 0. \tag{3.99}$$

Thus, the integral of the vorticity over some cross-sectional area of the fluid is unchanging, although both the vorticity and area of the fluid may individually change. Using Stokes' theorem, it may be written as

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t}\oint \boldsymbol{u} \cdot \,\mathrm{d}\boldsymbol{l}, \tag{3.100}$$

where the line integral is around the boundary of A. This is an example of Kelvin's circulation theorem, which we shall meet again in a more general form in Chapter 4, where we also consider the rotating case.

A slight generalization of (3.99) is possible. Consider the integral $I = \int F(Q)h \,\mathrm{d}A$ where again F is any differentiable function of its argument. It is clear that

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_A F(Q)h \,\mathrm{d}A = 0. \tag{3.101}$$

If the area of integration in (3.86) or (3.101) is the whole domain (enclosed by frictionless walls, for example) then it is clear that the integral of $hF(Q)$ is a constant, including as a special case the integral of ζ.

3.7.2 Energy Conservation: an Integral Invariant

Since we have made various simplifications in deriving the shallow water system, it is not self-evident that energy should be conserved, or indeed what form the energy takes. The kinetic energy density (KE), meaning the kinetic energy per unit area, is $\rho_0 h\boldsymbol{u}^2/2$. The potential energy density of the fluid is

$$\mathrm{PE} = \int_0^h \rho_0 g z \,\mathrm{d}z = \frac{1}{2}\rho_0 g h^2. \tag{3.102}$$

The factor ρ_0 appears in both kinetic and potential energies and, because it is a constant, we will omit it. For algebraic simplicity we also assume the bottom is flat, at $z = 0$.

Using the mass conservation equation (3.14b) we obtain an equation for the evolution of potential energy density, namely

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{gh^2}{2} + gh^2 \nabla \cdot \boldsymbol{u} = 0 \tag{3.103a}$$

or

$$\frac{\partial}{\partial t}\frac{gh^2}{2} + \nabla \cdot \left(\boldsymbol{u}\frac{gh^2}{2}\right) + \frac{gh^2}{2}\nabla \cdot \boldsymbol{u} = 0. \tag{3.103b}$$

From the momentum and mass continuity equations we obtain an equation for the evolution of kinetic energy density, namely

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{h\boldsymbol{u}^2}{2} + \frac{\boldsymbol{u}^2 h}{2}\nabla\cdot\boldsymbol{u} = -g\boldsymbol{u}\cdot\nabla\frac{h^2}{2} \tag{3.104a}$$

or

$$\frac{\partial}{\partial t}\frac{h\boldsymbol{u}^2}{2} + \nabla\cdot\left(\boldsymbol{u}\frac{h\boldsymbol{u}^2}{2}\right) + g\boldsymbol{u}\cdot\nabla\frac{h^2}{2} = 0. \tag{3.104b}$$

Adding (3.103b) and (3.104b) we obtain

$$\frac{\partial}{\partial t}\frac{1}{2}\left(h\boldsymbol{u}^2 + gh^2\right) + \nabla\cdot\left[\frac{1}{2}\boldsymbol{u}\left(gh^2 + h\boldsymbol{u}^2 + gh^2\right)\right] = 0, \tag{3.105}$$

or

$$\frac{\partial E}{\partial t} + \nabla\cdot\boldsymbol{F} = 0, \tag{3.106}$$

where $E = \mathrm{KE} + \mathrm{PE} = (h\boldsymbol{u}^2 + gh^2)/2$ is the density of the total energy and $\boldsymbol{F} = \boldsymbol{u}(h\boldsymbol{u}^2/2 + gh^2)$ is the energy flux. If the fluid is confined to a domain bounded by rigid walls, on which the normal component of velocity vanishes, then on integrating (3.105) over that area and using Gauss's theorem, the total energy is seen to be conserved; that is

$$\frac{\mathrm{d}\widehat{E}}{\mathrm{d}t} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_A \left(h\boldsymbol{u}^2 + gh^2\right)\mathrm{d}A = 0. \tag{3.107}$$

Such an energy principle also holds in the case with bottom topography. Just as we found in the case for a compressible fluid in Chapter 2, the energy flux in (3.106) is not just the energy density multiplied by the velocity; it contains an additional term $g\boldsymbol{u}h^2/2$, and this represents the energy transfer occurring when the fluid does work against the pressure force.

3.8 SHALLOW WATER WAVES

Let us now look at the gravity waves that occur in shallow water. To isolate the essence we will consider waves in a single fluid layer, with a flat bottom and a free upper surface, in which gravity provides the sole restoring force.

3.8.1 Non-rotating Shallow Water Waves

Given a flat bottom the fluid thickness is equal to the free surface displacement (Fig. 3.1), and taking the basic state of the fluid to be at rest we let

$$h(x,y,t) = H + h'(x,y,t) = H + \eta'(x,y,t), \tag{3.108a}$$
$$\boldsymbol{u}(x,y,t) = \boldsymbol{u}'(x,y,t). \tag{3.108b}$$

The mass conservation equation, (3.14b), then becomes

$$\frac{\partial \eta'}{\partial t} + (H + \eta')\nabla\cdot\boldsymbol{u}' + \boldsymbol{u}'\cdot\nabla\eta' = 0, \tag{3.109}$$

and neglecting squares of small quantities this yields the linear equation

$$\frac{\partial \eta'}{\partial t} + H\nabla\cdot\boldsymbol{u}' = 0. \tag{3.110}$$

Similarly, linearizing the momentum equation, (3.8) with $\boldsymbol{f} = 0$, yields

$$\frac{\partial \boldsymbol{u}'}{\partial t} = -g\nabla\eta'. \tag{3.111}$$

Eliminating velocity by differentiating (3.110) with respect to time and taking the divergence of (3.111) leads to

$$\frac{\partial^2\eta'}{\partial t^2} - gH\nabla^2\eta' = 0, \tag{3.112}$$

which may be recognized as a wave equation. We can find the dispersion relationship for this by substituting the trial solution

$$\eta' = \mathrm{Re}\,\widetilde{\eta}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}, \tag{3.113}$$

where $\widetilde{\eta}$ is a complex constant, $\boldsymbol{k} = \mathbf{i}k + \mathbf{j}l$ is the horizontal wavenumber and Re indicates that the real part of the solution should be taken. If, for simplicity, we restrict attention to the one-dimensional problem, with no variation in the y-direction, then substituting into (3.112) leads to the dispersion relationship

$$\omega = \pm ck, \tag{3.114}$$

where $c = \sqrt{gH}$; that is, the wave speed is proportional to the square root of the mean fluid depth and is independent of the wavenumber — the waves are dispersionless. The general solution is a superposition of all such waves, with the amplitudes of each wave (or Fourier component) being determined by the Fourier decomposition of the initial conditions.

Because the waves are dispersionless, the general solution can be written as

$$\eta'(x,t) = \frac{1}{2}\left[F(x-ct) + F(x+ct)\right], \tag{3.115}$$

where $F(x)$ is the height field at $t = 0$. From this, it is easy to see that the shape of an initial disturbance is preserved as it propagates both to the right and to the left at speed c.

3.8.2 Rotating Shallow Water (Poincaré) Waves

We now consider the effects of rotation on shallow water waves. Linearizing the rotating, flat-bottomed f-plane shallow water equations, (SW.1) and (SW.2) on page 109, about a state of rest we obtain

$$\frac{\partial u'}{\partial t} - f_0 v' = -g\frac{\partial \eta'}{\partial x}, \qquad \frac{\partial v'}{\partial t} + f_0 u' = -g\frac{\partial \eta'}{\partial y}, \qquad \frac{\partial \eta'}{\partial t} + H\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = 0. \tag{3.116a,b,c}$$

To obtain a dispersion relationship we let

$$(u, v, \eta) = (\widetilde{u}, \widetilde{v}, \widetilde{\eta})\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}, \tag{3.117}$$

and substitute into (3.116), giving

$$\begin{pmatrix} -\mathrm{i}\,\omega & -f_0 & \mathrm{i}\,gk \\ f_0 & -\mathrm{i}\,\omega & \mathrm{i}\,gl \\ \mathrm{i}\,Hk & \mathrm{i}\,Hl & -\mathrm{i}\,\omega \end{pmatrix} \begin{pmatrix} \widetilde{u} \\ \widetilde{v} \\ \widetilde{\eta} \end{pmatrix} = 0. \tag{3.118}$$

This homogeneous equation has non-trivial solutions only if the determinant of the matrix vanishes, and that condition gives

$$\omega(\omega^2 - f_0^2 - c^2K^2) = 0, \tag{3.119}$$

where $K^2 = k^2 + l^2$ and $c^2 = gH$. There are two classes of solution to (3.119). The first is simply $\omega = 0$, i.e., time-independent flow corresponding to geostrophic balance in (3.116). Because

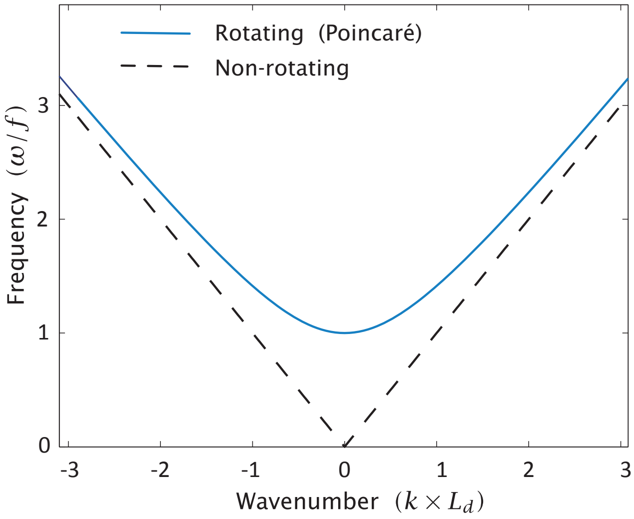

Fig. 3.8 Dispersion relation for Poincaré waves and non-rotating shallow water waves. Frequency is scaled by the Coriolis frequency f, and wavenumber by the inverse deformation radius $\sqrt{gH}/f$. For small wavenumbers the frequency of the Poincaré waves is approximately f, and for high wavenumbers it asymptotes to that of non-rotating waves.

geostrophic balance gives a divergence-free velocity field for a constant Coriolis parameter the equations are satisfied by a time-independent solution. (If the Coriolis parameter varies in space then the $\omega = 0$ solution morphs into a non-trivial dispersion relation for *Rossby waves,* considered in Chapter 5.) The second set of solutions gives the dispersion relation

$$\omega^2 = f_0^2 + c^2(k^2 + l^2), \tag{3.120}$$

or

$$\boxed{\omega^2 = f_0^2 + gH(k^2 + l^2).} \tag{3.121}$$

The corresponding waves are known as *Poincaré* waves,[6] and the dispersion relationship is illustrated in Fig. 3.8. Note that the frequency is always greater than the Coriolis frequency f_0. There are two interesting limits:

(i) The short wave limit. If

$$K^2 \gg \frac{f_0^2}{gH}, \tag{3.122}$$

where $K^2 = k^2 + l^2$, then the dispersion relationship reduces to that of the non-rotating case (3.114). This condition is equivalent to requiring that the wavelength be much shorter than the *deformation radius,* $L_d \equiv \sqrt{gH}/f$. Specifically, if $l = 0$ and $\lambda = 2\pi/k$ is the wavelength, the condition is

$$\lambda^2 \ll L_d^2(2\pi)^2. \tag{3.123}$$

The numerical factor of $(2\pi)^2$ is more than an order of magnitude, so care must be taken when deciding if the condition is satisfied in particular cases. Furthermore, the wavelength must still be longer than the depth of the fluid, otherwise the shallow water condition is not met.

(ii) The long wave limit. If

$$K^2 \ll \frac{f_0^2}{gH}, \tag{3.124}$$

that is if the wavelength is much longer than the deformation radius L_d, then the dispersion relationship is

$$\omega = f_0. \tag{3.125}$$

These are known as *inertial oscillations*. The equations of motion giving rise to them are

$$\frac{\partial u'}{\partial t} - f_0 v' = 0, \qquad \frac{\partial v'}{\partial t} + f_0 u' = 0, \tag{3.126}$$

which are equivalent to material equations for free particles in a rotating frame, unconstrained by pressure forces, namely

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} - f_0 v = 0, \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + f_0 u = 0. \tag{3.127}$$

3.8.3 Kelvin Waves

The Kelvin wave is a particular type of gravity wave that exists in the presence of both rotation and a lateral boundary. Suppose there is a solid boundary at $y = 0$; clearly harmonic solutions in the y-direction are not allowable, as these would not satisfy the condition of no normal flow at the boundary. Do any wave-like solutions exist? The affirmative answer to this question was provided by W. Thomson and the associated waves are now eponymously known as *Kelvin waves*.[7] We begin with the linearized shallow water equations, namely

$$\frac{\partial u'}{\partial t} - f_0 v' = -g\frac{\partial \eta'}{\partial x}, \qquad \frac{\partial v'}{\partial t} + f_0 u' = -g\frac{\partial \eta'}{\partial y}, \qquad \frac{\partial \eta'}{\partial t} + H\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = 0. \tag{3.128a,b,c}$$

The fact that $v' = 0$ at $y = 0$ suggests that we look for a solution with $v' = 0$ everywhere, whence these equations become

$$\frac{\partial u'}{\partial t} = -g\frac{\partial \eta'}{\partial x}, \qquad f_0 u' = -g\frac{\partial \eta'}{\partial y}, \qquad \frac{\partial \eta'}{\partial t} + H\frac{\partial u'}{\partial x} = 0. \tag{3.129a,b,c}$$

Equations (3.129a) and (3.129c) lead to the standard wave equation

$$\frac{\partial^2 u'}{\partial t^2} = c^2\frac{\partial^2 u'}{\partial x^2}, \tag{3.130}$$

where $c = \sqrt{gH}$, the usual wave speed of shallow water waves. The solution of (3.130) is

$$u' = F_1(x + ct, y) + F_2(x - ct, y), \tag{3.131}$$

with corresponding surface displacement

$$\eta' = \sqrt{H/g}\left[-F_1(x + ct, y) + F_2(x - ct, y)\right]. \tag{3.132}$$

The solution represents the superposition of two waves, one (F_1) travelling in the negative x-direction, and the other in the positive x-direction. To obtain the y dependence of these functions we use (3.129b) which gives

$$\frac{\partial F_1}{\partial y} = \frac{f_0}{\sqrt{gH}}F_1, \qquad \frac{\partial F_2}{\partial y} = -\frac{f_0}{\sqrt{gH}}F_2, \tag{3.133}$$

with solutions

$$F_1 = F(x + ct)\mathrm{e}^{y/L_d}, \qquad F_2 = G(x - ct)\mathrm{e}^{-y/L_d}, \tag{3.134}$$

where $L_d = \sqrt{gH}/f_0$ is the radius of deformation. If we consider flow in the half-plane in which $y > 0$, then for positive f_0 the solution F_1 grows exponentially away from the wall, and so fails to satisfy the condition of boundedness at infinity. It thus must be eliminated, leaving the general solution

$$u' = \mathrm{e}^{-y/L_d} G(x - ct), \qquad v' = 0,$$
$$\eta' = \sqrt{H/g}\,\mathrm{e}^{-y/L_d} G(x - ct). \tag{3.135a,b,c}$$

These are Kelvin waves, and they decay exponentially away from the boundary. In general, for f_0 positive the boundary is to the right of an observer moving with the wave. Given a constant Coriolis parameter, we could equally well have obtained a solution on a meridional wall, in which case we would find that the wave again moves such that the wall is to the right of the wave direction. (This is obvious once it is realized that f-plane dynamics are isotropic in x and y.) Thus, in the Northern Hemisphere the wave moves anticlockwise round a basin, and conversely in the Southern Hemisphere, and in both hemispheres the direction is cyclonic.

3.9 GEOSTROPHIC ADJUSTMENT

We noted in Chapter 2 that the large-scale, extratropical circulation of the atmosphere is in near-geostrophic balance. Why is this? Why should the Rossby number be small? Arguably, the magnitude of the velocity in the atmosphere and ocean is ultimately given by the strength of the forcing, and so ultimately by the differential heating between pole and equator (although even this argument is not satisfactory, since the forcing mainly determines the energy throughput, not directly the energy itself, and the forcing is itself dependent on the atmosphere's response). But even supposing that the velocity magnitudes are given, there is no a-priori guarantee that the forcing or the dynamics will produce length scales that are such that the Rossby number is small. However, there is in fact a powerful and ubiquitous process whereby a fluid in an initially unbalanced state naturally evolves toward a state of geostrophic balance, namely *geostrophic adjustment*. This process occurs quite generally in rotating fluids, whether stratified or not. To pose the problem in a simple form we consider the free evolution of a single shallow layer of fluid whose initial state is manifestly unbalanced, and we suppose that surface displacements are small so that the evolution of the system is described by the linearized shallow equations of motion. These are

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta, \qquad \frac{\partial \eta}{\partial t} + H\nabla\cdot\boldsymbol{u} = 0, \tag{3.136a,b}$$

where η is the free surface displacement and H is the mean fluid depth, and we omit the primes on the linearized variables.

3.9.1 Non-rotating Flow

We consider first the non-rotating problem set, with little loss of generality, in one dimension. We suppose that initially the fluid is at rest but with a simple discontinuity in the height field so that

$$\eta(x, t = 0) = \begin{cases} +\eta_0 & x < 0 \\ -\eta_0 & x > 0, \end{cases} \tag{3.137}$$

and $u(x, t = 0) = 0$ everywhere. We can realize these initial conditions physically by separating two fluid masses of different depths by a thin dividing wall, and then quickly removing the wall. What is the subsequent evolution of the fluid? The general solution to the linear problem is given by (3.115) where the functional form is determined by the initial conditions so that here

$$F(x) = \eta(x, t = 0) = -\eta_0 \,\mathrm{sgn}(x). \tag{3.138}$$

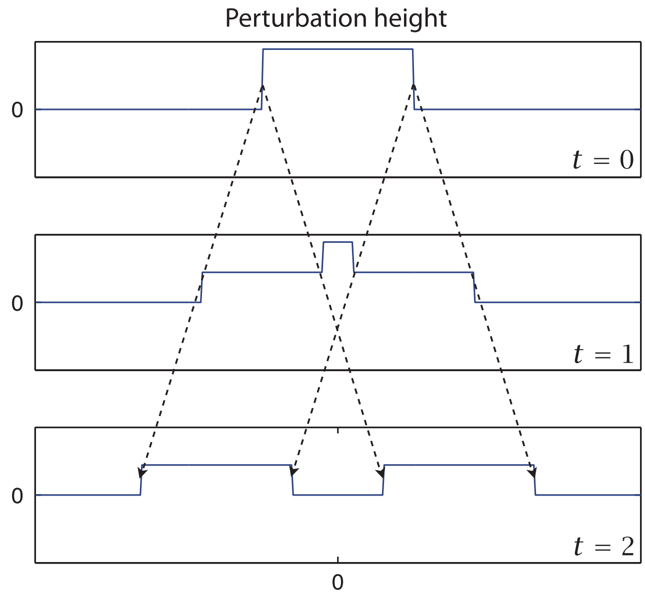

Fig. 3.9 The time development of an initial 'top hat' height disturbance, with zero initial velocity, in non-rotating flow. Fronts propagate in both directions, and the velocity is non-zero between fronts, but ultimately the disturbances are radiated away to infinity, and the fluid is left at rest with zero perturbation height.

Equation (3.115) states that this initial pattern is propagated to the right and to the left. That is, two discontinuities in fluid height move to the right and left at a speed $c = \sqrt{gH}$. Specifically, the solution is

$$\eta(x,t) = -\frac{1}{2}\eta_0[\mathrm{sgn}(x+ct) + \mathrm{sgn}(x-ct)]. \tag{3.139}$$

The initial conditions may be much more complex than a simple front, but, because the waves are dispersionless, the solution is still simply a sum of the translation of those initial conditions to the right and to the left at speed c. The velocity field in this class of problem is obtained from

$$\frac{\partial u}{\partial t} = -g\frac{\partial \eta}{\partial x}, \tag{3.140}$$

which gives, using (3.115),

$$u = -\frac{g}{2c}[F(x+ct) - F(x-ct)]. \tag{3.141}$$

Consider the case with initial conditions given by (3.137). At a given location, away from the initial disturbance, the fluid remains at rest and undisturbed until the front arrives. After the front has passed, the fluid surface is again undisturbed and the velocity is uniform and non-zero. Specifically:

$$\eta = \begin{cases} -\eta_0 \mathrm{sgn}(x) \\ 0 \end{cases} \qquad u = \begin{cases} 0 & |x| > ct \\ (\eta_0 g/c) & |x| < ct. \end{cases} \tag{3.142}$$

The solution with 'top-hat' initial conditions in the height field, and zero initial velocity, is a superposition of two discontinuities similar to (3.142) and is illustrated in Fig. 3.9. Two fronts propagate in either direction from each discontinuity and, in this case, the final velocity, as well as the fluid displacement, is zero after all the fronts have passed. That is, the disturbance is radiated completely away.

3.9.2 Rotating Flow

Rotation makes a profound difference to the adjustment problem of the shallow water system, because a steady, adjusted, solution can exist with non-zero gradients in the height field — the associated pressure gradients being balanced by the Coriolis force — and potential vorticity conservation provides a powerful constraint on the fluid evolution.[8] In a rotating shallow fluid that conservation is represented by

$$\frac{\partial Q}{\partial t} + \boldsymbol{u} \cdot \nabla Q = 0, \tag{3.143}$$

where $Q = (\zeta + f)/h$. In the linear case with constant Coriolis parameter, (3.143) becomes

$$\frac{\partial q}{\partial t} = 0, \qquad q = \left(\zeta - f_0 \frac{\eta}{H}\right). \tag{3.144}$$

This equation may be obtained either from the linearized velocity and mass conservation equations, (3.136), or from (3.143) directly. In the latter case, we write

$$Q = \frac{\zeta + f_0}{H + \eta} \approx \frac{1}{H}(\zeta + f_0)\left(1 - \frac{\eta}{H}\right) \approx \frac{1}{H}\left(f_0 + \zeta - f_0 \frac{\eta}{H}\right) = \frac{f_0}{H} + \frac{q}{H}, \tag{3.145}$$

having used $f_0 \gg |\zeta|$ and $H \gg |\eta|$. The term f_0/H is a constant and so dynamically unimportant, as is the H^{-1} factor multiplying q. Further, the advective term $\boldsymbol{u} \cdot \nabla Q$ becomes $\boldsymbol{u} \cdot \nabla q$, and this is second order in perturbed quantities and so is neglected. Thus, making these approximations, (3.143) reduces to (3.144). The potential vorticity field is therefore fixed in space! Of course, this was also true in the non-rotating case where the fluid is initially at rest. Then $q = \zeta = 0$ and the fluid remains irrotational throughout the subsequent evolution of the flow. However, this is rather a weak constraint on the subsequent evolution of the fluid; it does nothing, for example, to prevent the conversion of all the potential energy to kinetic energy. In the rotating case the potential vorticity is non-zero, and potential vorticity conservation and geostrophic balance are all we need to infer the final steady state, assuming it exists, without solving for the details of the flow evolution, as we now see.

With an initial condition for the height field given by (3.137), the initial potential vorticity is given by

$$q(x, y) = \begin{cases} -f_0\eta_0/H & x < 0 \\ \ \ f_0\eta_0/H & x > 0, \end{cases} \tag{3.146}$$

and this remains unchanged throughout the adjustment process. The final steady state is then the solution of the equations

$$\zeta - f_0 \frac{\eta}{H} = q(x, y), \qquad f_0 u = -g\frac{\partial \eta}{\partial y}, \qquad f_0 v = g\frac{\partial \eta}{\partial x}, \tag{3.147a,b,c}$$

where $\zeta = \partial v/\partial x - \partial u/\partial y$. Because the Coriolis parameter is constant, the velocity field is horizontally non-divergent and we may define a streamfunction $\psi = g\eta/f_0$. Equations (3.147) then reduce to

$$\left(\nabla^2 - \frac{1}{L_d^2}\right)\psi = q(x, y), \tag{3.148}$$

where $L_d = \sqrt{gH}/f_0$ is known as the *Rossby radius of deformation* or often just the 'deformation radius' or the 'Rossby radius'. It is a naturally occurring length scale in problems involving both rotation and gravity, and arises in a slightly different form in stratified fluids.

The initial conditions (3.146) admit of a nice analytic solution, for the flow will remain uniform in y, and (3.148) reduces to

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{L_d^2}\psi = \frac{f_0\eta_0}{H}\operatorname{sgn}(x). \tag{3.149}$$

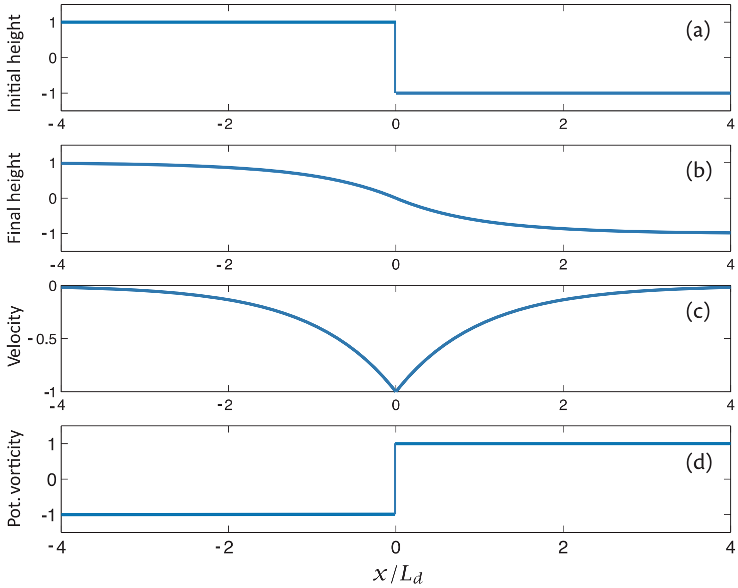

Fig. 3.10 Solutions of a linear geostrophic adjustment problem. (a) Initial height field, given by (3.137) with $\eta_0 = 1$. (b) Equilibrium (final) height field, η given by (3.150) and $\eta = f_0\psi/g$. (c) Equilibrium geostrophic velocity, normal to the gradient of height field, given by (3.151). (d) Potential vorticity, given by (3.146), and this does not evolve.

The distance, x is nondimensionalized by the deformation radius L_d and the velocity by $\eta_0(g/f_0L_d)$. Changes to the initial state occur within $\mathcal{O}(L_d)$ of the initial discontinuity.

We solve this separately for $x > 0$ and $x < 0$ and then match the solutions and their first derivatives at $x = 0$, also imposing the condition that the velocity decays to zero as $x \to \pm\infty$. The solution is

$$\psi = \begin{cases} -(g\eta_0/f_0)(1 - \mathrm{e}^{-x/L_d}) & x > 0 \\ +(g\eta_0/f_0)(1 - \mathrm{e}^{x/L_d}) & x < 0. \end{cases} \tag{3.150}$$

The velocity field associated with this is obtained from (3.147b,c), and is

$$u = 0, \qquad v = -\frac{g\eta_0}{f_0L_d}\mathrm{e}^{-|x|/L_d}. \tag{3.151}$$

The velocity is perpendicular to the slope of the free surface, and a jet forms along the initial discontinuity, as illustrated in Fig. 3.10.

The important point of this problem is that the variations in the height and field are not radiated away to infinity, as in the non-rotating problem. Rather, potential vorticity conservation constrains the influence of the adjustment to within a deformation radius (we see now why this name is appropriate) of the initial disturbance. This property is a general one in geostrophic adjustment — it also arises if the initial condition consists of a velocity jump.

A snapshot of the time evolution of flow, obtained by a numerical integration of the shallow water equations for both rotating and non-rotating flow, is illustrated in Fig. 3.11. The initial conditions are a jump in the height field, as in Fig. 3.10. Fronts propagate away at a speed $\sqrt{gH} = 1$ in both cases, but in the rotating flow they leave behind a geostrophically balanced state with a non-zero meridional velocity.

3.9.3 ✦ Energetics of Adjustment

How much of the initial potential energy of the flow is lost to infinity by gravity wave radiation, and how much is converted to kinetic energy? The linear equations (3.136) lead to

$$\frac{1}{2}\frac{\partial}{\partial t}(H\boldsymbol{u}^2 + g\eta^2) + gH\nabla\cdot(\boldsymbol{u}\eta) = 0, \tag{3.152}$$

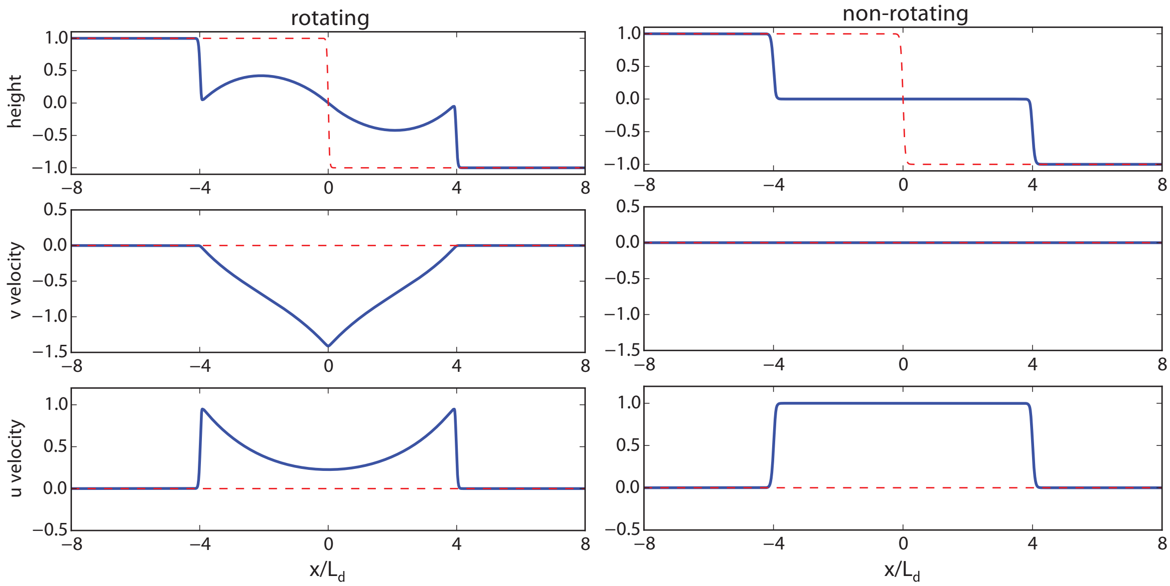

Fig. 3.11 The solutions of the shallow water equations obtained by numerically integrating the equations of motion with and without rotation. The panels show snapshots of the state of the fluid (solid lines) soon after being released from a stationary initial state (red dashed lines) with a height discontinuity. The rotating flow is evolving toward an end state similar to Fig. 3.10 whereas the non-rotating flow will eventually become stationary. In the non-rotating case L_d is defined using the rotating parameters.[9]

so that energy conservation holds in the form

$$E = \frac{1}{2}\int (H\boldsymbol{u}^2 + g\eta^2)\,\mathrm{d}\boldsymbol{x}, \qquad \frac{\mathrm{d}E}{\mathrm{d}t} = 0, \tag{3.153}$$

provided the integral of the divergence term vanishes, as it normally will in a closed domain. The fluid has a non-zero potential energy, $(1/2)\int_{-\infty}^{\infty} g\eta^2\,\mathrm{d}x$, if there are variations in fluid height, and with the initial conditions (3.137) the initial potential energy is

$$PE_I = \int_0^{\infty} g\eta_0^2\,\mathrm{d}x. \tag{3.154}$$

This is nominally infinite if the fluid has no boundaries, and the initial potential energy density is $g\eta_0^2/2$ everywhere.

In the non-rotating case, and with initial conditions (3.137), after the front has passed, the potential energy density is zero and the kinetic energy density is $Hu^2/2 = g\eta_0^2/2$, using (3.142) and $c^2 = gH$. Thus, all the potential energy is locally converted to kinetic energy as the front passes, and eventually the kinetic energy is distributed uniformly along the line. In the case illustrated in Fig. 3.9, the potential energy and kinetic energy are both radiated away from the initial disturbance. (Note that although we can superpose the solutions from different initial conditions, we cannot superpose their potential and kinetic energies.) The general point is that the evolution of the disturbance is not confined to its initial location.

In contrast, in the rotating case the conversion from potential to kinetic energy *is largely confined to within a deformation radius of the initial disturbance,* and at locations far from the initial disturbance the initial state is essentially unaltered. The conservation of potential vorticity has prevented the complete conversion of potential energy to kinetic energy, a result that is not sensitive to the precise form of the initial conditions.

In fact, in the rotating case, some of the initial potential energy is converted to kinetic energy, some remains as potential energy and some is lost to infinity; let us calculate these amounts. The final potential energy, after adjustment, is, using (3.150),

$$PE_F = \frac{1}{2} g\eta_0^2 \left[\int_0^\infty \left(1 - \mathrm{e}^{-x/L_d}\right)^2 \mathrm{d}x + \int_{-\infty}^0 \left(1 - \mathrm{e}^{x/L_d}\right)^2 \mathrm{d}x \right]. \tag{3.155}$$

This is nominally infinite, but the change in potential energy is finite and is given by

$$PE_I - PE_F = g\eta_0^2 \int_0^\infty (2\mathrm{e}^{-x/L_d} - \mathrm{e}^{-2x/L_d})\, \mathrm{d}x = \frac{3}{2} g\eta_0^2 L_d. \tag{3.156}$$

The initial kinetic energy is zero, because the fluid is at rest, and its final value is, using (3.151),

$$KE_F = \frac{1}{2} H \int \boldsymbol{u}^2 \, \mathrm{d}x = H \left(\frac{g\eta_0}{f L_d} \right)^2 \int_0^\infty \mathrm{e}^{-2x/L_d} \, \mathrm{d}x = \frac{g\eta_0^2 L_d}{2}. \tag{3.157}$$

Thus one-third of the difference between the initial and final potential energies is converted to kinetic energy, and this is trapped within a distance of the order of a deformation radius of the disturbance; the remainder, an amount $gL_d\eta_0^2$ is radiated away and lost to infinity. In any finite region surrounding the initial discontinuity the final energy is less than the initial energy.

3.9.4 ✦ General Initial Conditions

Because of the linearity of the (linear) adjustment problem a spectral viewpoint is useful, in which the fields are represented as the sum or integral of *non-interacting* Fourier modes. For example, suppose that the height field of the initial disturbance is a two-dimensional field given by

$$\eta(0) = \iint \widetilde{\eta}_{k,l}(0) \mathrm{e}^{\mathrm{i}(kx+ly)} \, \mathrm{d}k \, \mathrm{d}l, \tag{3.158}$$

where the Fourier coefficients $\widetilde{\eta}_{k,l}(0)$ are given, and the initial velocity field is zero. Then the initial (and final) potential vorticity field is given by

$$q = -\frac{f_0}{H} \iint \widetilde{\eta}_{k,l}(0) \mathrm{e}^{\mathrm{i}(kx+ly)} \, \mathrm{d}k \, \mathrm{d}l. \tag{3.159}$$

To obtain an expression for the final height and velocity fields, we express the potential vorticity field as

$$q = \iint \widetilde{q}_{k,l} \, \mathrm{d}k \, \mathrm{d}l. \tag{3.160}$$

The potential vorticity field does not evolve, and it is related to the initial height field by

$$\widetilde{q}_{k,l} = -\frac{f_0}{H} \eta_{k,l}(0). \tag{3.161}$$

In the final, geostrophically balanced state, the potential vorticity is related to the height field by

$$q = \frac{g}{f_0} \nabla^2 \eta - \frac{f_0}{H} \eta \qquad \text{and} \qquad \widetilde{q}_{k,l} = \left(-\frac{g}{f_0} K^2 - \frac{f_0}{H} \right) \widetilde{\eta}_{k,l}, \tag{3.162a,b}$$

where $K^2 = k^2 + l^2$. Using (3.161) and (3.162), the Fourier components of the final height field satisfy

$$\left(-\frac{g}{f_0} K^2 - \frac{f_0}{H} \right) \widetilde{\eta}_{k,l} = -\frac{f_0}{H} \widetilde{\eta}_{k,l}(0) \tag{3.163}$$

or

$$\widetilde{\eta}_{k,l} = \frac{\widetilde{\eta}_{k,l}(0)}{K^2 L_d^2 + 1}. \tag{3.164}$$

In physical space the final height field is just the spectral integral of this, namely

$$\eta = \iint \widetilde{\eta}_{k,l}\, \mathrm{e}^{\mathrm{i}(kx+ly)}\, \mathrm{d}k\, \mathrm{d}l = \iint \frac{\widetilde{\eta}_{k,l}(0)\, \mathrm{e}^{\mathrm{i}(kx+ly)}}{K^2 L_d^2 + 1}\, \mathrm{d}k\, \mathrm{d}l. \tag{3.165}$$

We see that at large scales ($K^2 L_d^2 \ll 1$) $\eta_{k,l}$ is almost unchanged from its initial state; the velocity field, which is then determined by geostrophic balance, thus adjusts to the pre-existing height field. At large scales most of the energy in geostrophically balanced flow is potential energy; thus, it is energetically easier for the velocity to change to come into balance with the height field than vice versa. At small scales, however, the final height field has much less variability than it did initially.

Conversely, at small scales the height field adjusts to the velocity field. To see this, let us suppose that the initial conditions contain vorticity but have zero height displacement. Specifically, if the initial vorticity is $\nabla^2\psi(0)$, where $\psi(0)$ is the initial streamfunction, then it is straightforward to show that the final streamfunction is given by

$$\psi = \iint \widetilde{\psi}_{k,l}\, \mathrm{e}^{\mathrm{i}(kx+ly)}\, \mathrm{d}k\, \mathrm{d}l = \iint \frac{K^2 L_d^2 \widetilde{\psi}_{k,l}(0)\, \mathrm{e}^{\mathrm{i}(kx+ly)}}{K^2 L_d^2 + 1}\, \mathrm{d}k\, \mathrm{d}l. \tag{3.166}$$

The final height field is then obtained from this, via geostrophic balance, by $\eta = (f_0/g)\psi$. Evidently, for small scales ($K^2 L_d^2 \gg 1$) the streamfunction, and hence the vortical component of the velocity field, are almost unaltered from their initial values. On the other hand, at large scales the final streamfunction has much less variability than it does initially, and so the height field is largely governed by whatever variation it (and not the velocity field) had initially. In general, the final state is a superposition of the states given by (3.165) and (3.166). The divergent component of the initial velocity field does not affect the final state because it has no potential vorticity, and so all of the associated energy is eventually lost to infinity.

Finally, we remark that just as in the problem with a discontinuous initial height profile, the change in total energy during adjustment is negative — this can be seen from the form of the integrals above, although we leave the specifics as a problem to the reader. That is, some of the initial potential and kinetic energy is lost to infinity, but some is trapped by the potential vorticity constraint.

3.9.5 A Variational Perspective

In the non-rotating problem, all of the initial potential energy is eventually radiated away to infinity. In the rotating problem, the final state contains both potential and kinetic energy. Why is the energy not all radiated away to infinity? It is because potential vorticity conservation on parcels prevents all of the energy being dispersed. This suggests that it may be informative to think of the geostrophic adjustment problem as a *variational problem:* we seek to minimize the energy consistent with the conservation of potential vorticity. We stay in the linear approximation in which, because the advection of potential vorticity is neglected, potential vorticity remains constant at each point.

The energy of the flow is given by the sum of potential and kinetic energies, namely

$$\text{energy} = \int (H\boldsymbol{u}^2 + g\eta^2)\, \mathrm{d}A, \tag{3.167}$$

(where $\mathrm{d}A \equiv \mathrm{d}x\, \mathrm{d}y$) and the potential vorticity field is

$$q = \zeta - f_0 \frac{\eta}{H} = (v_x - u_y) - f_0 \frac{\eta}{H}, \tag{3.168}$$

where the subscripts x, y denote derivatives. The problem is then to extremize the energy subject to potential vorticity conservation. This is a constrained problem in the calculus of variations, sometimes called an *isoperimetric* problem because of its origins in maximizing the area of a surface for a given perimeter.[10] The mathematical problem is to extremize the integral

$$I = \int \left\{ H(u^2 + v^2) + g\eta^2 + \lambda(x, y)\left[(v_x - u_y) - f_0\eta/H\right] \right\} \mathrm{d}A, \tag{3.169}$$

where $\lambda(x, y)$ is a Lagrange multiplier, undetermined at this stage. It is a function of space: if it were a constant, the integral would merely extremize energy subject to a given integral of potential vorticity, and rearrangements of potential vorticity (which here we wish to disallow) would leave the integral unaltered.

As there are three independent variables there are three Euler–Lagrange equations that must be solved in order to minimize I. These are

$$\begin{gathered} \frac{\partial L}{\partial \eta} - \frac{\partial}{\partial x}\frac{\partial L}{\partial \eta_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial \eta_y} = 0, \\ \frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\frac{\partial L}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial u_y} = 0, \qquad \frac{\partial L}{\partial v} - \frac{\partial}{\partial x}\frac{\partial L}{\partial v_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial v_y} = 0, \end{gathered} \tag{3.170}$$

where L is the integrand on the right-hand side of (3.169). Substituting the expression for L into (3.170) gives, after a little algebra,

$$2g\eta - \frac{\lambda f_0}{H} = 0, \qquad 2Hu + \frac{\partial \lambda}{\partial y} = 0, \qquad 2Hv - \frac{\partial \lambda}{\partial x} = 0, \tag{3.171}$$

and then eliminating λ gives the simple relationships

$$u = -\frac{g}{f_0}\frac{\partial \eta}{\partial y}, \qquad v = \frac{g}{f_0}\frac{\partial \eta}{\partial x}, \tag{3.172}$$

which are the equations of geostrophic balance. Thus, in the linear approximation, *geostrophic balance is the minimum energy state for a given field of potential vorticity.*

3.10 ISENTROPIC COORDINATES

We now return to the continuously stratified primitive equations, and consider the use of potential density as a vertical coordinate. In practice this means using potential temperature in the atmosphere and (for simple equations of state) buoyancy in the ocean; such coordinate systems are generically called *isentropic coordinates,* and sometimes *isopycnal coordinates* if density is used. This may seem an odd thing to do but for adiabatic flow the resulting equations of motion have an attractive form that aids the interpretation of large-scale flow. The thermodynamic equation becomes a statement for the conservation of the mass of fluid with a given value of potential density and, because the flow of both the atmosphere and the ocean is largely along isentropic surfaces, the momentum and vorticity equations have a quasi-two-dimensional form.

The particular choice of vertical coordinate is determined by the form of the thermodynamic equation in the equation-set at hand; thus, if the thermodynamic equation is $\mathrm{D}\theta/\mathrm{D}t = \dot{\theta}$, we transform the equations from (x, y, z) coordinates to (x, y, θ) coordinates. The material derivative in this coordinate system is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + u\left(\frac{\partial}{\partial x}\right)_\theta + v\left(\frac{\partial}{\partial y}\right)_\theta + \frac{\mathrm{D}\theta}{\mathrm{D}t}\frac{\partial}{\partial \theta} = \frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla_\theta + \dot{\theta}\frac{\partial}{\partial \theta}, \tag{3.173}$$

where the last term on the right-hand side is zero for adiabatic flow.

3.10.1 A Hydrostatic Boussinesq Fluid

In the simple Boussinesq equations (see the table on page 74) the buoyancy is the relevant thermodynamic variable. With hydrostatic balance the horizontal and vertical momentum equations are, in height coordinates,

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla\phi, \qquad b = \frac{\partial\phi}{\partial z}, \tag{3.174}$$

where b is the buoyancy, the variable analogous to the potential temperature θ of an ideal gas. The thermodynamic equation is

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b}, \tag{3.175}$$

and because $b = -g\delta\rho/\rho_0$, isentropic coordinates are analogous to isopycnal coordinates.

Using (2.142) the horizontal pressure gradient may be transformed to isentropic coordinates:

$$\left(\frac{\partial\phi}{\partial x}\right)_z = \left(\frac{\partial\phi}{\partial x}\right)_b - \left(\frac{\partial z}{\partial x}\right)_b \frac{\partial\phi}{\partial z} = \left(\frac{\partial\phi}{\partial x}\right)_b - b\left(\frac{\partial z}{\partial x}\right)_b = \left(\frac{\partial M}{\partial x}\right)_b, \tag{3.176}$$

where

$$M \equiv \phi - zb. \tag{3.177}$$

Thus, the horizontal momentum equation becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla_b M. \tag{3.178}$$

where the material derivative is given by (3.173), with b replacing θ. Using (3.177) the hydrostatic equation becomes

$$\frac{\partial M}{\partial b} = -z. \tag{3.179}$$

The mass continuity equation may be derived by noting that for a Boussinesq fluid the mass element may be written as

$$\delta m = \rho_0 \frac{\partial z}{\partial b} \delta b\, \delta x\, \delta y. \tag{3.180}$$

The mass continuity equation, $\mathrm{D}\delta m/\mathrm{D}t = 0$, becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{\partial z}{\partial b} + \frac{\partial z}{\partial b}\nabla_3 \cdot \boldsymbol{v} = 0, \tag{3.181}$$

where $\nabla_3 \cdot \boldsymbol{v} = \nabla_b \cdot \boldsymbol{u} + \partial\dot{b}/\partial b$ is the three-dimensional derivative of the velocity in isentropic coordinates. Equation (3.181) may thus be written

$$\frac{\mathrm{D}\sigma}{\mathrm{D}t} + \sigma\nabla_b \cdot \boldsymbol{u} = -\sigma\frac{\partial\dot{b}}{\partial b}, \tag{3.182}$$

where $\sigma \equiv \partial z/\partial b$ is a measure of the thickness between two isentropic surfaces and the material derivative is given by (3.173) with θ replaced by b. Equations (3.178), (3.179) and (3.182) comprise a closed set, with dependent variables $\boldsymbol{u}$, M and z in the space of independent variables x, y and b.

3.10.2 A Hydrostatic Ideal Gas

Deriving the equations of motion for this system requires a little more work than in the Boussinesq case but the idea is the same. For an ideal gas in hydrostatic balance we have, using (1.110),

$$\frac{\delta\theta}{\theta} = \frac{\delta T}{T} - \kappa\frac{\delta p}{p} = \frac{\delta T}{T} + \frac{\delta\Phi}{c_pT} = \frac{1}{c_pT}\delta M, \tag{3.183}$$

where $\delta\Phi = g\delta z$ and $M \equiv c_pT + \Phi$ is the 'Montgomery potential', equal to the dry static energy. (We use some of the same symbols as in the Boussinesq case to facilitate comparison, but their meanings are slightly different.) From this

$$\frac{\partial M}{\partial\theta} = \Pi, \tag{3.184}$$

where $\Pi \equiv c_pT/\theta = c_p(p/p_R)^{R/c_p}$ is the 'Exner function'. Equation (3.184) represents the hydrostatic relation in isentropic coordinates. Note also that $M = \theta\Pi + \Phi$.

To obtain an appropriate form for the horizontal pressure gradient force first note that, in the usual height coordinates, it is given by

$$\frac{1}{\rho}\nabla_z p = \theta\nabla_z\Pi, \tag{3.185}$$

where $\Pi = c_pT/\theta$. Using (2.142) gives

$$\theta\nabla_z\Pi = \theta\nabla_\theta\Pi - \frac{\theta}{g}\frac{\partial\Pi}{\partial z}\nabla_\theta\Phi. \tag{3.186}$$

Then, using the definition of Π and the hydrostatic approximation to help evaluate the vertical derivative, we obtain

$$\frac{1}{\rho}\nabla_z p = c_p\nabla_\theta T + \nabla_\theta\Phi = \nabla_\theta M. \tag{3.187}$$

Thus, the horizontal momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} = -\nabla_\theta M. \tag{3.188}$$

Much as in the Boussinesq case, the mass continuity equation may be derived by noting that the mass element may be written as

$$\delta m = -\frac{1}{g}\frac{\partial p}{\partial\theta}\delta\theta\,\delta x\,\delta y. \tag{3.189}$$

The mass continuity equation, $\mathrm{D}\delta m/\mathrm{D}t = 0$, becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{\partial p}{\partial\theta} + \frac{\partial p}{\partial\theta}\nabla_3\cdot\boldsymbol{v} = 0 \quad \text{or} \quad \frac{\mathrm{D}\sigma}{\mathrm{D}t} + \sigma\nabla_\theta\cdot\boldsymbol{u} = -\sigma\frac{\partial\dot\theta}{\partial\theta}, \tag{3.190a,b}$$

where now $\sigma \equiv \partial p/\partial\theta$ is a measure of the (pressure) thickness between two isentropic surfaces. Equations (3.184), (3.188) and (3.190b) form a closed set, analogous to (3.179), (3.178) and (3.182).

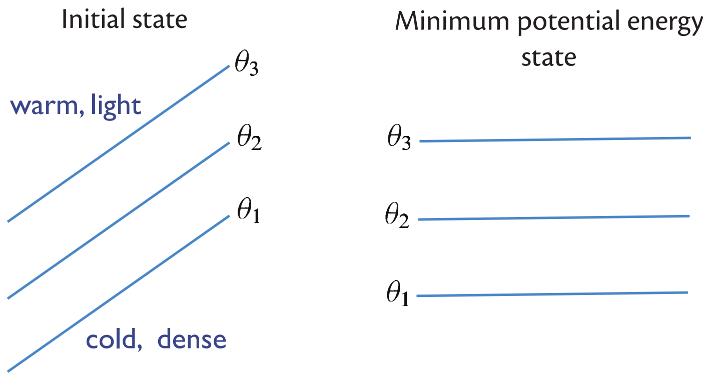

Fig. 3.12 If a stably stratified initial state with sloping isentropes (left) is adiabatically rearranged then the state of minimum potential energy has flat isentropes, as on the right, but the amount of fluid contained between each isentropic surface is unchanged. The difference between the potential energies of the two states is the *available potential energy*.

3.10.3 ✦ Analogy to Shallow Water Equations

The equations of motion in isentropic coordinates have an analogy with the shallow water equations, and we may think of the shallow water equations as a finite-difference representation of the primitive equations written in isentropic coordinates, or think of the latter as the continuous limit of the shallow water equations as the number of layers increases. For example, consider a two-isentropic-level representation of (3.184), (3.188) and (3.190), in which the lower boundary is an isentrope. A natural finite differencing gives

$$-M_1 = \Pi_0 \Delta\theta_0, \qquad M_1 - M_2 = \Pi_1 \Delta\theta_1, \tag{3.191a,b}$$

where the $\Delta\theta$s are constants, and the momentum equations for each layer become

$$\frac{\mathrm{D}\boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_1 = -\Delta\theta_0 \nabla \Pi_0, \qquad \frac{\mathrm{D}\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_2 = -\Delta\theta_0 \nabla \Pi_0 - \Delta\theta_1 \nabla \Pi_1. \tag{3.192}$$

Together with the mass continuity equation for each level these are similar to the two-layer shallow water equations (3.49).

3.11 AVAILABLE POTENTIAL ENERGY

We now revisit the issue of the internal and potential energy in stratified flow, motivated by the following remarks. In adiabatic, inviscid flow the total amount of energy is conserved, and there are conversions between internal energy, potential energy and kinetic energy. In an ideal gas the potential energy and the internal energy of a column extending throughout the atmosphere are in a constant ratio to each other — their sum is called the total potential energy. In a simple Boussinesq fluid, energetic conversions involve only the potential and kinetic energy, and not the internal energy. Yet, plainly, in neither a Boussinesq fluid nor an ideal gas can *all* the total potential energy in a fluid be converted to kinetic energy, for then all of the fluid would be adjacent to the ground and the fluid would have no thickness. Given a state of the atmosphere or ocean, how much of its total potential energy is available for conversion to kinetic energy? In particular, because energy is conserved only in adiabatic flow, we may usefully ask: how much potential energy is available for conversion to kinetic energy under an adiabatic rearrangement of fluid parcels?

Suppose that at any given time the flow is stably stratified, but that the isentropes (or more generally the surfaces of constant potential density) are sloping, as in Fig. 3.12. The potential energy of the system would be reduced if the isentropes were flattened, for then heavier fluid would be moved to lower altitudes, with lighter fluid replacing it at higher altitudes. In an adiabatic rearrangement the amount of fluid between the isentropes would remain constant, and a state with flat isentropes (meaning parallel to the geopotential surfaces) evidently constitutes a state of minimum total potential energy. The difference between the total potential energy of the fluid and the

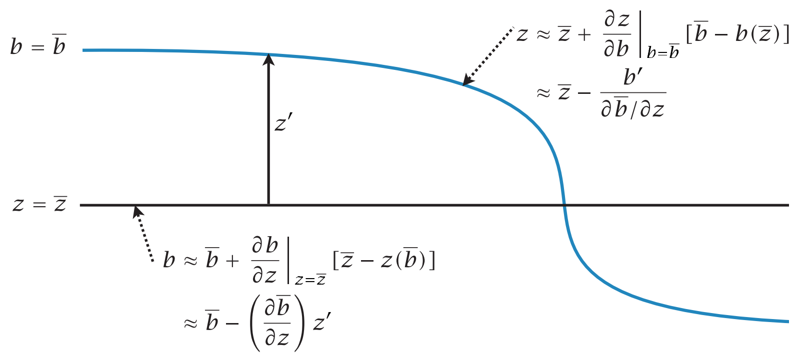

Fig. 3.13 An isopycnal surface, $b = \overline{b}$, and the constant height surface, $z = \overline{z}$, where $\overline{z}$ is the height of the isopycnal surface after a rearrangement to a minimum potential energy state, equal to the average height of the isopycnal surface. The values of z on the isopycnal surface, and of b on the constant height surface, can be obtained by the Taylor expansions shown. For an ideal gas in pressure coordinates, replace z by p and b by θ.

total potential energy after an adiabatic rearrangement to a state in which the isentropic surfaces are flat is called the *available potential energy*, or APE.[11]

3.11.1 A Boussinesq Fluid

The potential energy of a column of a Boussinesq fluid of unit area is given by

$$P = -\int_0^H bz\,\mathrm{d}z = -\int_0^H \frac{b}{2}\,\mathrm{d}z^2. \tag{3.193}$$

and the potential energy of the entire fluid is given by the horizontal integral of this. The minimum potential energy of the fluid arises after an adiabatic rearrangement in which the isopycnals are flattened, and the resulting buoyancy is only a function of z. The available potential energy is then the difference between the energy of the initial state and of this minimum state, and to obtain an approximate expression for this we first integrate (3.193) by parts to give

$$P = \frac{1}{2}\int_0^{b_m} z^2\,\mathrm{d}b - \left[\frac{bz^2}{2}\right]_0^H = \frac{1}{2}\int_0^{b_m} z^2\,\mathrm{d}b - \frac{b_m H^2}{2}, \tag{3.194}$$

where b_m is the maximum value of b in the domain, and we may formally take the upper boundary to have this value of b without affecting the final result. The minimum potential energy state arises when z is a function only of b, $z = Z(b)$ say. Because mass is conserved in the rearrangement, Z is equal to the horizontally averaged value of z on a given isopycnal surface, $\overline{z}$, and the surfaces $\overline{z}$ and $\overline{b}$ thus define each other completely. The average available potential energy, per unit area, is then given by

$$\mathrm{APE} = \frac{1}{2}\int_0^{b_m} (\overline{z^2} - \overline{z}^2)\,\mathrm{d}b = \frac{1}{2}\int_0^{b_m} \overline{z'^2}\,\mathrm{d}b, \tag{3.195}$$

where $z = \overline{z} + z'$; that is, z' is the height variation of an isopycnal surface, and the last term on the right-hand side of (3.194) has cancelled with an identical term in the expression for the potential energy of the re-arranged state. The available potential energy is thus proportional to the integral of the variance of the altitude of such a surface, and it is a positive-definite quantity. To obtain an expression in z-coordinates, we express the height variations on an isopycnal surface in terms of buoyancy variations on a surface of constant height by Taylor-expanding the height about its value on the isopycnal surface. Referring to Fig. 3.13 this gives

$$z(\overline{b}) = \overline{z} + \left.\frac{\partial z}{\partial b}\right|_{b=\overline{b}} [\overline{b} - b(\overline{z})] = \overline{z} - \left.\frac{\partial z}{\partial b}\right|_{b=\overline{b}} b', \tag{3.196}$$

where $b' = b(\overline{z}) - \overline{b}$ is corresponding buoyancy perturbation on the $\overline{z}$ surface and $\overline{b}$ is the average value of b on the $\overline{z}$ surface. Furthermore, $\partial z/\partial b|_{z=\overline{b}} \approx \partial\overline{z}/\partial b \approx (\partial\overline{b}/\partial z)^{-1}$, and (3.196) thus becomes

$$z' = z(\overline{b}) - \overline{z} \approx -b'\left(\frac{\partial\overline{z}}{\partial b}\right) \approx -\frac{b'}{(\partial\overline{b}/\partial z)}, \tag{3.197}$$

where $z' = z(b) - \overline{z}$ is the height perturbation of the isopycnal surface, from its average value. Using (3.197) in (3.195) we obtain an expression for the APE per unit area, to wit

$$\boxed{\mathrm{APE} \approx \frac{1}{2}\int_0^H \frac{\overline{b'^2}}{\partial\overline{b}/\partial z}\,\mathrm{d}z.} \tag{3.198}$$

The total APE of the fluid is the horizontal integral of the above, and so is proportional to the variance of the buoyancy on a height surface. We emphasize that APE is not defined for a single column of fluid, for it depends on the variations of buoyancy over a horizontal surface. Note too that the derivation neglects the effects of topography; this, and the use of a basic-state stratification, effectively restrict the use of (3.198) to a single ocean basin, and even for that the approximations used limit the accuracy of the expressions.

3.11.2 An Ideal Gas

The expression for the APE for an ideal gas is obtained, *mutatis mutandis,* in the same way as for a Boussinesq fluid and the trusting reader may skip directly to (3.206). The internal energy of an ideal gas column of unit area is given by

$$I = \int_0^\infty c_v T\rho\,\mathrm{d}z = \int_0^{p_s} \frac{c_v}{g}T\,\mathrm{d}p, \tag{3.199}$$

where p_s is the surface pressure, and the corresponding potential energy is given by

$$P = \int_0^\infty \rho g z\,\mathrm{d}z = \int_0^{p_s} z\,\mathrm{d}p = \int_0^\infty p\,\mathrm{d}z = \int_0^{p_s} \frac{R}{g}T\,\mathrm{d}p. \tag{3.200}$$

In (3.199) we use hydrostasy, and in (3.200) the equalities make successive use of hydrostasy, an integration by parts, hydrostasy and the ideal gas relation. Thus, the total potential energy (TPE) is given by

$$\mathrm{TPE} \equiv I + P = \frac{c_p}{g}\int_0^{p_s} T\,\mathrm{d}p. \tag{3.201}$$

Using the ideal gas equation of state we can write this as

$$\mathrm{TPE} = \frac{c_p}{g}\int_0^{p_s}\left(\frac{p}{p_s}\right)^\kappa \theta\,\mathrm{d}p = \frac{c_p p_s}{g(1+\kappa)}\int_0^\infty \left(\frac{p}{p_s}\right)^{\kappa+1}\mathrm{d}\theta, \tag{3.202}$$

after an integration by parts. (We omit a term proportional to $p_s\theta_s$ that arises in the integration by parts, because it cancels in a similar fashion to the boundary term in the Boussinesq derivation; or take $\theta_s = 0$.) The total potential energy of the entire fluid is equal to a horizontal integral of (3.202). The minimum total potential energy arises when the pressure in (3.202) is a function only of θ, $p = P(\theta)$, where by conservation of mass P is the average value of the original pressure on the isentropic surface, $P = \overline{p}$. The average available potential energy per unit area is then given by the difference between the initial state and this minimum, namely

$$\mathrm{APE} = \frac{c_p p_s}{g(1+\kappa)}\int_0^\infty \left[\overline{\left(\frac{p}{p_s}\right)^{\kappa+1}} - \left(\frac{\overline{p}}{p_s}\right)^{\kappa+1}\right]\mathrm{d}\theta, \tag{3.203}$$

which is a positive-definite quantity. A useful approximation to this is obtained by expressing the right-hand side in terms of the variance of the potential temperature on a pressure surface. We first use the binomial expansion to expand $p^{\kappa+1} = (\overline{p} + p')^{\kappa+1}$. Neglecting third- and higher-order terms (3.203) becomes

$$\mathrm{APE} = \frac{R\overline{p}_s}{2g} \int_0^\infty \left(\frac{\overline{p}}{\overline{p}_s}\right)^{\kappa+1} \overline{\left(\frac{p'}{\overline{p}}\right)^2}\, \mathrm{d}\theta. \tag{3.204}$$

The variable $p' = p(\theta) - \overline{p}$ is a pressure perturbation on an isentropic surface, and is related to the potential temperature perturbation on an isobaric surface by [cf. (3.197)]

$$p' \approx -\theta' \frac{\partial \overline{p}}{\partial \theta} \approx -\frac{\theta'}{\partial\overline{\theta}/\partial p}, \tag{3.205}$$

where $\theta' = \theta(p) - \theta(\overline{p})$ is the potential temperature perturbation on the $\overline{p}$ surface. Using (3.205) in (3.204) we finally obtain

$$\boxed{\mathrm{APE} = \frac{R\overline{p}_s^{-\kappa}}{2} \int_0^{p_s} p^{\kappa-1} \left(-g\frac{\partial\overline{\theta}}{\partial p}\right)^{-1} \overline{\theta'^2}\, \mathrm{d}p.} \tag{3.206}$$

The APE is thus proportional to the variance of the potential temperature on the pressure surface or, from (3.204), proportional to the variance of the pressure on an isentropic surface.

3.11.3 Use and Interpretation

The potential energy of a fluid is reduced when the dynamics acts to flatten the isentropes. Consider, for example, Earth's atmosphere, with isentropes sloping upwards toward the pole (as in the left panel of Fig. 3.12 with the pole on the right). Flattening these isentropes amounts to a sinking of dense air and a rising of light air, and this reduction of potential energy leads to a corresponding production of kinetic energy. Thus, if the dynamics is such as to reduce the temperature gradient between equator and pole by flattening the isentropes then APE is converted to KE by that process. A statistically steady state is achieved because the heating from the Sun continually acts to restore the horizontal temperature gradient between equator and pole, thus replenishing the pool of APE, and to this extent the large-scale atmospheric circulation acts like a heat engine.

It is a useful exercise to calculate the total potential energy, the available potential energy and the kinetic energy of atmosphere and the ocean. One finds

$$\mathrm{TPE} \gg \mathrm{APE} > \mathrm{KE} \tag{3.207}$$

with, very approximately, TPE $\sim$ 100 APE and APE $\sim$ 10 KE. The first inequality should not surprise us (as it was this that led us to define APE in the first instance), but the second inequality is not obvious (and in fact the ratio is larger in the ocean). It is related to the fact that the instabilities of the atmosphere and ocean occur at a scale smaller than the size of the domain, and are unable to release all the potential energy that might be available. Understanding this more fully is the topic of Chapters 9 and 12.

Notes

1 The algorithm to solve these equations numerically differs from that of the free-surface shallow water equations because the mass conservation equation can no longer be stepped forward in time. Rather, an elliptic equation for p_{lid} must be derived by eliminating time derivatives between (3.21) using (3.20), and this is then solved at each timestep.

2 This correspondence was known to Matsuno (1966). Gill & Clarke (1974), McCreary (1985) and others also provide derivations of various kinds.

3 Chelton *et al.* (1998), who also provide maps of the first deformation radius and related quantities for the world's oceans.

4 After Margules (1903). Margules sought to relate the energy of fronts to their slope. In this same paper the notion of available potential energy arose.

5 'Form stress' is an expression derived from 'form drag', an expression commonly used in aerodynamics. In aerodynamics, form drag is the force due to the pressure difference between the front and rear of an object, or any other 'form', moving through a fluid. Aerodynamic form drag may, albeit uncommonly, also include frictional effects between the wind and the surface itself.

6 (Jules) Henri Poincaré (1854–1912) was a prodigious French mathematician, physicist and philosopher, certainly one of the greatest mathematicians living at the turn of the twentieth century. He is remembered for his original work in (among other things) algebra, topology, dynamical systems and celestial mechanics, obtaining many results in what would be called nonlinear dynamics and chaos when these fields re-emerged some 60 years later — the notion of 'sensitive dependence on initial conditions', for example, is present in his work. He obtained a number of the results of special relativity independently of Einstein, and worked on the theory of rotating fluids — hence the Poincaré waves of this chapter. He also wrote extensively and successfully for the general public on the meaning, importance and philosophy of science. Among other things he discussed whether scientific knowledge was an arbitrary convention, a notion that remains discussed and controversial to this day. (His answer: 'convention', in part, yes; 'arbitrary', no.) He was a proponent of the role of intuition in mathematical and scientific progress, and did not believe that mathematics could ever be wholly reduced to pure logic.

7 Thomson (1869). William Thompson later became Lord Kelvin.

8 As was considered by Rossby (1938).

9 The code (available from the author's web site) will also reproduce Fig. 3.9.

10 An introduction to variational problems may be found in Weinstock (1952) and a number of other textbooks. Applications to many traditional problems in mechanics are discussed by Lanczos (1970).

11 Margules (1903) introduced the concept of potential energy that is available for conversion to kinetic energy, Lorenz (1955) clarified its meaning and derived useful, approximate formulae for its computation, and there has since been a host of papers on the subject. Thus, for example, Shepherd (1993) showed that the APE is just the non-kinetic part of the pseudoenergy, Huang (1998) looked at some of the limitations of the approximate expressions in an oceanic context, and on the atmospheric side Pauluis (2007) looked at the effects of moisture.

In addition to his formulation of available potential energy, Edward Lorenz (1917–2008) made enormous contributions to the atmospheric sciences over the course of a long career spent almost entirely at MIT. He is perhaps most famous for being one of the modern founders of chaos theory as it emerged in the 1960s, and his paper *Deterministic non-periodic flow*, published in the *Journal of the Atmospheric Sciences* in 1963, was not only a watershed in meteorology but it changed the way we think about irregular systems — see also the endnotes on page 443. Lorenz also introduced the idea of empirical orthogonal functions to meteorology and wrote with clarity and insight about the atmospheric general circulation in a monograph in 1967.

In common with a number of meteorologists of his generation (Eric Eady was another), Lorenz was first educated in mathematics and then, because of World War II, was trained as a weather forecaster. After the war he moved to MIT in 1946 for a PhD, and in 1953 was hired on the MIT faculty, apparently on Jule Charney's recommendation, and stayed there for the rest of his career. An avid hiker who often spent summers in the mountains at NCAR in Boulder, he was a quiet, modest man with nothing to be modest about. See Palmer (2009) and Emanuel (2011) for longer biographical memoirs.