5
Dynamics of Large-Scale Ocean Circulation

George Veronis

5.1 Introduction and Summary

The past 30 years have witnessed a rapid evolution of circulation theory. Much of the progress can be attributed to the intuition and physical balance that have emerged from the use of simple models that isolate important processes. Major contributions along these lines were made by Stommel, Welander, and others. An excellent presentation of the ideas together with a number of significant advances appears in Stern's (1975a) book. More recently numerical simulations have provided a different attack on the problem. Processes that are difficult to study with analytical models become accessible through the latter approach. Early, climatological-type studies by Bryan have now been supplemented by numerical models oriented toward the isolation of the effects of individual mechanisms. The papers of Rhines and Holland cited below have been especially instructive.

The development of the theory for the dynamics of large-scale oceanic flows is very recent. One has only to look at the chapter on dynamics in Sverdrup, Johnson, and Fleming (1942) to realize how primitive the theory was in the mid-1940s. Sverdrup's (1947) important demonstration of the generation of planetary vorticity by wind stress was the first step in obtaining explicit information about oceanic flow from a simple external observable. Until that time the dynamic method (i.e., geostrophic-hydrostatic balance) was used to obtain flow information, but this hardly constitutes a theory since one internal property must be used to determine another.

Ekman's (1905) theory for what we now call the Ekman layer was a significant early contribution, but its application to large-scale theory was not understood until Charney and Eliassen (1949) showed the coupling to large-scale flows via the spin-up mechanism. Actually, the generation of large-scale flow by Ekman suction in the laboratory was observed and described by Pettersson (1931), who repeated some of Ekman's (1906) early experiments with a stratified fluid to determine the inhibition of vertical momentum transport by stratification. Pettersson found the large-scale circulation to be an annoying interference, however, in his primary objective, determining vertical transfer of momentum by turbulence, and he discarded the approach as unpromising.

Shortly after Sverdrup's paper Stommel (1948) produced the first significant, closed-basin circulation model showing that westward intensification of oceanic flow is due to the variation of the Coriolis parameter with latitude. Hidaka (1949) proposed a closed set of equations for the circulation including the effects of lateral (eddy) dissipation of momentum. Munk (1950) continued the development by obtaining

a solution that resembled Stommel's except for details in the boundary layers near the eastern and western sides of the basin. He applied his solution to an idealized ocean basin with observed wind stresses and related a number of observed oceanic gyres to the driving wind patterns. The first nonlinear correction to these linearized models (Munk, Groves and Carrier, 1950) showed that inertia shifts positive vortices to the south and negative vortices to the north. Nonlinear effects thus introduce the observed north-south asymmetry into a circulation pattern that is predicted by steady linear theory to be symmetric about mid-latitude when the wind driving is symmetric.

Fofonoff (1954) approached the problem from the opposite extreme, treating a completely inertial, nondriven model. His solution exhibits the pure effect of inertia for steady westward flows. The circulation pattern is symmetric in the east-west direction and closes with the center of a cyclonic (anticyclonic) vortex at the south (north) edge of the basin. When linear, frictional effects perturb the nonlinear pattern (Niiler, 1966), the center of the vortex shifts westward. Niiler's model had been proposed independently by Veronis (1966b) after a numerical study of nonlinear effects in a barotropic ocean, and Niiler's solution had been suggested heuristically by Stommel (1965).

The theoretical models leading to these results for wind-driven circulation are discussed below in sections 5.5 and 5.6. More general considerations in section 5.2, based on conservation integrals for the nondissipative equations (Welander, 1971a), prepare the way for the ordered system of quasi-geostrophic equations that are presented in section 5.3. The latter are derived for a fluid with arbitrary stable stratification and for a two-layer approximation to the stratification.[1] A large portion of the remainder of the paper reports results obtained with the simpler two-layer system.[2]

Section 5.7 concludes the discussion of simple models of steady, wind-driven circulation with a suggested simple explanation of why the Gulf Stream and other western boundary currents leave the coast and flow out to sea (Parsons, 1969; Veronis, 1973a). Separation of the Gulf Stream from the coast occurs within an anticyclonic gyre at a latitude where the Ekman drift due to an eastward wind stress in the interior must be returned geostrophically in the western boundary layer. If the mean thermocline depth is sufficiently small, i.e., if the amount of upper-layer water is sufficiently limited, the thermocline surfaces on the onshore side of the Gulf Stream and separation occurs. The surfacing of the thermocline is enhanced by the poleward transport by the Gulf Stream of upper-layer water that eventually reaches polar latitudes and sinks.

A review of models of thermohaline circulation is given in section 5.8. The open models introduced by Welander (1959) and Robinson and Stommel (1959) and the subsequent developments by them as well as other authors are described. The section concludes with a description of a closed, two-layer model in which the heating and cooling processes are parameterized by an assumed upwelling of lower-layer water across the thermocline (Veronis, 1978). The closure of the model leads to an evaluation of the magnitude of upwelling of 1.5×10^{-7} m s^{-1}, in agreement with values obtained from chemical tracers and the estimated age of deep water.

The normal modes for a two-layer system are derived in section 5.9 and the free-wave solutions are obtained for an ocean of constant depth. The derivation is a generalization of the treatment by Veronis and Stommel (1956) but the method is basically the same. The results include barotropic and baroclinic modes of inertiogravity and quasi-geostrophic Rossby waves. Brief mention is made of observations of these waves and the roles they play in developed flows.

Topography introduces a new class of long-period wave motions. Quasi-geostrophic analysis leads to the three types of waves described by Rhines (1970, 1977) as topographic-barotropic Rossby waves, fast baroclinic (bottom-trapped) waves, and slow baroclinic (surface-trapped) waves. The properties of slow baroclinic waves are independent of topography, yet the creation of these waves may be facilitated by steep topography that inhibits deep motions. For purposes of comparison the analysis is carried out with stratification approximated by two layers and by a vertically uniform density gradient.

Baroclinic instability in a two-layer system is described in section 5.11. The model (Phillips, 1951; Bretherton, 1966a) has convenient symmetries (equal layer depths and equal and opposite mean flows in the two layers) that simplify the analysis and show the nature of the instability more clearly. The stabilizing effect of β is evident after the simpler model has been analyzed. After a discussion of the energetics and of the relative phase of the upper- and lower-layer motions required for instability, the study of linear processes ends with a brief review of the stability study made by Gill, Green, and Simmons (1974) for a variety of mean oceanic conditions.

The last section extends the discussion to include the effects of turbulence and strong nonlinear interactions. Batchelor's (1953a) argument that two-dimensional turbulence leads to a red cascade in wavenumber space is followed by a description of several of Rhines's (1977) numerical experiments exhibiting the red cascade for barotropic quasi-geostrophic flow and the inhibition of the red cascade by lateral boundaries and topography. An initially turbulent flow in a two-layer fluid will evolve toward a barotropic state followed by the red cascade when nonlinear interactions or baro-

clinic instability generate motions on the scale of the internal radius of deformation. The latter scale is the window leading to barotropic behavior. Rough topography can inhibit the tendency toward barotropy by scattering the energy of the flow away from the deformation scale.

The generation of deep motions in wind-driven flows by upper-layer eddies that evolve from barotropic and baroclinic instabilities leads to a mean flow that is very different from the one predicted by the linear theories of the earlier sections. The closed-basin circulation obtained in a two-layer quasi-geostrophic numerical experiment by Holland (1978) and analyzed by Holland and Rhines (1980) shows how many of the processes described earlier come together to generate the mean flow. Simple balances for some of the results are suggested. A significant result of this experiment (and others mentioned) is the enhancement of the mean transport by the circulation resulting from the eddy interactions. A similar enhancement is made possible when topography and baroclinic effects are present (Holland, 1973). A brief discussion of several other numerical studies concludes the review.

Most of the emphasis in this paper is on linear processes and on the remaining features of the dynamics that can be used as building blocks to synthesize the involved, interactive flows observed in the ocean. Only a selected few of the many numerical studies that have emerged in the past few years are discussed, and even for those only some of the generalizable results are mentioned. Some important topics, such as the use of diagnostic models (Sarkisyan, 1977) and the generation of mean circulation by fluctuating winds (Pedlosky, 1964a; Veronis, 1970; Rhines, 1977), are omitted only because time limits forced me to draw the line somewhere. Most of the references are to the literature in the English language because that is the literature with which I am most familiar.

5.2 The Equations for Large-Scale Dynamics

The complete equations for conservation of momentum, heat, and salt are never used for studies of large-scale oceanic dynamics because they are much too complicated, not only for analytical studies but even for numerical analyses. Justification for use of an appropriate set of simplified equations requires a much more extensive argument than is feasible here so we shall confine ourselves to a short discussion with references to publications that discuss the different issues. It is appropriate, however, to mention a general result for a fluid with a simple equation of state.

If dissipative processes are ignored, the conservation of momentum for a fluid in a rotating system can be written as

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{1}{\rho}\nabla p - \nabla\Phi, \tag{5.1}$$

or equivalently as

$$\frac{\partial \mathbf{v}}{\partial t} + (2\boldsymbol{\Omega} + \nabla \times \mathbf{v}) \times \mathbf{v}$$

$$= -\frac{1}{\rho}\nabla p - \nabla\left(\frac{1}{2}\mathbf{v}\cdot\mathbf{v}\right) - \nabla\Phi \tag{5.2}$$

where $\mathbf{v}$ is the three-dimensional velocity vector, $\boldsymbol{\Omega}$ is the angular rotation vector of the system, ρ the density, p the pressure, and $\nabla\Phi$ the total gravity term (Newtonian plus rotational acceleration).

Conservation of mass is described by

$$\frac{d\rho}{dt} + \rho\nabla\cdot\mathbf{v} = 0, \qquad \frac{d}{dt} \equiv \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla. \tag{5.3}$$

Furthermore, if a state variable $s(p,\rho)$ is conserved along a trajectory, it satisfies the equation

$$\frac{ds}{dt} = 0. \tag{5.4}$$

These equations can be combined to yield the conservation of potential vorticity (Ertel, 1942):

$$\frac{d}{dt}\left[\frac{(2\boldsymbol{\Omega} + \nabla \times \mathbf{v})\cdot\nabla s}{\rho}\right] = 0. \tag{5.5}$$

This general result for a dissipation-free fluid does not apply precisely to sea water where the density is a function not only of temperature and pressure but also of the dissolved salts. The effect of salinity on density is very important in the distribution of water properties. However, for most dynamic studies the effect of the extra state variable is not significant and (5.5) is valid.

Circulation of waters in the world ocean involves trajectories from the surface to the deep sea and from one ocean basin to another. The relative densities of two parcels of water formed at the surface in different locations can be inverted when the parcels sink to great depths. Thus, surface water in the Greenland Sea is denser than surface water in the Weddell Sea; yet when these water masses sink and flow to the same geographic location, the latter (Antarctic Bottom Water) is denser and lies below the former (North Atlantic Deep Water). This inversion is due in large part to the different amounts of thermal expansion of waters of different temperatures and salinities.[3]

Neither compressibility nor individual effects of temperature and salinity on the density are included in the treatment that follows. Use of potential density (not only in the equations but in boundary conditions as well) together with the Boussinesq approximation (Spiegel and Veronis, 1960) makes it possible to treat the dynamic effects of buoyancy forces in a dynami-

cally consistent fashion. Comparison of observed motions (especially long- and short-period waves) with those deduced when potential density is used yields good qualitative, and often quantitative, agreement. But it is clear that some phenomena, such as the relative layering of water masses and small-scale mixing related to double-diffusive processes, cannot be analyzed without the use of a more extended thermodynamic analysis. Therefore, although the present discussion allows a treatment of inertially controlled flows, it does not admit the interesting array of phenomena associated with tracer distributions, except in the crudest sense. By implication, motions related to the largest time and space scales are not accessible either.

In those cases where a homogeneous fluid model is invoked the effects of stratification are implicitly present since the basic equations would be different for a truly homogeneous fluid (where the direction of the rotation axis could be more important than the local vertical). The fluid is sometimes assumed to be homogeneous only because the feature that is being emphasized is independent of stratification or because the simplified analytical treatment is a helpful preliminary for the more complicated stratified system.

The effects of rotation and Newtonian gravitation lead to an equilibrium shape for the earth that is nearly a planetary ellipsoid. For earth parameters the ellipticity is small (1/298) and an expansion in the ellipticity yields a spherical system with a mean (rather than variable) radius to lowest order (Veronis, 1973b). An additional simplification is to neglect the horizontal component of the earth's rotation. This assumption is not entirely separate from the use of a mean radius (N. A. Phillips, 1966a). It is normally valid for the types of motion treated here, though the effect of the neglected term is discussed for certain physical situations by Needler and LeBlond (1973) and by Stern (1975a). Grimshaw (1975) has reexamined the β-plane approximation and gives a procedure in which the horizontal rotation is retained.

With all these simplifications the foregoing equations simplify to

$$\frac{d\mathbf{v}}{dt} + \mathbf{f} \times \mathbf{v} = -\frac{1}{\rho_m} \nabla p - g\frac{\rho}{\rho_m} \hat{\mathbf{k}}, \tag{5.6}$$

$$\frac{d\rho}{dt} = 0, \tag{5.7}$$

$$\nabla \cdot \mathbf{v} = 0, \tag{5.8}$$

$$\frac{d}{dt}\left[(\mathbf{f} + \nabla \times \mathbf{v}) \cdot \nabla \rho\right] \equiv \frac{dq}{dt} = 0, \tag{5.9}$$

where $\mathbf{f} = 2\Omega \sin\phi \hat{\mathbf{k}}$ is twice the locally vertical (direction $\hat{\mathbf{k}}$) component of the earth's rotation, ϕ is the latitude, g is gravity, ρ_m is a mean (constant) density, and ρ is the deviation of density from the mean. The hydrostatic pressure associated with the mean density has been subtracted from the system. Equations (5.7) and (5.8) describe the incompressible nature of this Boussinesq fluid. The quantity s in (5.4) can then be replaced by ρ, and the potential vorticity q in (5.9) is simplified accordingly [note the change of dimensions of potential vorticity as defined in (5.5) and (5.9)].

For steady or statistically steady flows we can multiply (5.6) by $\mathbf{v}$ to obtain a kinetic energy equation which can be written as

$$\mathbf{v} \cdot \nabla \left(\frac{\rho_m}{2} \mathbf{v} \cdot \mathbf{v} + p + g\rho\right) \equiv \mathbf{v} \cdot \nabla B = 0, \tag{5.10}$$

where B is the Bernoulli function. In this case, since q, ρ, and B are each conserved along flow paths, any one of them can be expressed in terms of the other two and we obtain

$$\rho = \rho(B,q), \qquad B = B(\rho,q), \qquad q = q(B,\rho). \tag{5.11}$$

Even though the distributions of the surfaces cannot be determined without knowledge of the flow field, the relationship between ρ, B and q is conceptually useful.

The quantities B, q, and ρ are specified by their values in certain source regions where dissipation, mixing, and other physical processes are important. (Obvious source regions are Ekman layers, areas of convective overturning, and boundary layers near coasts.) Having acquired values of B, q, and ρ at the sources, fluid particles will retain these values along their flow paths. If particles from different sources and with different values of B, q, and ρ converge to the same geographical location, regions of discontinuity will develop, and mixing, dissipation or some other non-ideal fluid process will be required. The locations of these discontinuous regions can be determined only from a solution to the general problem, and, in general, we may anticipate new sources of B, q, and ρ to develop there. Hence, the system becomes a strongly implicit one and the closure of the problem is very complicated.

Even though a solution to the general problem may be impossible, these general considerations are important. We should be prepared for the likelihood that the solution at a particular location will not be simply determined by values at solid boundaries that are easily specified. The ocean is more likely a collection of dynamically self-contained pools (some subsurface) that interact along open-ocean boundaries where they join. Perhaps only the most persistent of these are statistically steady features. It is possible that locally the flow is relatively laminar. In that case the solution would be accessible once the source regions were identified and the values of B, q, and ρ in these regions could be specified.

5.3 The Quasi-Geostrophic Equations and the β-Plane

Even with the simplifications made in the previous section the equations are more general than required for a study of large-scale dynamics. We shall therefore simplify them further by invoking geostrophic and hydrostatic balances at lowest order and by restricting attention to spatial scales of interest. In so doing we shall derive an appropriate β-plane approximation for the study of oceanic waves and mesoscale motions. A similar procedure is followed by N. A. Phillips (1963).[4]

5.3.1 Continuous Stratification

The spherical components of (5.6) take the form

$$\frac{du}{dt} - \frac{uv\tan\phi}{a} + \frac{uw}{a} - 2\Omega\sin\phi\, v = -\frac{1}{a\cos\phi}\frac{dP}{d\lambda}, \tag{5.12}$$

$$\frac{dv}{dt} + \frac{u^2\tan\phi}{a} + \frac{vw}{a} + 2\Omega\sin\phi\, u = -\frac{1}{a}\frac{\partial P}{\partial\phi}, \tag{5.13}$$

$$\frac{dw}{dt} - \frac{u^2+v^2}{a} = -\frac{\partial P}{\partial z} - g\frac{\rho}{\rho_m}, \tag{5.14}$$

$$\frac{1}{a\cos\phi}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial}{\partial\phi}(v\cos\phi)\right] + \frac{\partial w}{\partial z} + \frac{2w}{a} = 0, \tag{5.15}$$

$$\frac{d\rho}{dt} = 0, \tag{5.16}$$

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \frac{u}{a\cos\phi}\frac{\partial}{\partial\lambda} + \frac{v}{a}\frac{\partial}{\partial\phi} + w\frac{\partial}{\partial z}, \tag{5.17}$$

where (λ, ϕ, z) are longitude, latitude, and upward and have respective velocities (u, v, w); P is p/ρ_m, a is the mean radius of the earth, and ρ is the total density minus ρ_m.

Center attention on a latitude ϕ_0, write $\phi = \phi_0 + \phi'$, and consider flows with north-south scale L substantially smaller than a. Then with $a\phi' = y$, we can expand the trigonometric functions in y, keeping only terms of $O(L/a)$, to obtain

$$\sin\phi \simeq \sin\phi_0(1 + \cot\phi_0 y/a),$$
$$\cos\phi \simeq \cos\phi_0(1 - \tan\phi_0 y/a), \tag{5.18}$$
$$f_0 = 2\Omega\sin\phi_0,$$

$$\frac{\partial}{\partial x} \equiv \frac{1}{a\cos\phi_0}\frac{\partial}{\partial\lambda}, \qquad \frac{\partial}{\partial y} \equiv \frac{1}{a}\frac{\partial}{\partial\phi}. \tag{5.19}$$

To first order in y/a the equations become

$$\frac{du}{dt} + \frac{y}{a}\tan\phi_0 u\frac{\partial u}{\partial x} + \frac{uw}{a} - \frac{uv}{a}\tan\phi_0\left(1 + \frac{2}{\sin 2\phi_0}\frac{y}{a}\right) - f_0 v\left(1 + \frac{y}{a}\cot\phi_0\right) = -\frac{\partial P}{\partial x}\left(1 + \frac{y}{a}\tan\phi_0\right), \tag{5.20}$$

$$\frac{dv}{dt} + \frac{y}{a}\tan\phi_0 u\frac{\partial v}{\partial x} + \frac{vw}{a} + \frac{u^2}{a}\tan\phi_0\left(1 + \frac{2}{\sin 2\phi_0}\frac{y}{a}\right) + f_0 u\left(1 + \frac{y}{a}\cot\phi_0\right) = -\frac{\partial P}{\partial y}, \tag{5.21}$$

$$\frac{dw}{dt} + \frac{y}{a}\tan\phi_0 u\frac{\partial w}{\partial x} - \frac{u^2+v^2}{a} = -\frac{\partial P}{\partial z} - \frac{g\rho}{\rho_m}, \tag{5.22}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} - \frac{\partial}{\partial y}\left(\frac{y}{a}v\tan\phi_0\right) - \frac{y}{a}\tan\phi_0\frac{\partial w}{\partial z} + \frac{2w}{a} = 0, \tag{5.23}$$

$$\frac{d\rho}{dt} + \frac{y}{a}\tan\phi_0 u\frac{\partial\rho}{\partial x} = 0, \tag{5.24}$$

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}. \tag{5.25}$$

Flows with a primary geostrophic balance will satisfy

$$f_0 v \sim \frac{\partial P}{\partial x}, \qquad -f_0 u \sim \frac{\partial P}{\partial y}. \tag{5.26}$$

Hydrostatic balance yields

$$\frac{\partial P}{\partial z} \sim -g\frac{\rho}{\rho_m}. \tag{5.27}$$

Variations over the depth H of the ocean are described by

$$\frac{\partial}{\partial z} \sim \frac{1}{H}, \tag{5.28}$$

so the "pressure" scale derived from (5.27) is

$$P \sim \frac{gH\Delta\rho}{\rho_m}. \tag{5.29}$$

Geostrophic balance then suggests the velocity scale

$$V \sim \frac{gH}{f_0 L}\frac{\Delta\rho}{\rho_m}. \tag{5.30}$$

If these scales are used as orders of magnitudes for the respective variables and if we also take

$$\frac{\partial}{\partial x} \sim \frac{1}{L}, \qquad \frac{\partial}{\partial y} \sim \frac{1}{L}, \qquad \frac{\partial}{\partial t} \sim f_0,$$

$$\frac{y}{a} \sim \frac{L}{a}, \quad \delta = \frac{H}{L}, \qquad w \sim V\delta,$$

we note the following.

Relative to the lowest order (in y/a) Coriolis terms, the nonlinear terms in d/dt in (5.19) and (5.20) are $O(Ro)$ where $Ro = V/f_0L$. The remaining nonlinear terms are

$$O\left(\frac{L}{a}Ro\right) \quad \text{or} \quad O\left(\delta\frac{L}{a}Ro\right).$$

In the vertical equation of motion the acceleration terms are

$$O(\delta^2), \quad O(\delta^2 Ro), \quad \text{or} \quad O\left(\delta\frac{L}{a}Ro\right)$$

when compared to the terms on the right. Observations of the flows of interest support the inequalities

$$Ro \ll 1, \quad L/a \ll 1, \quad \delta \ll 1. \tag{5.31}$$

Rather than expand the equations in powers of the small parameters we shall simply make use of (5.31) and drop all terms which involve products of Ro, δ and L/a. Also, rather than give a relative ordering of these three parameters we keep all terms up to first order in Ro, δ, and L/a, a procedure that yields the following general system of equations

$$\frac{du}{dt} - f_0 v\left(1 + \frac{y}{a}\cot\phi_0\right) = -\frac{\partial P}{\partial x}\left(1 + \frac{y}{a}\tan\phi_0\right), \tag{5.32}$$

$$\frac{dv}{dt} + f_0 u\left(1 + \frac{y}{a}\cot\phi_0\right) = -\frac{\partial P}{\partial y}, \tag{5.33}$$

$$\frac{\partial P}{\partial z} = -g\rho/\rho_m, \tag{5.34}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} - \frac{\partial}{\partial y}\left(\frac{y}{a}v\tan\phi_0\right) - \frac{y}{a}\tan\phi_0\frac{\partial w}{\partial z} = 0, \tag{5.35}$$

$$\frac{d\rho}{dt} + \frac{y}{a}\tan\phi_0 u\frac{\partial\rho}{\partial x} = 0, \tag{5.36}$$

keeping in mind that the nonlinear terms in (5.32) and (5.33) are $O(Ro)$ compared to the lowest-order Coriolis terms.

Now write

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_1, \quad P = P_0 + P_1, \quad \rho = \rho_0 + \rho_1, \tag{5.37}$$

where $(\mathbf{v}_1, P_1, \rho_1)$ are $O(Ro)$ or $O(L/a)$. We shall also assume that time variations appear at first order, i.e., $\partial/\partial t = O(Ro)$ or $O(L/a)$. Then at lowest order we obtain the expected geostrophic hydrostatic system:

$$f_0 v_0 = \frac{\partial P_0}{\partial x}, \tag{5.38}$$

$$f_0 u_0 = -\frac{\partial P_0}{\partial y}, \tag{5.39}$$

$$\frac{\partial P_0}{\partial z} = -g\rho_0/\rho_m, \tag{5.40}$$

$$\frac{\partial u_0}{\partial x} + \frac{\partial v_0}{\partial y} = 0, \quad \frac{\partial w_0}{\partial z} = 0, \tag{5.41}$$

If w_0 vanishes at any level or if it is required to satisfy inconsistent (with $\partial w_0/\partial z = 0$) boundary conditions at top and bottom, it will vanish everywhere. One or the other of these two conditions is satisfied for all of the flows that we shall consider, so we obtain the result

$$w_0 = 0. \tag{5.42}$$

This means that the scaling $w \sim V\delta$ suggested by the geometry is inappropriate and that a factor L/a or Ro should be included on the right-hand side. In other words, quasi-geostrophic flows are quasi-horizontal and the convective derivative in (5.32) reduces to

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u_0\frac{\partial}{\partial x} + v_0\frac{\partial}{\partial y}. \tag{5.43}$$

The restriction to flows with less than global scales precludes a treatment leading to the basic stratification. Since vertical density changes $\Delta\rho$ are generally much larger than the horizontal changes, say $\Delta\rho'$, generated by the motion field, we must account for the difference in (5.36). In particular, we write $\rho = \bar{\rho}(z) + \rho'(x, y, z, t)$ so that

$$\frac{\partial\rho'}{\partial t} + u\frac{\partial\rho'}{\partial x} + v\frac{\partial\rho'}{\partial y} + w\frac{\partial\rho'}{\partial z} + w\frac{\partial\bar{\rho}}{\partial z} + \frac{y}{a}\tan\phi_0 u\frac{\partial\rho'}{\partial x} = 0. \tag{5.44}$$

The considerations leading to (5.43) apply here as well for the terms involving ρ'. Accordingly, at lowest order we can drop the terms $w\,\partial\rho'/\partial z$ and $(y/a)\tan\phi_0 u(\partial\rho'/\partial x)$ to end up with

$$\frac{d\rho'}{dt} + w\frac{\partial\bar{\rho}}{\partial z} = 0, \tag{5.45}$$

where $\Delta\rho'$ is assumed to be $O(Ro)$ or $O(L/a)$ relative to $\Delta\rho$. Since w is correspondingly smaller than u or v, the two terms balance. In terms of our ordering, therefore, we can write

$$\frac{d\rho_0}{dt} + w_1\frac{\partial\bar{\rho}}{\partial z} = 0, \tag{5.46}$$

where we have used the fact that the density used in the hydrostatic equation is really ρ' (since the balance $\partial\bar{P}/\partial z = -g\bar{\rho}$ is valid when there is no motion and hence can be subtracted from the system).

At next order we have

$$\frac{du_0}{dt} - f_0 v_1 - \frac{y}{a} f_0 \cot\phi_0 v_0$$

$$= -\frac{\partial P_1}{\partial x} - \frac{y}{a}\tan\phi_0 \frac{\partial P_0}{\partial x}, \tag{5.47}$$

$$\frac{dv_0}{dt} + f_0 u_1 + \frac{y}{a} f_0 \cot\phi_0 u_0 = -\frac{\partial P_1}{\partial y}, \tag{5.48}$$

$$\frac{\partial u_1}{\partial x} + \frac{\partial v_1}{\partial y} + \frac{\partial w_1}{\partial z} - \frac{\partial}{\partial y}\left(\frac{y}{a}\tan\phi_0 v_0\right) = 0. \tag{5.49}$$

These equations, in addition to providing the balances for first-order quantities, serve the important function of closing the zero-order system when first-order terms are eliminated. Thus, cross differentiating (5.47) and (5.48) and making use of (5.38) to (5.41) and (5.49), we obtain

$$\frac{d\zeta_0}{dt} + \beta v_0 = f_0 \frac{\partial w_1}{\partial z}, \qquad \beta = \frac{f_0 \cot\phi_0}{a},$$
$$\zeta_0 = \frac{\partial v_0}{\partial x} - \frac{\partial u_0}{\partial y}. \tag{5.50}$$

From (5.46) we observe

$$\frac{\partial w_1}{\partial z} = -\frac{\partial}{\partial z}\left[\frac{1}{\partial\bar{\rho}/\partial z}\frac{d\rho_0}{dt}\right] = -\frac{d}{dt}\frac{\partial}{\partial z}\left[\rho_0 \Big/ \frac{\partial\bar{\rho}}{\partial z}\right]. \tag{5.51}$$

Also,

$$\beta v_0 = \frac{df}{dt}, \tag{5.52}$$

where $f = f_0(1 + \cot\phi_0 (y/a))$. Then using (5.38) to (5.41) to express u_0, v_0, ρ_0 in terms of p_0 we obtain the lowest-order closure

$$\frac{d}{dt}\left[\nabla^2 P_0 + f f_0 + \frac{\partial}{\partial z}\left(\frac{f_0^2 \partial P_0/\partial z}{N^2}\right)\right] = 0, \tag{5.53}$$

where $N^2 = -g(\partial\bar{\rho}/\partial z)/\rho_m$ is the square of the buoyancy frequency. Equation (5.53) describes the conservation of quasi-geostrophic potential vorticity. It is sometimes written in terms of the stream functions $\psi = P_0/f_0$,

$$\frac{d}{dt}\left[\nabla^2\psi + f + \frac{\partial}{\partial z}\left(\frac{f_0^2 \partial\psi/\partial z}{N^2}\right)\right] = 0. \tag{5.54}$$

The derivation given here has been carried out in dimensional form. It is as rigorous, though not as formal, as derivations with nondimensional variables (e.g., Pedlosky, 1964a) and has the advantage of including the intermediate equations in dimensional form. Obviously, the equations are valid only for those motions (smaller than global scale, low frequency, etc.) that satisfy the assumptions.

5.3.2 Equations More Commonly Encountered

Instead of the set (5.47)–(5.49) one more often encounters the equations with rectangular cartesian coordinates, no subscripts, and with $f = f_0 + \beta y$, i.e.,

$$\frac{du}{dt} - fv = -\frac{\partial P}{\partial x}, \tag{5.55}$$

$$\frac{\partial v}{\partial t} + fu = -\frac{\partial P}{\partial y}, \tag{5.56}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \tag{5.57}$$

This system is often used even when the flow is not quasi-geostrophic. For quasi-geostrophic flows, particularly if one makes use principally of the vorticity equation and the fact that w is really a higher order quantity, one can avoid serious errors.

For flows at low latitudes (small ϕ_0) the neglected terms ($\sim\tan\phi_0$) are small and (5.55) to (5.57) may be adequate. But errors notwithstanding, a large part of the literature deals with this more approximate system, and we shall have to refer to it frequently.

5.3.3 Layered Stratification

Continuous density stratification is frequently approximated by a series of discrete layers each of uniform density. The derivation parallels the one just given but it is easier to make use of what we have done and to note the following.

Number the layers sequentially downward from the top so that the ith layer has thickness $h^{(i)}$, density $\rho^{(i)}$, and mean thickness (for linearized cases) $H^{(i)}$. Furthermore, write $h^{(i)} = \eta^{(i)} + H^{(i)} - \eta^{(i+1)}$ so that $\eta^{(i)}$ and $\eta^{(i+1)}$ are the deviations of the top and bottom surfaces of the layer from the mean. Integrate the hydrostatic relation downward from the top surface to layer i to derive the horizontal pressure gradient in terms of gradients of thicknesses

$$\nabla p^{(i)} = g\sum_{n=1}^{i-1}\rho^{(n)}\nabla h^{(n)} + g\rho^{(i)}\nabla\eta^{(i)}. \tag{5.58}$$

Conservation of mass for each homogeneous layer is $\nabla_3 \cdot \mathbf{v}^{(i)} = 0$, where ∇_3 is the three-dimensional operator. The horizontal velocities are independent of z because the flow is hydrostatic. Therefore, integrating over the depth of the layer yields

$$h^{(i)}\nabla\cdot\mathbf{v}^{(i)} + \frac{dh^{(i)}}{dt} = 0, \tag{5.59}$$

where we have used the free surface conditions

$$w^{(i)}(x, y, \eta^{(i)}, t) = \frac{d\eta^{(i)}}{dt},$$
$$w^{(i)}(x, y, \eta^{(i+1)}, t) = \frac{d\eta^{(i+1)}}{dt}. \tag{5.60}$$

Also, since $H^{(i)}$ is constant

$$\frac{d}{dt}(\eta^{(i)} - \eta^{(i+1)}) = \frac{dh^{(i)}}{dt}.$$

We can thus integrate (5.50) over the depth of each layer to obtain the conservation of potential vorticity for the layered system

$$\frac{d}{dt}\left(\frac{\zeta_0 + f}{h_0^{(i)}}\right) = 0. \tag{5.61}$$

The velocity in the convective derivative is $(u_0^{(i)}, v_0^{(i)})$.

In subsequent treatments of the two-layer, β-plane, inviscid, momentum equations, we shall use the approximate form (5.55) and (5.56) together with the vertically integrated form of (5.57). The equations are

$$\frac{du_1}{dt} - fv_1 = -g\frac{\partial \eta_1}{\partial x}, \tag{5.62}$$

$$\frac{dv_1}{dt} + fu_1 = -g\frac{\partial \eta_1}{\partial y}, \tag{5.63}$$

$$\frac{dh_1}{dt} + h_1\left(\frac{\partial u_1}{\partial x} + \frac{\partial v_1}{\partial y}\right) = 0, \tag{5.64}$$

$$\frac{du_2}{dt} - fv_2 = -g\left[(1-\epsilon)\frac{\partial \eta_1}{\partial x} + \epsilon\frac{\partial \eta_2}{\partial x}\right], \tag{5.65}$$

$$\frac{dv_2}{dt} + fu_2 = -g\left[(1-\epsilon)\frac{\partial \eta_1}{\partial y} + \epsilon\frac{\partial \eta_2}{\partial y}\right], \tag{5.66}$$

$$\frac{dh_2}{dt} + h_2\left(\frac{\partial u_2}{\partial x} + \frac{\partial v_2}{\partial y}\right) = 0, \tag{5.67}$$

where, $\epsilon\rho_2 = \rho_2 - \rho_1$, $h_2 = \eta_2 + H_2 - \eta_3$, and η_3 is the height of the bottom above an equilibrium level. The subscripts in (5.62) to (5.67) identify the layer rather than the order of L/a or Ro.

For linear steady flows the above system is sometimes used with spherical coordinates.

5.4 Ekman Layers

The equations derived above do not contain friction explicitly. However, when the variables are written in terms of a mean (ensemble, time average, etc.) plus a fluctuation and the equations are averaged, Reynolds stresses emerge and these are often parameterized in frictional form through the use of Austausch or eddy coefficients. Though this procedure is often questionable, it may not be a bad approximation near the top surface where wind stresses impart momentum to the ocean and near the bottom where frictional retardation brakes the flow. This was the view taken by Ekman (1905), who introduced the model for what is now called the Ekman layer.[5]

5.4.1 Pure Ekman Layers

Ekman first applied the theory to the wind-driven layer near the surface of the ocean. It is preferable to introduce the subject by investigating how a horizontally uniform geostrophic flow given by $f\hat{\mathbf{k}} \times \mathbf{v}_g = -\nabla P$ in a fluid occupying the region $z > 0$ is brought to rest by frictional processes acting near the bottom.[6] Assuming that horizontal variations of the stress are small compared to vertical variations (easily verified a posteriori), we can write

$$-fv = -fv_g + \nu u_{zz}, \tag{5.68}$$

$$fu = fu_g + \nu v_{zz}, \tag{5.69}$$

where subscript z corresponds to $\partial/\partial z$ and the pressure gradient is written in terms of the geostrophic velocity. The velocity vanishes at the (flat) bottom

$$\mathbf{v} = 0 \qquad \text{at} \quad z = 0. \tag{5.70}$$

The method of solution is well-known (Lamb, 1932, p. 593). Combining u and v as $u + iv$, $i = \sqrt{-1}$, the equations (5.68) and (5.69) take the form

$$(u+iv)_{zz} = \frac{if}{\nu}(u+iv) - \frac{if}{\nu}(u_g + v_g) \tag{5.71}$$

and the solution satisfying (5.70) with $\mathbf{v} \to 0$ as $z \to \infty$ is

$$u + iv = (u_g + iv_g)(1 - e^{-\sqrt{i}\delta z}), \tag{5.72}$$

where $\delta = \sqrt{f/\nu}$. Accordingly, the flow vanishes at $z = 0$, tends to $\mathbf{v}_g$ for large z and is predominantly to the left of $\mathbf{v}_g$ in between.

The vertically integrated transport of the exponentially decaying part of (5.71) is $(-1 + i)(u_g + iv_g)(\nu/2f)^{1/2}$, which suggests $h_e = (\nu/2f)^{1/2}$ as the scale of the Ekman layer. If we integrate the geostrophic part over the depth, h_e, we obtain the transport $(u_g + iv_g)h_e$. Hence, the *net* transport is $i(u_g + iv_g)h_e$, which is to the left of the geostrophic current, i.e., down the pressure gradient required to support $\mathbf{v}_g$, as we would expect. In vector form the net transport is $(-v_g, u_g)h_e$.

Next consider Ekman's problem, with fluid occupying the region $z < 0$ and with the flow driven by the spatially uniform wind stress (divided by the density) given by (τ^x, τ^y) acting at $z = 0$. With $\mathbf{v}_g = 0$, the solution is

$$u + iv = \frac{e^{\delta z}}{\sqrt{f\nu}}\{\tau^x \sin(\delta z + \pi/4) - \tau^y \sin(\delta z - \pi/4) + i[\tau^x \sin(\delta z - \pi/4) + \tau^y \sin(\delta z + \pi/4)]\}. \tag{5.73}$$

In the hodograph (u, v)-plane the solution has the form of a spiral (called the Ekman spiral). Just as rotation generates a velocity component to the right (for $f > 0$) of the (pressure) force for geostrophically balanced flow, a flow to the right of the tangential-stress force is generated in the Ekman spiral solution. In contrast to geostrophic flow, however, the present system is dissipative, and a velocity component parallel to the force is also present. At the surface the magnitudes of the components are equal so the flow is directed 45° to the right of the wind stress. The velocity component par-

allel to $\boldsymbol{\tau}$ decreases with depth but near the surface the normal component does not (it cannot since $\boldsymbol{\tau}$ has the same direction as $\partial\boldsymbol{\tau}/\partial z$). But below that the stress veers to the right as does the velocity vector.

Though Ekman's solution provided a satisfactory explanation of Nansen's observation of surface velocity, the spiral is not normally observed in the field. Ekman failed to observe it in spite of repeated attempts. Hunkins (1966) reported measuring a well-defined Ekman spiral (ironically, in the Arctic Ocean, where Nansen's first observations were made). The spiral structure depends on the form of the stress term, and since the stresses near the surface are turbulent (due to thermal convection, surface waves, and other small-scale processes) and therefore not necessarily of Navier-Stokes form, it is not surprising that the observed current structure differs from the theoretical one. Also, the mixed layer at the surface sits on a stably stratified fluid and the depth h of the former often does not exceed h_e when a turbulent eddy viscosity is used. Gonella (1971) showed that when a stress-free condition is applied at the base of the mixed-layer the solution is a function of h_e/h. For shallow ($h \ll h_e$) mixed layers there is essentially no spiral. Csanady (1972) reported that field measurements in the mixed layer in Lake Huron support Gonella's findings. He also reformulated the problem in terms of external parameters of the system instead of using an eddy viscosity.

In contrast to the detailed velocity structure, the vertically integrated transport of the wind-driven Ekman layer is independent of the form of vertical variation of the stress. If the stress terms in (5.68) and (5.69) are written as $\partial\boldsymbol{\tau}/\partial z$ and if we integrate the equations vertically, the transports are given by $(\tau^y, -\tau^x)/f$. Thus, the total transport is to the right of the wind stress irrespective of the form of $\boldsymbol{\tau}$ and subject only to these conditions: $\boldsymbol{\tau} = (\tau^x, \tau^y)$ at the surface and $\boldsymbol{\tau} = 0$ at the bottom. In vector form, with $V_e = \int_{-h}^{0} \mathbf{v}\, dz$ (where h is a depth—finite or infinite—at which $\boldsymbol{\tau}$ vanishes), the result (called the Ekman drift) is

$$\mathbf{V}_e = (\boldsymbol{\tau} \times \hat{\mathbf{k}})/f. \tag{5.74}$$

where $\boldsymbol{\tau}$ is now the wind stress vector and $\hat{\mathbf{k}}$ is the vertical unit vector.

5.4.2 Effect of Ekman Layers on Interior Flows

Although the pure Ekman layer theory given above requires horizontally uniform conditions, the theory is valid with horizontal variations as long as the horizontal scale is substantially larger than h_e. The neglected horizontal variations of the stress terms are smaller than $\partial\boldsymbol{\tau}/\partial z$ by the ratio of the squares of vertical to horizontal scales. Furthermore, for the mixed layer near the surface the vertical pressure gradient in the vertical equation of motion vanishes (as long as we consider scales larger than the small-scale turbulence which generates the mixed layer). Hence, the horizontal pressure gradients associated with Ekman layer processes are negligible at lowest order, and the original equations, and therefore the results given by (5.73), are still applicable.

Accordingly, suppose that $\boldsymbol{\tau}$ in (5.74) varies horizontally. When the continuity equation $\nabla\cdot\mathbf{v} = 0$ is integrated in the vertical over the depth of the Ekman layer and the boundary condition ($w = 0$ at the top) is applied, we find (Charney, 1955a)

$$w_e = \nabla\cdot\mathbf{V}_e, \tag{5.75}$$

where w_e is the vertical velocity at the base of the Ekman layer. With (5.74) this becomes

$$w_e = \frac{\partial}{\partial x}\left(\frac{\tau^y}{f}\right) - \frac{\partial}{\partial y}\left(\frac{\tau^x}{f}\right) = \hat{\mathbf{k}}\cdot\nabla \times \left(\frac{\boldsymbol{\tau}}{f}\right). \tag{5.76}$$

Thus, horizontal variations in $\boldsymbol{\tau}$ generate vertical motions which penetrate into the fluid below. Since the Ekman layer is thin relative to the depth of the ocean, this forced vertical velocity (called *Ekman pumping*) can be applied as a boundary condition (approximately at the surface) for the underlying inviscid fluid.

The same analysis can be applied to the bottom (subscript b) Ekman layer, where the vertically integrated transport was found to be $\mathbf{V}_b = (-v_g, u_g)h_b$. If the bottom is flat, so that $w = 0$ there, the vertically integrated continuity equation yields

$$w_b = -\nabla\cdot\mathbf{V}_b,$$

where w_b is the vertical velocity induced at the top of the Ekman layer. Substituting for $\mathbf{V}_b$ we obtain

$$w_b = \frac{\partial}{\partial x}(v_g h_b) - \frac{\partial}{\partial y}(u_g h_b)$$

or taking h_b constant,

$$w_b = h_b \hat{\mathbf{k}}\cdot\nabla \times \mathbf{v}_g. \tag{5.77}$$

This value for w serves as a boundary condition (approximately at the bottom of the ocean) for the overlying inviscid fluid.

5.4.3 Additional Considerations

Only the simplest results of Ekman layer theory have been given here. A number of important extensions are discussed by Stern (1975a, chapters 7 and 8). Horizontal momentum is imparted to the ocean by the wind stress acting at the surface; yet the momentum vanishes at the base of the Ekman layer. Stern answers the question where the momentum goes by analyzing the angular momentum balance about the axis of rotation for a cylindrical system. The analysis is carried out in an inertial frame of reference where the torque of the wind stress is balanced by the absolute angular momentum of the fluid. The latter is proportional to the absolute

vorticity of the undisturbed (no wind stress) vortex flow (in our case, solid-body rotation). The correspondence between the cylindrical problem and the rectilinear system [with (5.74) as the result] is that for large radii the angular momentum argument is equivalent to saying that the rate of momentum imparted by the wind stress is balanced by the divergence of the radial flux of absolute azimuthal momentum.

Though the Ekman layer depth h_e is clearly defined for laminar boundary layers, the value for turbulent boundary layers is not. Caldwell, van Atta, and Holland (1972) formed the boundary layer scale $\tau^{1/2}/f$ from the (only) external parameters τ and f. Assuming that the molecular scale $(\nu/f)^{1/2}$ is not likely to affect the turbulent scale, they suggest that $\tau^{1/2}/f$ is the turbulent Ekman boundary layer thickness. Stern (1975a, §8.1) carried out a crude stability analysis to conclude that a layer thicker than $h_e \sim \tau^{1/2}/f$ will radiate energy to the deep water. He surmised that nonlinear modifications will show that the turbulent energy is thereby reduced as the thickness shrinks to $\tau^{1/2}/f$, where the system will stabilize. For typical values of τ, the value of h_e (so defined) is $O(100\text{ m})$ at mid-latitudes. These considerations are based on the assumption of a homogeneous fluid. For a stratified fluid like the ocean the stratification may be decisive in determining the boundary layer thickness as Csanady's (1972) report of observed velocities in Lake Huron indicates.

As we saw from the simple analysis presented above, the effect of the top Ekman layer on the underlying water is determined completely by the wind stresses, whereas in the bottom Ekman layer the condition is expressed in terms of the velocity of the overlying water. More generally there will be a nonlinear coupling between the Ekman layer and the interior which can alter the results significantly. Fettis (1955) carried out the analysis for a laboratory model of a nonlinear Ekman layer to show that the results can be approximated by (5.74) but with the absolute vertical vorticity replacing f. Stern (1966; 1975a, §8.3) and Niiler (1969) have investigated the effect of coupling of Ekman layer flow with geostrophic vorticity (eddies) and have shown that the latter can have a dominant influence since coupling with the interior can occur even for a uniform wind stress.

5.5 Steady Linear Models of the Wind-Driven Circulation

For steady, linear flow of moderate scale we have $Ro \ll L/a$ so the term $d\zeta_0/dt$ in (5.50) can be neglected. The resulting equation is

$$\beta v_0 = f_0 \frac{\partial w_1}{\partial z}. \tag{5.78}$$

When integrated vertically from $z = -h$ to $z = 0$ this yields

$$\beta V = f_0 w_1 \Big|_{-h}^{0}, \qquad V = \int_{-h}^{0} v_0 \, dz$$

or

$$\beta V = \hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau} - f_0 w_1(x, y, -h), \tag{5.79}$$

where the variation of f in $\hat{\mathbf{k}} \cdot \nabla \times (\boldsymbol{\tau}/f)$ is (consistently) neglected at lowest order.

5.5.1 Sverdrup Transport

If the stratification is strong enough so that distortion of the density surfaces is negligible at some depth above the bottom, the last term in (5.79) vanishes and we obtain the Sverdrup transport

$$\beta V = \hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau}. \tag{5.80}$$

Thus, the vertically integrated north–south transport is determined by the curl of the wind stress. Sverdrup (1947) introduced this relation to estimate transports in the eastern equatorial Pacific (see chapter 6). Physically, the interpretation of (5.80) is straightforward. With βV written as $h\, df/dt$ we see that a column of fluid moves to a new latitude (new value of planetary vorticity f) with a speed that compensates for the rate at which the wind stress imparts vorticity to the ocean.

The continuity equation (5.41) can be integrated in the vertical and in x to yield

$$U = -\int \frac{\partial V}{\partial y} \, dx + F(y)$$

or

$$U = -\int \frac{1}{\beta} \frac{\partial}{\partial y} (\hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau}) dx + F(y), \tag{5.81}$$

where $F(y)$ is arbitrary. The most common procedure for theoretical analyses is to assume that the foregoing is valid eastward to a meridional boundary $x = L$, where U must vanish. Then

$$U = \int_x^L \frac{1}{\beta} \frac{\partial}{\partial y} (\hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau}) dx, \tag{5.82}$$

and the transport is determined in the entire region in which the assumptions are valid. In general, the theory does not determine the flow in a basin bounded on the west as well since it is not possible to satisfy the zero normal flow condition there.

5.5.2 Stommel's Frictional Model

If the fluid motion penetrates to the (flat) bottom, the last term in (5.79) is given by (5.75) with $\mathbf{v}_g = \mathbf{v}_0|_{z=-h}$ and (5.79) becomes

$$\beta V = \hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau} - f_0 h_b \left(\frac{\partial v_0}{\partial x} - \frac{\partial u_0}{\partial y} \right)_{z=-h}. \tag{5.83}$$

Thus, one must supplement this equation with additional ones that determine the vertical structure of the velocity field. However, if the fluid is assumed to be homogeneous so that v_0 is independent of z and if one then writes

$$(U, V) = (u_0, v_0)h,$$

the system closes with

$$\beta V = \hat{\mathbf{k}}\cdot\nabla \times \boldsymbol{\tau} - \frac{f_0 h_b}{h}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right). \tag{5.84}$$

Introducing the transport stream function $\mathbf{V} = \mathbf{k} \times \nabla\psi$, yields

$$\beta\psi_x = \hat{\mathbf{k}}\cdot\nabla \times \boldsymbol{\tau} - K\nabla^2\psi, \tag{5.85}$$

where $K = f_0 h_b/h$. Stommel (1948) obtained (5.85) by assuming a bottom drag law for the friction term. The derivation using Ekman layer theory makes the assumptions more evident.

The solution with $\tau^x = -T\cos(\pi y/M)$, $\tau^y = 0$ and with $\psi = 0$ at $x = 0, L$ and $y = 0, M$ is

$$\psi = \frac{MT}{\pi K}\left\{1 - \frac{(1 - e^{D_2 L})e^{D_1 x} - (1 - e^{D_1 L})e^{D_2 x}}{e^{D_1 L} - e^{D_2 L}}\right\} \times \sin\frac{\pi y}{M}, \tag{5.86}$$

where

$$D_1 = -\frac{\beta}{2K} + \sqrt{\left(\frac{\beta}{2K}\right)^2 + \left(\frac{\pi}{M}\right)^2},$$

$$D_2 = -\frac{\beta}{2K} - \sqrt{\left(\frac{\beta}{2K}\right)^2 + \left(\frac{\pi}{M}\right)^2}.$$

Values of ψ versus x are shown in figure 5.1 for the case with $L = 6000$ km, $M = 3000$ km, $\beta = 2 \times 10^{-11}$ m^{-1} s^{-1}, and $K = 2 \times 10^{-6}$ s^{-1}. Stommel's model was the first to exhibit the westward intensification of the oceanic response to a symmetric wind-stress curl.

With $K/\beta L \ll 1$, (5.85) is a boundary-layer problem, where the highest derivative term (the bottom frictional effect) is important only in a narrow region near the western boundary where the flow is northward. In the remainder of the basin the Sverdrup balance (5.80) is approximately valid (but see below), the flow is slow and southward, and friction is unimportant. The negative vorticity injected into the ocean by the wind is eventually dissipated in the western boundary layer, where the induced northward flow deposits columns of fluid at their original latitudes with the original planetary vorticity restored. Detailed balances and a fairly comprehensive discussion are given by Veronis (1966a).

The westward intensification is normally explained in terms of the vorticity balance, but a qualitative discussion in terms of momentum balance is also possible. Thus, we note that the Ekman wind drift in the northern half-basin is southward whereas that of the southern half-basin is northward. Water piles up at mid-latitude, raising the free surface level and creating a high pressure ridge at mid-latitude (H in figure 5.2). The induced eastward geostrophic flow in the northern half-basin requires a low pressure along the northern boundary. In the southern half-basin a westward flow of the same magnitude requires less of a north-south pressure difference (because the Coriolis parameter is smaller) so the low pressure (HL in figure 5.2) at the south is higher than the low pressure (LL in figure 5.2) at the north. The solid boundaries at the east and west will divert the flow. A narrow frictional boundary layer at the east would require flow from the low low pressure at the north to the high low pressure at the south, i.e., flow up the (gross) pressure gradient. On the western side, on the other hand, a narrow frictional boundary layer supports flow from high to low pressure. Hence, if a thin frictional boundary layer exists, it must be on the western side. This "explanation" ignores a lot of important details, but the reasoning is consistent with the roles that rotation and friction play in balancing the pressure gradient.

If a system without meridional boundaries (a zonal channel) were subjected to a zonal stress, a zonal flow would be generated (apart from the Ekman drift). Hence, the Sverdrup transport of Stommel's model must depend on the presence of meridional boundaries. Yet it seems likely that if the meridional boundaries are far enough apart, the system should resemble a zonal channel more than an enclosed ocean except in relatively narrow regions near the east and west where meridional flow takes place. Welander (1976) showed that that is the case. With the zonal wind stress given above one can substitute $\psi = \Phi(x)\sin(\pi y/M)$ to derive

$$K\Phi'' - K\frac{\pi^2}{M^2}\Phi + \beta\Phi' = -\frac{\pi T}{M}. \tag{5.87}$$

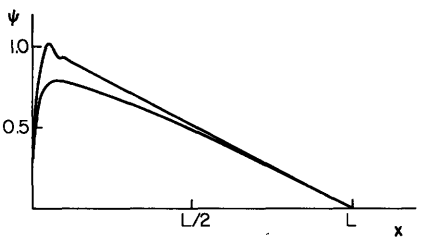

Figure 5.1 The transport streamfunction, normalized with respect to the Sverdrup transport and divided by $\sin\pi y/M$, is shown for Munk's solution with lateral diffusion (top curve) and Stommel's solution with bottom friction. The nominal boundary layer thickness is $L/60$. Stommel's solution shows the decreased transport because of the effect of friction in a basin with $\pi L/M \gg 1$. Munk's solution oscillates near the western boundary, giving rise to a weak countercurrent to the east of the main northward flow.

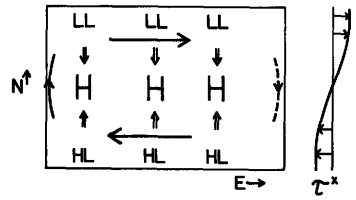

Figure 5.2 A cosine wind stress τ^x causes an Ekman drift (double arrows) toward mid-latitude where the free surface is elevated and a high pressure region (H) is created. A geostrophically balanced current flows eastward in the north half-basin and westward in the south. Because of the larger Coriolis parameter a lower low pressure (LL) is required along the north boundary than along the south (HL) to support the same transport geostrophically. If the zonal transport is deflected southward in a frictional boundary layer near the eastern side (dashed curve), the flow must go against the gross pressure difference (from LL to HL). If the flow is in a western boundary layer (solid curve), the gross pressure difference drives the flow against frictional retardation. The latter is a consistent picture.

As we have seen, the second-derivative term is important only in the western boundary layer where the scale of variation is $K/\beta = 100$ km. North-south diffusion (the undifferentiated Φ term) is unimportant when the geometry is square. But when the zonal separation is large ($\pi L/M \gg 1$), the balance is between wind-stress curl and north-south diffusion, $\Phi \simeq MT/\pi K$), and the flow is zonal. The Sverdrup transport relation holds in an eastern boundary layer with the east-west scale $\beta M^2\pi^2/K$. Bye and Veronis (1979) pointed out that the northward transport in the western boundary layer is much smaller than the transport calculated by the Sverdrup balance if the aspect ratio $\pi L/M$ is large, as is the case for nearly all wind-driven oceanic gyres. Of course, these results are contained in the complete solution of the simple model discussed here. But when relatively modest refinements are introduced (e.g., spherical geometry), a complete solution is no longer possible and boundary layer methods must be used. It is then necessary to recognize the correct approximate balance in the different regions of the basin.

5.5.3 Topography and Lateral Friction

The principal result of the foregoing analysis, viz., the westward intensification of an oceanic gyre, is verified both by observations and by much more sophisticated analyses. Hence, it is a feature that appears to be insensitive to the drastic simplifications that were made. But it is a simple matter to change the result by relaxing one of the simplifications and then restoring the result with a second, seemingly unrelated, assumption. In other words, the simple model is not as crude as it appears to be.

For example, introduce realistic topography (Holland, 1967; Welander, 1968). Then on vertical integration, we see from (5.60) that, in addition to the effects of wind stress and bottom friction, the vertical divergence term in (5.78) will also contribute the term $f\,dh/dt$ to the right-hand side of (5.84). If the latter is combined with the β term, the result is

$$h^2\frac{d}{dt}\left(\frac{f}{h}\right) = \hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau} - K\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right). \tag{5.88}$$

Hence, the driving and dissipative forces on the right will cause a fluid column to respond by moving to points determined by the value of f/h rather than f as before. Since the contours of f/h are sometimes strongly inclined to latitude circles (Gill and Parker, 1970), the transport pattern is very different from (in fact, less realistic than) Stommel's. Thus, the effect of topography is exaggerated in a homogeneous model.

Stratification can reduce the topographic effect. In fact, if the density surfaces adjust so that the pressure gradient in (5.55) vanishes at and below a given level, there will be no driving force to support a flow. If topography does not project above this level of density compensation, it has no effect on the flow. In an intermediate situation, the density distribution can compensate for part of the pressure gradient so that at the level where it interacts with the bottom the velocity is considerably weaker than the surface velocity. A treatment of the latter case would necessarily incorporate convective processes in some form.

When complete compensation takes place in a steady model, the topographic influence is eliminated, but our derivation of bottom friction is no longer valid because it is no longer possible to parameterize the frictional processes at the bottom in terms of the mean velocity. The essential results of the model can be preserved, however, by parameterizing frictional effects in terms of an assumed lateral eddy diffusion. The last term in (5.83) is then replaced by a lateral frictional term so that the vorticity equation, in terms of the transport stream function becomes

$$\beta\frac{\partial\psi}{\partial x} = \hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau} + A\,\nabla^4\psi, \tag{5.89}$$

where A is the magnitude of eddy viscosity based on the intensity of eddy processes at scales smaller than those being analyzed. Hidaka (1949) introduced this equation together with the vertically integrated continuity equation

$$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = 0. \tag{5.90}$$

A convenient set of boundary conditions where the wind stress curl is proportion to $\sin(\pi y/M)$ is

$$\begin{aligned} U = 0 = V \quad &\text{at} \quad x = 0, L, \\ V = 0 = \frac{\partial U}{\partial y} \quad &\text{at} \quad y = 0, M. \end{aligned} \tag{5.91}$$

The solution is easily obtained (Munk, 1950) and is included in figure 5.1. It contains a Sverdrup transport in the interior; a narrow eastern boundary layer in which V decreases to zero at the eastern wall; and a western boundary layer with no tangential velocity at the western wall, a northward flow near the wall and a weak, narrow countercurrent just east of the northward flow. Because frictional processes are now associated with higher derivatives, the effect of friction in the interior is considerably weaker than in Stommel's model, and the Sverdrup balance is valid throughout the interior. Accordingly, in this case the aspect ratio of the basin has little influence on the magnitude of the transport. Because the zonal velocity increases linearly with distance from the eastern boundary, for broad ocean basins the flow has a strongly zonal appearance. In Stommel's model the north-south flow essentially vanishes in the western portions of the basin and the flow is truly zonal there.

Although these formal models are steady, the application is to flows that are transient but statistically steady. Transient motions can have a strong barotropic component even when the statistically steady flow is largely baroclinic. With that in mind we may still use a bottom frictional drag for the stratified steady model, though the connection to the mean flow will then be not through the coupling to a steady Ekman layer but through a time averaging of interacting transient motions. Rooth (1972) has made such an estimate for K and obtains a value considerably smaller than the one normally used.

5.5.4 Laboratory Models

Though these steady, linear models can provide only the crudest approximation to real oceanic flows, they have served an important function in the development of oceanic theory. Stommel (1957b) put together the important components (Ekman suction and β-effect) to construct a comprehensive picture of ocean current theory as determined by these simple processes. The ideas were tested in a laboratory model of ocean circulation (Stommel, Arons, and Faller, 1958) in which the β-effect was simulated by the paraboloidal depth of a homogeneous layer of water in a pie-shaped basin rotating about the apex (see chapter 16). The equivalence of β and variable depth is suggested by the linearized form of potential vorticity,

$$(\zeta + f)/h \simeq (\zeta + f)/H_0 - \frac{f_0 \eta}{H_0^2},$$

where η is the deviation of the free surface from its mean value H_0, so that a change in $f_0\eta/H_0^2$ is equivalent to a change in f, i.e., to β. When water is being added at the apex, the free surface in the interior rises not by a direct vertical motion but by a radially uniform inward movement of columns of fluid (figure 5.3). The circulation generated in this way simulates the Sverdrup transport, the inward radial direction corresponding to north (increasing f or decreasing depth).

In the experiment, boundary layers near the "western" boundary and the rim and apex are required to complete the circulation pattern (figure 5.3). The azimuthal flow and the rising free surface needed to feed the interior radial flow are generated in the rim boundary layer. Near the apex the flow is diverted to the western boundary layer to join the fluid being injected. It is interesting to note that the radially inward flow that causes the free surface to rise is *toward* the source of fluid. Thus, the transport in the western boundary layer is twice that of the source. Half of the former goes to raise the free surface; the other half serves as the vehicle for the indirect circulation. (Also see Figures 16.1 and 16.2 and the accompanying discussion.)

Additional experiments and a rigorous analysis using rotating-fluid theory to treat the various boundary layers were subsequently provided by Kuo and Veronis (1971), who showed that for different parametric ranges the experiment could be used to simulate Stommel's model with a bottom frictional boundary layer or the Hidaka-Munk model with a lateral frictional boundary layer. Veronis and Yang (1972) provided a perturbation treatment of the nonlinear effects and verified the results with a series of experiments. Pedlosky and Greenspan (1967) proposed an alternative laboratory model with the depth variation provided by an inclined boundary at the top and/or bottom of a rotating cylinder. The flow was driven by the differential rotation of the top plate. For this model Beardsley (1969, 1972) carried out a comprehensive set of experiments and

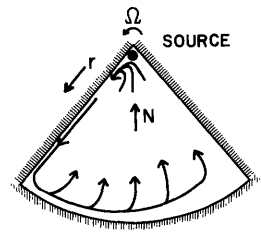

Figure 5.3A A weak source of fluid at the apex of a rotating pie-shaped basin will cause flow toward the rim in a "western" boundary layer. Fluid flows from the rim boundary layer radially inward toward the apex as shown.

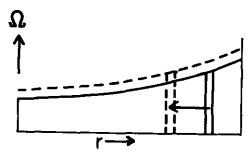

Figure 5.3B A vertical cross section through the apex. The basin is filled in the interior by the inward movement of columns of fluid as shown.

extended the theory analytically and numerically to include inertial effects.

The foregoing experiments and theories are more appropriate areas of application than the real ocean is for the ideas introduced by Sverdrup, Stommel, and Munk. At the time that they were introduced, however, these ideas were remarkable advances into unknown territory. They have provided a framework for further development and some of them persist as important elements in more extensive theories.

5.6 Preliminary Nonlinear Considerations

The first perturbation analysis of nonlinear effects in a wind-driven gyre was by Munk, Groves, and Carrier (1950), but it is easier to see the qualitative changes by looking at Stommel's model (Veronis, 1966a). From the linear problem we saw that the vorticity and its zonal variation are largest in the western boundary layer, so we expect the largest nonlinearities there. The wind stress is not important in that region, and we start with the vorticity equation, including inertial terms but not the wind-stress curl:

$$\frac{d}{dt}(\zeta + f) = \mathbf{v}\cdot\nabla\zeta + \beta v = -\epsilon\zeta. \tag{5.92}$$

In the southern half of the basin the flow is westward ($u < 0$) into the boundary layer where it is diverted northward. Thus, a fluid particle is carried from the interior, where ζ vanishes, into the boundary layer, where ζ is large and negative, so $d\zeta/dt < 0$. Northward flow implies $df/dt > 0$. Hence, the convective term balances part of the β-effect and $-\epsilon\zeta$ must consequently decrease in size. Since the vorticity is essentially $\partial v/\partial x$, it will decrease if v decreases or if the horizontal scale increases. But from $v = \partial\psi/\partial x$ we see that a decrease in v also corresponds to an increase in the horizontal scale. Therefore, we conclude that inertial effects weaken the flow by broadening the scale. This effect will also decrease the dissipation in the inflow region. The same considerations apply to the case with lateral friction.

In the northern half of the basin where the flow emerges ($u > 0$) from the western boundary layer, fluid is carried from a region of negative vorticity to the interior where ζ vanishes, so $d\zeta/dt > 0$. Since the flow is northward in the boundary layer, df/dt is also positive. Therefore, the amplitude of the vorticity must be larger since the dissipation $-\epsilon\zeta$ must be larger than in the linear case. Hence, the horizontal scale of variation must decrease.

The net effect of inertial processes is thus to broaden the boundary layer thickness and to reduce the dissipation in the region of inflow, and to sharpen the boundary layer thickness and increase the dissipation in the region of outflow.

For more nonlinear flows the dissipation takes place largely in the northern half of the boundary layer. Furthermore, the excess inertia of the particles causes them to overshoot their original (interior) latitudes so there must be an additional region where inertial processes and friction restore the particles (southward) to their starting points. The effect is to spread the region of inertial and frictional control first to the north and eventually eastward from the northwest corner of the basin. A discussion of the successively stronger effects of nonlinear processes and a division of the basin into regions where different physical balances obtain is given by Veronis (1966b).

This argument strongly suggests that it may be possible to analyze the region of formation of western boundary currents in terms of a frictionless inertial model. Stommel (1954) proposed such an analysis which he subsequently included in his book (Stommel, 1965).

Fofonoff (1954) focused his attention on nonlinear processes by treating the steady circulation in a frictionless, homogeneous ocean. The starting point is the conservation of potential vorticity in a basin of constant depth, viz.,

$$\frac{d}{dt}(\zeta + f) = 0, \tag{5.93}$$

together with the two-dimensional continuity equation. These equations are satisfied by $u = -\partial\psi/\partial y =$ constant or

$$\psi = -uy, \tag{5.94}$$

but boundary conditions are not, so it is necessary to add boundary layers at the eastern and western sides of the basin.

A first integral of (5.93) is

$$\nabla^2\psi + f = F(\psi) \tag{5.95}$$

and in the interior where the relative vorticity vanishes

$$F(\psi) = f = f_0 + \beta y. \tag{5.96}$$

But (5.94) yields $y = -\psi/u$ there, so that $F(\psi) = f_0 - \beta\psi/u$ and (5.95) becomes

$$\nabla^2\psi + \frac{\beta\psi}{u} = -\beta y. \tag{5.97}$$

This equation is satisfied nearly everywhere by $\psi = \phi(x)y$ so that

$$\phi'' + \frac{\beta\phi}{u} = -\beta. \tag{5.98}$$

The north-south flow near the meridional boundaries is thus geostrophic. A boundary layer solution with $\phi = 0$ at $x = 0, L$ is possible for $u < 0$ if $\epsilon = (-\beta/u)^{1/2} \gg L$. It is

$$\phi = \frac{u}{\sinh \epsilon L} [\sinh \epsilon x - \sinh \epsilon L + \sinh \epsilon (L - x)]$$
$$\times [y - Me^{-(M-y)\epsilon}]. \quad (5.99)$$

This yields a uniform, westward flow in the interior, and boundary layers of thickness ϵ^{-1} with northward flow at the west, southward flow at the east, and a jet across the northern edge (figure 5.4).

It is possible to have the eastward jet at any latitude by adding an appropriate constant to ψ in (5.94). With $u > 0$ the system does not have a boundary layer solution but oscillates across the basin (Fofonoff, 1962a).

Although Fofonoff's solution appears to be very artificial, it is one of the survivors of the earlier theories. The strongly nonlinear version of Stommel's model leads to a solution that looks remarkably like Fofonoff's (Veronis, 1966b; Niiler, 1966). The recirculation region just south of the Gulf Stream after the latter has separated from the coast has the appearance of a local inertial circulation. Thus, it is likely that some version of the latter will be part of any successful model of large-scale ocean circulation.

Shortly after Fofonoff's analysis and following Stommel's (1954) suggestions, Charney (1955b) and Morgan (1956) produced models of the Gulf Stream as an inertial boundary layer. By using observed or simulated conditions at the inflow edge of the Gulf Stream to fix the form of $F(\psi)$, and working with a two-layer model with potential vorticity $(f + \zeta)/h$ and geostrophic balance for the northward flow, they were able to calculate the streamfunction pattern and the thermocline depth distribution in the formation region of the Gulf Stream. Charney showed that in a two-layer ocean inertial forces can cause the thermocline to rise to the surface at a latitude corresponding to Cape Hatteras. His solution could not extend beyond that point.

Morgan began his analysis by dividing the ocean into an interior with a Sverdrup balance, a formation region for the western boundary current, which he analyzed using the same model that Charney did, and a northern region. He speculated that friction and inertial and transient processes would interact in the north, but he did not attempt to analyze that region. He was one of the first to point out that pressure torques at the bottom and sides of the ocean can help to balance the torque exerted by the wind stress about a mid-ocean axis.

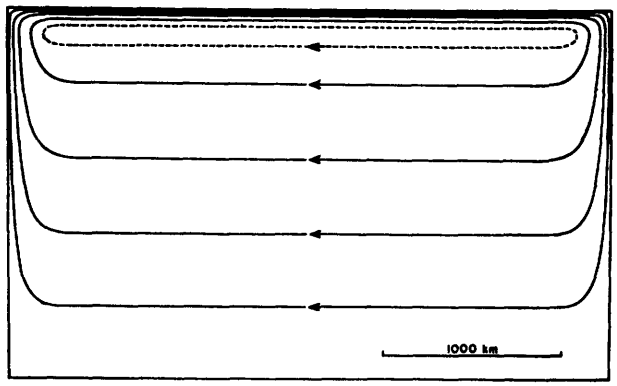

Figure 5.4 Fofonoff's (1954) inertial flow pattern for steady westward flows in the interior. An inertial boundary layer at the west diverts the flow northward and an eastward jet is formed. The latter feeds into an inertial boundary layer on the east that supplies the steady westward flow of the interior.

In contrast to the demonstration following (5.92) that inertial effects are consistent with the formation of a western boundary layer by the interior flow, a similar argument for the formation of an eastern boundary layer is not possible. For example, consider an anticyclonic gyre when an eastward interior flow generates an eastern boundary layer with southward flow. The vorticity in the boundary layer is negative, so $-\epsilon\zeta$ is positive. For southward flow df/dt is negative and therefore $d\zeta/dt$ must be positive. But that is not possible since ζ must change from a nearly zero value in the interior to a large negative value in the boundary layer. An analysis of the various possibilities for both cyclonic and anticyclonic gyres shows that it is generally not possible to form eastern boundary layers from eastward interior flows (Veronis, 1963). The actual existence of eastern boundary layers means that the necessary physical processes (in my opinion horizontal advection of density must be included) are missing from these simple models.

In an important model of a steady wind-driven gyre in a homogeneous ocean of constant depth, Derek Moore (1963) produced a complete circulation pattern with contributions from frictional and inertial processes in both inflow and outflow regions of the western boundary layer. Moore combined boundary-layer arguments from classical fluid mechanics with most of the features given above. Using a Navier-Stokes form for friction, he proved that frictional and inertial processes cannot be combined consistently to produce a boundary layer confined to the eastern side. In the vorticity equation of his model inertia is included as an east-west convection of the vorticity with a zonal velocity, $U(y) = U_0 \cos(\pi y/M)$, consistent with the form of the wind stress. In the southern half-basin (figure 5.5) the incoming (westward) flow forms an inertially controlled western boundary current. In the northern half-basin the emerging flow oscillates eastward and has the appearance of standing, damped Rossby waves imbedded in an eastward current. The center of the gyre is north of mid-latitude, consistent with the effects of inertia mentioned earlier. His results depend on the magnitude of a Reynolds number defined by $Re = U_0^{3/2}/\nu\beta^{1/2}$, which can be looked upon as the ratio of the inertial boundary layer scale $(U_0/\beta)^{1/2}$ to the viscous scale ν/U_0. The result is shown for $Re = 5$. As Re is decreased, the flow tends toward the Munk pat-

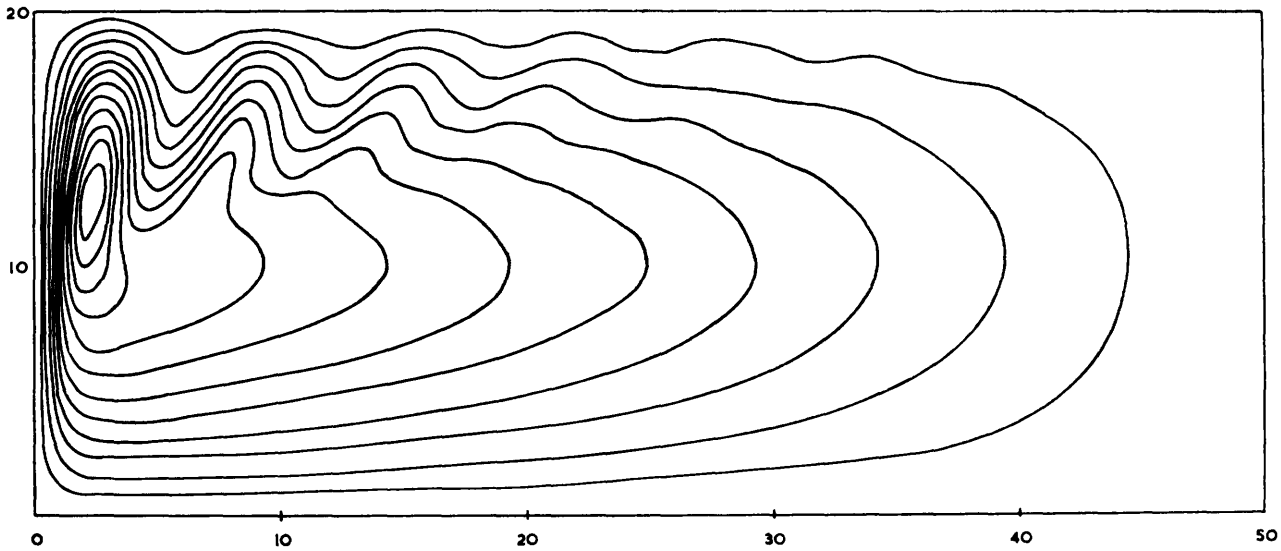

Figure 5.5 Contours of the streamfunction in a homogeneous ocean driven by a wind stress of the form $-\cos\pi y/M$ as derived by Moore (1963). An Oseen approximation for the nonlinear terms with a mean current $U(y) \propto \cos\pi y/M$ was used. The wavy contours in the north half-basin are standing Rossby waves imbedded in the mean velocity field.

tern. With larger *Re* the oscillations extend farther to the east and eventually fill the northern half of the basin. In the latter case there is a rapid transition across mid-latitude in the interior and the oscillatory flow becomes unstable. Qualitatively this homogeneous model contains a remarkably realistic array of features of oceanic flow, though the observed recirculation in the northwest corner is missing.

We turn to a discussion of stronger nonlinear effects in Stommel's model. As *Ro* is increased (Veronis, 1966b), the western boundary layer in the southern half-basin broadens and dissipative effects are more confined to the north. Inertial effects also intensify in the north so that a particle overshoots the northernmost latitude that it had in the interior. Hence, a new boundary layer region must be generated (offshore of the original one) where friction and inertia force the particle southward to its original latitude. In this latter region the relative vorticity is actually positive because the return flow to the south is stronger close to the boundary layer than it is farther to the east. The overshoot can be seen in figure 5.6A.

With even stronger driving the overshoot is larger and eventually the particle is driven close to the northern boundary and then eastward before it starts its southward return to its original latitude (figure 5.6B). Thus, the frictional-inertial region is broadened. In an extreme case (figure 5.6C) fluid particles move eastward in a jet at the north and reach the eastern boundary before turning south. In the latter case, there is essentially no Sverdrup interior, and the flow pattern resembles Fofonoff's free inertial flow with a mild east-west asymmetry as the only evidence that the flow is wind driven. An interesting fact here is that the northward transport in the western boundary layer does not increase beyond the Sverdrup transport until the eastward moving inertial jet reaches the eastern boundary. In the calculations cited, that happens when the inertial scale $(U_0/\beta)^{1/2}$ $(\equiv Ro^{1/2}L)$ exceeds the viscous scale K/β by a factor of 2 or so. Here, U_0 is a measure of the Sverdrup velocity. Qualitatively, at least, the observed recirculation to the south and east of the Gulf Stream after it has separated from the coast is simulated by this model. The separation from the coast is not. An analytic model of the highly nonlinear case was suggested by Veronis (1966b) and independently carried out by Niiler (1966). The resulting pattern is consistent with the one shown in figure 5.6C. Stommel (1965) guessed a similar pattern.

Bryan (1963) carried out an extensive set of numerical calculations in a rectangular basin for the nonlinear Hidaka-Munk model with $\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau} \sim \sin\pi y/M$, zero velocity boundary conditions at east and west, and zero-shear conditions at north and south. He presented his results in terms of a Reynolds number *Re* essentially the same as Moore's, and the Rossby number, *Ro*. The results differ greatly from those with bottom friction because for $Re > 60$ a barotropic (Rayleigh-type) instability can occur near the western boundary where the tangential velocity must vanish. Figure 5.7 illustrates his results for three values of *Re*, with $Ro = 1.28 \times 10^{-3}$ for figures 5.7A and 5.7B and $Ro = 3.2 \times 10^{-4}$ for figure 5.7C. The first two cases, with $Re = 20$ and $Re = 60$, show the development of the flow with increasing nonlinearity. Only a mild, steady, oscillatory pattern is present with $Re = 20$, whereas with $Re = 60$ the oscillations are more intense and a closed eddy (recirculation) is present near the northwest corner. For $Re = 100$ the flow is transient with a barotropic

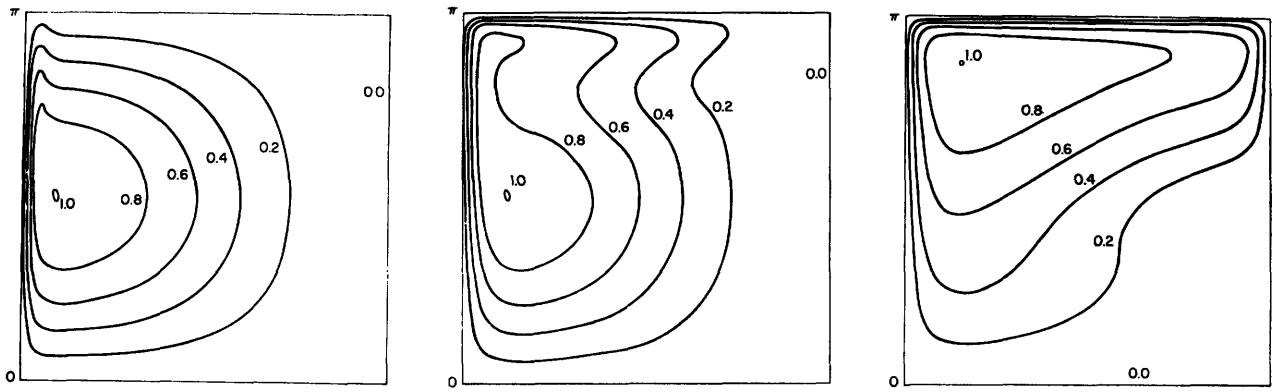

Figure 5.6 Three streamfunction patterns by Veronis (1966b) for an ocean basin with varying degrees of intensity of wind stress. (A) shows the perturbation effect of nonlinearity with fluid particles in the western boundary layer overshooting their equilibrium latitudes. (B) shows a much stronger inertial effect. In (C) inertia dominates the system, creating an eastward jet along the north reminiscent of Fofonoff's solution.

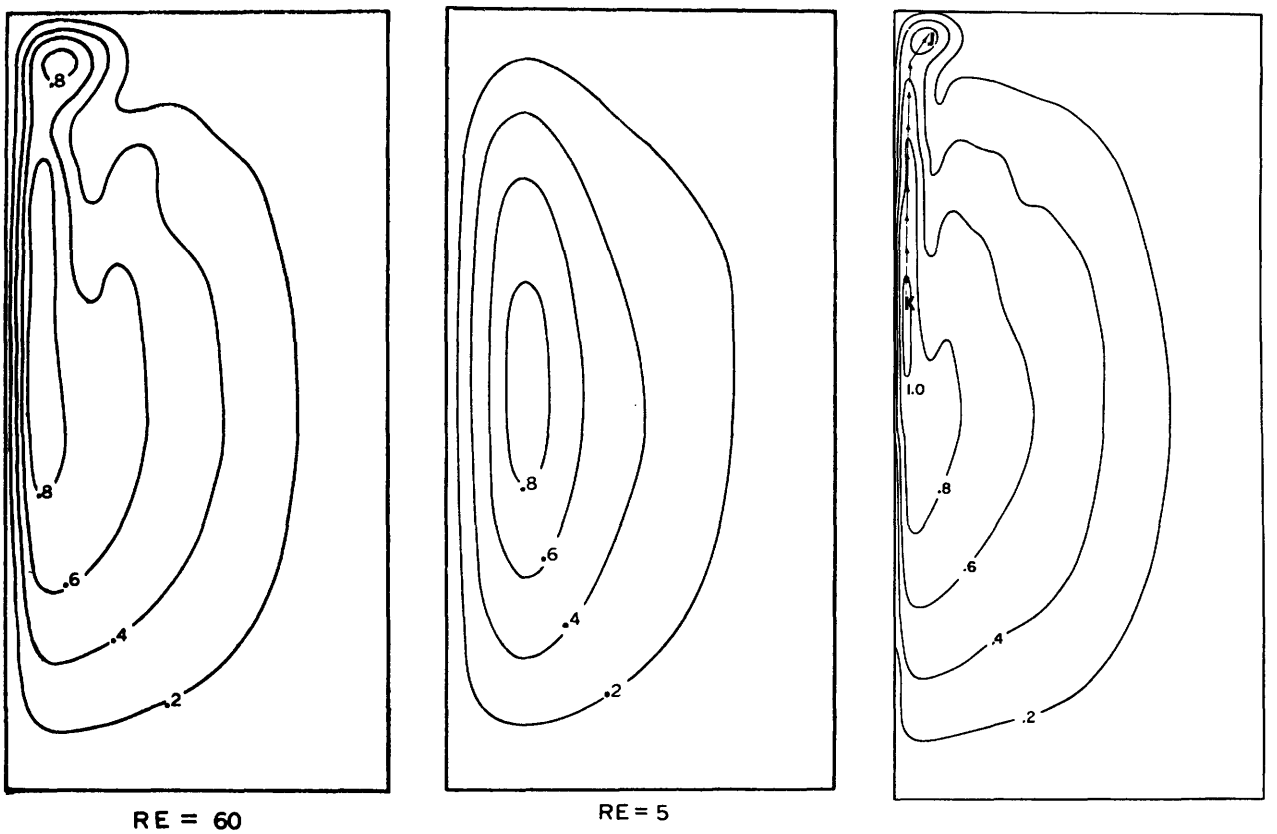

Figure 5.7 Bryan's (1963) streamfunction contours for a homogeneous ocean with lateral friction. The circulation in (A) is nearly linear; that of (B) is near the limit of forcing that still leads to a steady circulation. With even more intense driving a barotropic instability occurs as in (C), where a time-average field is shown after the system approaches a statistically steady state. See also Figure 3.13 and discussion there. *Re* is 5 for (A), 60 for (B), and 100 for (C).

instability induced in the northern half of the intense northward jet. Figure 5.7C shows the time-averaged flow after the transients have settled down. In this case also there is an offshore region in the north with positive vorticity where particles return southward to their starting latitudes. It is not possible to obtain an intense recirculation with this model because of the barotropic instability.

Bryan also calculated the flow for a basin with a western boundary directed north, then due east, and then north again. The break was north of mid-latitude. The object was to see whether the break in the boundary would force the western boundary current out to sea. The flow pattern was modified mildly, but the stream turned the corner and hugged the coast.

5.7 Why Does the Gulf Stream Leave the Coast?

The Gulf Stream flows along the coast from Florida to Cape Hatteras, where it parts from the coast and flows slightly north of eastward out to sea (see chapter 4). The Kuroshio and all other western boundary currents also separate. The phenomenon is explained here by a very simple argument. Although processes more complicated than the ones discussed below are also present, I believe that the argument given here contains the essential features even though the local dynamical details are not included.

Consider a two-layer system with the lower layer at rest. Then from equations (5.65) and (5.66) it follows that

$$\nabla\eta_2 = -\frac{\rho_1}{\Delta\rho}\nabla\eta_1, \qquad \nabla\eta_1 = \frac{\Delta\rho}{\rho_2}\nabla h_1. \tag{5.100}$$

If the motion is geostrophic ($Ro \ll 1$) except for the vertical stress term near the surface, equation (5.62) upon vertical integration over the depth h_1 of the top layer becomes, with $g' = g\,\Delta\rho/\rho_2$,

$$-fV_1 = -\frac{g'h_1}{a\cos\phi}\frac{\partial h_1}{\partial\lambda} + \tau, \tag{5.101}$$

where spherical coordinates have been retained so there is no geometrical distortion. Here the stress at the interface is assumed negligible and τ corresponds to the zonal wind stress. Multiply (5.101) by $a\cos\phi$ and apply the operator $\int_{\lambda_e}(\,)d\lambda$, where λ_e is the meridian of the eastern boundary, to obtain

$$h_1^2 = h_{1e}^2 - \frac{2f}{g'}T_1 - \frac{2f}{g'}T_E, \tag{5.102}$$

where subscript e denotes a value at λ_e, $T_1 = \int_{\lambda_e} a\cos\phi\, V_1\, d\lambda$ is the meridional transport, and $T_E = \int_{\lambda_e} a\cos\phi\,\tau\, d\lambda/f$ is the Ekman drift.

In all of the calculations reported in the previous section, the downstream velocity in the western boundary layer is geostrophic to a very good approximation. Hence, (5.102) is valid not only for interior flow but for the entire basin from west to east. Therefore, if we evaluate (5.102) at the western edge λ_w, T_1 represents the total meridional transport. If the ocean basin is enclosed to the north of the latitude in question, T_1 must vanish in the steady state and (5.102) becomes

$$h_{1w}^2 = h_{1e}^2 - \frac{2f}{g'}T_E. \tag{5.103}$$

Now, for $\tau > 0$ the Ekman drift, T_E is toward the south (positive as defined above) and the depth of the upper layer at the western boundary h_{1w} will be less than h_{1e}. For sufficiently large T_E, h_{1w} will vanish, i.e., the thermocline (interface) rises to the surface. With observed values for $\Delta\rho/\rho_2$, τ, and h_{1e} for the North Atlantic, h_{1w} vanishes at about the latitude of Cape Hatteras.

North of that latitude τ is even larger and (5.103) cannot be satisfied because T_E is too large. However, the solution can be extended northward by setting h_{1w} equal to zero at a new longitude ($>\lambda_w$) which is chosen to reduce T_E so that the terms on the right of (5.103) balance. This new longitude marks the westernmost edge of the warm-water mass and is the longitude of the Gulf Stream. But $\lambda > \lambda_w$ means that the Gulf Stream must separate from the coast and extend out to sea. This argument alone does not suffice for higher latitudes where τ eventually becomes negative. We shall return to that issue presently.

Before doing so, however, we discuss the simple physical balances given above. The meridional flow in the interior is a combination of geostrophically balanced motion and Ekman drift. If the flow were completely geostrophic, vanishing T_1 would require equal values of h_1 at the eastern and western edges. But the Ekman wind drift, which does not involve a pressure gradient, accounts for part of the southward transport when $\tau > 0$. Therefore, since the total transport vanishes, there is a net northward geostrophic transport, of magnitude T_E, which requires $h_{1w}^2 < h_{1e}^2$. Thus, the Ekman drift causes the thermocline to rise to the surface. Separation of the Gulf Stream from the coast simply moves the western edge of the warm-water mass (upper layer) eastward so that the smaller Ekman drift acting on that water mass of more limited east–west width can just balance the geostrophic flow determined by h_{1e}^2 (since h_{1w}^2 vanishes).

It is also interesting to note that the Coriolis parameter does not appear in (5.103). In fact, the result is exactly the one obtained for a nonrotating lake where the wind blows the warm water to the leeward edge and causes the thermocline to rise on the windward side. The principal difference between the two phenomena is that the induced pressure gradient drives a vertical circulation in the lake, whereas it is geostrophically balanced in the rotating ocean, thereby generat-

ing a horizontal cell. But the leeward piling up of water is the same in the two cases.

Returning to the problem at high latitudes, we note first that the analysis given above must be supplemented by the remaining dynamic balances. The reader is referred to Veronis (1973a) for the details for the wind-driven model. The qualitative discussion given here is simpler and clearer than in the original paper.

The first problem is that the Sverdrup transport for the interior vanishes with $\hat{\mathbf{k}}\cdot\nabla \times \boldsymbol{\tau}$, and without adding to the simple argument there is no way of supplying warm water to the north of the latitude (≃40°N in the North Atlantic) where the curl vanishes. Second, even supposing that warm water has somehow been supplied to the north, the Sverdrup transport there is northward ($\hat{\mathbf{k}}\cdot\nabla \times \boldsymbol{\tau} > 0$), so the southward return of the flow by a western boundary current would require that the thermocline be deeper on the western side of the boundary layer. That is not possible with the boundary current in mid-ocean.

Both of these issues can be resolved by considering what happens even farther to the north where warm water flows northward and impinges on the northern boundary. In the real ocean and in a model including thermal driving (Veronis, 1978), this water will sink and give rise to a deep circulation and an overturning cell. In a wind-driven model the water travels counterclockwise as an isolated warm boundary current and rejoins the stream at the point of separation. In the analysis given above, this recirculating current represents an excess transport in the separated boundary current. Because its transport does not depend on local winds, it can transport water past the latitude of vanishing wind-stress curl and supply warm water to the interior at high latitudes. When it is included in the analysis, a revised longitude for the separated boundary current is obtained. The calculation, which can be made consistent and quanitative for both the wind-driven model and the one including thermal driving, is contained in the two papers cited above. The path of the separated Gulf Stream is reproduced in figure 5.8. It is especially interesting to note that the vestigial current in the northeastern corner of the basin corresponds to the Norwegian Current (the Alaskan Current in the Pacific) and that its transport is important for the separation of the Gulf Stream and also for the determination of the longitude of the current after it has separated.

The analysis leading to the separation of the Gulf Stream from the coast is contained in a quasi-geostrophic model by Parsons (1969). It was derived independently by Veronis (1973a) as part of a study of the circulation of the World Ocean. The extension poleward of the latitude where the wind-stress curl vanishes is contained in the latter paper. Kamenkovich and Reznik (1972) included a (bottom friction) analysis of the deep circulation induced by the separated current.

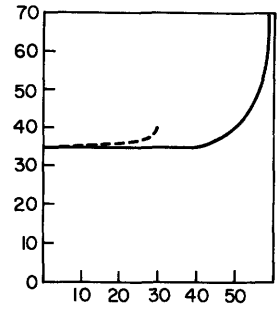

Figure 5.8 The path (solid curve) of the Gulf Stream after it has separated from the coast [from a reduced gravity model by Veronis (1973a)]. The zonal wind stress that drives the system is taken from observations and has zero curl at 40°N. The Norwegian Current is the narrow jet in the northeast. The dashed curve is the prediction for an isolated anticyclonic wind gyre (Parsons, 1969). The latter solution cannot be extended north of the latitude of zero wind-stress curl. Axes are latitude and longitude.

All of the above make use of a steady, linear, quasi-geostrophic model, and it is certain that the details (e.g., the longitude of the separated current) will be altered when a more complete dynamic model is used. The key elements of the argument, however, are the geostrophic balance of downstream velocity in the western boundary current, the Ekman wind drift, and a limited amount of upper-layer water. As long as a different dynamic model does not drastically change those three features (they are pretty rugged and can withstand a lot of battering) the more complicated dynamics can be incorporated to change the details of the results, leaving the main argument unchanged.

By the same token, the present analysis suggests that an explanation of the separation of western boundary currents from the coast must necessarily include the surfacing of the thermocline (with a possible mixed layer at the surface). Western boundary currents can be forced out to sea between wind-driven gyres of opposite sign, but that occurs at low latitudes as well where the phenomenon is qualitatively different because the thermocline does not surface.

In addition, the argument given here depends on properties of global scale. A more precise dynamic treatment based on local properties can lead to a better understanding of the detailed mechanistic balances of the separated current, but the cause of separation seems to be based on global properties.

5.8 Thermohaline Circulation

The physical processes that are involved in the formation of the thermocline have been studied as a separate part of the general circulation. The models incorporate geostrophic dynamics and steady convection

of density, the latter often including vertical diffusion. Though the analyses sometimes make use of the β-plane, the scales are really global and spherical coordinates are more appropriate. The real difficulty is the nonlinearity in the convection of density, and as it turns out, the limited successes of the analyses have been achieved as often in the spherical system as in the β-plane. None of the nonlinear investigations treats a closed basin, though a single eastern boundary is sometimes included. A closed two-layer basin is tractable (see section 5.8.2).

5.8.1 Continuous Models for an Open Basin

The starting point for these studies is the simplified set of equations in spherical coordinates

$$fv = \frac{1}{a\cos\phi}\frac{\partial P}{\partial \lambda}, \tag{5.104}$$

$$fu = -\frac{1}{a}\frac{\partial P}{\partial \phi}, \tag{5.105}$$

$$\frac{\partial P}{\partial z} = -g\frac{\dot{\rho}}{\rho_m}, \tag{5.106}$$

$$\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \phi}(v\cos\phi) + a\cos\phi\frac{\partial w}{\partial z} = 0, \tag{5.107}$$

$$\frac{u}{a\cos\phi}\frac{\partial \rho}{\partial \lambda} + \frac{v}{a}\frac{\partial \rho}{\partial \phi} + w\frac{\partial \rho}{\partial z} = K\frac{\partial^2 \rho}{\partial z^2}, \tag{5.108}$$

where the last term in (5.108) contains the only dissipative process, vertical diffusion of density. These equations cannot be used to analyze the balances for a closed basin because there is not enough flexibility to satisfy even the condition of no normal flow through the boundaries. Essentially all past efforts have been restricted to this open system.

In principle, there is enough flexibility to satisfy four boundary conditions in the vertical (three if diffusion is omitted). These must be chosen to be consistent with the form of the solution that is obtained; therefore, much of the flexibility is lost. Still, it is possible to obtain interesting, if limited, information about the thermal structure.

Results based on linearized models by Lineikin (1955) and Stommel and Veronis (1957) were superseded by the nonlinear models of Welander (1959), who treated the ideal fluid system ($K = 0$), and Robinson and Stommel (1959), who obtained a similarity solution with K included. Stommel and Webster (1962) made use of the latter model to determine the dependence of the vertical structure of w and T on the value of K and on surface boundary values of w and T. Exact solutions were obtained by Fofonoff (1962a), Blandford (1965), Kozlov (1966), Needler (1967, 1972) and Welander (1959, 1971a). More recently the problem has been reformulated by Hodnett (1978) with density instead of vertical distance as an independent coordinate. A review of the earlier papers is given by Veronis (1969).

The variables u, v, and ρ are given in terms of P by (5.104) to (5.106). These can be substituted in (5.108) to give w in terms of P and the continuity equation then yields Needler's pressure equation

$$\kappa\sin\phi\cos\phi(P_{zz}P_{zzzz} - P_{zzz}^2)$$

$$= P_{zzz}\frac{\partial(P_z,P)}{\partial(\lambda,\phi)} + P_{zz}\frac{\partial(P,P_{zz})}{\partial(\lambda,\phi)} + \cot\phi P_\lambda P_{zz}^2, \tag{5.109}$$

where $\kappa = 2\Omega K a^2$.

Welander (1959, 1971b) defined the variable

$$M = \int_0^z P\,dz + a^2 f\sin\phi\int_0^\lambda w(\lambda,\phi,0)\,d\lambda \tag{5.110}$$

(so that $P = M_z$), to obtain the simpler equation

$$\kappa\sin\phi\cos\phi M_{zzzz} + \frac{\partial(M_{zz},M_z)}{\partial(\lambda,\phi)}$$

$$-\cot\phi M_\lambda M_{zzz} = 0. \tag{5.111}$$

By integrating (5.104) from λ to 0 and setting $P(0,\phi,z) = 0$, it is easy to see that M/f is the geostrophic, wind-driven, meridional transport between λ and 0 and below level z. $P(0,\phi,z) \neq 0$ means that the reference pressure is not passive (there must be density anomalies at $\lambda = 0$) and it gives rise to an added transport. In interpreting the system, however, it is best to think in terms of $P(0,\phi,z) = 0$.

Needler (1967) derived a solution that had been obtained previously by Blandford (1965) under more restrictive conditions and by Welander (1959), who ignored K. In his analysis Needler proposed the following form with three arbitrary functions of λ and ϕ:

$$P(\lambda,\phi,z) = A(\lambda,\phi) + B(\lambda,\phi)e^{zC(\lambda,\phi)}. \tag{5.112}$$

This is a solution to (5.109) provided that A, B, and C are independent of λ *or* that C is given by

$$C(\lambda,\phi) = c/\sin\phi, \tag{5.113}$$

where c is a constant. Only the latter case seems to have received much attention, even though the case with $A_\lambda = B_\lambda = C_\lambda = 0$ could be a zero-order solution to which necessary corrections could be made (away from the coasts the oceans exhibit a quasi-zonal distribution of properties).

With (5.113) the remaining variables are given by

$$\frac{\rho}{\rho_m} = -\frac{cB}{g\sin\phi}e^{zc/\sin\phi}, \tag{5.114}$$

$$u = -\frac{1}{fa}\left\{A_\phi + \left[B_\phi - \frac{zc\cos\phi}{\sin^2\phi}B\right]e^{zc/\sin\phi}\right\}, \tag{5.115}$$

$$v = \frac{1}{fa\cos\phi}\left[A_\lambda + B_\lambda e^{zc/\sin\phi}\right], \tag{5.116}$$

$$w = \frac{1}{fa^2}\left[\frac{B_\phi}{c}e^{zc/\sin\phi} + \frac{\tan\phi}{cB}\frac{\partial(A,B)}{\partial(\lambda,\phi)} + \frac{A_\phi}{c} + \frac{zA_\phi}{\sin\phi}\right] + \frac{Kc}{\sin\phi}. \tag{5.117}$$

The general functions A and B and the constant c are available to satisfy boundary conditions.

Before proceeding further, it is worth noting that the diffusivity K appears only in the last term of w. In fact, with $w_d \equiv Kc/\sin\phi$ it is evident that

$$w_d\frac{\partial\rho}{\partial z} = K\frac{\partial^2\rho}{\partial z^2}. \tag{5.118}$$

If we write the vertical velocity as $w = w_a + w_d$ (the subscripts a and d correspond to *advective* and *diffusive*, respectively) we see that w_d absorbs the diffusive effect and w_a satisfies the ideal fluid system with $K = 0$. Hence, with this solution diffusion plays a minor role in the dynamic balances, and the essential balances coincide with those in Welander's (1959) solution.

It would appear at first sight that the general functions A and B and the constant c are available to satisfy boundary conditions. We note, however, that only B multiplies the exponential and that the properties described by A penetrate undiminished to the bottom. Since we expect neither the surface density nor the Ekman pumping to generate effects that penetrate undiminished to the bottom, we can discard A for the time being and concentrate on $B(\lambda,\phi)$. Evaluating (5.114) at $z = 0$ yields

$$\frac{\rho(\lambda,\phi,0)}{\rho_m} = -\frac{cB}{g\sin\phi}, \tag{5.119}$$

so that the product cB is determined by the surface density distribution. The constant c can be evaluated by matching the e^{-1} decay depth with the middle of the thermocline at one latitude. This is not really a boundary condition, but it is forced on us if we want the solution to generate a realistic vertical density profile. With $c^{-1} = 1500$ m, a surface temperature (a linear measure of density by the Boussinesq approximation) proportional to $\cos(\phi + 10°)$ and a reference temperature of 2.45°C at 5000 m depth at $\phi = 10°$, Needler constructed the isotherm pattern in a vertical (ϕ versus z) section shown in figure 5.9. The choice of c corresponds to a thermocline depth of 750 m at $\phi = 30°$.

It is not possible to satisfy any more conditions. Hence, the Ekman pumping velocity is determined by the surface density. One could specify w_E instead and then the surface density would be determined. We shall return to this point at the end of this section.

The pattern shown in figure 5.9 reproduces the observed density minimum at mid-latitude for levels near the thermocline. The same feature appears in the solutions of Welander (1959) [from the M equation (5.111)] and of Robinson and Stommel (1959). The pattern lacks the abrupt variations with latitude of the observed system, though that could be remedied by choosing the surface conditions more realistically, a procedure that is equivalent to introducing higher-order dynamic effects through the boundary conditions. But we should remember that this is an open system and that introducing more realistic surface conditions to obtain a more pleasing pattern requires that fluid leave the region governed by the simple, assumed balances and reenter it after substantial changes in the properties have taken place. This would amount to rationalizing the data without really learning anything about the processes that bring about the desired change.

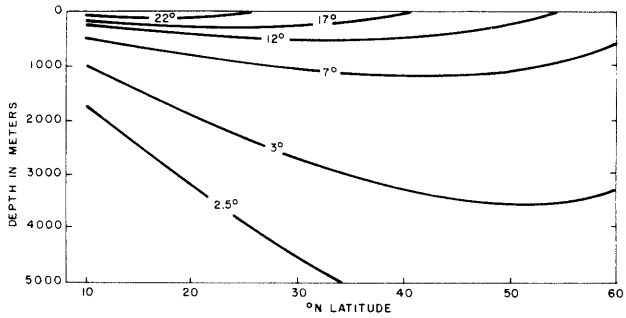

Figure 5.9 Isotherms, or isopycnals, from Needler's (1967) thermocline model with vertical dependence $\propto \exp(cz/\sin\phi)$. The observed maximum vertical penetration of warm water at mid-latitude is a feature of almost all of the thermocline models.

The vertical velocity w_a decays exponentially with depth so this advective solution is a surface boundary layer solution. The analysis described is valid as long as the boundary layer thickness (determined by c) is small compared to the depth of the ocean. Below the boundary layer the vertical velocity is given by w_d and is induced by vertical diffusion. Once c is chosen w_d is determined by K. However, the value (and even the form!) of K is as unknown and as unmeasurable as w_d. In fact, the balance given by (5.118) is often used to obtain an estimate of the scale depth K/w_d, and when it is used in conjunction with measured vertical profiles of tracers with a known decay rate, individual estimates of w_d and K can be made. But the whole procedure makes use of purely vertical balances and is itself questionable. In the present context we can only conclude that the model is not sufficiently constrained by K to enable us to determine its effects. Typical values used for K and w_d in deep water (Munk, 1966) are $K = 10^{-4}\ \mathrm{m^2\,s^{-1}}$ and $w_d = 10^{-7}\ \mathrm{m\,s^{-1}}$.

The finite depth of the ocean requires that the vertical velocity w_d at the base of the thermocline region match the vertical velocity at the top of the layer below. Within the framework of the present approach

that means that a barotropic mode must be included to satisfy the boundary condition of zero normal velocity at the bottom. Needler (1967, 1972) gives a thorough discussion of the issue. In his second paper he seeks the conditions under which $P_1(\lambda, \phi, z) = P(\lambda, \phi, z) + D(\lambda, \phi)$ is a solution where P itself is a solution, i.e., what are the restrictions on P for an arbitrary barotropic mode $D(\lambda, \phi)$ to be added as part of the solution? It is easy to see from (5.109) that, since D_z vanishes and P must satisfy the equation, one is left with

$$D_\phi \frac{\partial(P_{zz}, P_z)}{\partial(\lambda, z)} - D_\lambda \left[\frac{\partial(P_{zz}, P_z)}{\partial(\phi, z)} + \cot\phi P_{zz}^2 \right] = 0. \quad (5.120)$$

Furthermore, with P independent of D, and D an arbitrary function, each of the P expressions must vanish. A straightfoward argument then shows that P must be of the form

$$P = \sin\phi \, \Phi(\eta) + E(\lambda, \phi), \quad (5.121)$$

$$\eta = \frac{cz}{\sin\phi} + F(\lambda, \phi), \quad (5.122)$$

where E, F, and Φ are arbitrary functions of their arguments. Hence $E(\lambda, \phi)$ can be absorbed into $D(\lambda, \phi)$. With $K \neq 0$ the solution reduces essentially to the exponential one given earlier. For $K = 0$ it can be shown that the conditions given by (5.120) are equivalent to the statement that the density and potential vorticity ($f\rho_z$ in this case) are functions of each other. N. A. Phillips (1963) had already shown that Welander's (1959) (hence, Needler's) exponential solution satisfied $f\rho_z = 2\Omega c\rho$, a special case of the above. We shall return to this point shortly when we discuss Welander's more general solutions for an ideal fluid thermocline.

Needler (1972) satisfies the bottom boundary condition of zero normal flow by using the arbitrary barotropic mode introduced above with $K = 0$. In addition, he shows that the consistency conditions required in order to add an arbitrary barotropic mode make it possible to satisfy only two of the three independent conditions: $w(\lambda, \phi, 0)$, $T(\lambda, \phi, 0)$, and zero normal flow at the bottom. Once two of them are satisfied, the third is determined.

Needler's two papers are highly recommended reading. He discusses both the possibilities and the inadequacies of this approach to the thermocline circulation and he gives a sound analysis of some very difficult problems.

Welander has spearheaded perhaps the most significant advances in the theory of the thermohaline circulation. His first paper on the problem contained the exponential solution given above with an arbitrary function available to satisfy a general surface boundary condition. The next paper (Robinson and Welander, 1963) merged his approach with that of Robinson and Stommel (1959). In his third paper for steady ideal flows (Welander, 1971a) he first derived, and then applied, the general relationship (5.11) between potential vorticity q, density ρ, and the Bernoulli function B to the geostrophic, hydrostatic system of equations (5.104) to (5.108). The latter yields the simplified forms $q = f\rho_z$ and $B = p + g\rho z$ so that equation (5.11) reduces to

$$\sin\phi \, \rho_z = F(\rho, p + g\rho z), \quad (5.123)$$

where F is an arbitrary function.

Linearization of F yields

$$\sin\phi \, \rho_z = a\rho + b(p + g\rho z) + c, \quad (5.124)$$

where a, b, and c are arbitrary constants. Upon differentiation with respect to z and use of the continuity equation, we obtain

$$\sin\phi \, \rho_{zz} = (a + bgz)\rho_z, \quad (5.125)$$

and two integrations yield

$$\rho(\lambda, \phi, z) = \rho(\lambda, \phi, 0) - C(\lambda, \phi) \int_z^0 e^{bg(\zeta + z_0^2)/(2\sin\phi)} d\zeta, \quad (5.126)$$

where $z_0 = a/bg$ and C is an arbitrary function of λ and ϕ. It is evident that b must be negative; otherwise the integral grows indefinitely with z. Furthermore, z is negative, so $a > 0$ implies a monotonic profile. With $a < 0$ an inflection point occurs at $z = -ab/g$. Thus, the constants a and b can be chosen to give an inflection point at a desired depth and a desired thickness to the thermocline. The latter varies inversely as $\sin^{1/2}\phi$. Welander fitted the constants to match the observed density profile along 160°W in the South Pacific (Reid, 1965) shown in figure 5.10A. His theoretical solution (figure 5.10B) captures the general structure of the observed profile, though it is smoother, as one would expect. In the construction Welander used the observed surface density for $\rho(\lambda, \phi, 0)$, and $C(\lambda, \phi)$ was chosen to give a deep constant density. He gives no other details for the construction.

This solution is a remarkable step forward. It takes advantage of only the simplest of the possibilities that the general conservation integrals contain and it justifies Welander's faith in the use of ideal-fluid theory to obtain realistic results. Welander also presented more general solutions to the system, but the latter are quite formal and no detailed results from them have been reported. Making use of this first integral to the general system is very promising and it is surprising that this path has not been pursued more actively.

In a subsequent paper on this topic Welander (1971b) explored the possible balances in the M equation (5.111) by means of a scale analysis. His conclusions can be summarized without detailed analysis by making use of the results already found. In regions of Ekman suction ($w > 0$) diffusive processes adjust the density to surface values, a simple possible balance

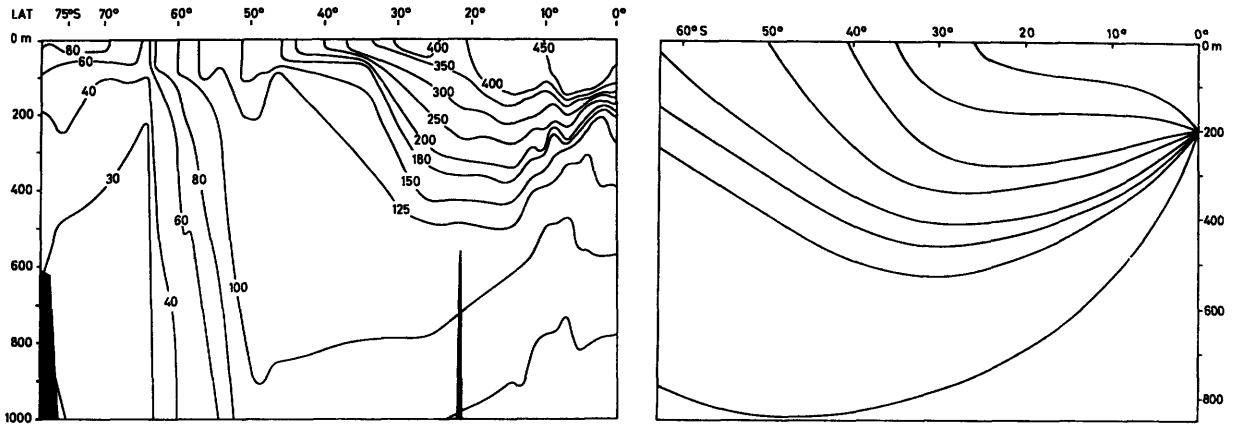

Figure 5.10 (A) Contours of thermosteric anomaly (units of $10^{-5}\,cm^3\,g^{-1}$) in the upper kilometer of the South Pacific along 160°W from Reid (1965). (B) Isopycnal contours from Welander's (1971a) ideal-fluid thermocline model.

being given by (5.118) with a scale depth $H \sim K/w$. With K fixed this diffusive depth decreases with increasing w. Ekman pumping ($w < 0$), on the other hand, forces lighter water into the oceanic interior so that the surface value of ρ extends to some depth. With ρ constant, (5.118) is satisfied trivially and advection must be important so that the order-of-magnitude balance is $V/a \sim W/H_a$, where we use the global scale a in the horizontal and H_a is the vertical (advective) scale. W and V are velocity scales. Geostrophic balance yields $fV/H_a \sim g\rho/a\rho_m$ and eliminating V then yields $H_a \sim (fWa^2\rho_m/g\rho)^{1/2}$, which increases with W. Therefore, more intense surface forcing gives rise to a deep advective layer where $w < 0$ and an advective layer under a thin diffusive layer where $w > 0$. The geostrophic transport is carried by the advective layer, the diffusive process serving simply to adjust the density to surface values. Welander gives a more detailed analysis of the possibilities to show that an advective layer must be present. He also points out that a deep diffusive layer with a balance like (5.118) but with w_d different from W_E is also likely.

5.8.2 Layered Models

Stommel (1957b), noting that upwelling suggested by (5.118) would produce a vertical divergence from the (level) bottom to the base of the thermocline, assumed that the deep ocean is homogeneous and used the planetary divergence relation $v = a \tan\phi\, \partial w/\partial z$ to determine the meridional velocity. With a uniform upwelling at the base of the thermocline v is poleward everywhere in the interior. Then with $u = 0$ at all eastern boundaries (taken along meridians) he calculated the zonal velocities by integrating the continuity equation with respect to longitude to obtain a trajectory pattern for the interior of the world ocean.

Interior upwelling of deep water requires that sources of deep water be present somewhere. Stommel chose sources of equal strength in the North Atlantic and in the Weddell Sea (South Atlantic). He assumed that these source waters flowed along western boundary layers and then eastward to supply the upwelling flow in the interior. The transports in the western boundary layers were obtained by requiring mass conservation for a basin bounded by two meridians, a northern boundary and the latitude in question. The pattern of flow that results from these considerations is shown in figure 5.11.

Veronis (1978) has combined this reasoning with an analysis similar to that of section 5.7 to construct a two-layer model of the thermohaline circulation in the world ocean with wind stress acting on the surface. The intensity of the upwelling and the locations and intensities of the sources of deep water can be deduced from the model. The reasoning is as follows.

On the basis of an expected balance like that of (5.118) in deep water, assume a vertical flux of water through the interface from the lower to the upper layer. The amplitude of upwelling is taken to be horizontally uniform but of unknown magnitude. The height h_2 of the interface above the level bottom is determined by the wind stress acting on the surface and the upwelling through the interface. The two-layer, steady, linear system of equations (5.62) to (5.67) on a sphere can be manipulated to yield a first-order partial differential equation for h_2 with coefficients depending on h_2. This quasi-linear equation can be integrated along characteristics from (assumed) known values on the eastern boundary to give h_2 throughout the interior.

Here, too, the assumed upwelling will require sources of deep water that will flow along the western boundaries to supply the oceanic interior. The downstream flow in the western boundary layers is assumed to be geostrophic. Mass conservation of water in both layers is required in the region bounded by boundaries at the sides and along the north and by the latitude in question. As in section 5.7 this will lead to an expression for the depth of the thermocline (or the height h_2)

at the western boundary. This expression will depend not only on the value of h_2 at λ_e, however, but will be a function also of the unknown amplitude of upwelling and the unknown sources of deep water. If the latter quantities were known, it would be possible to determine h_2 and, in particular, the latitude at which the thermocline rises to the surface. North of this latitude the western boundary current will flow eastward and poleward across the open ocean. Since it represents the boundary between upper- and lower-layer water, it will also determine the area covered by upper-layer water, and therefore, the total amount of upwelling (w times the area) that occurs north of any latitude. So we have an implicit problem with h_2, w, and the strengths and locations of the sources interrelated.

As stated earlier, obtaining an estimate for the upwelling is not straightforward, depending as it does on complex, turbulent, convective processes. Therefore, the problem is inverted. Instead of assuming values for w and for the strengths and locations of the sources to determine the surfacing latitude, the latter, a simple observable, is taken from observation and the former quantities are determined by the model. It turns out that to evaluate w *and* the sources requires more than one piece of information. For example, in the Pacific the surfacing latitudes of the Kuroshio (35°N) and the East Australian Current (31°S) are specified and these yield an upwelling velocity w of magnitude 1.5×10^{-7} m s^{-1} and a distribution of sources of deep water at the northern boundary (along the Alaskan-Aleutian current system), at the latitude of separation of the Kuroshio, and along the Australian coast from 31 to 35°S. The circulation patterns and the details of the calculations are given in the paper cited.

Some major features (e.g., deepest penetration of light water at mid-latitude) are consistent with those obtained by the continuous thermocline models discussed earlier. However, the present model also allows one to close the circulation with boundary layers, and in particular, to determine the open-ocean path of the separated boundary current. For the continuous model that possibility would enable one to adjust surface boundary conditions as part of the analysis in order to obtain a more realistic vertical density distribution with latitude.

Most noteworthy of the results obtained with this two-layer model is the deduced magnitude of the assumed upwelling. Most estimates for w are made from observed tracer distributions by assuming that vertical advective and diffusive processes balance locally. They yield values between 10^{-7} and 2×10^{-7} m s^{-1} (see chapter 15). The present value lies midway in the range cited and is based on global circulation processes with no reference to the vertical diffusive process.

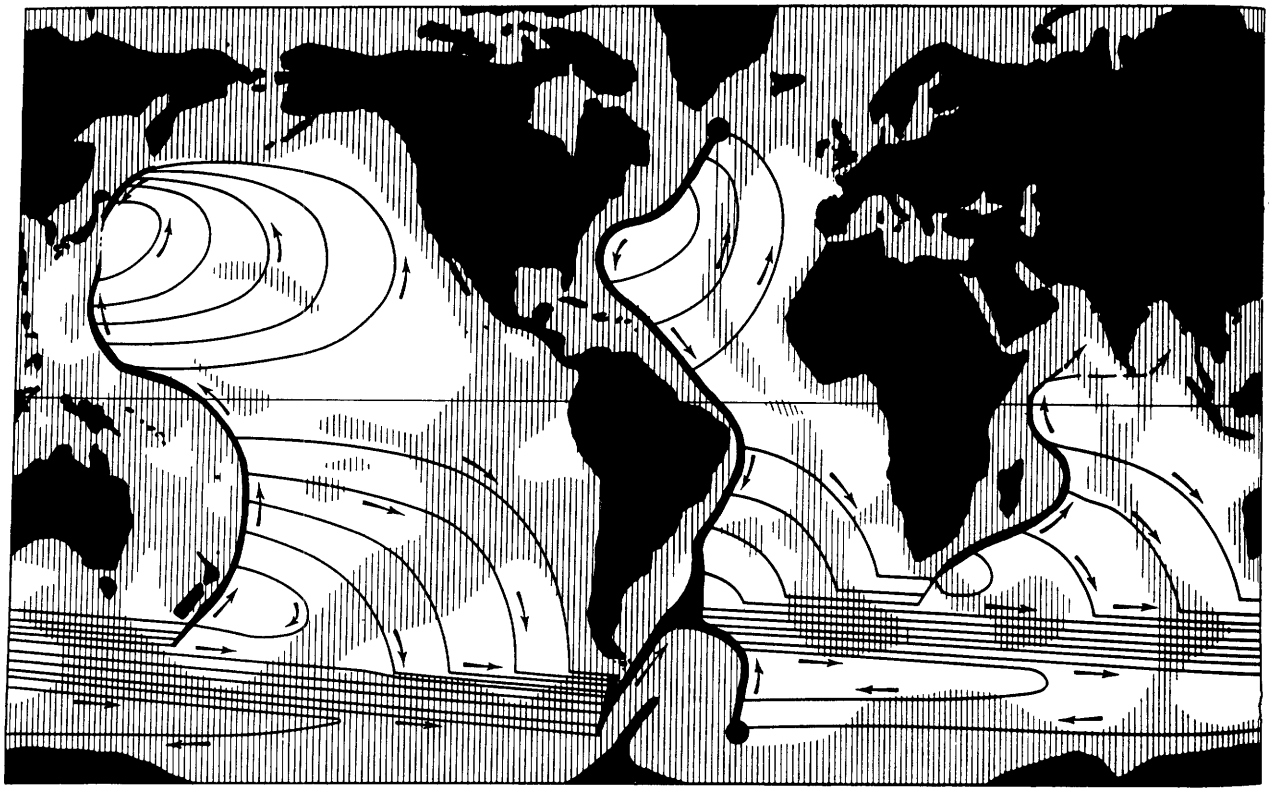

Figure 5.11 The abyssal circulation obtained by Stommel (1958) and generated by equal sources in the North Atlantic and in the Weddell Sea with uniform upwelling elsewhere.

5.9 Free Waves for a Constant-Depth Two-Layer Ocean on the β-Plane

The linear equations for a two-layer ocean on the β-plane with constant depth are (5.62) to (5.67) (with $Ro \ll 1$):

$$u_{1t} - fv_1 = -g\eta_{1x}, \tag{5.127}$$

$$v_{1t} + fu_1 = -g\eta_{1y}, \tag{5.128}$$

$$\frac{1}{H_1}(\eta_1 - \eta_2)_t + u_{1x} + v_{1y} = 0, \tag{5.129}$$

$$u_{2t} - fv_2 = -g[(1-\epsilon)\eta_1 + \epsilon\eta_2]_x, \tag{5.130}$$

$$v_{2t} + fu_2 = -g[(1-\epsilon)\eta_1 + \epsilon\eta_2]_y, \tag{5.131}$$

$$\frac{1}{H_2}\eta_{2t} + u_{2x} + v_{2y} = 0, \tag{5.132}$$

where $\epsilon = \Delta\rho/\rho_2$ and H_1, H_2 are the constant mean depths of the two layers. Elimination of all but one variable leads to a sixth-order equation. However, it is possible to simplify the mathematics to a third-order system by introducing normal modes (Veronis and Stommel, 1956). We do so by multiplying (5.130) to (5.132) by a constant α and adding to (5.127)–(5.129), respectively, to derive

$$(u_1 + \alpha u_2)_t - f(v_1 + \alpha v_2)$$
$$= -g\{[1 + \alpha(1-\epsilon)]\eta_1 + \alpha\epsilon\eta_2\}_x, \tag{5.133}$$

$$(v_1 + \alpha v_2)_t + f(u_1 + \alpha u_2)$$
$$= -g\{[1 + \alpha(1-\epsilon)]\eta_1 + \alpha\epsilon\eta_2\}_y, \tag{5.134}$$

$$\left[\frac{\eta_1}{H_1} + \left(\frac{\alpha}{H_2} - \frac{1}{H_1}\right)\eta_2\right]_t$$
$$+ (u_1 + \alpha u_2)_x + (v_1 + \alpha v_2)_y = 0. \tag{5.135}$$

The velocities appear in the same combination $\mathbf{v}_1 + \alpha\mathbf{v}_2$, everywhere, but the surface deviations appear in two different forms. If the latter are the same except for a multiplicative constant, the three equations will involve only three variables (plus parameters). So let us define

$$\mathbf{V}_1 = \mathbf{v}_1 + \alpha\mathbf{v}_2, \tag{5.136}$$

$$\Phi = [1 + \alpha(1-\epsilon)]\eta_1 + \alpha\epsilon\eta_2, \tag{5.137}$$

$$\frac{\Phi}{h} = \frac{\eta_1}{H_1} + \left(\frac{\alpha}{H_2} - \frac{1}{H_1}\right)\eta_2, \tag{5.138}$$

where h is a constant to be determined.

Equating Φ in (5.137) and (5.138) yields

$$[1 + \alpha(1-\epsilon)]\eta_1 + \alpha\epsilon\eta_2$$
$$= h\left[\frac{\eta_1}{H_1} + \left(\frac{\alpha}{H_2} - \frac{1}{H_1}\right)\eta_2\right]. \tag{5.139}$$

Since this must be valid for arbitrary η_1 and η_2, the coefficients of η_1 and η_2 must be the same, i.e.,

$$1 + \alpha(1-\epsilon) = \frac{h}{H_1}, \qquad \alpha\epsilon = \frac{h\alpha}{H_2} - \frac{h}{H_1}. \tag{5.140}$$

Eliminating h yields

$$\frac{H_1}{H_2}(1-\epsilon)\alpha^2 + \left(\frac{H_1}{H_2} - 1\right)\alpha - 1 = 0. \tag{5.141}$$

For small ϵ the two values for α are

$$\alpha_1 \simeq \frac{H_2}{H_1}, \qquad \alpha_2 \simeq -1. \tag{5.142}$$

Corresponding values of h and the variables are

$$h_1 = H_1 + H_2, \qquad h_2 = \epsilon H_1 H_2/(H_1 + H_2), \tag{5.143}$$

$$\mathbf{V}_1 = \mathbf{v}_1 + \frac{H_2}{H_1}\mathbf{v}_2, \qquad \mathbf{V}_2 = \mathbf{v}_1 - \mathbf{v}_2, \tag{5.144}$$

$$\Phi_1 = \frac{H_1 + H_2}{H_1}\eta_1 + \epsilon\frac{H_2}{H_1}\eta_2,$$

$$\Phi_2 = \frac{\epsilon H_2}{H_1 + H_2}\eta_1 - \epsilon\eta_2. \tag{5.145}$$

Then the six equations can be reduced to the following two independent sets ($i = 1,2$)

$$U_{it} - fV_i = -g\Phi_{ix}, \tag{5.146}$$

$$V_{it} + fU_i = -g\Phi_{iy}, \tag{5.147}$$

$$\Phi_{it} + h_i(U_{ix} + V_{iy}) = 0, \tag{5.148}$$

which describe the linear, time-dependent motions on a β-plane for a barotropic ocean of depth $H = H_1 + H_2$ when $i = 1$ and for a baroclinic ocean of depth $\epsilon H_1 H_2/H$ when $i = 2$. When Φ_2 is replaced by Φ_2/ϵ and g by $g' = \epsilon g$, the internal (baroclinic) case is then called the reduced-gravity system with depth H_1H_2/H. If U_i and Φ_i are eliminated from (5.146) to (5.148) the resulting equation in V_i is

$$(\partial^2_{ttt} + f^2\partial_t - gh_i\partial^3_{xxt}$$
$$+ gh_i\beta\partial_x - gh_i\partial^3_{yyt})V_i = 0. \tag{5.149}$$

If f is now replaced by its reference value, f_0 (this is lowest order in L/a), all coefficients are constant. By substituting $V_i \sim e^{i(-\omega t + kx + ly)}$, we obtain the frequency equation for free waves,

$$\omega_i^3 - (f_0^2 + gh_iK^2)\omega_i^2 - gh_i\beta k = 0,$$
$$K^2 = k^2 + l^2. \tag{5.150}$$

This yields the (approximate) dispersion relations[7]

$$\omega_{i1} = f_0\sqrt{1 + \lambda_i^2K^2}, \qquad \omega_{i2} = -\omega_{i1},$$
$$\omega_{i3} = -\frac{\beta k\lambda_i^2}{1 + \lambda_i^2K^2}. \tag{5.151}$$

A plot of ω versus K is shown in figure 5.12 for $k = l$ at mid-latitude ($f_0 = 10^{-4}\,s^{-1}$, $\beta = 2 \times 10^{-11}\,m^{-1}\,s^{-1}$) and for the λ values given below.

The first two of these waves for each mode reduce to long gravity waves for $\lambda_i K \gg 1$ (small wavelength) and to inertial waves for $\lambda_i K \ll 1$ (large wavelength). The dividing scale is λ_i, which is about 2000 km for the barotropic mode and about 36 km for the baroclinic. The third wave for each case is a westward-traveling Rossby wave that arises because of conservation of potential vorticity $(\zeta + f)/h$ on the β-plane. As a fluid column changes latitude or moves into a region with different depth, a relative vorticity is generated to keep $(\zeta + f)/h$ constant. The distortion of the free surface or of the density interface has an effect for scales larger than λ_i and the waves become nondispersive ($\omega_{i3} \sim \beta k \lambda_i^2$). For the baroclinic mode the period of this wave at mid-latitude is of the order of years for scales of the size of an ocean basin, making the linear baroclinic response of the ocean very slow (Veronis and Stommel, 1956; see chapters 10 and 11).

Lighthill (1967, 1969) applied wave theory in the limit of vanishing frequency to study the development of forced, steady flows. In his analysis of the responses of the equatorial Indian Ocean he used the fact that λ_i becomes very large near the equator to conclude that the baroclinic response there is much faster—of the order of weeks—than it is at mid-latitudes.

A comprehensive account of barotropic Rossby waves is given by Longuet-Higgins (1964, 1966). Here we shall make use only of a simple property that relates directly to large-scale circulation. The dispersion relation (5.151) in the (ω_{i3}, k)-plane in figure 5.12 shows that the phase velocity ω_{i3}/k is westward and that a given value of ω_{i3} corresponds to two wavelengths, a short wave with small velocity and a long wave with a fast velocity. The zonal group velocity $\partial\omega_{i3}/\partial k$, which transports energy zonally, is westward for the fast waves and eastward for the slow ones. Thus, at a meridional boundary, where the energy flux must vanish, the energy of an incoming wave will be reflected quickly at an eastern boundary and will accumulate at a western boundary.

Pedlosky (1965b) offered this as an alternative explanation of westward intensification, and N. A. Phillips (1966a) used these reflected properties to account for the more frequent observation of intense eddy motions near the western versus the eastern regions of the North Atlantic. Ibbetson and Phillips (1967) report a laboratory confirmation of this east-west distribution of energetic eddy motions. Observed long barotropic Rossby waves in the ocean are suggested in the bottom pressure records in the MODE area by Brown et al. (1975). Freeland, Rhines, and Rossby (1975) show longitude versus time plots of the streamfunction inferred from objective maps of currents at 1500 m depth in the

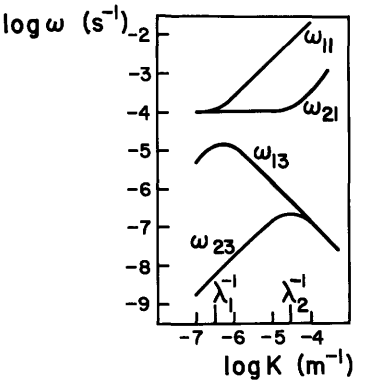

Figure 5.12 Frequency-wavenumber diagram for waves in a two-layer, constant-depth ocean in the β-plane. The upper two curves are inertiogravity waves, the two lower are Rossby waves. Transitions in the dispersion curves occur at the deformation scales shown on abscissa.

MODE area (figure 10.6). There is a definite tendency for constant phase lines to move westward, with phase speeds ranging from 0.02 to 0.12 $m\,s^{-1}$ (average 0.05 $m\,s^{-1}$). With $l = k$ this suggests wavelengths clustering around 400 km. A time versus latitude plot shows no definite north-south propagation.

The slow phase velocities deduced for baroclinic Rossby waves make the linear theory less reliable because particle motions equal to and exceeding the wave speeds occur in all parts of the ocean. Rhines (1977) identifies thermocline eddies (intense baroclinic modes, principally confined to the waters above the thermocline) with these baroclinic Rossby waves and offers evidence of their existence in observed records from open ocean regions. These noisy, nearly stationary modes make the determination of the slow mean flow in the open ocean a difficult task.

5.10 Effect of Bottom Topography on Quasi-Geostrophic Waves

The results of the previous section were extended by Rhines (1970) to include simple bottom topography. Though the general normal-mode procedure does not work in this case, Rhines modified it for quasi-geostrophic wave motions when topography varies linearly in y. The multiplicative constant α is a function of wavelength in that case and the method is difficult to interpret when topography varies with x as well. Because the equations in terms of the surface height lead to a quadratic dispersion relation in a straightforward manner even in the latter case, we shall not use normal modes.

5.10.1 Two-Layer Model

The linearized, potential vorticity equation for each layer becomes

$$\zeta_t + \beta v - \frac{fh_t}{H} + \frac{f\mathbf{v}\cdot\nabla h}{H} = 0. \tag{5.152}$$

For quasi-geostrophic motions we can substitute the geostrophic velocity in $\zeta = v_x - u_y$ and in $\mathbf{v}$. Furthermore, we have $h_{2t} = \eta_{2t}$, $h_{1t} = (\eta_1 - \eta_2)_t$, $\mathbf{v}_1\cdot\nabla h_1 \simeq 0$ (for linear flows), $\mathbf{v}_2\cdot\nabla h_2 \simeq -\mathbf{v}_2\cdot\nabla\eta_3$ where bottom topography η_3 is defined by a linear function of x and y and $\eta_3 \ll H_2$ so that the constant depth H_2 can be used in the coefficients. Then we obtain

$$\frac{f^2}{gH_1}(\eta_1 - \eta_2)_t - \nabla^2\eta_{1t} - \beta\eta_{1x} = 0, \tag{5.153}$$

and

$$\frac{f^2}{gH_2}\eta_{2t} - \nabla^2 p_{2t} - \beta p_{2x} + \frac{f}{H_2}\frac{\partial(\eta_3, p_2)}{\partial(x,y)} = 0, \tag{5.154}$$

where $p_2 = (1 - \epsilon)\eta_1 + \epsilon\eta_2$. The substitution $e^{i(-\omega t + kx + ly)}$ for η_1 and η_2 then leads to the quadratic frequency equation

$$(1 + \lambda^2K^2 + \lambda_1^2\lambda_2^2K^4)\omega^2 + \{(\lambda^2 + 2\lambda_1^2\lambda_2^2K^2)k\beta$$

$$+ (1 + \lambda_1^2K^2)\lambda_2^2kb\}\omega + \lambda_1^2\lambda_2^2k^2\beta(b + \beta) = 0, \tag{5.155}$$

where $\lambda^2 = gH/f^2$, $\lambda_2^2 = gH_2/f^2$, $\lambda_1^2 = \epsilon gH_1/f^2$, $b = f(k\eta_{3y} - l\eta_{3x})/H_2k$. The three λ's correspond, respectively, to the radii of deformation for a barotropic fluid with the total depth, a barotropic fluid of depth H_2, and a reduced gravity fluid with depth H_1. The quantity b is a "topographic β-effect" that simply reenforces β when the depth shallows northward and $\eta_{3x} = 0$. More generally, it combines with β [through $(d/dt)(f/h)$] to determine a new, pseudonorth direction.

For the ocean λ is large (2000 km), λ_2 is nearly as large (~1600 km) and λ_1 is small (~40 km). Hence, for large wavelengths (small K) (5.155) yields the barotropic solution

$$\omega_1 = -\frac{k\beta + (kbH_2/H)}{K^2 + (1/\lambda^2)}, \tag{5.156}$$

where kbH_2/H is independent of H_2. This dispersion relation reduces to the one for ordinary barotropic Rossby waves when b vanishes. With $b \neq 0$, it represents barotropic Rossby waves with both direction and frequency modified by topography. Where the topography is strong, it yields topographic Rossby waves with the direction of propagation to the left of upslope. For negative b (depth decreasing southward) it is possible for the two restoring mechanisms to cancel each other almost completely.

The second solution for large wavelengths is the baroclinic mode

$$\omega_2 = -\frac{k\beta\lambda_1^2\lambda_2^2}{\lambda^2}\frac{\beta + (bH/H_2)}{\beta + b}$$

$$= -\frac{k\beta g'H_1H_2}{f^2H}\frac{\beta + (bH/H_2)}{\beta + b}. \tag{5.157}$$

When topography is weak ($b \ll \beta$) this reduces to the nondispersive wave

$$\omega_2 = -\frac{k\beta g'H_1H_2}{f^2H} \tag{5.158}$$

[limit of (5.151) for small K]. In this case, both layers are in motion. With strong topography ($b \gg \beta$) the bottom depth drops out of the dispersion relation and (5.157) becomes

$$\omega_2 = \frac{k\beta g'H_1}{f^2}. \tag{5.159}$$

The point here is that the bottom slope is so large that only a small excursion by a column of fluid in the bottom layer is required to bring about vortex stretching, so that most of the motion is confined to the upper layer where the only restoring force is (the relatively weak) β. Thus, strong topography acts to decouple the layers and to increase the frequency, hence the phase speed, in the upper layer by a factor of H/H_2 over that with weak topography. Rhines (1977) emphasized this detuning effect of topography and showed that it applies for flows of much larger amplitudes.

For small wavelengths, $\lambda_1^2K^2 \gg 1$, (5.155) reduces to

$$\omega^2 + \frac{k}{K^2}(2\beta + b)\omega + \frac{k^2}{K^2}\beta(\beta + b) = 0, \tag{5.160}$$

which yields a nondivergent baroclinic Rossby wave confined to the upper layer with (low) frequency $\omega = -k\beta/K^2$. The second solution has the frequency

$$\omega = -\frac{k}{K^2}(\beta + b) \tag{5.161}$$

and is a barotropic mode that feels both β and the topography. In the limit of strong topography it reduces to $\omega = -kb/K^2$, and the motion is confined to the lower layer alone. This is a new type of motion, bottom trapped by topography, and does not occur in the flat-bottom case. Rhines calls it a fast baroclinic mode since it appears as an evanescent mode of relatively high frequency in the continuously stratified case, which is presented below.

5.10.2 Uniform Stratification

We noted earlier that effects from upper and lower boundaries are transmitted throughout the respective layers in the two-layer model. The vertical structure of the modes in the real ocean is represented somewhat more realistically in a model with uniform stratification, which represents the opposite extreme in mod-

eling the stratification; the density gradient is smeared uniformly over the depth instead of being squeezed into a layer of infinitesimal thickness.

Dropping the subscript zero we can write the linearized potential vorticity equation (5.53) as

$$\nabla^2 p_t + \frac{1}{S^2} p_{zzt} + \beta p_x = 0, \tag{5.162}$$

where $S = N/f_0$.

At the upper boundary we assume

$$w = 0 \quad \text{or} \quad p_z = 0 \qquad \text{at} \quad z = 0. \tag{5.163}$$

This is a "rigid-lid" condition that makes the barotropic radius of deformation infinite. The lower boundary is taken with a uniform slope α in the y direction only, and we write $w = \alpha v$ there, or in terms of p,

$$p_{zt} = -HbS^2 p_x \qquad \text{at} \quad z = -H, \tag{5.164}$$

where $b = f_0\alpha/H$ is the topographic β-effect.

Substitution of $p \sim e^{i(-\omega t + kx + ly)}$ in (5.162) leads to

$$\omega p_{zz} - S^2(\omega K^2 + k\beta)p = 0, \tag{5.165}$$

where $K^2 = k^2 + l^2$, and (5.164) becomes

$$\omega p_z = HbS^2 kp \qquad \text{at} \quad z = -H. \tag{5.166}$$

Solutions with sinusoidal and with exponential vertical variation are both possible and are discussed below.

(1) $p \sim \cos mz/H$: This is the form for pure Rossby waves ($\beta \neq 0$, $b = 0$) but is not a solution for pure topographic waves ($\beta = 0$, $b \neq 0$). It satisfies (5.163) and then (5.165) and (5.166) yield

$$\omega = -\frac{k\beta}{K^2 + (m^2/H^2S^2)} = -\frac{k\beta\lambda_1^2}{\lambda_1^2 K^2 + m^2} \tag{5.167}$$

and

$$m \tan m = \lambda_1^2 kb/\omega \tag{5.168}$$

where[8]

$$S^2H^2 = -\frac{g\,\partial\bar{\rho}/\partial z\, H^2}{\rho_m f_0^2} = \frac{g\,\Delta\rho H}{\rho_m f_0^2} = \lambda_1^2. \tag{5.169}$$

and $\Delta\rho$ is the density difference from bottom to top.

Eliminating ω from (5.167) and (5.168) leads to

$$-\gamma m \tan m = m^2 + \lambda_1^2 K^2, \tag{5.170}$$

where $\gamma = \beta/b$ measures the relative roles of β and topography as restoring mechanisms when fluid moves to different latitudes or depths.

If the depth is constant ($\gamma = \infty$), the solution to (5.170) is $m = n\pi$ ($n = 0,1,2,\ldots$). This yields pure Rossby waves with the lowest mode ($m = 0$) describing the barotropic wave with infinite deformation radius and $\omega = -k\beta/K^2$. Higher modes correspond to baroclinic Rossby waves, which have low frequencies and are nondispersive for $\lambda_1 K \ll 1$ (weak stratification or wavelengths much larger than λ_1). As in the two-layer case, with $\lambda_1 K \ll 1$, the β-effect can be taken up by vertical divergence ($\beta v \simeq f w_z$) and only a small change in the relative vorticity is required; hence ω is small. In the opposite extreme, with $\lambda_1 K \gg 1$, vertical motions are inhibited and the frequency is approximately the barotropic value $-k\beta/K^2$ for the lower range of m.

For finite positive slopes ($\gamma > 0$), when topography enhances β, the lowest mode solutions to (5.170) are modified forms of the first baroclinic Rossby wave, which in its pure form has $m = \pi$ and a node at mid-depth for the horizontal velocity. When a sloping bottom is present, fluid flowing north or south will have to move vertically against the constraint of stratification. If the slope is small (large γ), the induced vertical motion is small and the solution is a modified baroclinic Rossby wave. If the slope is large, the required vertical motion may be larger than the stratification will allow.

How does the system enable the fluid to negotiate a large slope? It does so by moving the node in the horizontal velocity from mid-depth to the bottom ($m \to \pi/2$). Thus with vanishing v at the bottom no vertical velocity is required. The mode is then simply the upper half of the first baroclinic mode for a constant-depth ocean with twice the depth H and a corresponding higher frequency. For moderate slopes the node is between mid-depth and the bottom, so that the required vertical velocity at the bottom is reduced to a value that can be sustained. The values of m and ω together with sketches of the vertical structure of the horizontal velocity are shown in figure 5.13 for the lowest-order mode with $l = k$ and for a range of values of γ. *For $\gamma > 0$, ω is normalized with respect to the frequency of the first baroclinic mode* (to which it tends as $\gamma \to \infty$). Strong stratification (large λ) or small wavelength (large K) have the same effect as a large slope.

It should be noted that when topography reenforces β there is no vertically oscillatory solution that yields a barotropic Rossby wave in the limit $\gamma \to \infty$. The reason is that the slope has its strongest effect on water near the bottom. These cosine solutions have a maximum amplitude at the surface and can do no better than extend that maximum to the bottom (if $m = 0$), which gives no enhancement. On the other hand, in the baroclinic mode the flow at the bottom is reversed (negative maximum) and the bottom slope can help by making the flow there less negative, i.e., somewhat more barotropic, though as we have seen, only to the point where the node is at the bottom. For $\gamma > 0$ the barotropic limit is described by evanescent modes [see (2) below].

When the depth decreases to the south ($b < 0$, hence $\gamma < 0$), the solution for large $|\gamma|$ is a barotropic Rossby

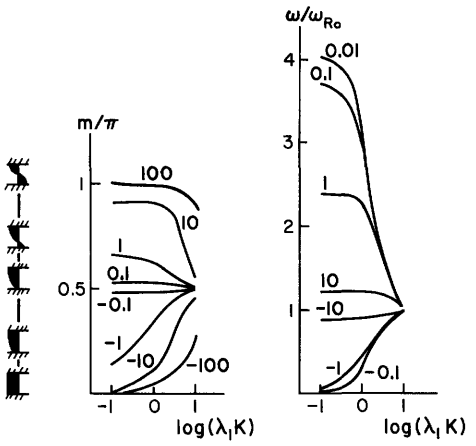

Figure 5.13A Values of the vertical wavenumber m for the gravest quasi-geostrophic mode in a uniformly stratified ocean as functions of horizontal wavenumber. Each curve is marked with value of γ. Topographic effects dominate at small γ; β dominates at large γ. The vertical structure of the modes is sketched at the left for the different values of m.

Figure 5.13B Ratio of frequency to pure barotropic Rossby wave frequency for $\gamma < 0$ and to baroclinic Rossby wave frequency for $\gamma > 0$.

wave modified by topography. In this case the sloping bottom works against β and tends to reduce the frequency. The effect is achieved with a partial cosine wave of one sign throughout the depth. The results for m and ω are shown in figure 5.13. *For* $\gamma < 0$ figure 5.13B shows ω *normalized with respect to the barotropic Rossby wave frequency* (to which it tends as $\gamma \to -\infty$). A larger slope gives more vertical structure until the node reaches the bottom for small $|\gamma|$ and the form is again a half-baroclinic mode.

It is interesting to note that the frequency is always negative, i.e., the phase velocity is westward even when the bottom slope is large and works against β. These oscillatory modes are not possible without β so the most that the bottom slope can do is to change the magnitude but not the sign of the frequency. When it helps β, it makes the response less baroclinic. When it opposes β, it makes the response less barotropic.

(2) $p \sim \cosh \mu z/H$: This is the form for pure topographic waves ($\beta = 0$, $b \neq 0$) but is not a solution for pure Rossby waves ($\beta \neq 0$, $b = 0$). It satisfies (5.163), and then (5.165) and (5.166) yield

$$\omega = -\frac{k\beta}{K^2 - (\mu^2/S^2H^2)} = -\frac{k\beta\lambda_1^2}{\lambda_1^2K^2 - \mu^2} \tag{5.171}$$

and

$$-\mu \tanh \mu = \frac{\lambda_1^2 bk}{\omega}. \tag{5.172}$$

Eliminating ω gives

$$\gamma\mu \tanh \mu = \lambda_1^2K^2 - \mu^2. \tag{5.173}$$

For pure topographic waves ($\gamma = 0$) we have $\mu = \lambda_1 K$ and (5.172) becomes

$$\omega = -\frac{\lambda_1^2 bk}{\lambda_1 K \tanh \lambda_1 K}. \tag{5.174}$$

Strong stratification or short wavelength gives $\tanh \lambda_1 K \to 1$ and

$$\omega = -\frac{\lambda_1 bk}{K} = -\frac{\alpha N k}{K}. \tag{5.175}$$

The frequency has a maximum value for waves traveling parallel to bottom contours since the associated particle velocity is transverse, i.e., up or down the slope, and has maximum vertical displacement; hence it is subjected to maximum restoring force. The wave amplitude decays exponentially upward from the bottom.

For weak stratification or long waves ($\lambda_1 K \ll 1$), $\tanh \lambda_1 K \simeq \lambda_1 K$ and

$$\omega = -\frac{bk}{K^2}. \tag{5.176}$$

These are topographic, barotropic Rossby waves with b replacing β.

All of these evanescent waves, with or without β, have phase velocity always to the left of upslope, since they are basically topographic waves and can be only modified by β. When β and b work together, the phase velocity is westward. When they are opposed, the phase velocity is eastward.

With $\gamma > 0$ we have already seen that the oscillatory solution has no barotropic mode. That function is taken over by this evanescent form, which reenforces β and tends always to increase the frequency above the value for barotropic Rossby waves. For $\gamma \gg 1$ the change is small, but as the slope increases, the frequency does too. As $\lambda_1 K$ increases, the wave is confined closer to the bottom (figure 5.14A) and the frequency rises even more (figure 5.14B). This is Rhines's fast baroclinic mode referred to earlier. The frequencies in figures 5.14B and 5.14C are normalized with respect to the barotropic Rossby wave frequency.

When the slope is negative and small ($-\gamma \gg 1$), the mode is confined to a shallow layer near the bottom where the effect of the slope dominates (figure 5.14A). The frequency is small and tends to zero as $-\gamma$ increases. When the slope is large ($-\gamma \ll 1$), β is a perturbation and the frequency ratio exceeds unity (figure 5.14C). Strong stratification or short wavelength ($\lambda_1 K \gg 1$) also serves to confine the mode to the bottom and increase the frequency.

Rhines (1977) summarized the observational evidence of the existence of these bottom-trapped waves in current-meter records taken at site D (39°10'N, 70°W) by Luyten, Schmitz, and Thompson of the Woods

Hole Oceanographic Institution. Rhines (1971a) and Thompson and Luyten (1976) used spectral analysis to show that the kinetic energy increases toward the bottom and that the horizontal velocities are negatively correlated, as they should be for these waves. The evidence is particularly striking in the high-pass filtered records by Luyten that Rhines showed.

McWilliams and Flierl (1976) have described a number of mesoscale features observed during the MODE-0 and MODE-1 programs in terms of the linear, quasi-geostrophic waves that can exist given the mean properties in the observational regions. Although nonlinear effects seem to have been present, modified forms of the linear wave features can be identified (see chapter 11).

5.11 Baroclinic Instability

The wave solutions in the foregoing two sections are valid when the phase velocity is large compared to the ambient fluid velocity. The condition is not really satisfied either by baroclinic waves or by short barotropic waves. Furthermore, the steady wind-driven and thermohaline velocities derived earlier are equilibrium solutions but they are not necessarily stable. Here, we shall extend our study by considering quasi-geostrophic perturbations on a mean flow and exploring the question of stability. (Also see the discussion in chapter 18.)

In pursuing this approach one should allow for both horizontal and vertical variations in the basic velocity field. Although both types of variations can lead to instabilities, it is known from earlier studies, particularly for atmospheric motions (Kuo, 1949), that the observed fluctuating motions generally have a structure characteristic of barotropically stable modes. In other words, fluctuations do not seem to draw their energy from the horizontal variation in the basic velocity field. For that reason we shall explore a mean velocity that has a variation only in the vertical direction and refer the reader to the published literature for studies involving horizontal variations (Kuo, 1949; Lipps, 1963).

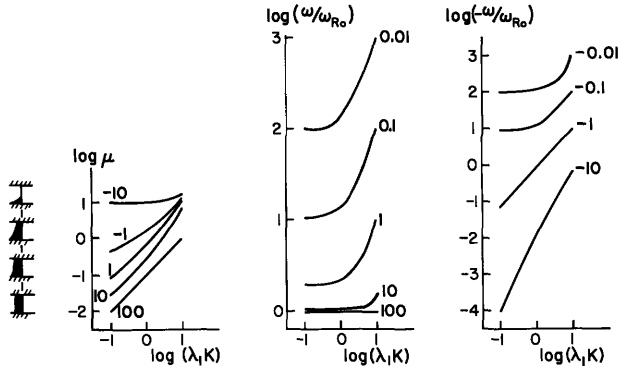

Figure 5.14A Vertical scale μ for the evanescent mode $\cosh \mu z/H$ as a function of horizontal wavenumber. Each curve is marked with value of γ. The vertical structure of the modes is sketched at the left for the different values of μ.

Figure 5.14B Ratio of frequency with $\gamma > 0$ to pure barotropic Rossby wave frequency. This mode tends to the limit of pure barotropic Rossby wave for vanishing topography. Decreasing γ corresponds to increasing bottom slope and serves to increase frequency.

Figure 5.14C Ratio of $-\omega$ to frequency of pure barotropic Rossby wave frequency. Frequency decreases as bottom slope decreases because mode is not possible without topography.

5.11.1 Linear Theory

The problem that we consider is baroclinic instability of the basic velocity field. The topic has an extensive literature in meteorology, beginning with the work of Charney (1947) and of Eady (1949) and extended, *inter alios*, by Kuo (1952), Phillips (1951, 1954), Green (1960), Charney and Stern (1962), Pedlosky (1964a), and Bretherton (1966a,b). As is the case with the development of a topic, the earliest papers explore the basic concepts and the later ones help to illuminate the issues raised. I have found Bretherton's two articles to be especially enlightening, though each of the ones mentioned, and some others too, discuss important aspects of the problem. The mathematical development given below follows Bretherton (1966a).

Since our intent is to discuss the dynamical balances, we do so with the simplest model that contains the essential elements. Additional effects are mentioned later. So consider the very special case of a uniformly rotating channel (f = constant) with two layers of equal thickness, $H_1 = H_2$, and a total depth $H = H_1 + H_2 = 2H_1$. The upper layer has a uniform basic current U, and the lower layer an equal but opposite current $-U$. Lateral walls at $y = 0, M$ are sufficiently close so that the depth of each layer in the coefficients can be considered constant. Geostrophic balance of the basic state yields

$$\begin{aligned} fU_1 &= fU = -gH_y, \\ fU_2 &= -fU = -g[(1-\epsilon)H_y + \epsilon H_{2y}], \end{aligned} \tag{5.177}$$

and this basic state is perturbed by infinitesimal quasi-geostrophic motions.

We first linearize the potential vorticity $(\zeta + f)/h$ of each layer by writing

$$\frac{\zeta+f}{h} = \frac{\zeta+f}{H(1+\eta/H)} \simeq \frac{\zeta+f}{H}\left(1-\frac{\eta}{H}\right) = \frac{f}{H} + \frac{\zeta}{H} - \frac{f\eta}{H^2},$$

where f/H is the mean and $\zeta/H - (f\eta/H^2)$ the perturbation potential vorticity, and η the fluctuating disturbance of the layer depth.

The linearized equation for conservation of potential vorticity becomes

$$(\partial_t + \bar{u}\,\partial_x)\left(\zeta - \frac{f\eta}{H}\right) - \frac{vf}{H} H_y = 0, \tag{5.178}$$

where we have multiplied through by H which is taken as a constant in the coefficients and where $\bar{u}$ is the mean velocity of the layer.

We use the following definitions and relations for the two layers:

Upper	Lower
H_1	$H_2 (= H_1)$
$\eta = \eta_1 - \eta_2$	$\eta = \eta_2$
$\bar{u}_1 = U$	$\bar{u}_2 = -U$
$Q_{1y} \equiv -\frac{f}{H_1} H_{1y} = \frac{2U}{\lambda_1^2}$	$Q_{2y} \equiv -\frac{f}{H_2} H_{2y} = -\frac{(2-\epsilon)}{\lambda_1^2} U \simeq -\frac{2U}{\lambda_1^2}$
$q_1 \equiv \zeta_1 - f(\eta_1 - \eta_2)/H_1$	$q_2 \equiv \zeta_2 - f\eta_2/H_2$

(5.179)

Here,

$$\lambda_1^2 = \frac{g\epsilon H_1}{f^2}$$

is the reduced gravity radius of deformation based on H_1. For convenience we have defined Q and q as H_1 times the potential vorticities. Using (5.179) in (5.178) we obtain

$$\begin{aligned} &(\partial_t + U\,\partial_x)q_1 + v_1 Q_{1y} = 0, \\ &(\partial_t - U\,\partial_x)q_2 - v_2 Q_{2y} = 0, \end{aligned} \tag{5.180}$$

where we have neglected the $O(\epsilon)$ term in Q_2 to write $Q_{2y} = -Q_{1y}$.

The q's can be related to the η's, or more conveniently to "pressure" potentials ϕ, though the geostrophic relations

$$v_1 = \phi_{1x}, \qquad u_1 = -\phi_{1y}, \qquad v_2 = \phi_{2x}, \qquad u_2 = -\phi_{2y}, \tag{5.181}$$

where

$$\phi_1 \equiv g\eta_1/f, \qquad \phi_2 \equiv g[(1-\epsilon)\eta_1 + \epsilon\eta_2]/f. \tag{5.182}$$

Then

$$\begin{aligned} &\zeta_1 = v_{1x} - u_{1y} = \nabla^2\phi_1, \qquad \zeta_2 = v_{2x} - u_{2y} = \nabla^2\phi_2, \\ &\frac{f}{H_1}(\eta_1 - \eta_2) = \frac{1}{\lambda_1^2}(\phi_1 - \phi_2), \\ &\frac{f\eta_2}{H_2} = \frac{1}{\lambda_1^2}(\phi_2 - \phi_1 - \epsilon\phi_1) \simeq \frac{1}{\lambda_1^2}(\phi_2 - \phi_1), \\ &q_1 = \nabla^2\phi_1 - \frac{1}{\lambda_1^2}(\phi_1 - \phi_2), \qquad q_2 = \nabla^2\phi_2 + \frac{1}{\lambda_1^2}(\phi_1 - \phi_2), \end{aligned} \tag{5.183}$$

where we have neglected the $O(\epsilon)$ term in the expression for $f\eta_2/H_2$. The neglect of the $O(\epsilon)$ terms is equivalent to assuming a rigid lid at the top since it corresponds to neglecting the barotropic radius of deformation gH/f^2.

Equations (5.180) take the form

$$\begin{aligned} &(\partial_t + U\,\partial_x)q_1 + \phi_{1x}Q_{1y} = 0, \\ &(\partial_t - U\,\partial_x)q_2 - \phi_{2x}Q_{1y} = 0, \end{aligned} \tag{5.184}$$

with ϕ and q given by (5.182) and (5.183). We now add and subtract the two equations in (5.184) to obtain the symmetric set

$$(q_1 + q_2)_t + U(q_1 - q_2)_x + Q_{1y}(\phi_1 - \phi_2)_x = 0, \tag{5.185a}$$

$$(q_1 - q_2)_t + U(q_1 + q_2)_x + Q_{1y}(\phi_1 + \phi_2)_x = 0. \tag{5.185b}$$

From (5.183)

$$\begin{aligned} &q_1 + q_2 = \nabla^2(\phi_1 + \phi_2), \\ &q_1 - q_2 = \nabla^2(\phi_1 - \phi_2) - \frac{2}{\lambda_1^2}(\phi_1 - \phi_2), \end{aligned} \tag{5.186}$$

and we have a system in which only sums and differences of the variables appear. The sum is related to the vertical mean, the difference to the baroclinic contribution. Thus, the potential vorticity for the entire depth is just the sum of relative vorticities because the height adjustments are equal and opposite. The differential potential vorticity involves twice the effect of the interface. The symmetry of the set is made possible by the equal mean depths, and equal and opposite mean velocities and mean potential vorticities.

Since all coefficients are constant, the system has solutions of the form

$$e^{-i\omega t + ikx}\sin ly, \qquad l = \frac{m\pi}{M}, \qquad m = 1,2,..., \tag{5.187}$$

which satisfies the lateral boundary condition of zero normal flow at $y = 0, M$. Then (5.186) can be used to express the amplitudes of $q_1 \pm q_2$ in terms of $\phi_1 \pm \phi_2$ as

$$\begin{aligned} &q_1 + q_2 = -K^2(\phi_1 + \phi_2), \qquad K^2 = k^2 + l^2, \\ &q_1 - q_2 = -(K^2 + 2/\lambda_1^2)(\phi_1 - \phi_2). \end{aligned} \tag{5.188}$$

For the moment we keep the x and t derivatives on the ϕ's but use (5.188) to write

$$K^2(\phi_1 + \phi_2)_t + [U(K^2 + 2/\lambda_1^2) - Q_{1y}](\phi_1 - \phi_2)_x = 0, \tag{5.189a}$$

$$(K^2 + 2/\lambda_1^2)(\phi_1 - \phi_2)_t + (UK^2 - Q_{1y})(\phi_1 + \phi_2)_x = 0. \tag{5.189b}$$

Then substituting for Q_{1y} yields

$$K^2(\phi_1 + \phi_2)_t + K^2U(\phi_1 - \phi_2)_x = 0, \tag{5.190a}$$

$$(K^2 + 2/\lambda_1^2)(\phi_1 - \phi_2)_t + UK^2(\phi_1 + \phi_2)_x$$

$$-\frac{2U}{\lambda_1^2}(\phi_1 + \phi_2)_x = 0. \tag{5.190b}$$

Use of (5.187) in (5.189) and in (5.190) leads to the frequency equation

$$\omega^2 = \frac{k^2}{K^2(K^2 + 2/\lambda_1^2)}[U(K^2 + 2/\lambda_1^2) - Q_{1y}][UK^2 - Q_{1y}]$$

$$= kU^2\frac{K^2 - (2/\lambda_1^2)}{K^2 + (2/\lambda_1^2)}. \tag{5.191}$$

When ω^2 is negative, the perturbations grow exponentially and instability occurs. In terms of potential vorticity, instability requires

$$U(K^2 + 2/\lambda_1^2) > Q_{1y} > UK^2. \tag{5.192}$$

The right-hand side in (5.191) shows that long waves, with $K^2 < 2/\lambda_1^2$, are unstable. For shorter wavelengths the system is oscillatory with no growth. For $l = k$ the maximum growth rate is $U(3 - 2\sqrt{2})^{1/2}/\lambda_1$ and occurs for $\lambda_1^2K^2 = 2(\sqrt{2} - 1)$.

We can get some insight into the instability mechanism by noting the following. The local change of the average potential vorticity $q_1 + q_2$ is balanced by the *sum* of the mean advections of q and by the *sum* of the perturbation advections of Q in the two layers. Because both the mean advection velocity and mean potential vorticity have opposite signs in the two layers, these two balancing quantities involve the *differences rather than the sums* of the perturbation quantities, as we can see in (5.185a). But we would not expect internal adjustments of the interface to affect the average potential vorticity since the effect in one layer is cancelled by the effect in the other. This expectation is borne out when we evaluate the different terms to get (5.190a), where only the mean advection of the relative vorticity is left to balance $(\phi_1 + \phi_2)_t$. Thus, if at some instant of time we had $q_1 - q_2 \sim \sin kx$ and $q_1 + q_2 \sim 0$, the quantity $q_1 + q_2$ would be $\sim -\cos kx$ at the next instant.

Now consider the local change in the *differential* potential vorticity $q_1 - q_2$. Again, because of the equal but opposite U and Q_y in the two layers, $(q_1 - q_2)_t$ is balanced by terms involving the *sums* of the perturbation quantities, as we see from (5.185b). The mean convective part $U(q_1 + q_2)_x$ gives rise to a change in $q_1 - q_2$ like $-\sin kx$ after the initial instant and acts as a restoring force, tending to cancel the initial $q_1 - q_2$ ($\sim \sin kx$). The perturbation advection of Q has the opposite sign as we see from (5.190b) and gives rise to a change in $q_1 - q_2$ like $\sin kx$, i.e., it reenforces the initial distribution. The latter, destabilizing, effect will dominate for large wavelengths, i.e., for $K^2 < 1/\lambda^2$. We expect this to be the case because *internal* adjustments of the interface should cause significant changes in the *differential* potential vorticity.

A second way of looking at the problem is to observe that the perturbation advection of mean potential vorticity vQ_y has the form of mean advection of perturbation potential vorticity Uq_x when we substitute the expressions for v and Q. Because vQ_y involves advection of the layer thickness, only the part due to the interface adjustment appears in Uq_x. When the terms are combined as mean advections in (5.190), we see that the first involves advection in the positive direction by U but the second can have either sign depending on the sign of $K^2 - (2/\lambda_1^2)$. For small K the phase is appropriate for reenforcement and instability occurs.

We can understand the energetics of the instability by writing the disturbances in the more convenient form

$$\phi_1 + \phi_2 = Ae^{\sigma t}\sin kx \sin ly,$$
$$\phi_1 - \phi_2 = Be^{\sigma t}\cos kx \sin ly, \tag{5.193}$$

where the solution of the stability problem gives the growth rate σ as

$$\sigma = \pm kU\frac{(2/\lambda_1^2 - K^2)^{1/2}}{(2/\lambda_1^2 + K^2)^{1/2}}. \tag{5.194}$$

Then (5.190a) yields

$$\sigma A = UkB$$

or

$$B = \pm\left(\frac{2/\lambda_1^2 - K^2}{2/\lambda_1^2 + K^2}\right)^{1/2}A. \tag{5.195}$$

For very long waves ($K^2 \ll 2/\lambda_1^2$) we have $A \simeq B$ for the growing mode and $A \simeq -B$ for the decaying mode. For $k = l$ the corresponding results are $B = \pm(\sqrt{2}-1)^{1/2}A$.

Now consider the case with $U > 0$ so that the upper layer is thicker toward $y = 0$ (south). The quantity η_2 is a measure of the perturbation thickness of the lower layer so $-\eta_2$ is a measure of the excess heat. Then $-v_1\eta_2$ and $-v_2\eta_2$ correspond to northward transports of heat in the respective layers. Using an overbar for an average over x and y, we have $-\overline{(v_1 + v_2)\eta_2}$ as the northward heat transport in the two layers. Eqs. (5.181) and (5.183) can be used to write

$$\overline{-(v_1 + v_2)\eta_2} = \frac{H_2}{f\lambda_1^2}\overline{(\phi_1 + \phi_2)_x(\phi_1 - \phi_2)}$$

$$= \frac{kH_2}{f\lambda_1^2}AB, \tag{5.196}$$

so we have northward heat transport for the unstable mode $A \simeq B$ and southward heat transport for the stable mode $A \simeq -B$.

The vertical velocity at the interface is $w = (\partial_t + U\partial_x)\eta_2 + v_1H_{2y}$ and the upward heat transport is given

by $-\overline{w\eta_2}$. When the foregoing relations are used to evaluate the latter quantity, we find that the unstable mode transports heat upward and the stable mode transports heat downward. For the unstable mode the northward and upward heat fluxes combine to exchange a particle of fluid initially at the south just above the interface with a particle at the north that was initially just below the interface. Thus, the net effect is to level the interface and the instability grows by drawing on the potential energy of the sloping interface. It can also be shown that the stable mode tends to increase the slope of the interface, and there is no energy available for such a change. Furthermore, the oscillatory, neutral modes that occur for $K^2 > 2/\lambda_1^2$ have zero northward heat flux.

From (5.193) the values of ϕ_1 and ϕ_2 are

$$\begin{aligned}\phi_2 &= \tfrac{1}{2}(A\sin kx + B\cos kx)e^{\sigma t}\sin ly,\\ \phi_1 &= \tfrac{1}{2}(A\sin kx - B\cos kx)e^{\sigma t}\sin ly,\end{aligned} \tag{5.197}$$

Hence, with $K^2 \ll 2/\lambda_1^2$ the unstable mode $A \simeq B$ yields

$$\begin{aligned}\phi_1 &= \frac{Ae^{\sigma t}}{\sqrt{2}}\sin\left(kx + \frac{\pi}{4}\right)\sin ly,\\ \phi_2 &= \frac{Be^{\sigma t}}{\sqrt{2}}\sin\left(kx - \frac{\pi}{4}\right)\sin ly\end{aligned} \tag{5.198}$$

and we note that the upper-layer disturbance is displaced to the west of (lags) the lower-layer disturbance by $\pi/2$. For $k = l$ the phase lag of the upper layer is about 66°. Observed baroclinic instabilities in the atmosphere are characterized by such a phase shift. The phases are reversed for the stable mode.

At this point it is a simple matter to take into account the effect of variable f by noting

$$\frac{\partial}{\partial y}\left(\frac{f}{H}\right) = \frac{\beta}{H} - \frac{fH_y}{H^2}, \tag{5.199}$$

so that the term βv must be added to each layer in (5.180). The result is to add $\beta(\phi_1 + \phi_2)_x$ to (5.185a) and $\beta(\phi_1 - \phi_2)_x$ to (5.185b) and the system (5.190) becomes

$$\begin{aligned}&(\phi_1 + \phi_2)_t + \frac{\beta}{K^2}(\phi_1 + \phi_2)_x + U(\phi_1 - \phi_2)_x = 0,\\ &\left(K^2 + \frac{2}{\lambda_1^2}\right)(\phi_1 - \phi_2)_t - \beta(\phi_1 - \phi_2)_x\\ &+ U\left(K^2 - \frac{2}{\lambda_1^2}\right)(\phi_1 + \phi_2)_x = 0.\end{aligned} \tag{5.200}$$

The frequency equation is

$$\begin{aligned}&\omega^2 + \frac{2k\beta(K^2 + 1/\lambda_1^2)}{K^2(K^2 + 2/\lambda_1^2)}\omega + \frac{k^2\beta^2}{K^2(K^2 + 2/\lambda_1^2)}\\ &- \frac{k^2U^2}{K^2 + 2/\lambda_1^2}(K^2 - 2/\lambda_1^2) = 0\end{aligned} \tag{5.201}$$

and the condition for instability is changed to

$$\lambda_1^8K^8 - 4\lambda_1^4K^4 + \frac{\beta^2\lambda_1^4}{U^2} < 0 \tag{5.202}$$

or

$$2 - 2\sqrt{1 - \frac{\beta^2\lambda_1^4}{4U^2}} < \lambda_1^4K^4 < 2 + 2\sqrt{1 - \frac{\beta^2\lambda_1^4}{4U^2}}. \tag{5.203}$$

Thus, the wavelength is bounded at the long end as well and the unstable range is cut down. A necessary condition on U is

$$2|U| > \beta\lambda_1^2. \tag{5.204}$$

With $\lambda_1 = 36$ km this gives a velocity difference between the layers of 2.5 cm s^{-1}.

These results were first derived by Phillips (1951). The condition on $|U|$ is a particular example of the more general necessary condition for instability (Charney and Stern, 1962; Pedlosky, 1964a) that the mean potential vorticity gradient (including β) must change sign in the region. If β dominates, Q_y is positive everywhere and the flow is stable.

In two-layer models we note particularly that eastward flow in the upper layer requires $-H_{1y} > 0$ so that the stabilizing β-effect is reenforced in (5.199). Hence, instability can occur only if $-fH_{2y}/H_2$ (which is negative) is large enough to offset β. If the lower layer is very deep, the system is stable. On the other hand, westward flow in the upper layer is destabilizing ($-H_{1y} < 0$) and a shallow upper layer is more conducive to instability since fH_{1y}/H_1 increases with decreasing H_1. Two-layer oceanic models normally have a shallow upper layer so instability sets in first in regions where the surface velocity is to the west. Since geostrophic westward flow is associated with a thermocline that deepens to the north, the instability serves to transport heat to the south when the thermocline is flattened. Thus, the primary instability tends to resist the effect of the thermal driving.

A bottom slope also affects the stability. We can make use of the general condition in the present case to observe that if the bottom slope is in the same direction as, and exceeds, the slope of the interface it stabilizes the flow because Q_{2y} has the same sign as Q_{1y}. More generally, when β is included, a stable flow can be destabilized (and vice versa) by changing the sign of U and/or the slope.

These results can be generalized to flows with arbitrary, stable stratification (Charney and Stern, 1962; Pedlosky, 1964a). Necessary conditions for instability are that (1) Q_y change sign somewhere between the surface and the bottom, (2) $Q_yU_z < 0$ at $z = 0$, or (3) $Q_y(U_z + N^2H_y/f) > 0$ at $z = -H$.

Bretherton (1966a) compared the two-layer model with the continuously stratified (constant-N) model of

Eady (1949) to explain the similarity in the results of two systems that have obvious differences. In the two-layer case the vertical difference in Q_y is necessary for instability whereas in the Eady model Q is constant. The latter system becomes unstable because the sloping isopycnals intersect the bottom and introduce destabilizing boundary effects. When the boundary conditions are incorporated into the potential vorticity of the fluid as infinitesimally thin sheets adjacent to the boundaries, their effects can be interpreted as internal differences in Q_y and the similarity between the two models is evident. The two models are opposite extremes in the simulation of the variable stratification of the ocean. In the two-layer model the stratification is squeezed into a discontinuous layer and boundary effects are distributed uniformly throughout each layer. The constant-N model smears out vertical differences in the stratification, thereby exaggerating the stratification at the boundaries, but boundary effects appear properly as boundary conditions.

The question of critical layer instability, where the phase speed of the disturbance is equal but opposite to the convective velocity, is addressed by Bretherton (1966b), who showed that the existence of a critical layer in a fluid makes it highly unlikely that the flow will be stable. If Q_y vanishes or if certain other conditions are satisfied at or near the critical layer, stability is possible, but generally instability will occur. Bretherton used his analysis to show that the addition of β to Eady's problem (Green, 1960) yields $Q_y = \beta/H$, i.e., Q_y does not vanish, and since a critical layer exists, the flow is unstable. This destabilizing effect of β is in direct contrast to the result that we found for the two-layer case. For a two-layer system, however, there is no critical layer (it is buried in the discontinuity of the interface), and the corresponding mode is stable. The instability that does occur with two layers is due to the change of sign of Q_y between the two layers. We have already pointed out that the equivalent effect is due to boundary effects in the Eady problem without β.

Stommel (1965) suggested that the enormous amount of potential energy present in the stratification of the ocean and associated with the mean circulation may be a possible source of energy for instabilities. The obvious mechanism to tap that energy is baroclinic instability. Gill, Green, and Simmons (1974) have explored the stability of several combinations of vertical profiles of density and of vertical shear of horizontal velocity. Using exponential approximations to mean observed vertical profiles, they concluded that westward flows[9] (isopycnals sloping up toward the equator) are the most likely unstable ones, and they analyzed one such case with a velocity maximum at the surface and two with the maximum at 100 m depth. For one of the latter the slope of isopycnals reverses (up toward the pole) at 100 m depth. The most unstable modes have wavenumbers close to the reciprocal of the internal radius of deformation (wavelength 190 km) and e-folding time of 80 days with velocities significant in the upper kilometer or so. When U has a subsurface maximum where the isopycnals reverse slope, the growth rate decreases, suggesting that seasonal changes that can affect the isopycnal slope can also affect the stability. The maximum growth rate occurs for profiles with monotonic U_z. Secondary instabilities with minimum amplitude at 1000 m depth, smaller wavelength, and smaller growth rate also occur. Since the amplitudes of these secondary instabilities are large at depth, they are strongly affected by topography. The conversion of available potential energy to eddy energy for all of the primary instabilities is confined to the upper half-kilometer.

Because linear instability theory yields no information about absolute amplitudes, Gill, Green, and Simmons (1974) assumed that the eddies draw energy from the mean field at the rate at which it is supplied to the mean field by the wind. That sets the amplitudes for the disturbances. By assuming further that the primary and secondary instabilities draw equal amounts of available potential energy, they obtain maximum eddy velocities of 0.14 $m\,s^{-1}$ and wavelengths of 200 km in the upper kilometer and corresponding values of 0.05 $m\,s^{-1}$ and 300–500 km in deep water. The latter values are close to the observed but the amplitudes of the upper velocities are on the small side.

The analysis by Gill, Green, and Simmons (1974) supplies qualitative, and even some quantitative, support for the pertinence of linear baroclinic-instability theory to observed eddy motions in the ocean. Linear growth rates have the right magnitude and the scales and distributions of the disturbances are also approximately correct. However, more detailed features and quantitative results require additional considerations.[10] Nonlinear processes must be important, particularly in altering the structure of the mean field, which is assumed known in the linear theory. Friction, too, must have at least a quantitative effect over the lifetime of these flows.

5.11.2 Finite-Amplitude Effects

Phillips (1954) calculated the finite-amplitude effects of the baroclinically unstable modes to determine the lowest-order corrections to the assumed mean properties. He derived the heat fluxes given above, the induced meridional circulation, and the changes in the zonal momentum wrought by nonlinear corrections. The altered fields are in qualitative agreement with observed structures in the atmosphere.

This use of the most rapidly growing eigenfunctions of linear theory to determine nonlinear corrections to

the system has been applied and extended by a number of other investigators. In part, the purpose is to enable one to take into account the effects of these smaller-scale processes in larger-scale models by a consistent parameterization. Nongeostrophic processes can also be included to give an idea of the effects on the evolution of the flow.

In an excellent review article on finite-amplitude baroclinic instability Hart (1979a) gives an account of the different approaches to the treatment of the nonlinear system and the reader is referred to Hart's paper for a more detailed discussion and an extensive bibliography. The different treatments summarized by Hart include simple physical effects of the first nonlinear interaction (e.g., Phillips, 1954), weakly nonlinear interactions between the perturbation and the mean field (e.g., Stuart, 1960), a truncated set of normal modes with amplitudes that must be determined by the dynamical system (e.g., Lorenz, 1963a), and a mean-field, single-wave approach (e.g., Herring, 1963). Pedlosky (1975) gives a discussion appropriate to oceanic eddies.

These studies of finite-amplitude effects are important in providing an understanding of the interplay of different parts of the system and especially for determining which processes are pertinent to the behavior of the observed system. For example, observed asymmetries in the structure of the evolved eddies can be traced to nongeostrophic effects [see Hart (1979a) for references], and inclusion of those effects may be necessary to recover the asymmetries, even in strongly nonlinear models. But other features, such as the occlusion observed in fully developed cyclones or eddies and the greater intensity of eddy amplitudes as compared to mean-flow velocities, do not emerge from these local, finite-amplitude treatments, and recourse to more nonlinear (numerical) models is suggested if the results are to be applicable to oceanic flows.

One such treatment, a numerical study of baroclinic instability by Orlanski and Cox (1973), using Bryan's (1969) numerical model of the general circulation, focused on the stability of an intense, confined, stream along a coast. The application is to the Florida Current (Miami to Cape Hatteras). The effects of a bounding coast, bottom topography, nonlinearity, friction, and diffusion are all included in this study, which shows that an evolved eddy field is established 10 days after the current, is initially uniform in the downstream direction, and is randomly perturbed in density and velocity. The disturbance of maximum growth has a scale of about 20 km, which seems to be the deformation radius (not reported but estimated from the published temperature pattern). The growth rate decreases by about a factor of 10 from the initial value when nonlinear effects become important. This seems to be a characteristic effect of nonlinearity; similar results are reported for meteorological studies (Gall, 1976). The initially rapid growth of the disturbances is much faster than one obtains for a current in an unbounded ocean, so there is a strong suggestion that bottom topography and the nearby coast are important factors in determining the growth rate. An energy diagram shows that the flux of energy is from mean to eddy potential energy, then to eddy and finally mean kinetic energy. These transfers are all consistent with baroclinic instability and its evolution.

5.12 Effect of Nonlinearity and Turbulence

Linear theory provides at least a qualitative explanation of some observed dynamic features in the ocean, including westward intensification of boundary currents, the several types of long-period wave motions, and the existence of baroclinically unstable modes. It is also clear, however, that nonlinear processes provide more than merely quantitative corrections. For example, the observed Gulf Stream transport is several times the value predicted by linear theory, and the recirculating gyre associated with the increased transport (see chapter 4) not only involves flows that are baroclinically unstable, but the unstable modes seem to be fully developed as well. The effects of smaller-scale processes cannot generally be parameterized in terms of eddy coefficients, so an analysis of the evolution of the flow requires analysis of strong, interacting features and the use of concepts and relationships developed for turbulent flows.

Because of the difficulties inherent in nonlinear studies, many of the explorations have used numerical models, some of which have required extensive calculations. Rhines (1977) has made a significant effort to construct a comprehensive account of the emerging dynamic picture. Only an outline of that remarkable synthesis can be given here.

Most of the numerical experiments discussed in this section are oriented toward a study of processes. Except for the calculations by Bryan (1969) they were not intended to be simulations of ocean circulation. Therefore, the models are often very idealized (two-layer, rectangular basins with cyclic boundary conditions, etc.). At times the results can only be suggestive of what happens in nature, but they should be of great help in indicating what should be included in more realistic, predictive models.

The linear theories discussed in the earlier sections must be supplemented by a few additional observations. One is that baroclinically unstable motions evolve into fully developed, closed eddies. A description of the latter is not accessible with theories based on extensions from linear treatments and one must resort to numerical experiments to obtain the evolved fields. A second observation is that mature eddies are

often (but not always) in an occluded state in which the motion has penetrated vertically to form a relatively barotropic gyre. In horizontal wavenumber space the window for vertical penetration is at or near the internal radius of deformation. This tendency toward barotropy in a Boussinesq fluid means that certain features of the system can be described as if the flow were two-dimensional, a significant simplification for both analytic and numerical studies.

5.12.1 Nonlinear Effects with Constant Depth

Batchelor (1953a) points out that for nonlinear, two-dimensional flow in an inviscid fluid the time rates of change of integrated energy and enstrophy (squared relative vorticity) vanish:

$$\frac{\partial}{\partial t}\int_0^\infty E\,dk = 0, \qquad \frac{\partial}{\partial t}\int_0^\infty k^2 E\,dk = 0, \tag{5.205}$$

where E is the kinetic energy in wavenumber space. (For a fluid with small viscosity the same relations hold approximately for times short compared to the transfer time to high wavenumbers where viscous dissipation is significant.) If one assumes that an initially narrow spectrum spreads in time about its mean wavenumber,

$$\frac{\partial}{\partial t}\int_0^\infty (k - k_1)^2 E\,dk > 0, \tag{5.206}$$

where $k_1 = \int kE\,dk / \int E\,dk$, then use of (5.205) yields

$$\frac{\partial k_1}{\partial t} < 0. \tag{5.207}$$

Thus, the mean wave number k_1 decreases. Furthermore, since $\int k^2 E\,dk$ is conserved, a transfer of energy from, say, k_0 to $2k_0$ must be balanced by a transfer of four times that energy to $k_0/2$ or an equivalent change. In order to achieve this red cascade of energy, fluid elements with vorticity of like sign must coalesce into larger eddies. Hence, a small number of strong, isolated vortices will emerge and they will continue to tend to coalesce. Nonlinearity will generate small scales as well, and since enstrophy involves a weighting by k^2, enstrophy will peak at smaller scales (see chapter 18).

The red cascade relies on nonlinear processes to provide the transfer to different scales. Furthermore, the turbulent field must be spatially homogeneous and isotropic. A local concentration of turbulent eddies will spread spatially as well as in wavenumber and the conditions for the red cascade can be violated. For example, an initially isolated cluster of eddies will spread to the point where nonlinear transfers are so weak that the system no longer acts as turbulence and the evolution to larger scales stops.

The relations (5.205) apply to an unbounded fluid on the β-plane as well. Yet in the vorticity balance the planetary vorticity term βv becomes comparable to the nonlinear term $\mathbf{v}\cdot\nabla\zeta$ when the Rossby number is small enough. If U is the rms turbulent velocity, and if motions cluster around wave number k_1, the pertinent measure of the Rossby number is $2Uk_1^2/\beta$. As the red cascade proceeds, k_1 becomes smaller and $2Uk_1^2/\beta$ drops toward unity. Energy and vorticity can then be radiated away by Rossby waves, decreasing the intensity of the motion still further. Thus, the red cascade is halted as turbulence gives rise to waves. It is interesting to note that the larger scales generated by the turbulence are the ones that radiate away most quickly. Hence, the end state is toward a pattern of zonal flows with scale $k_\beta^{-1} \sim (\beta/2U)^{-1/2}$ (Rossby number ~ 1).

Rhines (1975) ran a numerical calculation for a basin with periodic boundary conditions in which an initially turbulent, barotropic fluid with energy in a narrow band of wavenumbers evolves toward the red cascade with and without β. His results are illustrated in figure 5.15, where contours of ψ as a function of longitude x are shown as time progresses. The relatively small initial scales expand in time in both cases. With β, propagation to the west occurs shortly after the start of the experiment and the phase speed increases until k approaches k_β toward the end of the run.

A schematic illustration of the results is shown in figure 5.16B. The initially turbulent field with energy in a narrow band of wavenumbers starts at point a with frequency kU and evolves toward lower wavenumbers, hence lower frequencies. When it reaches the frequency corresponding to barotropic Rossby waves, evolution of the turbulent field is halted and gives way to an evolving Rossby wave field with scale $(\beta/2U)^{-1/2}$.

Another obstacle to the red cascade is afforded by a western meridional boundary, where, as we have seen, reflection of Rossby waves favors the generation and accumulation of smaller scales. These modes tend to line up meridionally. Nonlinear flows also concentrate energy and enstrophy near the western boundary.

The qualitative effect of stratification can be introduced via a quasi-geostrophic two-layer model on the β-plane. Rhines carried out a series of initial-value calculations varying both the intensity and the scale of the initial turbulence. In these experiments the internal radius of deformation $\lambda_i = (g'H_1H_2/Hf^2)^{1/2}$ is a critical parameter. When the initial, turbulent scale L is small compared to λ_i, the two layers are effectively decoupled, as they are for linear waves. If the intensity of the turbulent flow is not large, the red cascade will develop in each layer until the frequency and scale of the barotropic Rossby wave for that layer is reached. This case is shown in figure 5.16A by the arrow from point b to a point on the wave dispersion curve where $\lambda_i k > 1$.

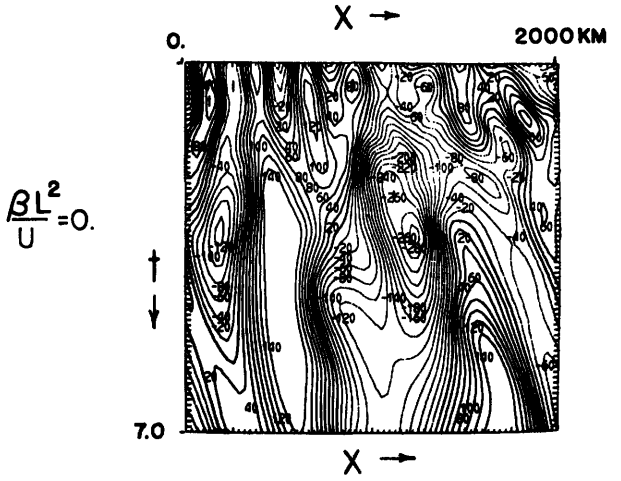

Figure 5.15A Contour diagrams for ψ as function of longitude and time for Rhines's numerical barotropic experiment with initially turbulent field without β. Cascade to large scales occurs as time increases.

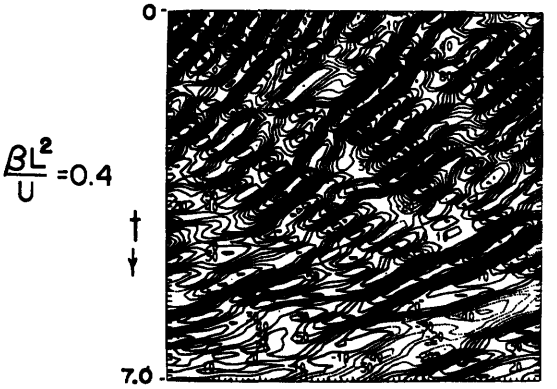

Figure 5.15B Same experiment with β. Westward propagation increases with time as cascade to larger scales occurs. Cascade is halted as $k \rightarrow \sqrt{\beta/2U}$.

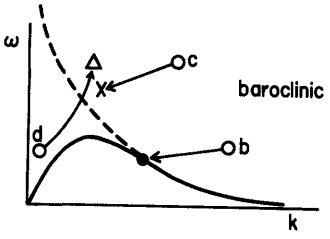

Figure 5.16A Frequency–wavenumber diagram for baroclinic modes. Solid curve shows linear baroclinic Rossby wave, dashed curve linear barotropic Rossby wave. For turbulent motions frequency is rms velocity U times k. Initial field of small-scale motions evolves by nonlinear interactions toward larger scales. If turbulent intensity is small (from point b), frequency arrives at linear wave values before deformation scale is reached and each layer experiences "barotropic" westward propagation. If intensity is larger (from point c), evolution to large scales proceeds until deformation scale is reached (point ×); then layers couple and barotropic development in figure 5.16B takes place. If initial turbulence has large scale (from point d), smaller scales are generated by nonlinear transfer and by baroclinic instability and system evolves to point Δ, where barotropic mode is generated and behavior shifts to figure 5.16B from point Δ.

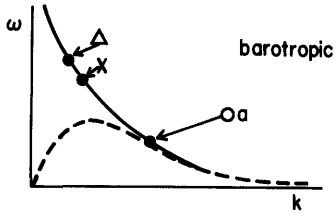

Figure 5.16B From point a nonlinear interactions of initially small-scale turbulent field produce red cascade until frequency reaches linear barotropic value (solid curve), when westward propagation disperses the energy to reduce nonlinear interactions. Same development from points × and Δ after barotropic mode is generated from initially baroclinic turbulent field.

If the intensity of the turbulence (in either layer) is increased to point c in figure 5.16A, the red cascade will take place again, but now the scale of the internal radius of deformation is reached (at point ×) before the wave steepness $U/\beta L^2$ has dropped to unity. As k decreases toward λ_i^{-1} the two layers will begin to interact with each other. The intensity of the flow allows a full evolution of the eddying motion to the occluded state described earlier and a barotropic response sets in. The latter is shown by the point × in figure 5.16B. From this point the system behaves like the barotropic system described earlier and the barotropic red cascade proceeds until the dispersion curve for the barotropic Rossby wave is reached. These results can be extended to the three-dimensional system.

If the initial, turbulent eddies have scales larger than λ_i, the flow can be treated locally as a large-scale basic flow field that becomes baroclinically unstable to disturbances with scale λ_i that derive their energy from the large-scale potential-energy field. Hence, the system will quickly develop *smaller* scales of motion. In this case the release of potential energy is responsible for the "blue" cascade. Once instability sets in at $k \simeq \lambda_i^{-1}$, the layers will couple, a barotropic response will be generated, and the barotropic red cascade will occur. The generation of smaller scales is shown in figure 5.16A from point d to Δ, where the layers couple; the barotropic red cascade from that point is shown in figure 5.16B. In Rhines's numerical experiment for this case the upper layer is shallow and the instability sets in where the large-scale eddy has a westward component, consistent with our observation that in a shallow upper layer westward flow is more unstable than eastward flow. At a later stage the eastward flow also becomes unstable. This particular case leads more quickly to the banded zonal flow described earlier be-

cause the baroclinic instability enhances the process by destabilizing the initially large-scale flow (particularly the large north-south scale) and by generating eddies of size λ_i, followed by a quick vertical coupling and then occlusion. If the red cascade leads to a zonal flow at an early stage, some of the potential energy in the initial, large-scale field remains to support a baroclinic zonal flow, i.e., one with a vertical shear. This is possible for $\Delta U/\lambda_i^2\beta < 1$ where ΔU is the velocity difference between the two layers.

Additional initial-value, run-down experiments were carried out by Rhines to study special features. One is the stability of a meridional baroclinic flow (essentially a westward-traveling Rossby wave in the upper layer). Eddies with scales slightly exceeding λ_i form and grow to larger amplitude, interacting laterally as they lock together vertically. The barotropic red cascade then sets in and a barotropic Rossby wave field again radiates energy away, leaving a zonal flow with scale k_β^{-1} as in the purely barotropic case discussed earlier.[11] Essentially all of the initial potential energy is released in the evolution of this flow.

Another experiment explores the instability of an eastward jet (an open-ocean "Gulf Stream") with $\beta = 0$. The flow is initially confined to the upper layer with a transverse scale of about $2\lambda_i$. A superposed weak disturbance develops rapidly via baroclinic instability and eddies are generated in both layers. The evolution of the flow involves an interaction between the (initially intense) mean flow and the growing waves as well as wave-wave interactions within each of the layers. The reader is referred to Rhines's paper for the details of the developing system, but the final result is mentioned here because of its pertinence to both observations and theories of the Gulf Stream. The flow that emerges has a strong barotropic component with a much increased "Gulf Stream transport," but there is little evidence in the flow field of an identifiable Gulf Stream. The density field (expressed as the interface height), however, preserves much more of the character of the observed Gulf Stream, with a meandering structure, detached rings, and a crowding of isopycnals along the axis. Thus, the dynamic picture involves two dissimilar modes, one barotropic, the other baroclinic, both of which are essential for a satisfactory description of the pertinent physical processes.

5.12.2 Effects of Topography in Two-Layer Flow

The conservation of potential vorticity requires a balance between $\partial\zeta/\partial t$, $\mathbf{v}\cdot\nabla\zeta$, df/dt, and $-f\,dh/h\,dt$. Nondimensional measures of these terms are a frequency ω/f, the Rossby number $Ro = Uk/f$, the β-effect β/fk, and the topographic effect $\delta = \Delta H\,k_T/Hk$. Here k and k_T are the horizontal scales of the flow and of the topography respectively, ΔH the amplitude of the topography, and H the depth.

For transient flows the local change ω/f will be balanced by a combination of the other parameters. If β/fk is larger than Ro or δ (weak flow over small topography), the system will respond with Rossby waves. In this case, if $k_T \gg k$ (small-scale topography), the Rossby waves will be weakly scattered by topography (Thompson, 1975). For large-scale topography ($k_T \ll k$) the response will be Rossby waves modified by topography as described in sections 5.9 and 5.10. For intermediate scales ($k_T \sim k$) the oscillations are irregular and will reflect the geometrical complexity of the topography (Rhines and Bretherton, 1974).

Nonlinear flows ($Ro > \delta$, β/fk) will exhibit the spectral broadening (red cascade) described above. Large topography ($\delta > Ro$, β/fk) of small horizontal scale ($k_T \gg k$) leads to generation of small scales through topographic scattering and refraction. If the topographic scale is large ($k_T \ll k$), a new "westward" direction is defined to the left of upslope.

With stratification the interplay of the three effects can be intricate. For example, if the flow is dominated by nonlinearity near the surface, spectral broadening will occur, and if the deep flow is weak, scattering will lead to small scales and the pseudowestward direction can generate fast baroclinic waves.

A series of initial-value numerical experiments by Rhines (1977) made use of random topography of rms amplitude $\delta = 0.053$ generated with a spectrum of scalar wavenumber $k^{-1.5}$ with scales $k \leq 8$ over a periodic domain of width 2000 km. The internal deformation radius λ_i was 40 km, so the topography cutoff scale was at the deformation radius.

A first experiment started off with turbulent eddies of scale $k \sim \lambda_i^{-1}$ and $Ro > \beta/fk$. The initial flow was confined to the upper layer. The deformation scale of the flow leads to coupling of the two layers and deep motion is generated. However, as soon as the deep water is set into motion, it interacts with the irregular topography which determines the direction of flow. Thus, topography detunes the layers and barotropic development is impeded.

When the initial flow includes deep motion as well so that $Ro > \delta$ there, the inertia of the flow eventually overcomes the relatively weak effect of the topography allowing coupling of the two layers to proceed. When the flow has evolved to expanded scales, the barotropic mode that emerges responds to the larger scales of the topography and the flow follows f/h contours. During the time that vertical adjustment is taking place, a more complicated, quasi-local relation among vertical structure, energy level (Ro) and topography (δ) may exist.

In both of these cases the initial flow has a small scale so that it contains only a limited amount of potential energy. If the initial eddies have a larger scale,

there is more potential energy available to the disturbances and faster growth via baroclinic instability is possible. When there is no topography, deformation scale disturbances will grow, the two layers will couple and occlude, and the barotropic red cascade will then take place. With rough topography, deformation-scale eddies are still generated and the growth is enhanced by the topography. This leads to an even larger release of the potential energy of the initial, larger-scale flow, though the locations of these enhanced deformation-scale eddies are determined by the topography so the two layers are detuned. To offset this detuning, continuous spectral broadening by topography takes place in deep water and the larger scales so generated can couple with the larger scales of the surface eddies. Hence, there is a tendency toward barotropy once again. If the flow is sufficiently nonlinear, topography is not so important and the development is more like that of a flat-bottom ocean.

What emerges from these considerations is that the response of the ocean depends in a complicated manner on the intensity and scale of the energy input and on the topographic structure in the region of response. However, the fact that the ratio of barotropic to baroclinic energy is a monotonic function of the ratio Ro/δ, as suggested by the calculations, seems to be a zero-order description of the observations. Rhines provides a much more detailed discussion of these issues as well as other considerations, such as the propagation of energy from a source area to distant regions with no input and the trapping of certain modes by topography. Several other investigators, including Bretherton and Karweit (1975) and Bretherton and Haidvogel (1976), have made important contributions to the study of the effects of different physical features on mid-ocean eddies. Holloway (1978), Salmon, Holloway, and Hendershott (1976), and Herring (1977) have applied closure modeling to the problem to enrich the story (see chapters 11 and 18).

5.12.3 Closed Basin Circulation

It is evident from the results reported above that the character of the circulation can undergo qualitative local changes because of nonlinear interactions and effects of bottom topography. When land boundaries are added, the circulation can be very different from that obtained by linear theory or even by nonlinear steady theory. Significant differences emerge if nonlinear, transient effects can develop fully. The latter are not easily treated analytically and most of the useful results have emerged from numerical experiments.

Some of the numerical studies (e.g., Robinson, Harrison, Mintz, and Semtner, 1977) use the primitive equations rather than the quasi-geostrophic set that we have been discussing. Although there are important nongeostrophic effects, many of the important features emerge from a study of the quasi-geostrophic system.

For a closed, forced system it is necessary to add dissipation. This can be done by introducing the terms $\partial\tau^x/\partial z + A\nabla^2 u$ and $\partial\tau^y/\partial z + A\nabla^2 v$ to the right-hand sides of (5.32) and (5.33), respectively, and $K_v\partial^2\rho_0/\partial z^2 + K_H\nabla^2\rho_0$ to the right-hand side of (5.44). The eddy coefficients A, K_v, and K_H are taken to be constant.[12] The vertical stress terms $\partial\tau^x/\partial z$ and $\partial\tau^y/\partial z$ contribute only near top and bottom boundaries, where they are normally evaluated to take into account the effect of wind stress and bottom drag via Ekman-layer processes.

Equation (5.54) is changed to

$$\frac{dq}{dt} = \frac{\partial}{\partial z}(\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}) + A\nabla^4\psi + K_H\nabla^2\frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\psi_z\right) + K_v\frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\psi_{zzz}\right), \tag{5.208}$$

where $q = \nabla^2\psi + f + \partial/\partial z(f_0^2\psi_z/N^2)$ is again the potential vorticity. If we write the variables in terms of an ensemble mean (overbar) and a perturbation (primed), we can write the equation for the mean potential vorticity as

$$\frac{\partial\bar{q}}{\partial t} + \bar{\mathbf{v}}\cdot\nabla\bar{q} + \overline{\mathbf{v}'\cdot\nabla q'} = \frac{\partial}{\partial z}(\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}) + A\nabla^4\bar{\psi} + K_H\nabla^2\frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\bar{\psi}_z\right) + K_v\frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\bar{\psi}_{zzz}\right). \tag{5.209}$$

For a two-layer model the equivalent system before averaging is

$$\begin{aligned}\frac{d}{dt}\left[\nabla^2\psi_1 + f - f_0\ln\left(\frac{h_1}{H_1}\right)\right] &= \frac{\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}_0}{h_1} + A\nabla^4\psi_1,\\ \frac{d}{dt}\left[\nabla^2\psi_2 + f - f_0\ln\left(\frac{h_2}{H_2}\right)\right] &= -K\frac{\nabla^2\psi_2}{h_2} + A\nabla^4\psi_2.\end{aligned} \tag{5.210}$$

Here, we have integrated vertically over each layer. The effect of topography is in the term η_3 (in h_2). Bottom friction is written explicitly as $-K\zeta_2 = -K\nabla^2\psi_2$ and $\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}_0$ is the wind-stress curl. The convective derivative for each layer involves the horizontal velocity components for that layer.

In (5.210) the thicknesses h_1 and h_2 of the two layers include the variations of the free surface, the interface, and the topography. In general, the h_i should be allowed to vanish, if need be, but no calculations have been made with variable h_i because following a material surface is very difficult to do numerically. The usual procedure is to set the h_i at their mean values in the coefficients. That means that surfacing of the thermocline is not permitted, so the separation mechanism of section 5.7 is not included. This is a serious omission for the complete problem. The numerical calculations

have been oriented toward determining the effects of eddies and topography on the vertical transfer of vorticity for the generation of abyssal circulation. The results are applicable so long as h_1 does not vanish.

Assuming a rigid lid at the top and a mean depth for the h_i in the coefficients, we can write (5.210) as

$$\frac{dq_1}{dt} = \frac{\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}_0}{H_1} + A\nabla^4\psi_1,$$
$$\frac{dq_2}{dt} = -K\frac{\nabla^2\psi_2}{H_2} + A\nabla^4\psi_2, \tag{5.211}$$

where $q_1 = \nabla^2\psi_1 + f + f_0\eta_2/H_1$, $q_2 = \nabla^2\psi_2 + f = -f_0(\eta_2 - \eta_3)/H_2$, where η_3 describes the topography. Again writing mean and fluctuation variables, we obtain for the mean equations

$$\frac{\partial\bar{q}_1}{\partial t} + \bar{\mathbf{v}}_1\cdot\nabla\bar{q}_1 + \overline{\mathbf{v}_1'\cdot\nabla q_1'} = \frac{\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}_0}{H_1} + A\nabla^4\bar{\psi}_1,$$
$$\frac{\partial\bar{q}_2}{\partial t} + \bar{\mathbf{v}}_2\cdot\nabla\bar{q}_2 + \overline{\mathbf{v}_2'\cdot\nabla q_2'} = -\frac{K\nabla^2\bar{\psi}_2}{H_2} + A\nabla^4\bar{\psi}_2, \tag{5.212}$$

where

$$\bar{q}_1 = \nabla^2\bar{\psi}_1 + f + \frac{f_0}{H_1}\bar{\eta}_2, \qquad q_1' = \nabla^2\psi_1' + f_0\eta_2'/H_1,$$
$$\bar{q}_2 = \nabla^2\bar{\psi}_2 + f - \frac{f_0(\bar{\eta}_2 - \eta_3)}{H_2}, \qquad q_2' = \nabla^2\psi_2' - f_0\eta_2'/H_2. \tag{5.213}$$

The mean properties in each layer depend on the correlations between fluctuating velocity and potential vorticity in addition to the remaining mean variables. We mention results from two numerical experiments.

The first is a numerical calculation by Holland (1973) of the continuous system [(5.209) plus the remaining equations of the system] for a rectangular basin 45° wide and extending from the equator (where symmetry is assumed) to about 65°N. A three-gyre wind system drives the circulation in a basin with topography shown in figure 5.17. Transport streamline patterns for a baroclinic ocean without topography, for a barotropic ocean with topography, and for a baroclinic ocean with topography are illustrated in figure 5.17. The first of the figures shows a structure not terribly different from the one obtained from linear theory with oceanic gyres determined by the wind pattern. The second exhibits the strong effect of topography on homogeneous water columns. The third contains response due to the combined effects of eddies and topography. The pattern is very different from the other two but the result of greatest significance is the enhanced transport generated by the eddies in conjunction with topography. The transport is increased by more than a factor of two over the case without topography.[13]

A more recent numerical experiment, analyzed by Holland and Rhines (1980), is based on the two-layer system (5.212) with a sinusoidal, cyclonic, wind stress in the north half-basin of amplitude 1 dyn cm^{-2} and an anticyclonic wind stress in the south. The basin is small, extending 1000 km eastward and 2000 km northward. The depth of the upper layer is 1 km, that of the lower 4 km, the bottom is flat, and the density difference is $\Delta\rho = 0.002\rho_2$. Instead of the Navier–Stokes form for lateral friction, a biharmonic form is used so that ∇^4 is replaced by $-\nabla^6$ in (5.212). At the sides normal flow and tangential stress vanish.

The experiment was run until it approached a statistically steady state, and time averages of the fields were then determined. For a statistically steady state equations (5.212) and (5.213) can be rewritten as

$$\bar{v}_1\cdot\nabla\bar{q}_1 + \overline{v_1'\cdot\nabla\nabla^2\psi_1'} + \frac{f_0}{H_1}\overline{v_1'\cdot\nabla\eta_2'} + A\nabla^6\bar{\psi}_1 - \frac{\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}_0}{H_1} = 0.$$
$$\bar{v}_2\cdot\nabla\bar{q}_2 + \overline{v_2'\cdot\nabla\nabla^2\psi_2'} - \frac{f_0}{H_2}\overline{v_2'\cdot\nabla\eta_2'} + A\nabla^6\bar{\psi}_2 + \frac{K\nabla^2\bar{\psi}_2}{H_2} = 0, \tag{5.214}$$

Statistically steady streamlines for the two layers are shown in figure 5.18. The upper layer consists of two gyres, one anticyclonic, one cyclonic, with their centers displaced toward mid-latitude as is characteristic of circulations with strong inertial effects. The smaller gyres C and D contain the nonlinear recirculation. The upper-layer pattern is not so different from the one for a homogeneous ocean with strong inertial effects (figure 5.6). The lower layer, on the other hand, exhibits entirely different behavior. Small intense gyres exist under, and have flows parallel to, the small-surface gyres C and D. Counterrotating gyres lie to the north and south of the center two.

Now suppose that equations (5.214) are integrated over the depth of each layer and over an area bounded by the streamlines of the respective gyres. Then the mean convective derivatives in (5.214) will vanish [each of them can be rewritten in divergent form $\nabla\cdot(\mathbf{v}q)$ and vanishes on horizontal integration because the boundary is a streamline]. The remaining terms will then balance and we have

$$H_1\overline{\overline{v_1'\cdot\nabla\nabla^2\psi_1'}} + f_0\overline{\overline{v_1'\cdot\nabla\eta_2'}} + \overline{H_1A\nabla^6\bar{\psi}_1} - \overline{\hat{\mathbf{k}}\cdot\nabla\times\boldsymbol{\tau}_0} = 0$$
$$U_1 + U_2 + U_3 + U_4 = 0,$$
$$H_2\overline{\overline{v_2'\cdot\nabla\nabla^2\psi_2'}} - \overline{\overline{f_0v_2'\cdot\nabla\eta_2'}} + \overline{H_2A\nabla^6\bar{\psi}_2} + \overline{K\nabla^2\bar{\psi}_2} = 0 \tag{5.215}$$
$$L_1 + L_2 + L_3 + L_4 = 0,$$

where the subscripted capital letters (U for upper layer, L for lower) identify the different terms. For the upper layer from left to right these vorticity balances are due

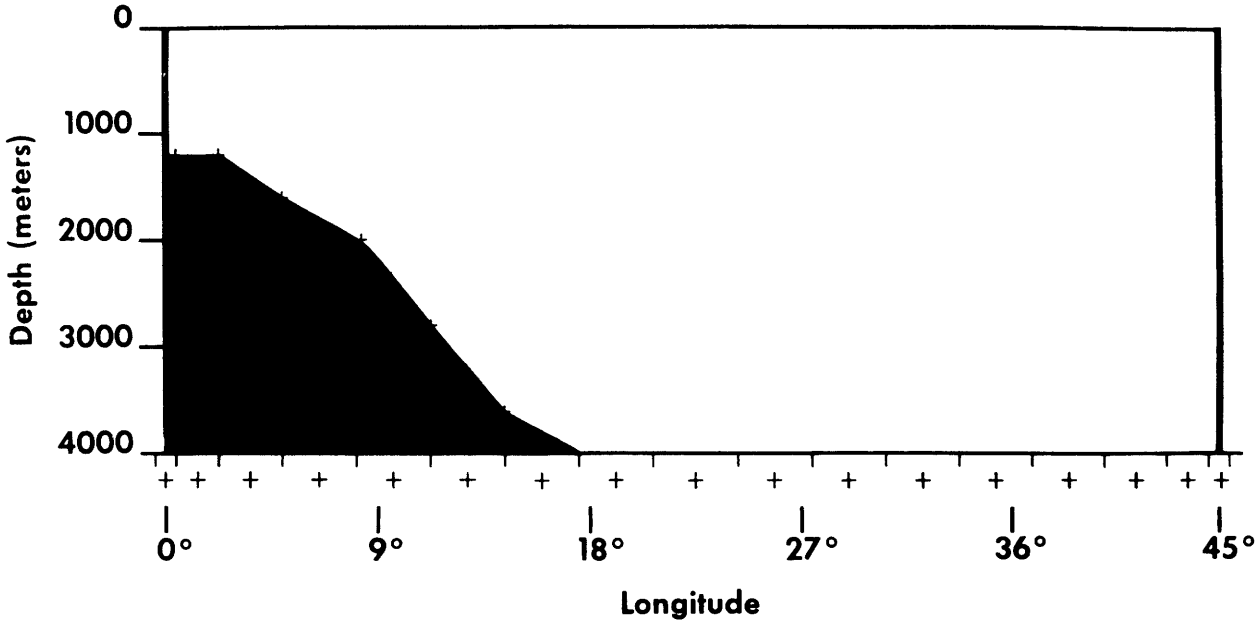

Figure 5.17A Topography used by Holland (1973) in ocean circulation calculation.

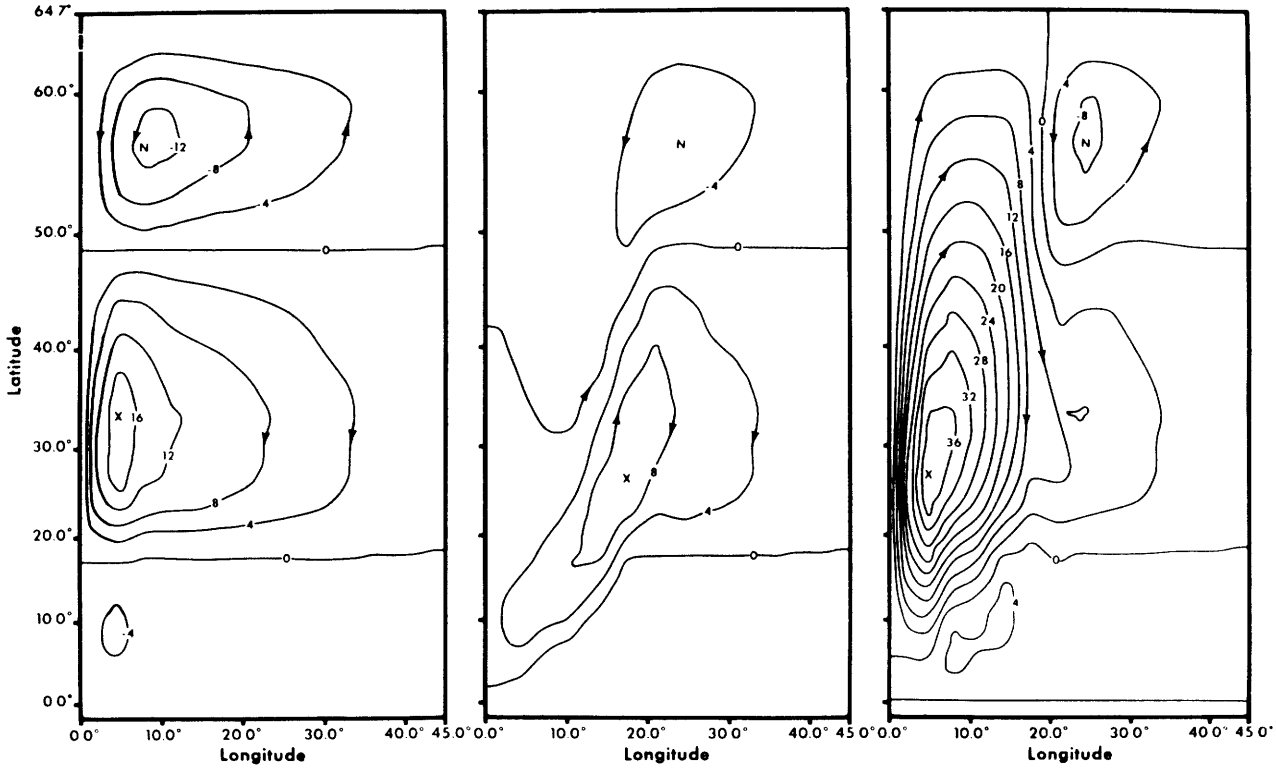

Figure 5.17B Horizontal transport streamfunction for (left) baroclinic ocean with no topography, (middle) barotropic ocean with topography, (right) baroclinic ocean with topography.

to Reynolds stress, interfacial stress, lateral friction, and wind stress, respectively. For the lower layer the first three are the same and the last is due to bottom friction.

The statistically steady streamlines fall between grid points in the numerical experiment, so the different quantities can be evaluated only approximately. Furthermore, the mean vorticities showed time dependence even for the long times of the averages. Hence, the numerical results can provide only an approximation to (5.215).

Holland and Rhines evaluated the integrals for each gyre. Their convention uses a positive sign for the integral if it serves to drive the vortex and a negative value if it opposes the vorticity. The results are shown in table 5.1. We try here to interpret the results in terms of the processes described in earlier sections, though such balances tell only part of the story.

In the upper layer the large gyres, *A* and *B*, are driven primarily by the wind. Resistance is effected via Reynolds stresses though there is a small retardation by the interfacial stresses. The assumed biharmonic lateral "friction" helps to drive the flow, but the magnitude of this driving is smaller than the errors due to the approximate evaluation of the integrals, so even the sign is not reliable.

The smaller gyres, *C* and *D*, are driven more by Reynolds stresses than by the winds (which is not surprising since these gyres are inertially controlled and have their centers in the wrong place for directly wind-driven gyres). The largest single contribution for these smaller gyres comes from the retarding effect of interfacial stress. The interacting fluctuations of velocity and interface height (temperature) in gyre *C* depress the interface, thus tending to stretch vortex lines and weaken the anticyclonic vortex. In the cyclonic gyre *D* the interface is raised by $v_1' \cdot \nabla \eta_2'$ and the vortex lines are compressed, thus generating anticyclonic vorticity. Hence, in both small gyres the mean of the fluctuating interactions serves to weaken the prevailing vorticity. The approximation errors for these smaller gyres are smaller but still annoying.

The difference between the integral values for the small and large gyres in the upper layer gives the corresponding values for the region between the small and large gyre in each basin—or it would if there were no error. It is especially important to note that U_2 for the small gyres is larger than the value for the large gyres. In other words, in the regions between the streamlines in each half-basin the interfacial stress term *drives* the gyre (against the retarding effect of Reynolds stresses). This serves to raise the thermocline in the south (between *C* and *A*) and lower it in the north (between *D* and *B*).

The lower layer, which would have no average flow in a real steady state or for linear transient motions, is

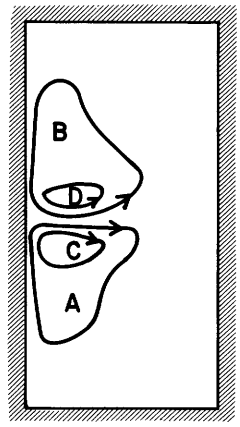

Figure 5.18A Circulation gyres generated in top layer of ocean model with cyclonic wind stress in north, equal-strength anticyclonic wind stress in south. Gyres *C* and *D* contain recirculating water.

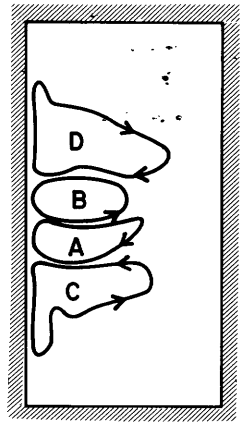

Figure 5.18B Lower-layer gyres near mid-latitude have same sign as those of upper layer. Other gyres opposite.

Table 5.1 Values of U_i and L_i in Equation (5.215): Units $m^3 s^{-2}$

Upper layer	U_1	U_2	U_3	U_4	Imbalance[a]
Gyre *A*	−61.2	−17.6	5.1	53.9	−19.8
Gyre *B*	−49.2	−10.2	5.7	57.9	5.2
Gyre *C*	11.0	−23.8	−0.3	7.9	−5.2
Gyre *D*	11.6	−21.2	0.4	3.3	−5.9

Lower layer	L_1	L_2	L_3	L_4	Imbalance[a]
Gyre *A*	−19.9	44.6	−0.7	−24.1	−0.1
Gyre *B*	−18.8	42.8	−0.6	−22.3	1.1
Gyre *C*	−11.4	44.5	−7.3	−25.0	0.8
Gyre *D*	−9.2	40.5	−6.5	−22.0	2.8

a. Imbalances are due to residual time dependence and approximation errors for evaluating integrals.

driven by the interfacial stresses (table 5.1). Thus, the depression of the interface under upper-level gyre *C* causes vortex shrinking in the lower layer and generates the anticyclonic gyre there. Similarly, the elevation of the interface under upper-level gyre *D* causes vortex stretching in the lower layer and generates the cyclonic gyre. We noted above that the interfacial stresses change sign in the region between the small and large gyre of each half-basin, so the same mechanism leads to vortices of opposite sign, i.e., in the lower layer *D* is opposed to *B* and *C* is opposed to *A*. These results are all consistent with the simple idea of vortex generation by vertical divergence—at least, in the mean. Local balances could well be different, especially if there are large horizontal variations in the different quantities, as there undoubtedly are.

The vertically integrated transport of the northward boundary current near mid-latitude exceeds the Sverdrup transport by a factor of three. Thus the effect of eddies and the corotating deep gyre contribute greatly to the transport. The transport of the counterrotating deep gyres can overwhelm the value due to the surface gyres above them.

An analysis of the results of the numerical experiment (Holland, 1978) indicates that there is no region where the Sverdrup balance is locally valid for the vertically integrated flow. However, the nature of eddy effects on the mean flow is such that spatial averaging tends to reduce their importance. Hence, Sverdrup balance may apply approximately to a spatially averaged region even when it does not apply at a point. Furthermore, the east-west scale of the basin is small (1000 km), and the recirculating gyres extend across almost the entire basin. In a wider basin the recirculating gyre is probably not much larger and a region of local Sverdrup balance can exist.

An interesting observation about this experiment is that if one calculates the northward transport from hydrographic data (the interface height) across the boundary current, assuming zero deep flow, one obtains a transport about 50% larger than the Sverdrup transport and about half of the actual transport in the two layers. Thus, the deeper thermocline gives a correction in the proper sense.

The Eulerian mean circulation in these models is driven by a combination of wind-stress curl and the divergence of the eddy flux of potential vorticity. In some circumstances this eddy effect can be rewritten in terms of an eddy diffusion coefficient based on Lagrangian dispersion of fluid particles. Taylor (1915) first proposed such an approach for nondissipative systems. Bretherton (1966b) made use of the argument and pursued the issue in subsequent papers. Rhines (1977) (see also Rhines and Holland, 1979) has been developing the theory for more general use. A parameterization of the effect of oceanic eddy processes is extremely important, but it is too early to try to summarize the theory.

Holland's (1978) paper describing the numerical solutions gives more details. An important point is that the free jet (Gulf Stream) is barotropically unstable in this case and the baroclinic instability of the westward return flows is weaker. The calculation is intended to study processes, however, rather than the observed phenomenon, so the relative strengths of the instabilities of the model are not necessarily a prediction of what happens in the ocean.

A linear stability analysis of the mean and instantaneous flows in the experiment was carried out by Haidvogel and Holland (1978). Many of the unstable features that emerge from the full experiment could be correctly accounted for by linear theory using the flows of the experiment. In some cases, however, use of the instantaneous velocity fields led to better results than did the use of time-averaged or mean flows. Since the time scale of the former was sometimes shorter than the time scale of the deduced unstable motions, the significance of the results is not clear. One point made by Haidvogel and Holland is that the Reynolds stresses deduced from linear theory have the wrong sign. Hence, finite-amplitude effects must change the interactions in the developed flow. We have already remarked on the importance of large amplitudes in the occlusion phenomenon.

Semtner and Holland (1978) compare the results of the quasi-geostrophic model with the results obtained from the more complete (and much more expensive) runs using a five-level primitive-equation model. The two-layer quasi-geostrophic model can simulate many of the features of the latter, particularly if the parameters, such as the depth ratio, are adjusted for optimum fit. Gill et al. (1974) and Flierl (1978) have also made such comparative studies but for more restricted physical models.

Large-scale numerical models to determine climatological oceanic behavior have been carried out by the NOAA group at Princeton [see Bryan, Manabe, and Pacanowski (1975) for earlier references]. These models are sometimes run together with meteorological models, allowing an interaction of ocean and atmosphere, and have the greatest potential scope of any of the numerical experiments. Because of the grand scope there is little opportunity to vary the parameters. As the simpler, process-oriented models map out the more realistic domain of parameter space, larger models based on more optimal parameters may achieve a predictive level.

Notes

1. Approximating the stratification by two layers is an idealization. Even a cursory examination of the data (Veronis, 1972) shows that it is not possible to simulate the real stratification with two layers each of constant density separated by an interface that corresponds to the thermocline. The idealization is useful, however, in developing an intuition about the effects of stratification, and the results obtained are suggestive as guides when one is working with more realistic models.

2. Flierl (1978) has analyzed a variety of flows using a two-layer fluid, two vertical normal modes and many vertical normal modes. In each case he has calibrated the parameters in the two-layer and two-mode systems to obtain the best fit with the results derived from the more complete system containing many normal modes. He has shown that the best-fit parameters for the two-layer system vary widely depending on the phenomenon under investigation. For example, a study of topographic effects in a two-layer model requires an optimal choice of mean depths for the upper and lower layers quite different from the optimal choice required for a study of nonlinear effects. The optimal parameters for a model using two normal modes are much less sensitive to the process being studied.

3. Fofonoff (1962b) and Kamenkovich (1973) give detailed discussions of the general thermodynamic properties of sea water. Lynn and Reid (1968) and Veronis (1972) discuss the vertical stability characteristics of water columns. Small-scale mixing in waters stratified by heat and salt is discussed by Turner (1973a) and by Stern (1975a). The latter raises the question of the possibly crucial role played by the salinity balance in global circulation.

4. Phillips first transforms the system to Mercator coordinates and then makes the β-plane approximation. That leads to a rectangular coordinate system with eastward distance $x = (a\cos\phi_0)\lambda$ and northward distance $y = (a\cos\phi_0)\phi$. Therefore, in polar latitudes, where linear distance between meridians decreases, consistency requires that a correspondingly smaller range of latitude be chosen for an equivalent measure of distance. The two expansions lead to the same vorticity equation obtained at first order, but his first-order momentum equations are symmetric whereas our set is not.

5. It is well known that Ekman developed the analysis to explain Nansen's observation that ice floes in the Arctic drifted at an angle of 45° to the right of the wind. Nansen, in fact, had a good physical argument to describe the process that Ekman subsequently quantified.

6. The present analysis is for an idealized case. In nature other processes, such as lateral penetration from distant topographic irregularities (Armi, 1978), may affect the vertical structure near the bottom.

7. $\lambda_i = (gh_i/f_0^2)^{1/2}$ is the radius of deformation for a rotating system with a free surface defined by Rossby (1938) as the distance over which a disturbance will be transmitted by longitudinal pressure forces before the (transverse) effect of rotation takes over. It emerges naturally from (5.146) to (5.148) for longitudinal (Kelvin) waves near a wall, say along the y direction, with no transverse (U) flow. Then (5.147) and (5.148) yield the dispersion relation $\omega^2 = gh_i l^2$ for long gravity waves and (5.146) gives a transverse decay scale of λ_i.

8. The parameter $B = NH/f_0L$, where L is the horizontal dimension, is frequently used with continuously stratified fluids (N. A. Phillips, 1963) and may be preferred because it contains only external parameters. The use of $\lambda_1^2K^2$ here means that we measure the wave scale K relative to the internal radius of deformation λ_1.

9. Baroclinic instability of a westward current implies equatorward heat transport by the finite-amplitude field. Thus, the required poleward heat transport in the ocean must be supplied by the mean circulation.

10. Gall (1976) shows that the vertical structure of finite-amplitude motions that appear to evolve from baroclinic instability differs significantly from the structure of the unstable infinitesimal motions.

11. Gareth Williams (1978, 1979) has carried out extensive numerical experiments to explain Jupiter's bands and has obtained strong zonal jets from both barotropic and baroclinic eddy fields.

12. The assumption of constant eddy coefficients is less serious here than in the linear theories discussed earlier if the numerical calculations use a grid that is fine enough to resolve the eddies that arise from baroclinically unstable modes. The effect of the eddies can be taken into account explicitly and the eddy coefficients then include processes on even smaller (and hopefully dynamically less important) scales.

13. As can be seen in the first of the figures, the driving in this case is not very strong. Perhaps stronger motions in the deeps could override some of the topographic effect, as Rhines (1977) showed in the calculations reported above. This experiment is one of the earlier eddy-resolving calculations and the dependence on the parameters was only marginally explored.