

Walter Munk

9.1 Introduction

Gravity waves in the ocean's interior are as common as waves at the sea surface—perhaps even more so, for no one has ever reported an interior calm.

Typical scales for the internal waves are kilometers and hours. Amplitudes are remarkably large, of the order of 10 meters, and for that reason internal waves are not difficult to observe; in fact they are hard *not* to observe in any kind of systematic measurements conducted over the appropriate space-time scales. They show up also where they are not wanted: as short-period fluctuations in the vertical structure of temperature and salinity in intermittent hydrocasts.

I believe that Nansen (1902) was the first to report such fluctuations;¹ they were subsequently observed on major expeditions of the early nineteenth century: the *Michael Sars* expedition in 1910, the *Meteor* expeditions in 1927 and 1938, and the *Snellius* expedition in 1929–1930. [A comprehensive account is given in chapter 16 of Defant (1961a)]. In all of these observations the internal waves constitute an undersampled small-scale noise that is then “aliased” into the larger space-time scales that are the principal concern of classical oceanography.

From the very beginning, the fluctuations in the hydrocast profiles were properly attributed to internal waves. The earliest theory had preceded the observations by half a century. Stokes (1847) treated internal waves at the interface between a light fluid overlaying a heavy fluid, a somewhat minor extension of the theory of surface waves. The important extension to the case of a vertical mode structure in *continuously* stratified fluids goes back to Rayleigh (1883). But the discreteness in the vertical sampling by hydrocasts led to an interpretation in terms of just the few gravest modes, with the number of such modes increasing with the number of sample depths (giving j equations in j unknowns). And the discreteness in sampling time led to an interpretation in terms of just a few discrete frequencies, with emphasis on tidal frequencies.

The development of the bathythermograph in 1940 made it possible to repeat soundings at close intervals. Ufford (1947) employed three vessels from which bathythermograph lowerings were made at 2-minute intervals! In 1954, Stommel commenced three years of temperature observations offshore from Castle Harbor, Bermuda, initially at half-hour intervals, later at 5-minute intervals.² Starting in 1959, time series of isotherm depths were obtained at the Navy Electronics Laboratory (NEL) oceanographic tower off Mission Beach, California, using isotherm followers (Lafond, 1961) installed in a 200-m triangle (Cox, 1962).

By this time oceanographers had become familiar with the concepts of continuous spectra (long before

routinely applied in the fields of optics and acoustics), and the spectral representation of surface waves had proven very useful. It became clear that internal waves, too, occupy a frequency continuum, over some six octaves extending from inertial to buoyant frequencies. [The high-frequency cutoff had been made explicit by Groen (1948).] With regard to the vertical modes, there is sufficient energy in the higher modes that for many purposes the discrete modal structure can be replaced by an equivalent three-dimensional continuum.

We have already referred to the measurements by Ufford and by Lafond at horizontally separated points. Simultaneous current measurements at vertically separated points go back to 1930 (Ekman and Helland-Hansen, 1931). In all these papers there is an expression of dismay concerning the lack of resemblance between measurements at such small spatial separations of oscillations with such long periods. I believe (from discussions with Ekman in 1949) that this lack of coherence was the reason why Ekman postponed for 23 years (until one year before his death) the publication of "Results of a Cruise on Board the 'Armauer Hansen' in 1930 under the Leadership of Björn Helland-Hansen" (Ekman, 1953). But the decorrelation distance is just the reciprocal of the bandwidth; waves separated in wavenumber by more than Δk interfere destructively at separations exceeding $(\Delta k)^{-1}$. The small observed coherences are simply an indication of a large bandwidth.

The search for an analytic spectral model to describe the internal current and temperature fluctuations goes back over many years, prompted by the remarkable success of Phillips's (1958) saturation spectrum for surface waves. I shall mention only the work of Murphy and Lord (1965), who mounted temperature sensors in an unmanned submarine at great depth. They found some evidence for a spectrum depending on scalar wavenumber as $k^{-5/3}$, which they interpreted as the inertial subrange of homogeneous, isotropic turbulence. But the inertial subrange is probably not applicable (except perhaps at very small scales), and the fluctuations are certainly not homogeneous and not isotropic.

Briscoe (1975a) has written a very readable account of developments in the early 1970s. The interpretation of multipoint coherences in terms of bandwidth was the key for a model spectrum proposed by Garrett and Munk (1972b). The synthesis was purely empirical, apart from being guided by dimensional considerations and by not violating gross requirements for the finiteness of certain fundamental physical properties. Subsequently, the model served as a convenient "straw-man" for a wide variety of moored, towed and "dropped" experiments, and had to be promptly modified [Garrett and Munk (1975), which became known

as GM75 in the spirit of planned obsolescence]. There have been further modifications [see a review paper by Garrett and Munk (1979)]; the most recent version is summarized at the end of this chapter.

The best modern accounts on internal waves are by O. M. Phillips (1966b), Phillips (1977a), and Turner (1973a). Present views of the time and space scales of internal waves are based largely on densely sampled moored, towed, and dropped measurements. The pioneering work with moorings was done at site D in the western North Atlantic (Fofonoff, 1969; Webster, 1968). Horizontal tows of suspended thermistor chains (Lafond, 1963; Charnock, 1965) were followed by towed and self-propelled isotherm-following "fishes" (Katz, 1973; McKean and Ewart, 1974). Techniques for dropped measurements were developed along a number of lines: rapidly repeated soundings from the stable platform FLIP by Pinkel (1975), vertical profiling of currents from free-fall instruments by Sanford (1975) and Sanford, Drever, and Dunlap (1978), and vertical profiling of temperature from a self-contained yo-yoing capsule by Cairns and Williams (1976). The three-dimensional IWEX (internal wave experiment) array is the most ambitious to date (Briscoe, 1975b). These experiments have served to determine selected parameters of model spectra; none of them so far, not even IWEX, has been sufficiently complete for a straightforward and unambiguous transform into the multi-dimensional (ω, \mathbf{k}) -spectrum. The FLIP measurements come closest, giving an objective spectrum in the two dimensions ω, k_z , with fragmentary information on k_x, k_y . Otherwise only one-dimensional spectra can be evaluated from any single experiment, and one is back to model testing. Yet in spite of these observational shortcomings, there is now evidence for some degree of universality of internal wave spectra, suggesting that these spectra may be shaped by a saturation process (the interior equivalent of whitecaps), rather than by external generation processes.

Internal waves have surface manifestations consisting of alternate bands of roughened and smooth water (Ewing, 1950; Hughes, 1978), and these appear to be visible from satellites (figure 9.1). High-frequency sonar beams are a powerful tool for measuring internal wave related processes in the upper oceans (figures 9.2, 9.3). The probing of the deep ocean interior by acoustics is ultimately limited by scintillations due to internal waves (Flatté et al., 1979; Munk and Wunsch, 1979) just as the "diffraction-limited" telescope has its dimensions set by the small-scale variability in the upper atmosphere.

It will be seen that internal waves are a lively subject. The key is to find the connections between internal waves and other ocean processes. The discovery of ever finer scales, down to the scale of molecular processes,

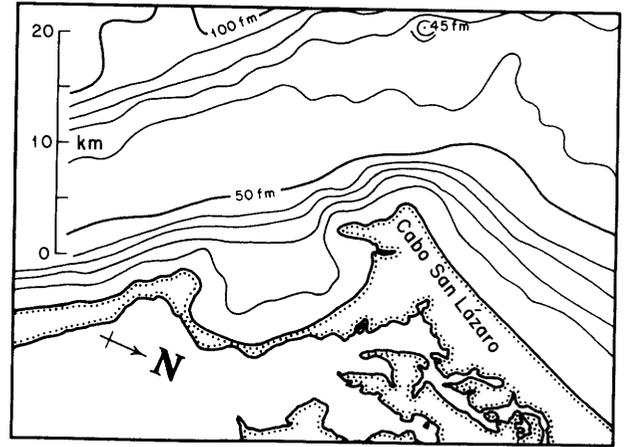
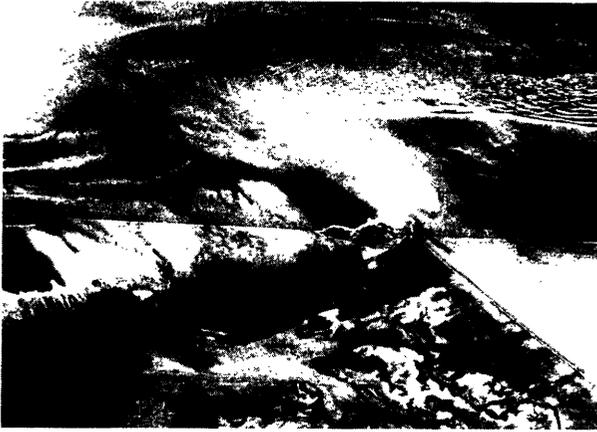


Figure 9.1 SEASAT synthetic aperture radar image off Cabo San Lázaro, Baja California (24°48'N, 112°18'W) taken on 7 July 1978. Scale of image nearly matches that of bathymetric

area. The pattern in the right top area is most likely formed by internal waves coming into the 50 fathom line. (I am indebted to R. Bernstein for this figure.)

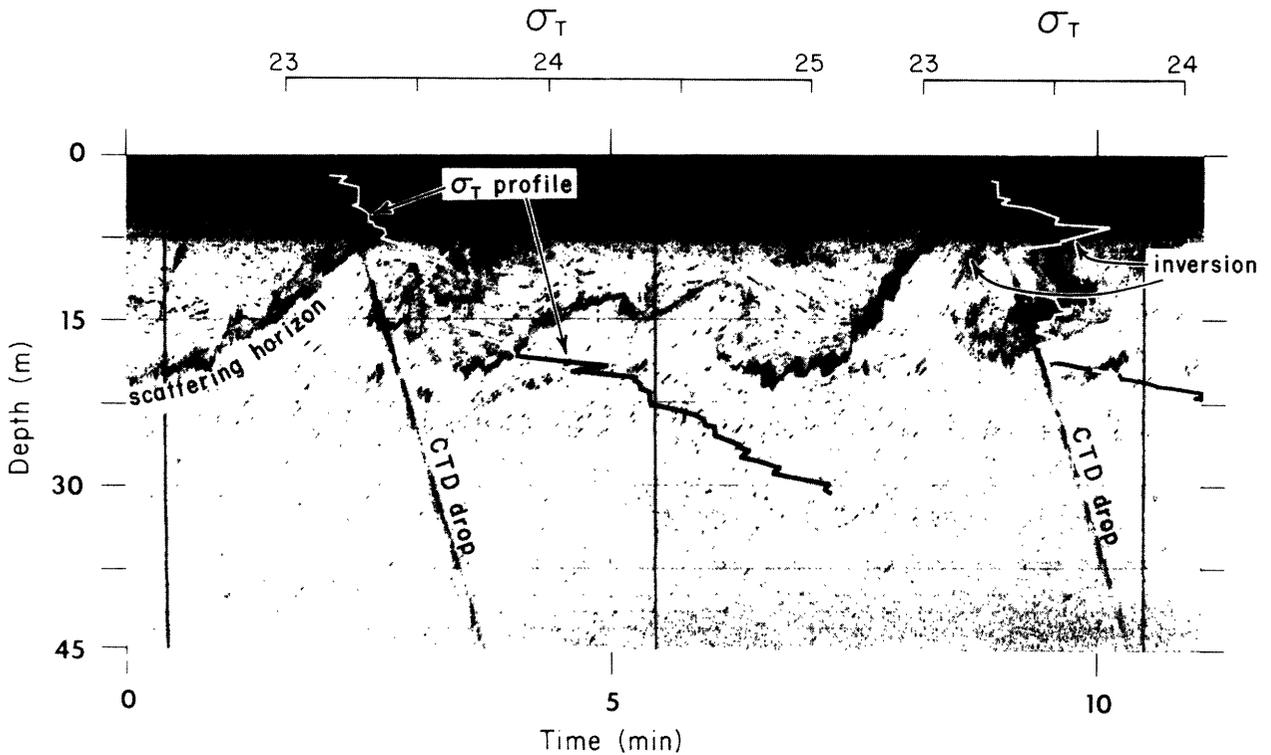


Figure 9.2 The water column is insonified with a narrow downward sonar beam of 200 kHz (wavelength 0.75 cm). The dark band is presumably a back-scattering layer convoluted by shear instabilities. In a number of places the instabilities have created density inversions. This is confirmed by the two σ_T -profiles. The acoustic reflection from the sinking CTD along the steeply slanting lines shows the depth-time history of the σ_T -profiles. The profiling sound source was suspended from a drifting ship. The horizontal distance between overturning events was estimated to be 60–70 m. (I am indebted to Marshall Orr of Woods Hole Oceanographic Institution for this figure; see Haury, Briscoe, and Orr, 1979.)

has been a continuing surprise to the oceanographic community for 40 years. Classical hydrographic casts employed reversing (Nansen) bottles typically at 100-m intervals in the upper oceans beneath the thermocline, and half-kilometer intervals at abyssal depths. Only the gross features can be so resolved. Modern sounding instruments (BT, STD, CTD) demonstrated a temperature and salinity³ fine structure down to meter scales. An early clue to microstructure was the steppy traces on the smoked slides of bathythermographs. These steps were usually attributed to "stylus stiction," and the instruments suitably repaired.

Free-fall apparatus sinking slowly ($\sim 0.1 \text{ m s}^{-1}$) and employing small, rapid-response ($\sim 0.01 \text{ s}$) transducers, subsequently resolved the structure down to centimeter scales and beyond. The evolving terminology

| | |
|------------------|----------------------------|
| gross structure: | larger than 100 m vertical |
| fine structure: | 1 m to 100 m vertical |
| microstructure: | less than 1 m vertical |

is then largely based on what could be resolved in a given epoch (see chapter 14). The fine-structure measurements of temperature and salinity owe much of their success to the evolution of the CTD (Brown, 1974). The pioneering microscale measurements were done by Woods (1968a) and by Cox and his collaborators (Gregg and Cox, 1972; Osborn and Cox, 1972). Measurements of velocity fine structure down to a few meters have been accomplished by Sanford (1975) and Sanford, Drever, and Dunlap (1978). Osborn (1974, 1980) has resolved the velocity microstructure between 40 and 4 cm. Evidently velocity and temperature structure have now been adequately resolved right down to the scales for which molecular processes become dominant. At these scales the dissipation of energy and mean-square temperature gradients is directly proportional to the *molecular* coefficients of viscosity and thermal diffusivity. The dissipation scale for salinity is even smaller (the haline diffusivity is much smaller than the thermal diffusivity) and has not been adequately resolved. The time is drawing near when we shall record the entire fine structure and microstructure scales of temperature, salinity and currents [and hence of the buoyancy frequency $N(z)$ and of Richardson number $Ri(z)$] from a single free-fall apparatus.

Perhaps the discovery of very fine scales could have been anticipated. There is an overall ocean balance between the generation and dissipation of mean-square gradients. Eckart (1948) refers to the balancing processes as *stirring* and *mixing*. Garrett (1979) has put it succinctly: "Fluctuations in ocean temperature produced by surface heating and cooling, and in salinity due to evaporation, precipitation, run-off and freezing, are stirred into the ocean by permanent current systems and large scale eddies." Mixing ultimately occurs

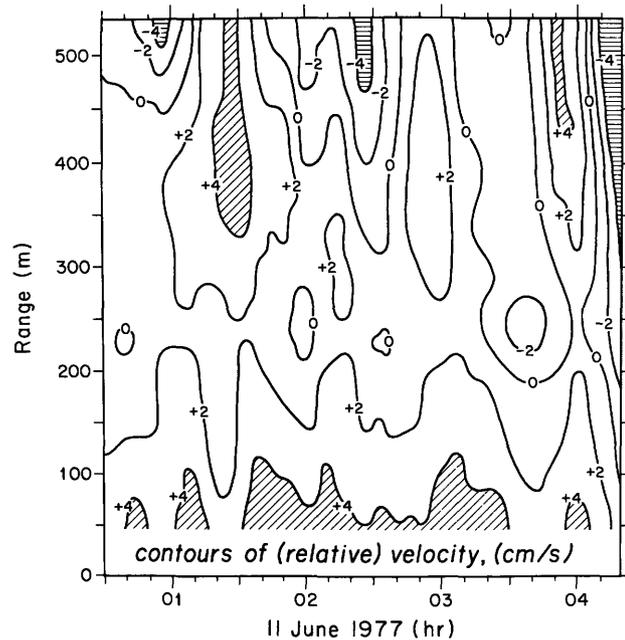


Figure 9.3 Measurements of Doppler vs. range were made at 2-minute intervals with a quasi-horizontal 88-kHz sound beam mounted on FLIP at a depth of 87 m. Bands of alternating positive and negative Doppler (in velocity contours) are the result of back scatter from particles drifting toward and away from the sound source (the mean drift has been removed). The velocities are almost certainly associated with internal wave-orbital motion. The range-rate of positive or negative bands gives the appropriate projection of phase velocity. The measurements are somewhat equivalent to successive horizontal tows at 3000 knots! (I am indebted to Robert Pinkel of Scripps Institution of Oceanography for this figure.)

through dissipation by "molecular action on small-scale irregularities produced by a variety of processes." The microstructure (where the mean-square gradients largely reside) are then a vital component of ocean dynamics. This leaves open the question whether mixing is important throughout the ocean, or whether it is concentrated at ocean boundaries and internal fronts, or in intense currents (an extensive discussion may be found in chapter 8).

What are the connections between internal waves and small-scale ocean structure? Is internal wave breaking associated with ocean microstructure? Is there an associated flux of heat and salt, and hence buoyancy? Does the presence of internal waves in a shear flow lead to an enhanced momentum flux, which can be parameterized in the form of an eddy viscosity? What are the processes of internal-wave generation and decay? I feel that we are close to having these puzzles fall into place (recognizing that oceanographic "break-throughs" are apt to take a decade), and I am uncomfortable with attempting a survey at this time.

Forty years ago, internal waves played the role of an attractive nuisance: attractive for their analytical ele-

gance and their accessibility to a variety of experimental methods, a nuisance for their interference with what was then considered the principal task of physical oceanography, namely, charting the "mean" density field. Twenty years from now I expect that internal waves will be recognized as being intimately involved with the vertical fluxes of heat, salt, and momentum, and so to provide a vital link in the understanding of the mean fields of mass and motion in the oceans.

9.1.1 Preview of This Chapter

We start with the traditional case of a two-layer ocean, followed by a discussion of continuous stratification: constant buoyancy frequency N , N decreasing with depth, a maximum N (thermocline), a double maximum. Conditions are greatly altered in the presence of quite moderate current shears. Short (compliant) internal waves have phase velocities that are generally slower than the orbital currents associated with the long (intrinsic) internal waves, and thus are subject to critical layer processes. There is further nonlinear coupling by various resonant interactions.

Ocean fine structure is usually the result of internal-wave straining, but in some regions the fine structure is dominated by intrusive processes. Microstructure is concentrated in patches and may be the residue of internal wave breaking. Little is known about the breaking of internal waves. Evidently, there are two limiting forms of instability leading to breaking: advective instability and shear instability.

The chapter ends with an attempt to estimate the probability of wave breaking, and of the gross vertical mixing and energy dissipation associated with these highly intermittent events. An important fact is that the Richardson number associated with the internal wave field is of order 1. Similarly the wave field is within a small numerical factor of advective instability. Doubling the mean internal wave energy can lead to a large increase in the occurrence of breaking events; halving the wave energy could reduce the probability of breaking to very low levels. This would have the effect of maintaining the energy level of internal waves within narrow limits, as observed. But the analysis is based on some questionable assumptions, and the principal message is that we do not understand the problem.

9.2 Layered Ocean

We start with the conventional discussion of internal waves at the boundary between two fluids of different density. The configuration has perhaps some application to the problem of long internal waves in the thermocline, and of short internal waves in a stepwise fine structure.

Following Phillips (1977a), this can be treated as a limiting case of a density transition from ρ_u above $z = -h$ to ρ_l beneath $z = -h$, with a transition thickness δh (figure 9.4). The vertical displacement $\zeta(z)$ has a peak at the transition, and the horizontal velocity $u(z)$ changes sign, forming a discontinuity (vortex sheet) in the limit $\delta h \rightarrow 0$. For the second mode (not shown), $\zeta(z)$ changes sign within the transition layer and $u(z)$ changes sign twice; this becomes unphysical in the limit $\delta h \rightarrow 0$. For higher modes the discontinuities are even more pathological, and so a two-layer ocean is associated with only the gravest internal mode.

For the subsequent discussion it is helpful to give a sketch of how the dependent variables are usually derived and related. The unknowns are u, v, w, p (after eliminating the density perturbation), where p is the departure from hydrostatic pressure. The four unknowns are determined by the equations of motion and continuity (assuming incompressibility). The linearized x, y equations of motion are written in the traditional f -plane; for the vertical equation it is now standard [since the work of Eckart (1960)] to display the density stratification in terms of the buoyancy (or Brunt-Väisälä) frequency

$$N(z) = \left\{ -\frac{g}{\rho} \left[\frac{d\rho}{dz} - \left(\frac{d\rho}{dz} \right)_{\text{adiabatic}} \right] \right\}^{1/2}, \quad (9.1)$$

thus giving

$$\frac{\partial w}{\partial t} = \frac{1}{\rho_0} \frac{\partial p}{\partial z} - N^2 \zeta = 0.$$

The last term will be recognized as the buoyancy force $-g \delta \rho / \rho_0$ of a particle displaced upwards by an amount $\zeta = \int w dt$.

For propagating waves of the form $\zeta(z) \exp i(kx - \omega t)$ the equations can be combined (Phillips, 1977a, §5.2 and §5.7) into

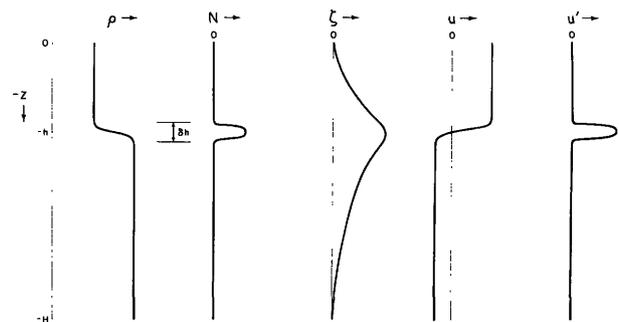


Figure 9.4 A sharp density transition from ρ_u to ρ_l takes place between the depths $-z = h - \frac{1}{2} \delta h$ and $-z = h + \frac{1}{2} \delta h$. This is associated with a delta-like peak in buoyancy frequency $N(z)$. Amplitudes of vertical displacement $\zeta(z)$, horizontal velocity $u(z)$, and shear $u'(z) = du/dz$ are sketched for the gravest internal wave mode.

$$\frac{d^2\zeta}{dz^2} + k^2 \frac{N^2(z) - \omega^2}{\omega^2 - f^2} \zeta = 0. \quad (9.2)$$

The linearized boundary conditions are $\zeta = 0$ at the surface and bottom.

A simple case is that of $f = 0$ and $N = 0$ outside the transition layer. We have then

$$\zeta_u = A \sinh kz,$$

$$\zeta_l = B \sinh k(z + H),$$

above and below the transition layer, respectively. The constants A and B are determined by patching the vertical displacement at the transition layer:

$$\zeta_u = \zeta_l = a \quad \text{at } z = -h.$$

The dispersion relation is found by integrating Eq. (9.2) across the transition layer:

$$\begin{aligned} \zeta'_u - \zeta'_l &= -k^2 \zeta \int_{\delta h} dz \frac{N^2(z) - \omega^2}{\omega^2} \\ &\approx -\frac{k^2 \zeta}{\omega^2} \left(g \frac{\delta \rho}{\rho} - \omega^2 \delta h \right) \quad \text{at } z = -h, \end{aligned}$$

where $\zeta' \equiv d\zeta/dz$. In the limit of small $k \delta h$, that is, for waves long compared to the transition thickness, the foregoing equations lead to the dispersion relation

$$\omega^2 = \frac{g[\delta \rho/\rho]k}{\coth kh + \coth k(H-h)}.$$

For a lower layer that is deep as compared to a wavelength, the denominator becomes $\coth k(h-H) + 1$. If the upper layer is also deep, it becomes $1 + 1$, and

$$\omega^2 = \frac{1}{2} gk \frac{\delta \rho}{\rho} = \frac{1}{2} gk \frac{\rho_l - \rho_u}{\frac{1}{2}(\rho_l + \rho_u)}.$$

As $\rho_u \rightarrow 0$, $\omega^2 \rightarrow gk$, which is the familiar expression for surface waves in deep water.

The case of principal interest here is that of an isolated density transition $\delta \rho \ll \rho$ and $k \delta h \ll 1$. Then $\omega^2 = \frac{1}{2} gk \delta \rho/\rho$. The vertical displacement is a maximum at the transition and dies off with distance δz from the transition as $a \exp(-k|\delta z|)$.

A question of interest is the variation of Richardson number across the transition layer. We know from the work of Miles and Howard [see Miles (1963)] that for a transition $\rho(z)$ and a steady $u(z)$ of the kind shown in figure 9.4, the flow becomes unstable to disturbances of length scale δh if $Ri < \frac{1}{4}$. I find it convenient to refer to the *root-reciprocal* Richardson number

$$Ri^{-1/2} = |u'/N|,$$

so that large values imply large instabilities (as for Reynolds numbers); the critical value is $|u'/N| = 2$. One would think offhand that $|u'/N|$ is a minimum at the

transition where N reaches a maximum, but just the opposite is true. To prove this, we use the condition of incompressibility, $iku(z) - i\omega \zeta' = 0$, and equation (9.2) to obtain

$$u'(z) = \frac{\omega}{k} \zeta'' = -\frac{N^2(z) - \omega^2}{\omega} ka, \quad (9.3)$$

and so $u' \sim N^2$ for small ω/N ; accordingly u'/N varies as N . Thus the layers of largest gravitational stability (largest N) are also the layers of largest shear instability (largest $|u'/N|$).

9.3 Continuously Stratified Ocean

The simplest case is that of constant N . The solution to (9.2) is

$$\zeta(z) = a \sin mz, \quad m^2 = k^2 \frac{N^2 - \omega^2}{\omega^2 - f^2} \quad (9.4)$$

with m so chosen that ζ vanishes at $z = -H$. Solving for ω^2 ,

$$\omega_j^2 = \frac{k^2 N^2 + m_j^2 f^2}{m_j^2 + k^2}, \quad m_j H = j\pi, \quad j = 1, 2, \dots \quad (9.5)$$

This dispersion relation is plotted in figure 9.5. The vertical displacements for the first and third mode are shown in figure 9.6. Very high modes (and the ocean is full of them) in the deep interior are many wavelengths removed from the boundaries, and we can expect the waves to be insensitive to the precise configuration of top and bottom. The discrete dispersion $\omega_j(k)$ is then replaced by an equivalent continuous dispersion $\omega(k, m)$.

The standard expressions for the particle velocities u, w and the group velocities \mathbf{c}_g with components $\partial\omega/\partial k, \partial\omega/\partial m$ as functions of the propagation vector

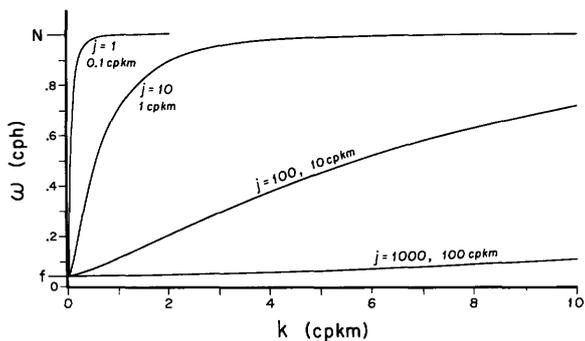


Figure 9.5 The dispersion $\omega_j(k)$ [equation (9.5)], for modes $j = 1, 10, 100, 1000$, corresponding to vertical wavenumbers $m = 0.1, 1, 10, 100$ cpkm in an ocean of depth 5 km. The inertial frequency is taken at $f = 0.0417$ cph (1 cpd), and the buoyancy frequency at $N = 1$ cph.

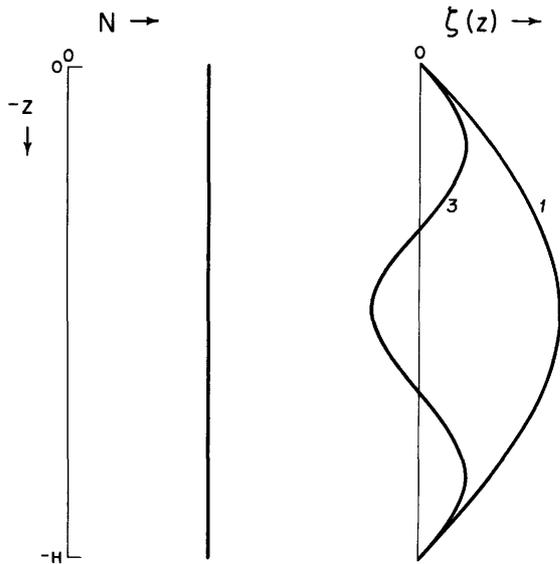


Figure 9.6 Vertical displacements $\zeta(z)$ in a constant- N ocean, for modes $j = 1$ and $j = 3$.

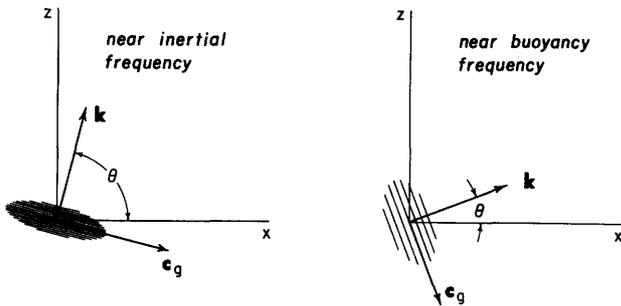


Figure 9.7 The wavenumber vector $\mathbf{k} = (k, m)$ and group velocity \mathbf{c}_g near the inertial frequency ($\omega = f + \epsilon$) and near the buoyancy frequency ($\omega = N - \epsilon$), respectively. A packet of wave energy is projected on the (x, z) -plane. Crests and troughs in the wave packet are in a plane normal to \mathbf{k} , and travel with phase velocity c in the direction \mathbf{k} . The wave packet travels with group velocity \mathbf{c}_g at right angles to \mathbf{k} , thus sliding sideways along the crests and troughs. The particle velocity \mathbf{u} (not shown) is in the planes at right angles to \mathbf{k} .

$\mathbf{k} = (k, m)$ are easy to derive, but hard to visualize. Consider a wave packet (figure 9.7) with crests and troughs along planes normal to the paper and inclined with respect to the (x, z) -axis as shown. The phase velocity is in the direction \mathbf{k} normal to the crests, but the group velocity \mathbf{c}_g is parallel to the crests, and the wave packet slides sideways. \mathbf{k} is inclined to the horizontal by

$$\tan \theta = \frac{m}{k} = \left(\frac{N^2 - \omega^2}{\omega^2 - f^2} \right)^{1/2} \quad (9.6a)$$

and so the angle is steep for inertial waves ($\omega = f + \epsilon$) and flat for buoyancy waves ($\omega = N - \epsilon$). The energy packet is propagated horizontally for inertial waves, and vertically for buoyancy waves, but the group velocity goes to zero at both limits.

The flow $\mathbf{u} = (u, w)$ takes place in the plane of the crest and troughs. For inertial waves, particles move in horizontal circles. The orbits become increasingly elliptical with increasing frequency, and for buoyancy waves the particle orbits are linear along the z -axis, in the direction of \mathbf{c}_g . The wavenumber \mathbf{k} is always normal to both \mathbf{c}_g and \mathbf{u} . [The nonlinear field accelerations $(\mathbf{u} \cdot \nabla)\mathbf{u}$ vanish for an isolated elementary wave train, leading to the curiosity that the linear solution is an exact solution.] Readers who find it difficult to visualize (or believe) these geometric relations should refer to the beautiful laboratory demonstrations of Mowbray and Rarity (1967).

It is not surprising, then, that internal waves will do unexpected things when reflected from sloping boundaries. The important property is that the inclination θ relative to the x -axis depends only on frequency [equation (9.6a)]. Since frequency is conserved upon reflection, incident and reflected θ must be symmetric with respect to a level surface rather than with respect to the reflecting surface. At the same time the flow \mathbf{u} for the combined incident and reflected wave must be parallel to the reflecting boundary. For a given ω , there is a special angle for which the orbital flow is parallel to the boundary. This requires that the boundary be inclined at a slope

$$\tan \beta(z) = \tan(90^\circ - \theta) = \left[\frac{\omega^2 - f^2}{N^2(z) - \omega^2} \right]^{1/2}. \quad (9.6b)$$

It can be shown that for slopes steeper than β , the energy of "shoreward" traveling internal waves is reflected "seaward": for slopes of less than β , the energy is forward reflected. Repeated reflections in a wedge-shaped region such as the ocean on the continental slope can lead to an accumulation of energy at ever smaller scales (Wunsch, 1969). For a given slope, we can expect an amplification of the internal waves at the frequency ω determined by (9.6b). Wunsch (1972b) has suggested that a peak in the spectrum of temperature fluctuations measured southeast of Bermuda

could be so explained. Pertinent values are $N = 2.6$ cph, $f = 0.045$ cph, and $\beta \approx 13^\circ$. Equation (9.6b) gives $\omega = 0.59$ cph, in agreement with the observed spectral peak at 0.5 cph.

9.4 Turning Depths and Turning Latitudes

Figure 9.8 shows the situation for an ocean with variable $N(z)$. For frequencies that are less than N throughout the water column, the displacements are similar to those for constant N (figure 9.6) except that the positions of the maxima and zeros are displaced somewhat upward, and that the relative amplitudes are somewhat larger at depth. The important modification occurs for frequencies that exceed $N(z)$ somewhere within the water column. At the depths z_T where $\omega = N(z_T)$, called turning depths, we have the situation shown to the right in figure 9.8. Equation (9.2) is locally of the form $\zeta'' + z\zeta = 0$ where z is now a rescaled vertical coordinate relative to z_T . The solution (called an *Airy function*) has an inflection point at the turning depth (here $z = 0$), is oscillatory above the turning depth, and is exponentially damped beneath. The amplitudes are somewhat larger just above the turning depth than at greater distance, but nothing very dramatic happens.

The refraction of a propagating wave packet is illustrated in figure 9.9. As the packet moves into depths of diminishing $N(z)$ the crests and troughs turn steeper, and the direction of energy propagation becomes more nearly vertical. The waves are totally reflected at the turning depth z_T where $\omega = N(z_T)$. Modal solutions $\zeta_j(z) \times \exp i(kx - \omega_j t)$ with $\zeta_j(z)$ as illustrated in figure 9.8 can be regarded as formed by superposition of propagating waves with equal upward and downward energy transport. The wave energy remains trapped between the surface and the turning depth.

The common situation for the deep ocean is the main thermocline associated with a maximum in $N(z)$. Internal waves with frequencies less than this maximum are in a waveguide contained between upper and lower turning depths. For relatively high (but still trapped) frequencies the sea surface and bottom boundaries play a negligible role, and the wave solutions can be written in a simple form (Eriksen, 1978). The bottom boundary condition (9.5) for a constant- N ocean, e.g., $m_j H = j\pi$, $j = 1, 2, \dots$, is replaced in the WKB approximation by

$$m_j b = j\pi \left(\frac{N^2 - \omega^2}{N_0^2 - \omega^2} \right)^{1/2} \approx j\pi N/N_0, \quad (9.7)$$

where b is a representative thermocline (or stratification) scale. Equation (9.7) assures an exponential attenuation outside the waveguide. For the case of a double peak in $N(z)$ with maxima N_1 and N_2 , the internal wave energy is concentrated first at one thermocline, then

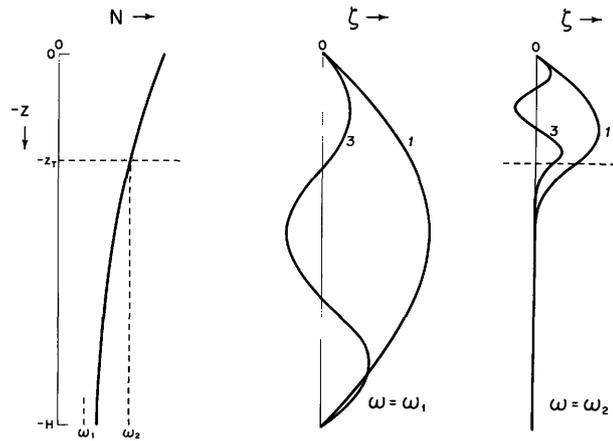


Figure 9.8 Vertical displacements $\zeta(z)$ in a variable- N ocean, for modes $j = 1$ and $j = 3$. ω_1 is taken to be less than $N(z)$ at all depths. ω_2 is less than $N(z)$ in the upper oceans above $z = -z_T$ only.

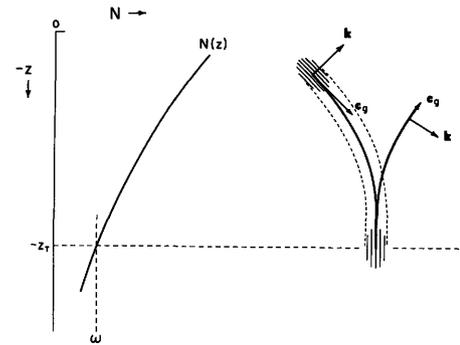


Figure 9.9 Propagation of a wave packet in a variable- $N(z)$ ocean without shear ($U = \text{constant}$). The turning depth z_T occurs when $\omega = N(z_T)$.

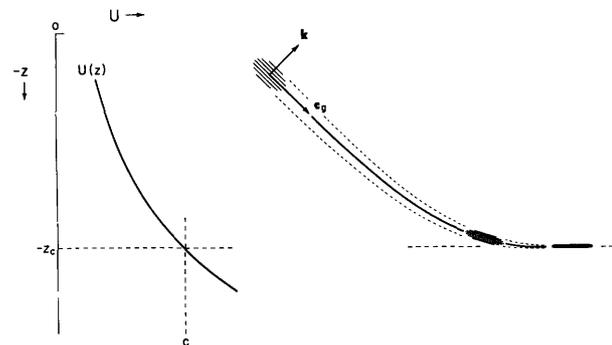


Figure 9.10 Propagation of a wave packet in a constant- N ocean with shear. The critical depth z_c occurs where $U = c(z_c)$.

the other, migrating up and down with a frequency $|N_1 - N_2|$ (Eckart, 1961). This is similar to the behavior of two loosely coupled oscillators. The quantum-mechanical analogy is that of two potential minima and the penetration of the potential barrier between them.

There is a close analogy between the constant- and variable- N ocean, and the constant- and variable- f ocean (the f -plane and β -plane approximations). For a fixed ω , the condition $\omega = f = 2\Omega \sin \phi_T$ determines the turning latitude ϕ_T . Eastward-propagating internal gravity waves have solutions of the form $\eta(y)\zeta(z) \exp i(kx - \omega t)$. The equation governing the local north-south variation is (Munk and Phillips, 1968)

$$\eta'' + y\eta = 0, \quad \eta'' = d^2\eta/dy^2,$$

where y is the poleward distance (properly scaled) from the turning latitude. This is in close analogy with the up-down variation near the turning depth, which is governed by

$$\zeta'' + z\zeta = 0, \quad \zeta'' = d^2\zeta/dz^2.$$

Thus $\eta(y)$ varies from an oscillatory to an exponentially damped behavior as one goes poleward across the turning latitude. Poleward-traveling wave packets are reflected at the turning latitude.

From an inspection of figure 9.7, it is seen that the roles of horizontal and vertical displacements are in-

terchanged in the $N(z)$ and $f(y)$ turning points. In the $N(z)$ case the motion is purely vertical; in the $f(y)$ case the motion is purely horizontal (with circular polarization).

It has already been noted that nothing dramatic is observed in the spectrum of vertical displacement (or potential energy) near $\omega = N$ —only a moderate enhancement, which can be reconciled to the behavior of the Airy function (Desaubies, 1975; Cairns and Williams, 1976). Similarly we might expect only a moderate enhancement in the spectrum of horizontal motion (or kinetic energy) near $\omega = f$. In fact, the spectrum is observed to peak sharply. If the horizontal motion is written as a sum of rotary components (Gonella, 1972), it is found that the peak is associated with negative rotation (clockwise in the northern hemisphere).

I have made a parallel derivation of the spectra at the two turning points (figure 9.11), assuming horizontally isotropic wave propagation within the entire equatorial waveguide. It turns out that the buoyancy peak is in fact much smaller than the inertial peak at moderate latitudes. But at very low latitude the inertial peak vanishes. This is in accord with the equatorial observations by Eriksen (1980). Fu (1980) gives an interesting discussion of the relative contributions to the spectral peak at the local inertial frequency $\omega = f_{\text{local}}$ from two processes: (1) local generation of resonant inertial

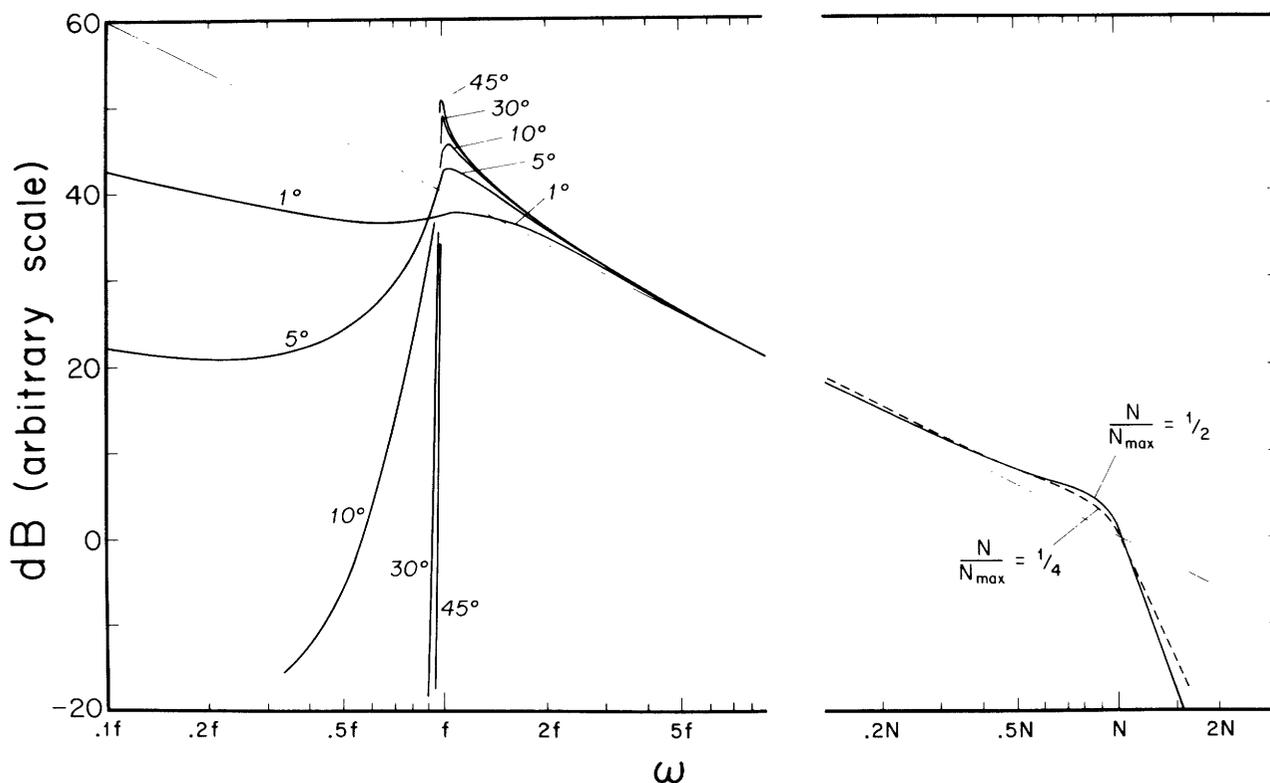


Figure 9.11 Enhancement of the kinetic-energy spectrum (left) and of the potential-energy spectrum (right) at the inertial and buoyancy frequencies, respectively. The inertial spec-

trum is drawn for latitudes 1°, 5°, 10°, 30°, 45°. The buoyancy spectrum is drawn for two depths, corresponding to $N = \frac{1}{2}, \frac{1}{4}$ times the maximum buoyancy frequency.

waves $\omega = f_{\text{local}}$; and (2) *remote* generation of waves of the same frequency $\omega = f_{\text{local}}$ at lower latitudes (where $f < f_{\text{local}}$). Figure 9.11 is drawn for case 2 under the assumption that the equatorial waveguide is filled with horizontally isotropic, freely propagating radiation. Take the curve marked 30° , say. Then for $\omega > f$ a station at lat. 30° is within the equatorial waveguide; for $\omega < f$ the spectrum is the result of evanescent extensions from a waveguide bounded by lower latitudes. Over rough topography and in regions of strong surface forcing, the case can be made for local generation of the inertial peak. It would seem that the buoyancy peak at mid-depth must always be associated with remote generation.

9.5 Shear

Internal waves are greatly modified by an underlying shear flow.⁴ A variable $U(z)$ can have a more traumatic effect on internal waves than a variable $N(z)$. For ready comparison with figure 9.9 showing the effect of a variable $N(z)$ on a traveling wave packet, we have sketched in figure 9.10 the situation for a wave packet traveling in the direction of an increasing $U(z)$. As the wave packet approaches the "critical depth" z_c where the phase velocity (in a fixed frame of reference) equals the mean flow, $c = U(z_c)$, the vertical wavenumber increases without limit (as will be demonstrated).

For the present purpose we might as well avoid additional complexities by setting $f = 0$. The theoretical starting point is the replacement of ∂_t by $\partial_t + U\partial_x + w\partial_z$ in the linearized equations of motion. The result is the Taylor-Goldstein equation [Phillips (1977a, p. 248)]:

$$\frac{d^2\zeta}{dz^2} + \left(\frac{N^2}{(U-c)^2} - \frac{U''}{U-c} - k^2 \right) \zeta = 0, \quad (9.8)$$

$$U'' = \frac{d^2U}{dz^2},$$

where c is the phase velocity in a fixed reference frame. (This reduces to

$$\frac{d^2\zeta}{dz^2} + k^2 \frac{N^2 - \omega^2}{\omega^2} \zeta = 0 \quad (9.9)$$

for $U = 0$.) The singularity at the critical depth where $U = c$ is in contrast with the smooth turning-point transition at $N = \omega$; this is the analytic manifestation of the relative severity of the effect of a variable $U(z)$ versus that of a variable $N(z)$.

Thorpe (1978c) has computed the wave function $\zeta(z)$ for (1) the case of constant N and U' and (2) the case where N and U' are confined to a narrow transition layer. The results are shown in figures 9.12 and 9.13. The profiles are noticeably distorted relative to the case

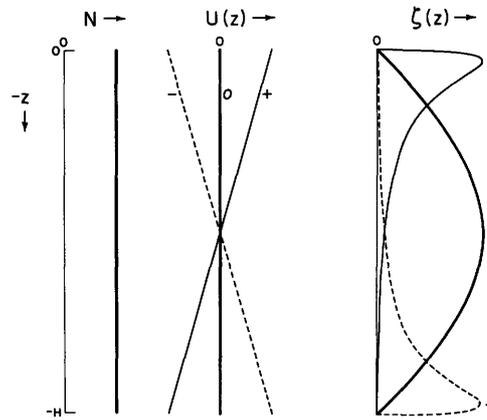


Figure 9.12 First mode vertical displacements $\zeta(z)$ in a Couette flow (constant U' and constant N), for $U'/N = 0, \pm 1$. Waves move from left to right, and U is positive in the direction of wave propagation. (Thorpe, 1978c.)

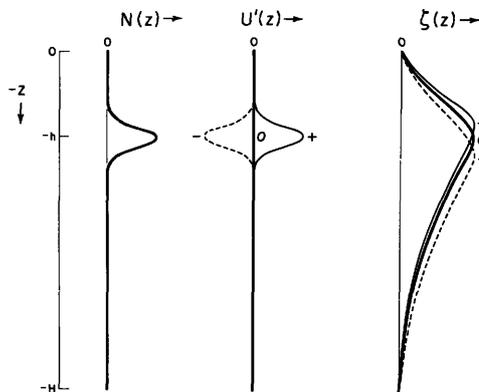


Figure 9.13 Similar to figure 9.12, but with U' and N confined to a narrow transition layer.

of zero shear, with the largest amplitudes displaced toward the level at which the mean speed (in the direction of wave propagation) is the greatest. Finite-amplitude waves have been examined for the case 2. Where there is a forward⁵ flow in the upper level (including the limiting case of zero flow), the waves have narrow crests and flat troughs, like surface waves; with backward flow in the upper layer, the waves have flat crests and narrow troughs. Wave breaking is discussed later.

9.5.1 Critical Layer Processes⁶

The pioneering work is by Bretherton (1966c), and by Booker and Bretherton (1967). Critical layers have been associated with the occurrence of clear-air turbulence; their possible role with regard to internal waves in the oceans has not been given adequate attention.

Following Phillips (1977a), let

$$\omega_F = kU + \omega, \quad \frac{\omega}{N} = \frac{k}{(k^2 + m^2)^{1/2}} = \cos \theta \quad (9.10)$$

designate the frequency in a *fixed* reference frame. $U(z)$ is the *mean current* relative to this fixed frame, and $\omega_F - kU = \omega$ is the *intrinsic* frequency [as in (9.5)], as it would be measured from a reference frame drifting with the mean current $U(z)$.

Bretherton (1966c) has given the WKB solutions for waves in an ocean of constant N and slowly varying U . [It is important to note the simplification to (9.8) when $U'' = 0$ at the critical layer.] Near the critical layer depth z_c , the magnitudes of w , u , and of the vertical displacement ζ vary as

$$w \sim |z - z_c|^{1/2}, \quad u \sim |z - z_c|^{-1/2}, \quad \zeta \sim |z - z_c|^{-1/2}.$$

The quantities ω_F and k are constant in this problem, but m and ω are not. The vertical wavenumber increases, whereas the intrinsic frequency decreases as a wave packet approaches its critical layer:

$$m \sim |z - z_c|^{-1}, \quad \omega \sim |z - z_c|.$$

A sketch of the trajectory is given in figure 9.10. Waves are refracted by the shear and develop large vertical displacements ζ (even though $w \rightarrow 0$), large horizontal velocities u , and very large induced vertical shears u' . This has implications for the dissipation and breaking of internal waves.

For $Ri > \frac{1}{4}$, Booker and Bretherton (1967) derived an energy transmission coefficient

$$\rho = \exp(-2\pi\sqrt{Ri - \frac{1}{4}}). \quad (9.11)$$

In the usual case, $U' \ll 2N$ so that $Ri \gg \frac{1}{4}$ and ρ is small. This is interpreted as wave energy and momentum being absorbed by the mean flow at z_c . As $Ri \rightarrow \infty$, $\rho \rightarrow 0$, consistent with the WKB prediction of Bretherton (1966c) that a wave packet approaches but never reaches the critical layer.

The small coefficient of transmission for Richardson numbers commonly found in the ocean implies that the critical layer inhibits the vertical transfer of wave energy. This effect has been verified in the laboratory experiments of Bretherton, Hazel, Thorpe, and Wood (1967). When rotation is introduced, the energy and momentum delivered to the mean flow may alternatively be transferred from high-frequency to low-frequency waves (if the time scales are appropriate). Thus it is possible that some sort of pumping mechanism may exist for getting energy into, for example, the high-mode, quasi-inertial internal waves. This mechanism can be compared with McComas and Bretherton's (1977) parametric instability, a weakly nonlinear interaction (section 9.6).

The work of Bretherton and of Booker and Bretherton has prompted a great number of critical-layer studies. One of the most interesting extensions was done by Jones (1968). Whereas Booker and Bretherton found the critical layer to be an absorber, not a reflector, when $Ri > \frac{1}{4}$, Jones found that reflection from the critical layer is possible when $Ri < \frac{1}{4}$; in fact, the reflected wave amplitude can exceed that of the incident wave. Jones called these waves "overreflected," their energy being enhanced at the expense of the mean flow. This is illustrated in figure 9.14, based on a solution for a hyperbolic-tangent profile intended to display the results of linear theory. Transmission and reflection ratios at $z = \pm\infty$ were derived using definitions of wave energy density appropriate to moving media. "Overtransmission" as well as overreflection occurs at very small Richardson numbers, with the internal waves gaining energy from the mean flow on both counts.

We shall now consider the condition for critical-layer absorption. Let ω_1 designate the intrinsic frequency of a wave packet at some depth z_1 with a mean flow U_1 in the direction of wave propagation. According to (9.10),

$$\omega_F = kU_1 + \omega_1.$$

Let U increase to some value U_2 at z_2 . Then since ω_F and k are conserved along the trajectory of the wave packet,

$$\omega_F = kU_2 + \omega_2.$$

For the special case that z_2 is to be a critical depth, we have $\omega_F = kU_2$, hence $\omega_2 = 0$, and so

$$\omega_1/k = U_2 - U_1.$$

The vertical wavenumber of internal waves is given by the dispersion relation

$$m = k \left(\frac{N^2 - \omega^2}{\omega^2 - f^2} \right)^{1/2} \approx kN/\omega \quad \text{for } f \ll \omega \ll N. \quad (9.12)$$

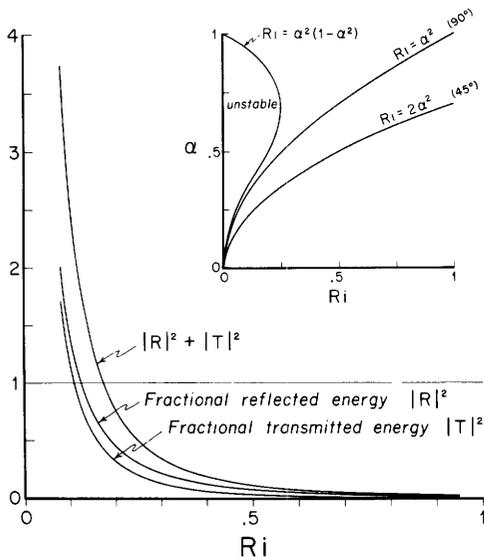


Figure 9.14 Fractional internal wave energy reflected and transmitted through a mean shear flow $U = U_0 \tanh(z/d)$ at constant N , as a function of the minimum Richardson number $Ri = N^2 d^2 / U_0^2$. Internal wave energy is lost to the mean flow for $|R|^2 + |T|^2 < 1$, or $Ri > 0.18$; internal wave energy is gained from the mean flow for $Ri < 0.18$. The plot is drawn for $Ri = 2\alpha^2$, where $\alpha = kd$ is the dimensionless horizontal wavenumber. This corresponds to a wave packet traveling at an inclination of 45° at $z = \pm\infty$. ($Ri = \alpha^2$ corresponds to the limiting case of vertical group velocity to $\pm\infty$.) [I am indebted to D. Broutman for this figure.]

For critical absorption within the interval Δz over which the mean flow varies by ΔU , we replace ω/k by ΔU , and obtain the critical vertical wavenumber

$$m_c = N/\Delta U. \quad (9.13)$$

R. Weller (personal communication) has analyzed a month of current measurements off California for the expected difference $\Delta U = |U_2 - U_1|$ in a velocity component (either of the two components) at two levels separated by $\Delta z = |z_2 - z_1|$. The observations are, of course, widely scattered, but the following values give representative magnitudes:

| | | | | | |
|----------------------------------|---|----|----|----|----------------------|
| Δz in m | 0 | 10 | 25 | 50 | 100 |
| ΔU in cm s^{-1} | 0 | 4 | 7 | 10 | 15 (upper 100 m) |
| | 0 | 7 | 10 | 12 | 15 (100–300 m depth) |

For $\Delta U = 10 \text{ cm s}^{-1}$ and $N = 0.01 \text{ s}^{-1}$ (6 cph), (9.13) gives $m_c = 10^{-3} \text{ cm}^{-1}$ (16 cpkm). Internal waves with vertical wavelengths of less than 60 m are subject to critical-layer interactions.

A large fraction of the measured velocity difference ΔU can be ascribed to the flow field $u(z)$ of the internal waves themselves, and deduced from the model spectra. The expected velocity difference increases to $\sqrt{2}$ times the rms value as the separation increases to the vertical coherence scale, which is of order 100 m. Here most of the contribution comes from low frequencies and low wavenumbers. I am tempted to interpret ΔU

for critical layer processes as rms u from the internal waves themselves. The internal wave spectrum is then divided into two parts: (1) the *intrinsic* part $m < m_c$, which contains most of the energy, and (2) the *compliant* part $m > m_c$, which is greatly modified by interaction with the intrinsic flow field. The phase speed for critical reflection is

$$c_c = \text{rms } u, \quad (9.14)$$

and the critical wavenumber is

$$m_c = N/\text{rms } u. \quad (9.15)$$

There is the separate question whether the internal waves at the critical layer will be underreflected, just reflected, or overreflected, and this depends on the ambient Richardson number. In the underreflected case there is a flux of energy from the compliant to the intrinsic waves. In the overreflected case the flow is the other way. For an equilibrium configuration, one may want to look for a transmission coefficient ρ near unity, and the exponential behavior of $\rho(Ri)$ will then set narrow bounds to the ambient spectrum. But this gets us into deep speculation, and had better be left to the end of this chapter.

9.6 Resonant Interactions

Up to this point the only interactions considered are those associated with critical layers. In the literature the focus has been on the resonant interaction of wave triads, using linearized perturbation theory. There are two ways in which critical layer interactions differ from resonant interactions: (1) compliant waves of *any* wavenumber and *any* frequency are modified, as long as c equals u somewhere in the water column; and (2) the modification is apt to be large (the ratio u/c being a very measure of nonlinearity). For the wave triads, the interaction is (1) limited to *specific* wavenumbers and frequencies, and (2) assumed to be small in the perturbation treatment.⁷ To borrow some words of O. M. Phillips (1966b), the contrast is between the “strong, promiscuous interactions” in the critical layer and the “weak, selective interactions” of the triads.

The conditions for resonance are

$$\mathbf{k}_1 \pm \mathbf{k}_2 = \mathbf{k}_3, \quad \omega_1 \pm \omega_2 = \omega_3,$$

where $\mathbf{k}_i = (k_i, l_i, m_i)$, and all frequencies satisfy the dispersion relation $\omega_i(\mathbf{k}_i)$. Resonant interactions are well demonstrated in laboratory experiments. For a transition layer (as in figure 9.4), Davis and Acrivos (1967) have found that a first-order propagating mode, which alternately raises and lowers the transition layer, was unstable to resonant interactions, leading to a rapid growth of a second-order mode, which alternately thickens and thins the transition layer like a

propagating link sausage. Martin, Simmons, and Wunsch (1972) have demonstrated a variety of resonant triads for a constant- N stratification.

Among the infinity of possible resonant interactions, McComas and Bretherton (1977) have been able to identify three distinct classes that dominate the computed energy transfer under typical ocean conditions. Figure 9.15 shows the interacting propagation vectors in (k, m) -space. The associated frequencies ω are uniquely determined by the tilt of the vectors, in accordance with (9.4). Inertial frequencies (between f and $2f$, say) correspond to very steep vectors, buoyancy frequencies (between $\frac{1}{2}N$ and N) to flat vectors, as shown.

Elastic scattering tends to equalize upward and downward energy fluxes for all but inertial frequencies. Suppose that \mathbf{k}_3 is associated with waves generated near the sea surface propagating energy downward (at right angles to \mathbf{k}_3 , as in figure 9.7). These are scattered into \mathbf{k}_1 , with the property $m_1 = -m_3$, until the upward energy flux associated with \mathbf{k}_1 balances the downward flux by \mathbf{k}_3 . The interaction involves a near-intertial wave \mathbf{k}_2 with the property $m_2 \approx 2m_3$. (The reader will be reminded of Bragg scattering from waves having half the wavelength of the incident and back-scattered radiation.) Similarly, for bottom-generated \mathbf{k}_1 waves with upward energy fluxes, elastic scattering will transfer energy into \mathbf{k}_3 waves.

Induced diffusion tends to fill in any sharp cutoffs at high wavenumber. The interaction is between two neighboring wave vectors of high wavenumber and frequency, \mathbf{k}_1 and \mathbf{k}_3 , and a low-frequency low-wavenumber vector \mathbf{k}_2 . Suppose the \mathbf{k}_2 waves are highly energetic, and that the wave spectrum drops sharply for wavenumbers just exceeding $|\mathbf{k}_3|$, such as $|\mathbf{k}_1|$. This interaction leads to a diffusion of action (energy/ ω) into the low region beyond $|\mathbf{k}_3|$, thus causing \mathbf{k}_1 to grow at the expense of \mathbf{k}_2 .

Parametric subharmonic instability transfers energy from low wavenumbers \mathbf{k}_2 to high wavenumbers \mathbf{k}_1 of half the frequency, $\omega_1 = \frac{1}{2}\omega_2$, ultimately pushing energy into the inertial band at high vertical wavenumber. The interaction involves two waves \mathbf{k}_1 and \mathbf{k}_3 of nearly opposite wavenumbers and nearly equal frequencies. The periodic tilting of the isopycnals by \mathbf{k}_2 varies the buoyancy frequency at twice the frequency of \mathbf{k}_1 and \mathbf{k}_3 . (The reader will be reminded of the response of a pendulum whose support is vertically oscillated at twice the natural frequency.)

The relaxation (or interaction) time is the ratio of the energy density at a particular wavenumber to the net energy flux to (or from) this wavenumber. The result depends, therefore, on the assumed spectrum. For representative ocean conditions, McComas (in preparation) finds the relaxation time for elastic scattering to be extremely short, of the order of a period, and so up- and downgoing energy flux should be in

balance. This result does not apply to inertial frequencies, consistent with observations by Leaman and Sanford (1975) of a downward flux at these frequencies. The relaxation time for induced diffusion is typically a fraction of a period! (This is beyond the assumption of the perturbation treatment.) Any spectral bump is quickly wiped out. The conclusion is that the resonant interactions impose strong restraints on the possible shapes of stable spectra.

In a challenging paper, Cox and Johnson (1979) have drawn a distinction between radiative and diffusive transports of internal wave energy. In the examples cited so far, energy in wave packets is radiated at group velocity in the direction of the group velocity. But suppose that wave-wave interactions randomize the direction of the group velocity. Then eventually the wave energy is spread by diffusion rather than radiation. The relevant diffusivity is $\kappa = \frac{1}{3}\langle c_g^2 \rangle \tau$, where τ is the relaxation time of the nonlinear interactions. Cox and Johnson have estimated energy diffusivities and momentum diffusivities (viscosities); they find that beyond 100 km from a source, diffusive spreading is apt to dominate over radiative spreading. There is an interesting analogy to crystals, where it is known that energy associated with thermal agitation is spread by diffusion rather than by radiation. The explanation lies in the anharmonic restoring forces between molecules, which bring about wave-wave scattering at room temperatures with relaxation times in the nanoseconds.

9.7 Breaking

This is the most important and least understood aspect of our survey. Longuet-Higgins has mounted a broadly based fundamental attack on the dynamics of breaking surface waves, starting with Longuet-Higgins and Fox (1977), and this will yield some insight into the internal-wave problems. At the present time we depend on laboratory experiments with the interpretation of the results sometimes aided by theoretical considerations.

Figure 9.16 is a cartoon of the various stages in an experiment performed by Thorpe (1978b). A density transition layer is established in a long rectangular tube. An internal wave maker generates waves of the first vertical mode. Before the waves have reached the far end of the tube, the tube is tilted through a small angle to induce a slowly accelerating shear flow. The underlying profiles of density, shear, and vertical displacement correspond roughly to the situation in figure 9.13.

For relatively steep waves in a weak positive⁸ shear, the waves have sharpened crests. At the position of the crest, the density profile has been translated upward and steepened (B_1). There is significant wave energy loss in this development (Thorpe, 1978c, figure 10).