

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 ($\approx 40^\circ\text{N}$ 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 quantitative 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) analy-

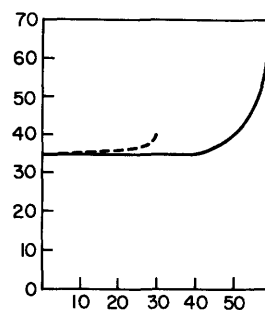


Figure 5.8 The path (solid curve) of the Gulf Stream after it has separated from the coast [from a reduced gravity model by Veronis (1973a)]. The zonal wind stress that drives the system is taken from observations and has zero curl at 40°N . The Norwegian Current is the narrow jet in the northeast. The 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.

sis of the deep circulation induced by the separated current.

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

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

$$fu = -\frac{1}{a} \frac{\partial P}{\partial \phi}, \quad (5.105)$$

$$\frac{\partial P}{\partial z} = -g \frac{\rho}{\rho_m}, \quad (5.106)$$

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

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

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 re-

view 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

$$\begin{aligned} & \kappa \sin \phi \cos \phi (P_{zz} P_{zzz} - P_{zzz}^2) \\ & = P_{zzz} \frac{\partial(P_{zz}, P)}{\partial(\lambda, \phi)} + P_{zz} \frac{\partial(P, P_{zz})}{\partial(\lambda, \phi)} + \cot \phi P_{\lambda} P_{zz}^2, \end{aligned} \quad (5.109)$$

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

Welander (1959, 1971b) defined the variable

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

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

$$\begin{aligned} & \kappa \sin \phi \cos \phi M_{zzz} + \frac{\partial(M_{zz}, M_z)}{\partial(\lambda, \phi)} \\ & - \cot \phi M_{\lambda} M_{zz} = 0. \end{aligned} \quad (5.111)$$

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 :

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

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

$$C(\lambda, \phi) = c/\sin \phi, \quad (5.113)$$

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

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

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

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

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

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

$$w_d \frac{\partial\rho}{\partial z} = K \frac{\partial^2\rho}{\partial z^2}. \quad (5.118)$$

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

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

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 $\cos(\phi + 10^\circ)$ and a reference temperature of 2.45°C at 5000 m depth at $\phi = 10^\circ$, 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 $\phi = 30^\circ$.

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 so-

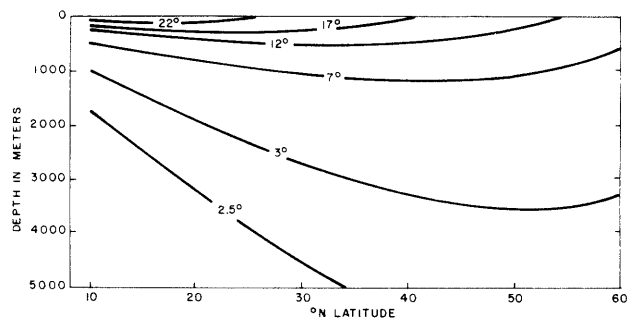


Figure 5.9 Isotherms, or isopycnals, from Needler's (1967) thermocline model with vertical dependence $\propto \exp(cz/\sin\phi)$. The observed maximum vertical penetration of warm water at mid-latitude is a feature of almost all of the thermocline models.

lutions 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.

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 $K = 10^{-4} \text{ m}^2 \text{ s}^{-1}$ and $w_d = 10^{-7} \text{ m s}^{-1}$.

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 straightforward 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

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

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

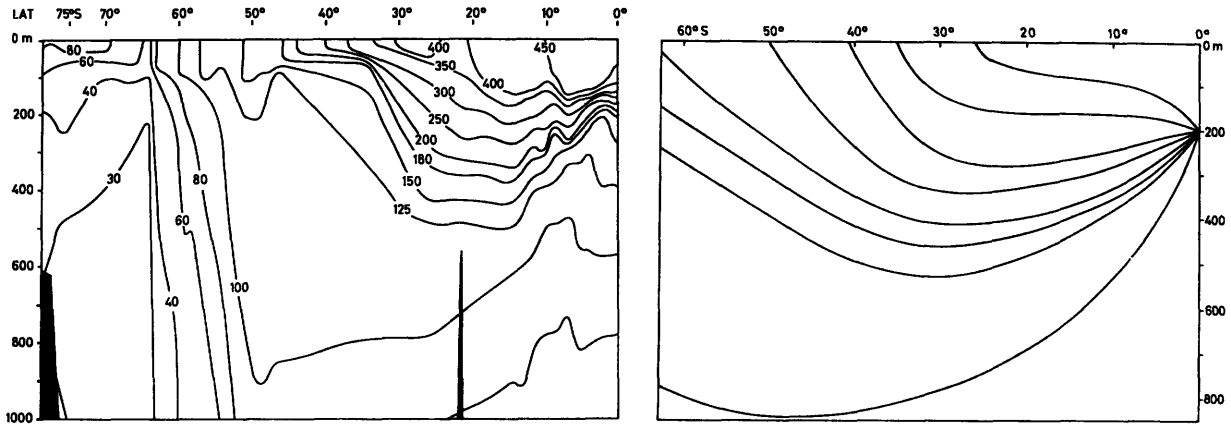


Figure 5.10 (A) Contours of thermocline anomaly (units of $10^{-5} \text{ cm}^3 \text{ g}^{-1}$) in the upper kilometer of the South Pacific along

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 $fV/H_a \sim g\rho/\rho_m$ 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 $1.5 \times 10^{-7} \text{ m s}^{-1}$ 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 $2 \times 10^{-7} \text{ m s}^{-1}$ (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.

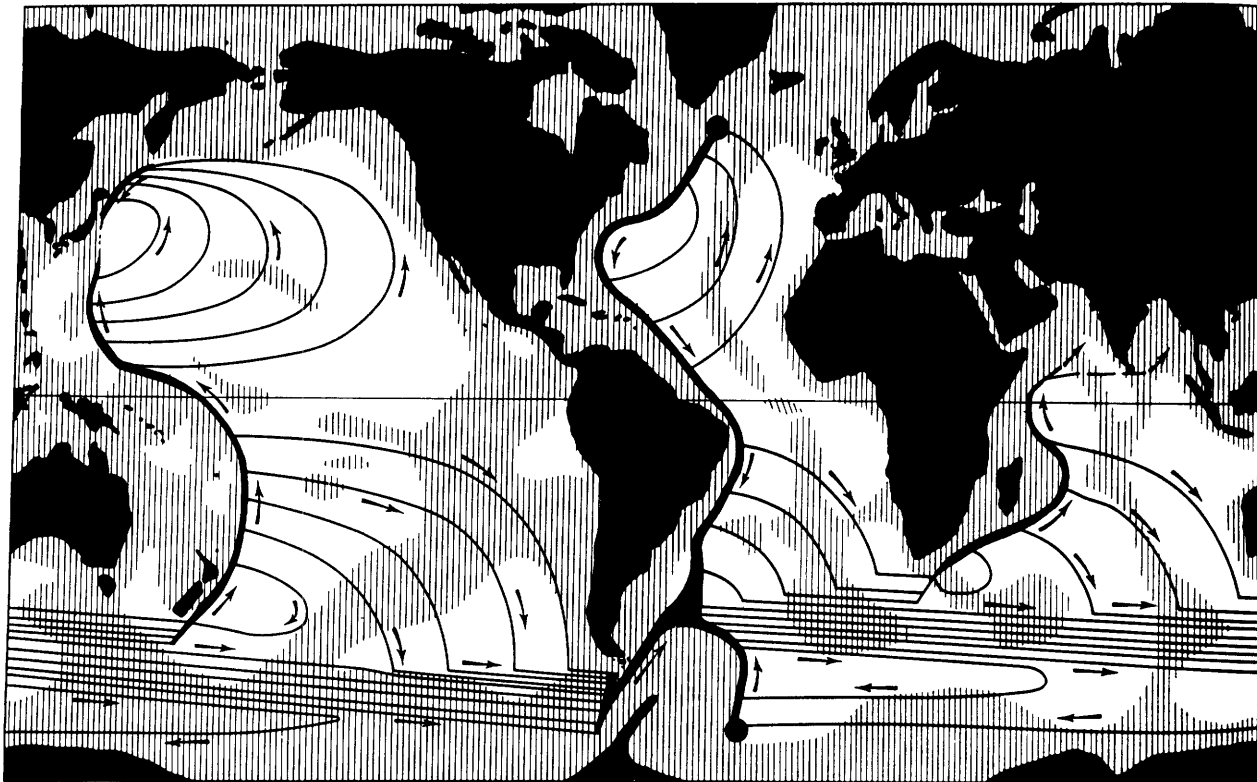


Figure 5.11 The abyssal circulation obtained by Stommel (1958) and generated by equal sources in the North Atlantic and in the Weddell Sea with uniform upwelling elsewhere.

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 $Ro \ll 1$):

$$u_{1t} - fv_1 = -g\eta_{1,x}, \quad (5.127)$$

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

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

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

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

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

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

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

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

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

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

$$\mathbf{V}_1 = \mathbf{v}_1 + \alpha\mathbf{v}_2, \quad (5.136)$$

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

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

where h is a constant to be determined.

Equating Φ in (5.137) and (5.138) yields

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

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.,

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

Eliminating h yields

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

For small ϵ the two values for α are

$$\alpha_1 \approx \frac{H_2}{H_1}, \quad \alpha_2 \approx -1. \quad (5.142)$$

Corresponding values of h and the variables are

$$h_1 = H_1 + H_2, \quad h_2 = \epsilon H_1 H_2 / (H_1 + H_2), \quad (5.143)$$

$$\mathbf{V}_1 = \mathbf{v}_1 + \frac{H_2}{H_1} \mathbf{v}_2, \quad \mathbf{V}_2 = \mathbf{v}_1 - \mathbf{v}_2, \quad (5.144)$$

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

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

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

$$U_{it} - fV_i = -g\Phi_{i,x}, \quad (5.146)$$

$$V_{it} + fU_i = -g\Phi_{i,y}, \quad (5.147)$$

$$\Phi_{it} + h_i(U_{i,x} + V_{i,y}) = 0, \quad (5.148)$$

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 $H_1 H_2 / H$. If U_i and Φ_i are eliminated from (5.146) to (5.148) the resulting equation in V_i is

$$(\partial_{tt}^2 + f^2 \partial_t - gh_i \partial_{xxt}^3 + gh_i \beta \partial_x - gh_i \partial_{yyt}^3) V_i = 0. \quad (5.149)$$

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,

$$\omega_i^3 - (f_0^2 + gh_i K^2) \omega_i^2 - gh_i \beta k = 0, \quad (5.150)$$

$$K^2 = k^2 + l^2.$$

This yields the (approximate) dispersion relations⁷

$$\omega_{i1} = f_0 \sqrt{1 + \lambda_i^2 K^2}, \quad \omega_{i2} = -\omega_{i1}, \quad (5.151)$$

$$\omega_{i3} = -\frac{\beta k \lambda_i^2}{1 + \lambda_i^2 K^2}.$$

A plot of ω versus K is shown in figure 5.12 for $k = l$ at mid-latitude ($f_0 = 10^{-4} \text{ s}^{-1}$, $\beta = 2 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$) 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 ($\omega_{13} \sim \beta k \lambda_i^2$). 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_{13}, k) -plane in figure 5.12 shows that the phase velocity ω_{13}/k is westward and that a given value of ω_{13} corresponds to two wavelengths, a short wave with small velocity and a long wave with a fast velocity. The zonal group velocity $\partial\omega_{13}/\partial k$, 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

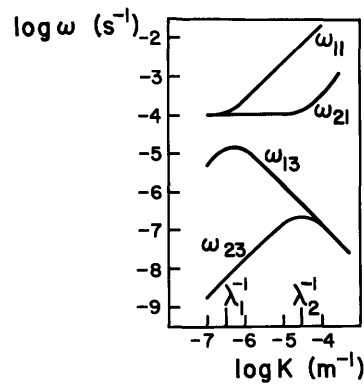


Figure 5.12 Frequency-wavenumber diagram for waves in a two-layer, constant-depth ocean in the β -plane. The upper two curves are inertio-gravity waves, the two lower are Rossby waves. Transitions in the dispersion curves occur at the deformation scales shown on abscissa.

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

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

For quasi-geostrophic motions we can substitute the geostrophic velocity in $\zeta = v_x - u_y$ and in \mathbf{v} . Furthermore, we have $h_{2t} = \eta_{2v}$, $h_{1t} = (\eta_1 - \eta_2)_t$, $\mathbf{v}_1 \cdot \nabla h_1 \approx 0$ (for linear flows), $\mathbf{v}_2 \cdot \nabla h_2 = -\mathbf{v}_2 \cdot \nabla \eta_3$ 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

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

and

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

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

$$(1 + \lambda^2 K^2 + \lambda_1^2 \lambda_2^2 K^4) \omega^2 + \{(\lambda^2 + 2\lambda_1^2 \lambda_2^2 K^2) k \beta + (1 + \lambda_1^2 K^2) \lambda_2^2 k b\} \omega + \lambda_1^2 \lambda_2^2 k^2 \beta (b + \beta) = 0, \quad (5.155)$$

where $\lambda^2 = gH/f^2$, $\lambda_2^2 = gH_2/f^2$, $\lambda_1^2 = \epsilon gH_1/f^2$, $b = f\{k\eta_{3y} - l\eta_{3x}\}/H_2 k$. 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 reinforces 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

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

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

$$\begin{aligned} \omega_2 &= -\frac{k\beta\lambda_1^2\lambda_2^2}{\lambda^2} \frac{\beta + (bH/H_2)}{\beta + b} \\ &= -\frac{k\beta g' H_1 H_2}{f^2 H} \frac{\beta + (bH/H_2)}{\beta + b}. \end{aligned} \quad (5.157)$$

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

$$\omega_2 = -\frac{k\beta g' H_1 H_2}{f^2 H} \quad (5.158)$$

[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

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

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

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

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

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

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

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

where $S = N/f_0$.

At the upper boundary we assume

$$w = 0 \quad \text{or} \quad p_z = 0 \quad \text{at} \quad z = 0. \quad (5.163)$$

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 ,

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

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

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

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

$$\omega p_z = HbS^2 k p \quad \text{at} \quad z = -H. \quad (5.166)$$

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

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

and

$$m \tan m = \lambda_1^2 k b / \omega \quad (5.168)$$

where⁸

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

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

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

$$-\gamma m \tan m = m^2 + \lambda_1^2 K^2, \quad (5.170)$$

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 $m = n\pi$ ($n = 0, 1, 2, \dots$). 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 ($\beta v \approx f w_z$) 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 ($m \rightarrow \pi/2$). 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 $\gamma \rightarrow \infty$). Strong stratification (large λ) or small wavelength (large K) have the same effect as a large slope.

It should be noted that when topography reinforces β there is no vertically oscillatory solution that yields a barotropic Rossby wave in the limit $\gamma \rightarrow \infty$. 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

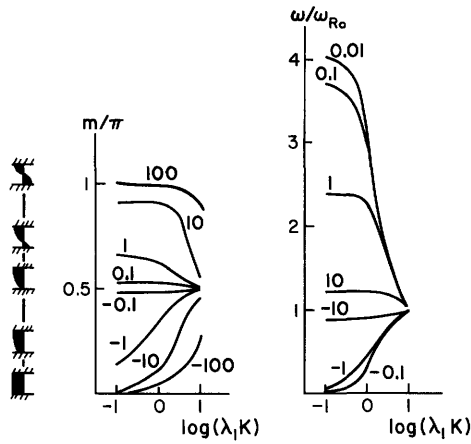


Figure 5.13A Values of the vertical wavenumber m for the gravest quasi-geostrophic mode in a uniformly stratified ocean as functions of horizontal wavenumber. Each curve is marked with value of γ . Topographic effects dominate at small γ ; β dominates at large γ . The vertical structure of the modes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

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

and

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

Eliminating ω gives

$$\gamma \mu \tanh \mu = \lambda_1^2 K^2 - \mu^2. \quad (5.173)$$

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

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

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

$$\omega = -\frac{\lambda_1 bk}{K} = -\frac{\alpha Nk}{K}. \quad (5.175)$$

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 ($\lambda_1 K \ll 1$), $\tanh \lambda_1 K \approx \lambda_1 K$ and

$$\omega = -\frac{bk}{K^2}. \quad (5.176)$$

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 reinforces 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

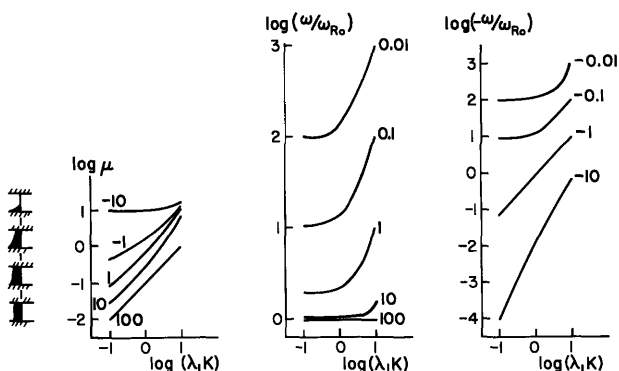


Figure 5.14A Vertical scale μ for the evanescent mode $\cosh \mu z/H$ as a function of horizontal wavenumber. Each curve is marked with value of γ . The vertical structure of the modes is sketched at the left for the different values of μ .

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.

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).

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 ($f = \text{constant}$) 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

$$fU_1 = fU = -gH_y, \tag{5.177}$$

$$fU_2 = -fU = -g[(1 - \epsilon)H_y + \epsilon H_{2y}],$$

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

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

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

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

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	$H_2 (= H_1)$
$\eta = \eta_1 - \eta_2$	$\eta = \eta_2$
$\bar{u}_1 = U$	$\bar{u}_2 = -U$

(5.179)

$$Q_{1\nu} \equiv -\frac{f}{H_1} H_{1\nu} = \frac{2U}{\lambda_1^2} \quad Q_{2\nu} \equiv -\frac{f}{H_2} H_{2\nu} = -\frac{(2-\epsilon)}{\lambda_1^2} U \approx -\frac{2U}{\lambda_1^2}$$

$$q_1 \equiv \zeta_1 - f(\eta_1 - \eta_2)/H_1 \quad q_2 \equiv \zeta_2 - f\eta_2/H_2$$

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

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

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

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

$$\begin{aligned} v_1 &= \phi_{1x}, & u_1 &= -\phi_{1y}, & v_2 &= \phi_{2x}, \\ u_2 &= -\phi_{2y}, \end{aligned} \quad (5.181)$$

where

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

Then

$$\begin{aligned} \zeta_1 &= v_{1x} - u_{1y} = \nabla^2 \phi_1, & \zeta_2 &= v_{2x} - u_{2y} = \nabla^2 \phi_2, \\ \frac{f}{H_1} (\eta_1 - \eta_2) &= \frac{1}{\lambda_1^2} (\phi_1 - \phi_2), \end{aligned} \quad (5.183)$$

$$\frac{f\eta_2}{H_2} = \frac{1}{\lambda_1^2} (\phi_2 - \phi_1 - \epsilon\phi_1) \approx \frac{1}{\lambda_1^2} (\phi_2 - \phi_1),$$

$$q_1 = \nabla^2 \phi_1 - \frac{1}{\lambda_1^2} (\phi_1 - \phi_2), \quad q_2 = \nabla^2 \phi_2 + \frac{1}{\lambda_1^2} (\phi_1 - \phi_2),$$

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

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

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

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

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

From (5.183)

$$q_1 + q_2 = \nabla^2 (\phi_1 + \phi_2), \quad (5.186)$$

$$q_1 - q_2 = \nabla^2 (\phi_1 - \phi_2) - \frac{2}{\lambda_1^2} (\phi_1 - \phi_2),$$

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

$$e^{-i\omega t + ikx} \sin ly, \quad l = \frac{m\pi}{M}, \quad m = 1, 2, \dots, \quad (5.187)$$

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

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

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

$$\begin{aligned} K^2 (\phi_1 + \phi_2)_t &+ [U(K^2 + 2/\lambda_1^2) - Q_{1\nu}] (\phi_1 - \phi_2)_x = 0, \end{aligned} \quad (5.189a)$$

$$\begin{aligned} (K^2 + 2/\lambda_1^2) (\phi_1 - \phi_2)_t &+ (UK^2 - Q_{1\nu}) (\phi_1 + \phi_2)_x = 0. \end{aligned} \quad (5.189b)$$

Then substituting for $Q_{1\nu}$ yields

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

$$(K^2 + 2/\lambda_1^2)(\phi_1 - \phi_2)_t + UK^2(\phi_1 + \phi_2)_x - \frac{2U}{\lambda_1^2}(\phi_1 + \phi_2)_x = 0. \quad (5.190b)$$

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

$$\begin{aligned} \omega^2 &= \frac{k^2}{K^2(K^2 + 2/\lambda_1^2)} [U(K^2 + 2/\lambda_1^2) - Q_{1y}][UK^2 - Q_{1y}] \\ &= kU^2 \frac{K^2 - (2/\lambda_1^2)}{K^2 + (2/\lambda_1^2)}. \end{aligned} \quad (5.191)$$

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

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

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 $\lambda_1^2 K^2 = 2(\sqrt{2} - 1)$.

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 reinforces 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 reinforcement and instability occurs.

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

$$\phi_1 + \phi_2 = Ae^{\sigma t} \sin kx \sin ly, \quad (5.193)$$

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

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

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

Then (5.190a) yields

$$\sigma A = UkB$$

or

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

For very long waves ($K^2 \ll 2/\lambda_1^2$) we have $A \approx B$ for the growing mode and $A \approx -B$ 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

$$\begin{aligned} -\overline{(v_1 + v_2)\eta_2} &= \frac{H_2}{f\lambda_1^2} \overline{(\phi_1 + \phi_2)_x(\phi_1 - \phi_2)} \\ &= \frac{kH_2}{f\lambda_1^2} AB, \end{aligned} \quad (5.196)$$

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

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

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

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

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

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

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

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

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

The frequency equation is

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

and the condition for instability is changed to

$$\lambda_1^2 K^4 - 4\lambda_1^4 K^4 + \frac{\beta^2 \lambda_1^4}{U^2} < 0 \quad (5.202)$$

or

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

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

$$2|U| > \beta\lambda_1^2. \quad (5.204)$$

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) $Q_y U_z < 0$ at $z = 0$, or (3) $Q_y(U_z + N^2 H_y/f) > 0$ 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_v 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_v 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_v 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 $Q_v = \beta/H$, i.e., Q_v 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_v 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 flows⁹ (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:

$$\frac{\partial}{\partial t} \int_0^\infty E dk = 0, \quad \frac{\partial}{\partial t} \int_0^\infty k^2 E dk = 0, \quad (5.205)$$

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,

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

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

$$\frac{\partial k_1}{\partial t} < 0. \quad (5.207)$$

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 $\lambda_1 = (g'H_1H_2/Hf^2)^{1/2}$ is a critical parameter. When the initial, turbulent scale L is small compared to λ_1 , 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 $\lambda_1 k > 1$.

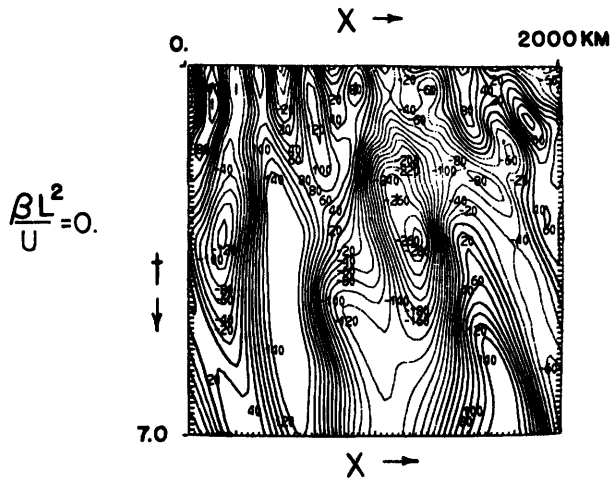


Figure 5.15A Contour diagrams for ψ as function of longitude and time for Rhines's numerical barotropic experiment with initially turbulent field without β . Cascade to large scales occurs as time increases.

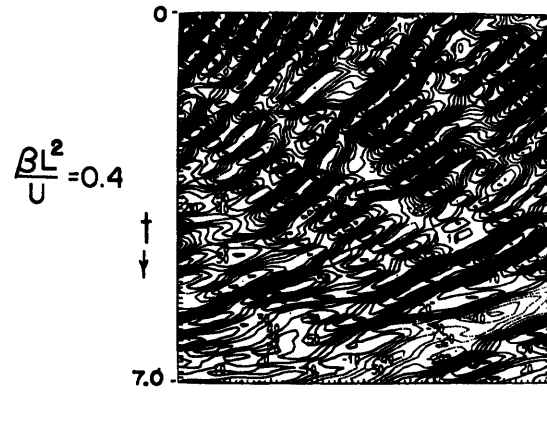


Figure 5.15B Same experiment with β . Westward propagation increases with time as cascade to larger scales occurs. Cascade is halted as $k \rightarrow \sqrt{\beta/2U}$.

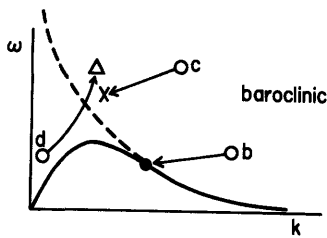


Figure 5.16A Frequency-wavenumber diagram for baroclinic modes. Solid curve shows linear baroclinic Rossby wave, dashed curve linear barotropic Rossby wave. For turbulent motions frequency is rms velocity U times k . Initial field of small-scale motions evolves by nonlinear interactions toward larger 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 \times); 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 .

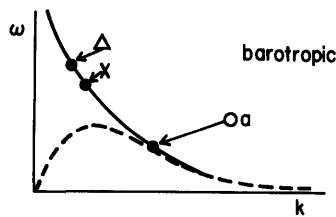


Figure 5.16B From point a nonlinear interactions of initially small-scale turbulent field produce red cascade until frequency reaches linear barotropic value (solid curve), when westward propagation disperses the energy to reduce nonlinear interactions. Same development from points \times and Δ after 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 \times) 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 \times 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 $k \approx \lambda_i^{-1}$, 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_1 , 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 $\Delta U/\lambda_1^2\beta < 1$ 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_1 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 $2\lambda_1$. 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 $-f dh/h dt$. Non-dimensional 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_1 was 40 km, so the topography cutoff scale was at the deformation radius.

A first experiment started off with turbulent eddies of scale $k \sim \lambda_1^{-1}$ 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

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

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

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

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

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

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

$$\begin{aligned} \frac{dq_1}{dt} &= \frac{\hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau}_0}{H_1} + A \nabla^4 \psi_1, \\ \frac{dq_2}{dt} &= -K \frac{\nabla^2 \psi_2}{H_2} + A \nabla^4 \psi_2, \end{aligned} \quad (5.211)$$

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

$$\begin{aligned} \frac{\partial \bar{q}_1}{\partial t} + \bar{\mathbf{v}}_1 \cdot \nabla \bar{q}_1 + \overline{\mathbf{v}'_1 \cdot \nabla q'_1} &= \frac{\hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau}_0}{H_1} + A \nabla^4 \bar{\psi}_1, \\ \frac{\partial \bar{q}_2}{\partial t} + \bar{\mathbf{v}}_2 \cdot \nabla \bar{q}_2 + \overline{\mathbf{v}'_2 \cdot \nabla q'_2} &= -\frac{K \nabla^2 \bar{\psi}_2}{H_2} + A \nabla^4 \bar{\psi}_2, \end{aligned} \quad (5.212)$$

where

$$\begin{aligned} \bar{q}_1 &= \nabla^2 \bar{\psi}_1 + f + \frac{f_0}{H_1} \bar{\eta}_2, & q'_1 &= \nabla^2 \psi'_1 + f_0 \eta'_2 / H_1, \\ \bar{q}_2 &= \nabla^2 \bar{\psi}_2 + f - \frac{f_0 (\bar{\eta}_2 - \eta_3)}{H_2}, & q'_2 &= \nabla^2 \psi'_2 - f_0 \eta'_2 / H_2. \end{aligned} \quad (5.213)$$

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

$$\begin{aligned} \bar{\mathbf{v}}_1 \cdot \nabla \bar{q}_1 + \overline{\mathbf{v}'_1 \cdot \nabla \nabla^2 \psi'_1} + \frac{f_0}{H_1} \overline{\mathbf{v}'_1 \cdot \nabla \eta'_2} + A \nabla^6 \bar{\psi}_1 - \frac{\hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau}_0}{H_1} &= 0, \\ \bar{\mathbf{v}}_2 \cdot \nabla \bar{q}_2 + \overline{\mathbf{v}'_2 \cdot \nabla \nabla^2 \psi'_2} - \frac{f_0}{H_2} \overline{\mathbf{v}'_2 \cdot \nabla \eta'_2} + A \nabla^6 \bar{\psi}_2 + \frac{K \nabla^2 \bar{\psi}_2}{H_2} &= 0, \end{aligned} \quad (5.214)$$

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

$$\begin{aligned} \overline{H_1 \mathbf{v}'_1 \cdot \nabla \nabla^2 \psi'_1} + \overline{f_0 \mathbf{v}'_1 \cdot \nabla \eta'_2} + \overline{H_1 A \nabla^6 \bar{\psi}_1} - \overline{\hat{\mathbf{k}} \cdot \nabla \times \boldsymbol{\tau}_0} &= 0 \\ U_1 + U_2 + U_3 + U_4 &= 0, \\ \overline{H_2 \mathbf{v}'_2 \cdot \nabla \nabla^2 \psi'_2} - \overline{f_0 \mathbf{v}'_2 \cdot \nabla \eta'_2} + \overline{H_2 A \nabla^6 \bar{\psi}_2} + \overline{K \nabla^2 \bar{\psi}_2} &= 0 \\ L_1 + L_2 + L_3 + L_4 &= 0, \end{aligned} \quad (5.215)$$

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

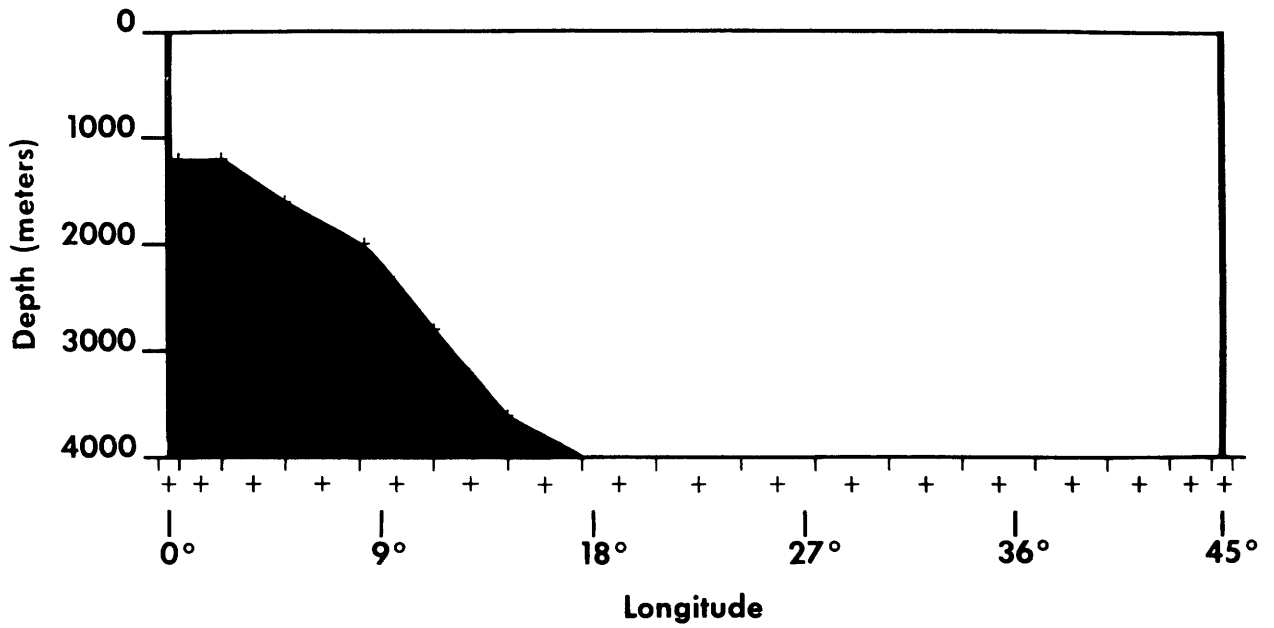


Figure 5.17A Topography used by Holland (1973) in ocean circulation calculation.

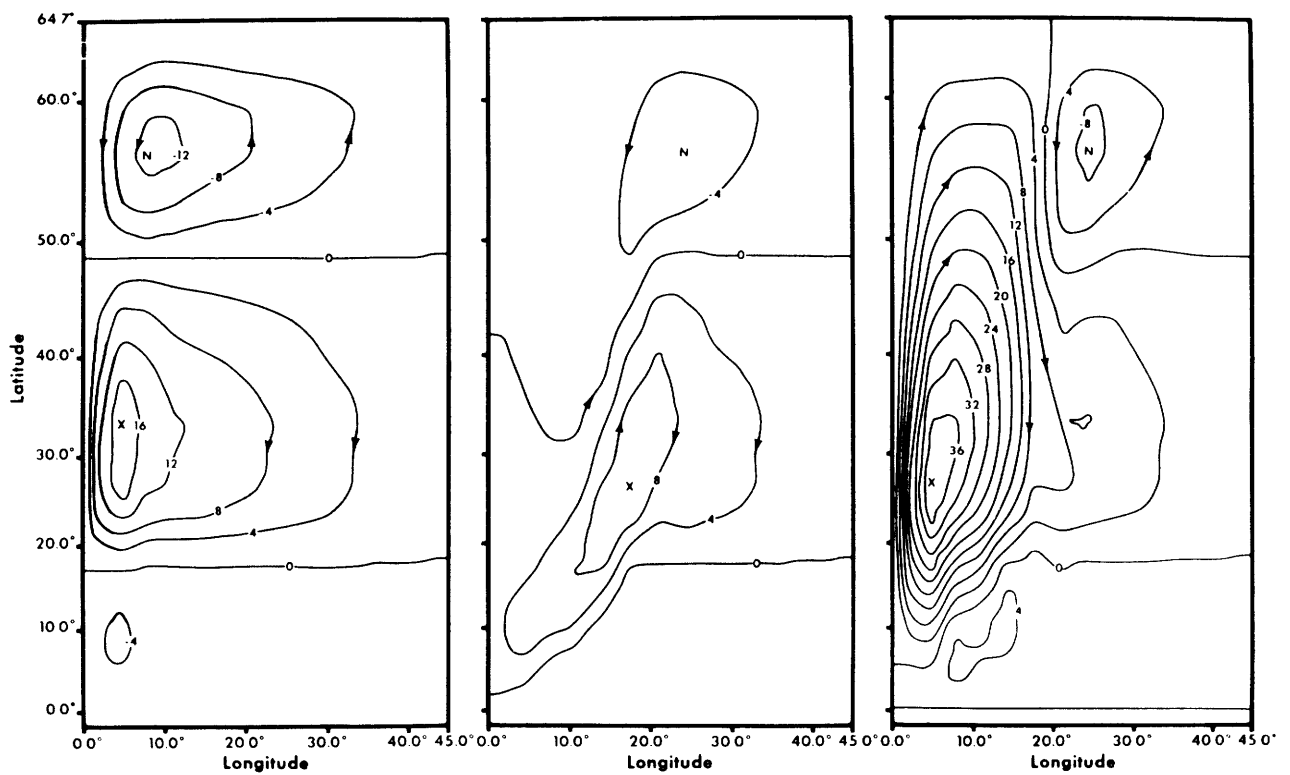


Figure 5.17B Horizontal transport streamfunction for (left) baroclinic ocean with no topography, (middle) barotropic ocean with topography, (right) baroclinic ocean with topography.

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

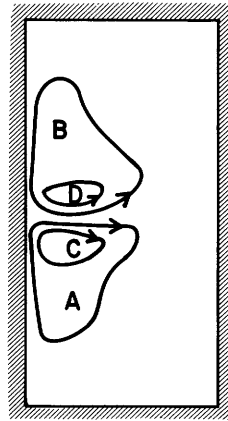


Figure 5.18A Circulation gyres generated in top layer of ocean model with cyclonic wind stress in north, equal-strength anticyclonic wind stress in south. Gyres *C* and *D* contain recirculating water.

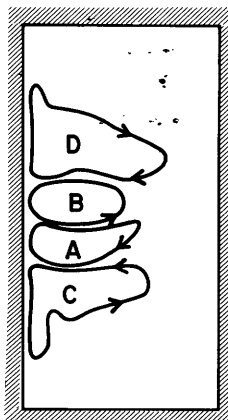


Figure 5.18B Lower-layer gyres near mid-latitude have same sign as those of upper layer. Other gyres opposite.

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 non-linear 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 $B = NH/f_0 L$, 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 $\lambda_i^2 K^2$ here means that we measure the wave scale K relative to the internal radius of deformation λ_i .

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.