An introduction to internal waves

T. Gerkema

J.T.F. Zimmerman

Lecture notes, Royal NIOZ, Texel, 2008

Por qué me preguntan las olas
lo mismo que yo les pregunto?

Quién puede convencer al mar
para que sea razonable?

Pablo Neruda, *El libro de las preguntas.*

Contents

Chapter 1

Introduction

Waves are all around us, but it is actually hard to say what a wave *is*. This is because it is an immaterial thing: a signal, a certain amount of energy propagating through a medium. The medium we consider is water, seawater in particular. Even though water waves as such are immaterial, they are supported by the oscillatory movement of the water parcels, and this indeed forms a way of detecting them. But it is always important not to confuse the water motion with the wave itself. The following analogies may help to clarify this point:

> "A bit of gossip starting in London reaches Edinburgh very quickly, even though not a single individual who takes part in spreading it travels between these two cities. There are two quite different motions involved, that of the rumour, London to Edinburgh, and that of the persons who spread the rumour.
>
> The wind, passing over a field of grain, sets up a wave which spreads out across the whole field. Here again we must distinguish between the motion of the wave and the motion of the separate plants, which undergo only small oscillations.
>
> We have all seen the waves that spread in wider and wider circles when a stone is thrown into a pool of water. The motion of the wave is very different from that of the particles of water. The particles merely go up and down. The observed motion of the wave is that of a state of matter and not of matter itself. A cork floating on the wave shows this clearly, for it moves up and down in imitation of the actual motion of the water, instead of being carried along by the wave." [16, pp. 104-105]

1.1 The ocean's inner unrest

Waves at the ocean's surface are a familiar sight. These lecture notes deal with waves that propagate *beneath* the surface; they are mostly hidden from eyesight. Occasionally, however, they produce a visible response at the ocean's surface.

An example is shown in Figure 1.1, a photograph taken from the Apollo-Soyuz spacecraft in 1975, when it passed over Andaman Sea, north of Sumatra.

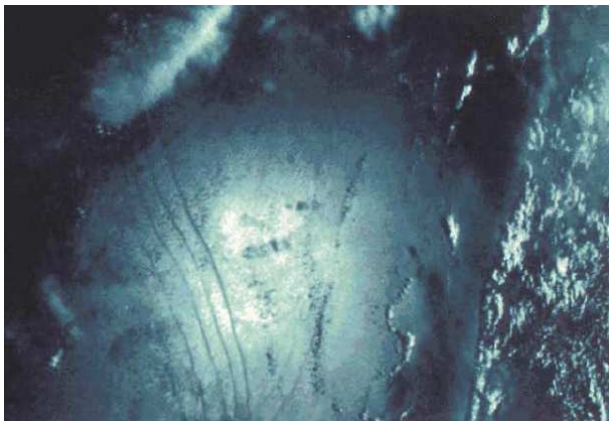

Fig. 1.1: A photograph from the Apollo-Soyuz spacecraft, made over Andaman Sea, showing stripes due to internal waves. The stripes stretch over 100 km or more, and have a mutual distance of the order of a few tens of kilometers; they propagate slowly (at a speed of about 2 $\mathrm{m\,s^{-1}}$) to the northeast.

The stripes can be observed from a ship as well; they appear as long bands of breaking waves, typically about 1 meter high. Spacecraft or satellite pictures showing such stripes have since been obtained from many other locations; an example from the Bay of Biscay is shown in Figure 8.11. Figure 1.2 provides a look into the ocean's interior, and brings us to the origin of the stripes, in this case in Lombok Strait. The echosounder signal shows the elevation and depression of levels of equal density (isopycnals); in the course of just 20 minutes, they descend more than 100 m, and rise again to their original levels.[1] The isopycnals closer to the surface, however, undergo a much smaller vertical displacement. This is the defining characteristic of *internal waves*: that their largest vertical amplitudes occur in the *interior* of the fluid.

Internal waves were discovered more than a century ago. One of the first observations is due to Helland-Hansen & Nansen [41]. They found that temperature profiles may change substantially within the course of just hours (Figure 1.3); they ascribed this to the presence of *"puzzling waves"*, an example of which is shown in Figure 1.4. They stressed the importance of this newly discovered phenomenon:

> "The knowledge of the exact nature and causes of these "waves" and their movements would, in our opinion, be of signal importance to Oceanography, and as far as we can see, it is one of its greatest problems that most urgently calls for a solution" [p. 88]

[1] The horizontal currents associated with these waves extend to the surface; these surface currents modify the roughness of the surface waves, thus rendering the internal waves (indirectly) detectable by satellite remote sensing imagery.

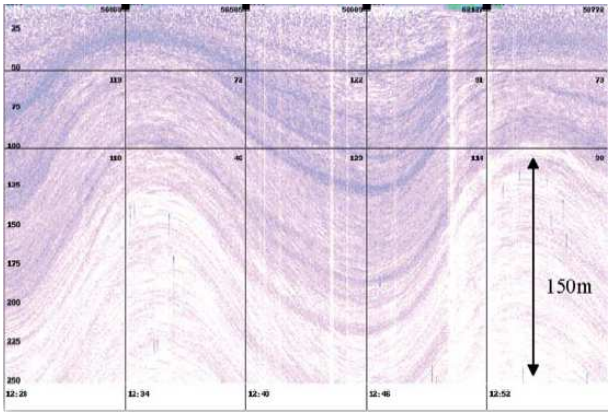

Fig. 1.2: Isopycnal movements associated with the passage of an internal wave, observed with an echosounder in Lombok Strait, covering the upper 250 m of the water column. Colours indicate levels of backscatter and can be used to distinguish levels of density. Horizontal is time; the spacing between two vertical lines corresponds to 6 minutes. From [82].

Although the internal waves shown in Figure 1.2 are perhaps unusually large, their presence as such is not at all unusual; they are a ubiquitous phenomenon in the ocean (and in the atmosphere as well). Internal waves provide the 'inner unrest' in the oceans, at time scales ranging from tens of minutes to a day. These oscillations are a nuisance when one attempts to establish the ocean's 'background state' (i.e. patterns of large-scale circulation, tracer distribution etc.). This was already recognized by Helland-Hansen & Nansen. The *puzzling waves*, they noted,

> "make it much more difficult than has hitherto generally been believed, to obtain trustworthy representations of the volumes of the different kinds of water; they certainly cannot be attained by observations at a small number of isolated stations, chosen more or less at random. Such irregularities, great or small, are seen in most vertical sections where the stations are sufficiently numerous and not too far apart. The equilines (isotherms, isohalines, as well as isopyknals) of the sections hardly ever have quite regular courses, but form bends or undulations, like waves, sometimes great, sometimes small." [p. 87]

From measurements made at any one moment it is thus impossible to deduce what the background isopycnal levels and current velocities are; the background state is continually being perturbed by internal-wave activity, which may produce isopycnal variations of the order of 100 m, and current velocities of tens of $\mathrm{cm\,s^{-1}}$. An example of such a variation is shown in Figure 1.5. Only prolonged measurements allow for a meaningful estimate of the 'mean background state'.

Before we continue our discussion on internal waves, we first take a closer look at the *medium* in which they propagate. Two properties are of primary

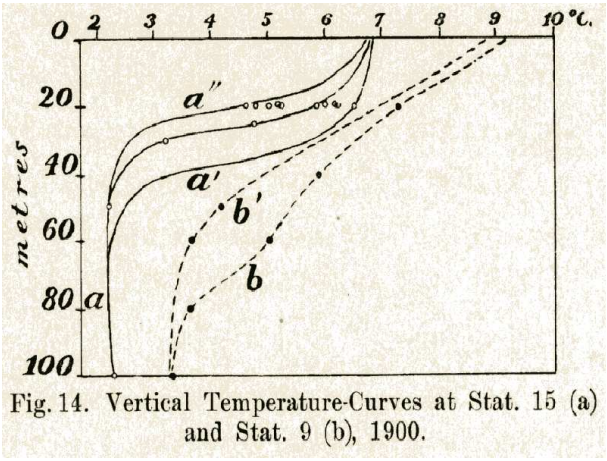

Fig. 14. Vertical Temperature-Curves at Stat. 15 (a) and Stat. 9 (b), 1900.

Fig. 1.3: Temporal changes in temperature profiles, at two different locations. The curves were constructed on the basis of the measurements shown in dots. Profiles *a'* and *a"*: northeast of Iceland, on August 5, 1900; variations at 20 m depth were measured during about 2 1/2 hours, but sometimes rapid changes occurred in just five minutes. Profiles *b* and *b'*: north of the Faeroes, on 25-26 July, 1900. From [41].

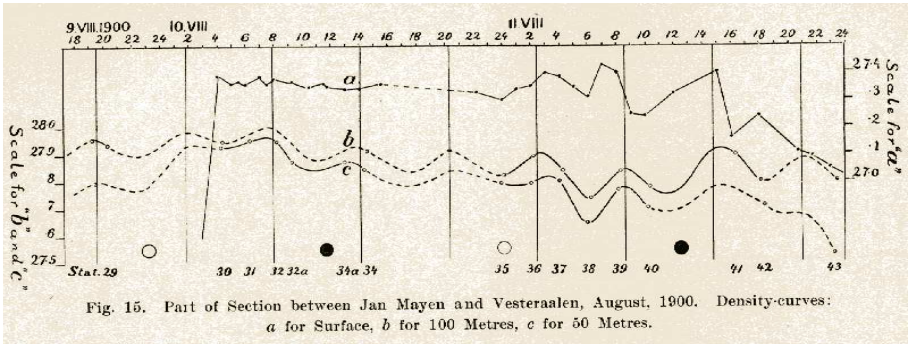

Fig. 15. Part of Section between Jan Mayen and Vesteraalen, August, 1900. Density-curves: *a* for Surface, *b* for 100 Metres, *c* for 50 Metres.

Fig. 1.4: Isopycnal variations with time. The dominant period is about half a semi-diurnal tidal period. For comparison, open and black circles have been added, denoting the spacing between high and low waters. From [41].

importance: 1) the vertical stratification in density, and 2) the diurnal rotation.

1.2 Restoring forces

Waves in fluids owe their existence to restoring forces; these forces push parcels that are brought out of their equilbrium position, back towards that position, thus bringing them into oscillation. Sound waves, for example, exist due to compression; here pressure (gradients) act as the restoring force. In internal waves, two restoring forces are at work: 1) buoyancy (i.e. reduced gravity in the

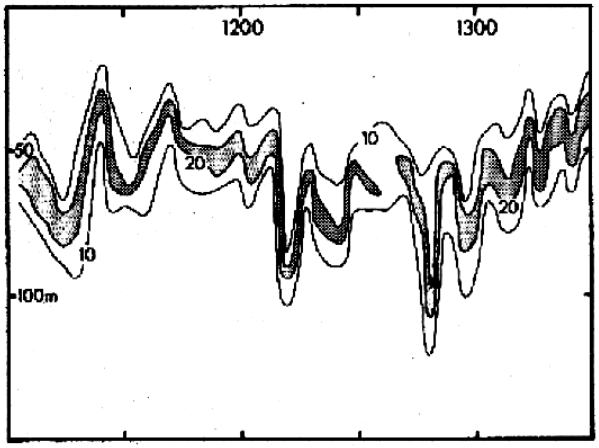

Fig. 1.5: Variation of the depth of the Chlorophyll maximum (shaded region) due to the presence of internal waves, in the Bay of Biscay. Horizontal is time (in hours), but the variations are both temporal and spatial, since the measurements were made from a moving ship. From [43].

ocean's interior), due to the ocean's stratification, and 2) the Coriolis force, due to the Earth's diurnal rotation.

1.2.1 The ocean's stratification

The bulk of the ocean is very cold; the ocean's mean temperature is only 3.5°C. The variation with depth, however, is large: below 1000 m depth, temperature is less than 5°C, but in the upper 200 m it rises strongly, especially in the tropics (see Figure 2.2a), and during summer at mid-latitudes. Together with variations in salinity (Figure 2.2b), this determines in-situ density ρ, being a function of pressure, temperature, and salinity (Figure 2.2e). The steady increase of in-situ density with depth does not in itself guarantee that the water column is gravitationally stable. As explained in Chapter 3, the stability of the water column is determined by

$$N^2 = g^2\Big(\frac{\partial\rho}{\partial p} - \frac{1}{c_s^2}\Big)\,, \tag{1.1}$$

where p is pressure, c_s the speed of sound (Figure 2.2f), and g the acceleration due to gravity. The water column is stably stratified if $N^2 > 0$. The quantity N is called the *Brunt-Väisälä* or *buoyancy frequency*; its unit is radians per second (but also common are cycles per hour, or cycles per day).

A typical distribution of N in the ocean is shown in Figure 1.6. It shows that N varies greatly: from $O(10^{-4})$ in the deepest layers to $O(10^{-2})$ rad s^{-1} in the upper 200 m. The latter region includes the *thermocline*, corresponding to a peak in N due to the rapid decrease of temperature with depth; the thermocline

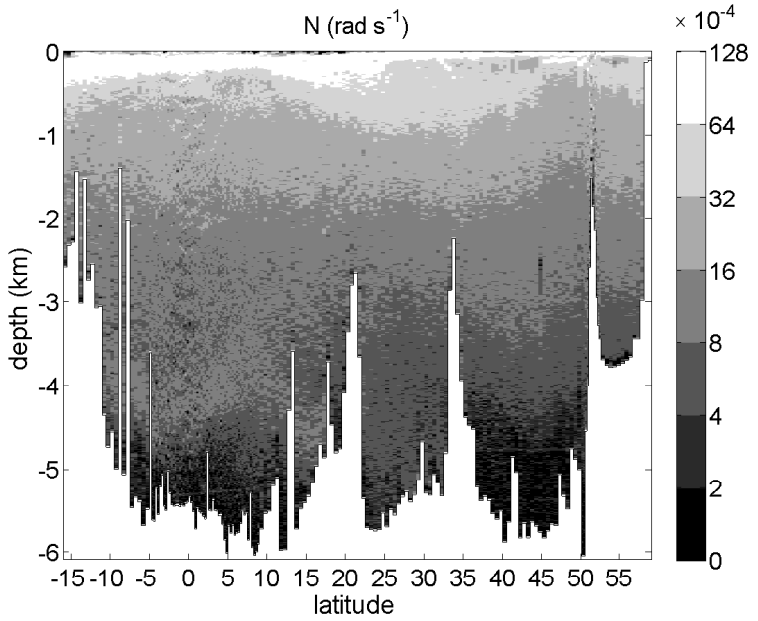

Fig. 1.6: The stratification N (in $\mathrm{rad\,s^{-1}}$), derived from temperature and salinity profiles in the Pacific Ocean, for a south-north section near 179°E (WOCE section P14, from the Fiji Islands into the Bering Sea, July/August 1993). Adapted from [30].

has a permanent character in the tropics and is seasonal at mid-latitudes. We see from Figure 1.6 that N decreases again in the upper 50 m or so, the *upper mixed layer*, which is due to the mixing by the wind.

In the atmosphere, values range from 0.01 in the troposphere to 0.02 $\mathrm{rad\,s^{-1}}$ in the stratosphere. Both in the ocean and atmosphere, N becomes locally very small in turbulently mixed, convective layers.

1.2.2 The Earth's diurnal rotation

The Earth undergoes a diurnal rotation on its axis. After one full rotational period it regains the same orientation with respect to the 'fixed stars'; this period of 23 h 56 min 4 s (=86164 s) is called a *sidereal day*, d_{sid}. It is distinct from the *solar* day (i.e. 24 hours) because, as the Earth traverses its path around the sun (in what as such is a *translational* motion), it takes slightly more[2] than the sidereal day to regain the same orientation with respect to the *sun* – which is what defines the solar day.

The Earth angular velocity thus is

$$\Omega = \frac{2\pi}{d_{sid}} = 7.292 \times 10^{-5}\,\mathrm{rad\,s^{-1}}\,.$$

[2]The Earth's diurnal rotation is prograde; if it were retrograde, the solar day would be *shorter* than the sidereal day.

(Note that the last two decimals would be different if one mistakenly uses the *solar* day.)

We can now express the vectorial character of the diurnal rotation as $\vec{\Omega}$, aligned to the axis of rotation (pointing northward), and with magnitude $|\vec{\Omega}| = \Omega$. To find the effects of rotation at a certain latitude ϕ, we can decompose the vector as indicated in Figure 1.7. Thus we find the Coriolis frequencies

$$f = 2\Omega \sin\phi\,; \quad \tilde{f} = 2\Omega \cos\phi\,. \tag{1.2}$$

These components determine the *Coriolis force*, which is formed by the outer product of $2\vec{\Omega}$ with velocity (see Chapter 2). The Coriolis force acts as a purely deflecting force: it never initiates a motion (the force does no work since it is perpendicular to velocity), it only deflects an existing motion.

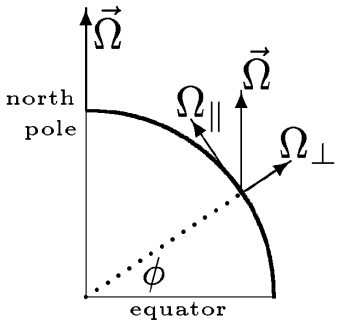

Fig. 1.7: Decomposition of the rotation vector $\vec{\Omega}$ at latitude ϕ, giving rise to the Coriolis components $f = 2\Omega_{\perp}$ and $\tilde{f} = 2\Omega_{\|}$.

The perpendicularity with velocity has still another consequence: the component $\tilde{f}$, being itself horizontal (see Figure 1.7), deflects downward moving parcels eastward (e.g. if you drop a stone from a tower, it will undergo a slight deflection to the east), and produces an upward force on eastward moving parcels (i.e. the weight of an eastward moving object is reduced, the so-called Eötvös effect). In either case, there is a *vertical* direction involved. Now, the currents in the ocean are predominantly *horizontal*, due to the fact that the ocean constitutes a thin layer compared to the Earth's radius. This diminishes the importance of $\tilde{f}$; the *effect* of f usually far exceeds that of $\tilde{f}$, despite the fact that f and $\tilde{f}$ as such are of similar magnitude at mid-latitudes. We continue this discussion in later chapters, but for the moment we assume that $\tilde{f}$ is negligible, so that the effects of the Earth's diurnal rotation are represented solely by f. Notice that f is negative in the Southern Hemisphere.

We have thus established two fundamental frequencies, N and f, each of them associated with a restoring force. These two restoring forces, gravity and the Coriolis force, lie at the heart of the phenomenon of internal waves; this is

reflected by the fact that N and f are key parameters in internal-wave theory. We note that in most parts of the ocean, N exceeds $|f|$.

Finally, a few words on nomenclature. Internal waves for which only gravity acts as the restoring force, are called internal gravity waves; this situation occurs for example in laboratory experiments on a non-rotating platform, or in the ocean for waves at frequencies much higher than $|f|$, in which case the Coriolis force can be neglected. Conversely, if only the Coriolis force is at work, they are called gyroscopic (or inertial) waves; this situation occurs in neutrally stratified layers ($N = 0$). Finally, if both forces are at work – as is commonly the case – they are called internal inertio-gravity waves.

1.3 Origins of internal waves

Where does the ubiquitous 'inner unrest' originate from? As it turns out, there are two principal sources of internal waves.

One is the atmospheric disturbance of the ocean's upper mixed layer by the wind; this was already recognized by Helland-Hansen & Nansen [41]:

> "It is a striking fact, and apparently not merely an accidental one, that by far the greatest "waves" of this kind in our sections, occurred in 1901, when the atmosphere was unusually stormy; and it appears probable that the "waves" in that year might have been due to stirring of the water masses, caused by disturbances in the atmosphere." [p. 88]

As the wind resides, variations of the base of the mixed layer slowly evolve towards equilibrium, in a process called *geostrophic adjustment* [33]. During this process, internal waves are emitted, predominantly at frequencies close to $|f|$, the inertial frequency. These waves are called *near-inertial waves*; they are usually clearly present in internal-wave spectra, as a peak centered around $|f|$. They form indeed the most energetic part of the internal-wave spectrum. Notwithstanding their importance, it would seem that a comprehensive understanding of their generation and propagation is still lacking.

This is very different for the other source of internal waves, also at low frequencies: the *internal tides*. They are formed by the flow of barotropic (i.e surface) tides over sloping bottom.

The origin of barotropic tides themselves lies in the astronomical tide-generating forces: the gravitational pull by the moon and, to a lesser extent, the sun. These forces, together with the diurnal rotation of the Earth, produce the barotropic tides, which traverse the oceans as surface waves (an example is shown in Figure 7.1). This movement acts as drag to the moon, and thus slows down its angular velocity. Conservation of angular momentum implies that the moon must recede from the Earth. This has been confirmed by observations: the distance between the moon and Earth increases by 3.8 cm per year.[3] From this, one can calculate

[3]Measured using laser beams reflecting from mirrors that were placed on the moon during the *Apollo 11* mission, in July 1969, and later missions. The drag not only retards the moon's

how much energy goes into the barotropic tides in the ocean (the amount going into tides in the atmosphere and the Earth's mantle is small by comparison): about 3.5 TW for all tidal components together (1 TeraWatt = 10^{12} Watt). The barotropic tide, in turn, loses its energy mostly by bottom friction in shallow seas, but also for a significant part, about 30% (1 TW), over ridges in the 'open ocean' (and for another, yet unknown part, over the continental slopes); here the energy is transferred to *internal tides*. This is illustrated in Figure 1.8.

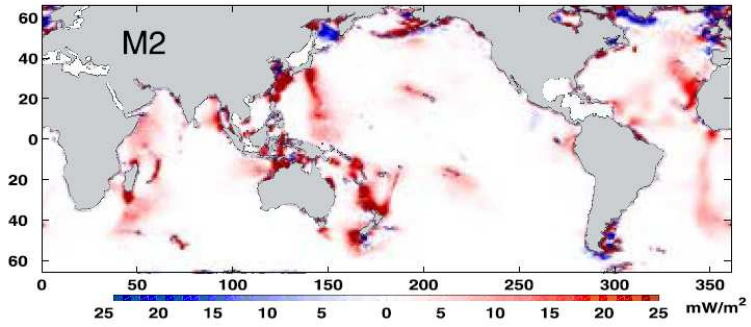

Fig. 1.8: Regions where dissipation of the semi-diurnal lunar barotropic tide (M_2) occurs, determined using data from satellite altimetry. There is a clear correspondence with bottom topography; noticeable dissipation occurs over, for example, the Mid-Atlantic Ridge and the Hawaiian Ridge. The results are less reliable in shallow regions (because of uncertainties in the estimates of tidal currents), where errors may lead to spots of negative values (in blue). From [15].

The idea behind this process is as follows. Barotropic tidal currents are predominantly horizontal (U), but over bottom slopes a *vertical* component must arise ($U\nabla h$, with bottom topography h), which, like the horizontal component, oscillates at the tidal frequency. This vertical tidal current brings isopycnal surfaces into oscillation; they are periodically lifted up and pulled down. These vertical oscillations act as a wavemaker, emitting waves at the forcing frequency: the internal tides. We may compare this process with that of wave generation in a stretched string or rope: if it is forced into vertical oscillation at one point, waves are generated which propagate away from that point.

But *how* do internal tides propagate away from the region of forcing? This brings us to what is perhaps the most remarkable (and in any case the most counter-intuitive) property of internal waves: their energy propagates at once horizontally *and vertically*, quite unlike surface waves, whose energy propagates only horizontally. The difference is due to the different nature of the stratification supporting the waves. Surface waves owe their existence to the sharp change in density between air and water, which is restricted to the surface – and so is their energy propagation. Internal waves, on the other hand, owe

movement around the Earth, but also lengthens the terrestrial day. The combined effect of lunar and solar tides amounts to an increase of 2.4 milliseconds per century (see [7], p. 249).

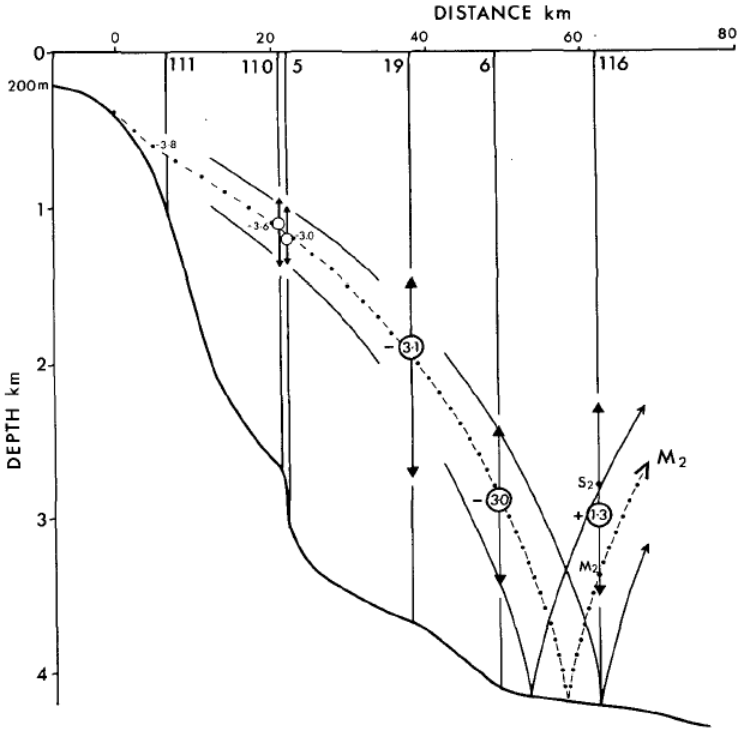

Fig. 1.9: The path of an internal tidal beam generated over the continental shelf break in the Bay of Biscay. The depth of maximum vertical isopycnal excursion was determined at various horizontal positions by CTD yoyoing; these depths are indicated by circles. They follow a path that coincides with the theoretical path of internal-tide propagation (dashed line). From [69].

their existence to the stratification of the ocean's interior, which is smoothly distributed over the vertical (see Figure 1.6), and so energy is carried from one depth level to the other. A vivid illustration of the path of energy propagation is shown in Figure 1.9: the internal tide generated over the continental slope propagates into the deep ocean, following a diagonal path. The arrows indicate the vertical extent of the energy, showing that the internal tide propagates in a beam-like manner.

To summarize, there are two main generation mechanisms: atmospheric forcing, and barotropic tidal flow over topography. Both generate low-frequency waves. However, interactions among these waves lead to internal waves at higher frequencies. As a result, internal waves are found at all frequencies between $|f|$ and N, although those at low frequencies dominate the spectrum.

1.4 Dissipation and mixing

In the previous section we have seen that the Earth-moon system loses energy to barotropic tides, which, in turn, lose part of their energy to internal tides. The question then arises where *their* energy goes. We have also seen that much of the internal-wave energy originates from the upper layer of the ocean (near-inertial waves generated by the wind, internal tides generated over the continental shelf break). However, the vertical component in their energy propagation opens up the possibility that their energy, though originating from the upper layer, may finally be dissipated in the abyssal ocean.

This, indeed, seems to be what is happening. The precise pathways to dissipation are yet to be established quantitatively, but the general picture has become clearer in recent years, see Figure 1.10. Internal waves can become unstable due to the presence of a background shear field, leading to internal-wave breaking and mixing.

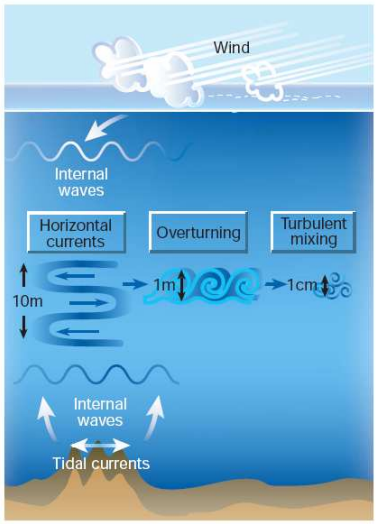

Fig. 1.10: Sketch of the pathway of internal-wave energy: from its origin, by the wind and by tidal flow over topography, to dissipation as small-scale mixing. From [21].

Figure 1.10 thus illustrates how energy is transferred to smaller scales. Surprisingly, this has important implications for the large-scale ocean circulation. For a large part, this circulation is wind-driven, but part of it consists in a sinking of cold water at high latitudes (deep convection), specifically in the Labrador, Greenland and Weddell Seas; this water spreads horizontally over the ocean basins, hence the low temperatures in the deep layers at all latitudes. If this were the only factor determining the ocean's vertical temperature distribution, one would find low temperatures extending upwards until the ocean's most upper layer, where direct warming by the sun takes place. In reality, the temperature gradient is much more gradual (see Figure 2.2a). This shows that

there must be downward mixing of heat (Figure 1.11). The combined effects of downward mixing and deep convection keep the ocean in a stationary state. The mixing is thought to be largely due to internal waves; at any rate, the numbers are consistent. The estimate of the required energy input into mixing is 2 TW; near-inertial waves and internal tides each contribute about 1 TW.

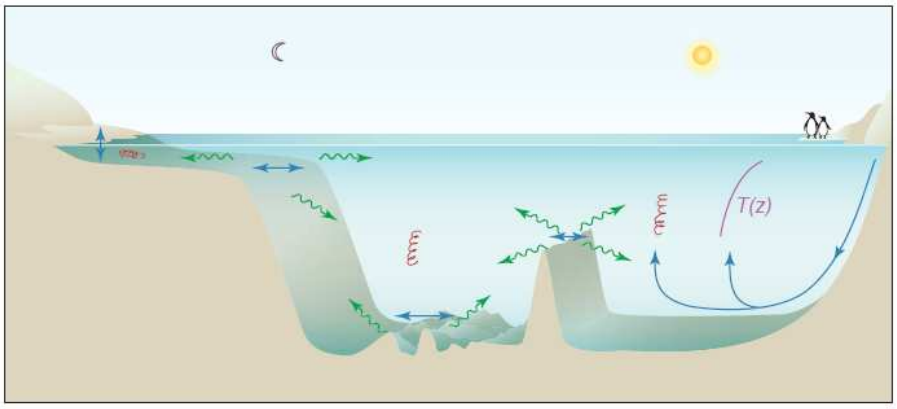

Fig. 1.11: Near-inertial waves (not depicted), along with internal tides generated over bottom topography by the barotropic tidal current, feature in the deep ocean. These internal waves can lead to turbulence and mixing. This mixing plays a role in maintaining a gradual transition between the sun-warmed surface layer of the ocean and the upwelling cold, dense water formed at high latitudes. $T(z)$ denotes the temperature profile as a function of depth z. From [20].

1.5 Overview

As a guide through later chapters, we may use Figure 1.9 and the questions it raises: "the facts which call for explanation". To answer most of these questions, it suffices to consider linear theory, i.e. the theory of small-amplitude internal waves (Chapters 5 and 6). This theory explains the remarkable diagonal propagation as well as the reflection from the bottom. Examining Figure 1.9 more closely, we see that the beam becomes slightly steeper in deeper waters, i.e. refraction occurs; this, too, is explained by linear theory. At the origin of the beam lies the barotropic tidal flow over a slope; this generation mechanism is studied in Chapter 7. Not visible in Figure 1.9 is what happens after the beam has reflected from the bottom. Other observations, to be discussed later, show that the beam, with upward energy propagation, finally impinges on the seasonal thermocline (in the upper 100 m of the water column); this generates high-frequency high-amplitude internal waves, called internal solitons. To describe these waves, which are beyond the assumption of small amplitudes, nonlinear theory is required (Chapter 8).

First of all, however, we need to establish the basic equations of internal-wave

theory (Chapter 2), put the notion of stratification in an exact form (Chapter 3), and discuss the approximations underlying internal-wave theory (Chapter 4). To do this properly, we also need to examine carefully the thermodynamic principles that form part of the governing equations.

Further reading

Although these lecture notes are meant to be self-contained, it is of course useful to consult other literature as well; here we give some suggestions for further reading. More references follow in later chapters as appropriate.

Most textbooks on ocean physics or dynamical meteorology pay some attention to internal waves. Three older textbooks deal exclusively with internal waves in the ocean: Krauss [47], Roberts [72] and Miropol'sky [58]. Of these, the third is the most advanced text. The second is probably the most accessible and also provides an admirably complete reference list of the literature up to 1975. A lot of useful material on internal waves can be found in the textbook by Leblond & Mysak [48]. Chapters on internal waves can be found in the books by Turner [84], Phillips [67], and Lighthill [51]; on gyroscopic (i.e. inertial) waves, see Greenspan [35]. See Vlasenko et al. [88] for a recent monograph on the modelling of internal tides. The review papers by Garrett & St. Laurent on deep-ocean mixing [24] and by Garrett & Kunze on internal tides [23] provide a valuable account of the current understanding of these subjects. On short internal waves and internal-wave spectra, see the review paper by Munk [61].

We focus on the ocean, and will only in passing discuss internal waves in the atmosphere. More on this subject can be found in the textbook by Gossard & Hooke [34], and the review paper by Fritts & Alexander [19].

Chapter 2

The equations of motion

2.1 Introduction

In early 1913, Vilhelm Bjerknes gave his inaugural lecture at the University of Leipzig, which was titled "Die Meteorologie als exakte Wissenschaft" (Meteorology as an exact science). In it, he drew attention to the fact that the physical laws governing the motions of the atmosphere, together form a closed set; i.e. there are as many equations as unknowns. Meteorology, Bjerknes argues, has thus become an exact science. This offers the prospect, at least in principle, that a solution to the equations may be obtained, which would provide a mathematical description, and even prediction, of the motions in the atmosphere [5].

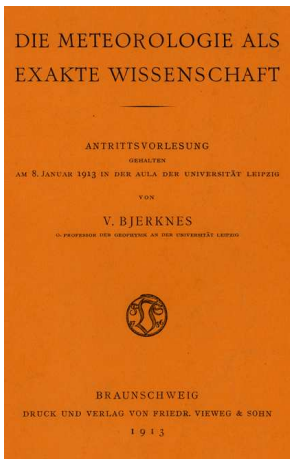

Fig. 2.1: Vilhelm Bjerknes (1862-1951), and the front page of his inaugural lecture.

The variables in question are the three velocity components, pressure, density, temperature and humidity (or, for the ocean: salinity). They feature in

the following laws, seven in total:

1-3: the three momentum equations;

4: conservation of mass;

5: the equation of state;

6-7: the two laws of thermodynamics.

If we consider the ocean instead of the atmosphere, the laws remain the same except the equation of state, which is specific for the medium in question.

We only briefly discuss the laws **1-4**, for they belong to the standard material in textbooks on (geophysical) fluid dynamics; they form the subject of Section 2.2. Much more attention needs to be paid to the equations relating to thermodynamics, **5-7**, for two reasons. First, thermodynamic aspects are usually dealt with cursorily in the oceanographic literature; as a result, neither the meaning nor the importance of thermodynamics is properly conveyed. Second, a recent development calls for a new approach. This development is the usage of the so-called Gibbs potential in ocean physics. At first, it may seem to make things more complicated, but once grasped, it brings out the structure of thermodynamics, and the way thermodynamics enters the equations of motion, more clearly than would otherwise be attainable. We discuss this in Sections 2.3 and 2.4. Finally, in Section 2.5, we arrive at the equations governing internal-wave dynamics.

2.2 Fluid mechanics

The forces that feature in the momentum equations governing fluid motions are pressure gradients, gravity, frictional forces, and external forces. Together they determine the acceleration that fluid parcels undergo in coordinate systems which are at rest or in uniform rectilinear motion with respect to 'absolute space', the 'fixed stars'. The Earth, however, spins on its axis, and thus rotates with respect to the 'fixed stars'. If we use a coordinate system that co-rotates with the Earth, we have to add two apparent forces: the Coriolis and the centrifugal force. The momentum equations then become

$$\frac{D\vec{u}}{Dt} = -\frac{1}{\rho}\nabla p - \nabla\Phi_g + \vec{F} - 2\vec{\Omega}\times\vec{u} - \vec{\Omega}\times(\vec{\Omega}\times\vec{r})\,. \tag{2.1}$$

Here we use a right-handed orthogonal Cartesian coordinate system which has its origin at the centre of the Earth, the x and y axes span the equatorial plane, and the z axis points towards the north pole. The Earth spins at angular velocity $\Omega = 7.292\ 10^{-5}$ rad s^{-1} on the z axis, and $\vec{\Omega} = (0, 0, \Omega)$; the position of the fluid parcel is denoted by $\vec{r} = (x, y, z)$. The velocity field is denoted by the vector $\vec{u}$; ρ is density; p pressure; Φ_g the gravitational potential; and $\vec{F}$ denotes

any frictional or external forces, which need not be specified here. D/Dt and ∇ denote the material derivative and gradient:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{u} \cdot \nabla \,, \qquad \nabla = \Big(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \Big) \,.$$

The centrifugal force, the last term on the right-hand side of (2.1), can be written as (minus) the gradient of the potential $\Phi_c = -\frac{1}{2}\Omega^2(x^2+y^2)$, and hence we can write (bringing the Coriolis force, the penultimate term in (2.1), to the left-hand side, as is customary)

$$\frac{D\vec{u}}{Dt} + 2\vec{\Omega} \times \vec{u} \quad = \quad -\frac{1}{\rho}\nabla p - \nabla(\Phi_g + \Phi_c) + \vec{F} \,. \tag{2.2}$$

The coordinate system adopted here is inconvenient in that we would rather have the origin at the surface of the Earth, and the axes oriented to it in a natural way. Before this can be done, we have to find an appropriate representation of the shape of the Earth.

On long geological time scales the Earth is not quite a solid object; rather like a fluid, it has adjusted itself to the state of rotation. It thus has taken an oblate shape such that the gradient of the geopotential Φ (i.e. the gravitational plus centrifugal potential: $\Phi = \Phi_g + \Phi_c$) has no components tangential to the surface. In other words, the surface of the Earth coincides with a level of constant geopotential. This is important for the dynamics of the atmosphere and ocean, since otherwise a fluid parcel would experience a tangential force due to the geopotential. This level of constant geopotential closely resembles an ellipsoid of revolution. The ellipticity is small (about 0.08), which suggests that we may pretend the surface of the earth to be spherical (with radius $R \approx 6371$ km). In this new representation, the Earth's surface (now a sphere) should act as a level of constant geopotential; otherwise the dynamics would become distorted.

The equations of motion can now be cast in terms of spherical coordinates. A simpler form can however be obtained if the phenomena of interest are so small that the curvature of the Earth's surface becomes insignificant; this yields the so-called f-plane approximation. The corresponding equations can be derived either by employing a local approximation to the equations in spherical coordinates (see [48]), or directly from (2.2) by moving the origin of the Cartesian coordinate system to the position of interest ($r = R$, $\phi = \phi_0$, say), and then tilting it such that the x,y plane becomes tangential to the Earth's surface (with x pointing eastward, y northward, and z positive in the outward radial direction). Since this new coordinate system is at rest with respect to the original system, the equations remain the same, except that $\vec{\Omega}$ now becomes $\vec{\Omega} = \Omega(0, \cos\phi, \sin\phi)$, and the gradient of the geopotential $\nabla\Phi = (0, 0, g)$. The momentum equations thus become

$$\boxed{\frac{D\vec{u}}{Dt} + 2\vec{\Omega} \times \vec{u} = -\frac{1}{\rho}\nabla p - g\hat{z} + \vec{F} \,,} \tag{2.3}$$

where $\hat{z}$ is the unit vector in the z direction (positive, outward). The f-plane derives its name from the common notation $2\vec{\Omega} = (0, \tilde{f}, f)$, in which $\tilde{f} = 2\Omega\cos\phi$ and $f = 2\Omega\sin\phi$ are regarded constant.

A following order of approximation would lead to the so-called β-plane, in which variations of the Coriolis parameter with latitude are taken into account while metric terms are still ignored. This amounts to replacing f by $f_0 + \beta y$, with f_0 constant and $\beta = (2\Omega\cos\phi)/R$; to ensure the conservation of angular momentum, $\tilde{f}$ should be taken constant, as on the f-plane, see [38].

To complete the mechanical part, we need an equation that expresses the conservation of mass; in local Cartesian coordinates it reads

$$\boxed{\frac{D\rho}{Dt} + \rho\nabla\cdot\vec{u} = 0\,.} \tag{2.4}$$

We now have four equations in total, but five unknowns: three velocity components, pressure, and density. Hence the set is not closed, unless density were simply assumed to be constant (incompressible fluid). However, variations in the density field are essential to the existence of internal waves, so we have to complete the set in a different way: by including thermodynamic principles, which provide relationships between thermodynamic state variables such as pressure and density. This will be elaborated on in the remainder of this chapter.

2.3 A brief introduction to thermodynamics

This section provides a résumé of thermodynamic principles, with a view to ocean physics and meteorology. It serves as a preparation for Section 2.4, where the set of governing equations is completed. In line with common usage in thermodynamics, we use here specific volume $\nu = 1/\rho$, instead of density ρ.

2.3.1 Fundamentals

Thermodynamics deals with transitions from one thermodynamic state to another. In the simplest case, both are *equilibrium states*; states, that is, which would remain unchanged if the system were isolated. Equilibrium states are defined by a certain number of thermodynamic *state variables* (such as pressure, temperature, density, internal energy or entropy).

For example, the equilibrium state of an ideal gas is defined by two such variables; pressure p and temperature T, say. All other state variables are then a function of those two; such a functional relationship is called an *equation of state*. Specific volume ν, for instance, is given by the following expression:

$$\nu(p,T) \quad = \quad \frac{R_d T}{p}\,, \tag{2.5}$$

in which $R_d = R_*/m$, where m is the mass (in kg) of 1 mol, and R_* the universal gas constant:[1] $R_* = 8.31$ J K^{-1} mol^{-1}. The constant R_d thus depends on the type of gas. For dry air[2] one finds $R_d = 287$ J K^{-1} kg^{-1}.

Despite the fact that the actual state of the ocean, or atmosphere, considered in its *entirety*, is far removed from thermodynamic equilibrium, the concept of thermodynamic equilibrium still proves very useful in the geophysical context. This is because we can adopt the 'local equilibrium assumption' [46, §15.1]. Its meaning is most easily grasped by supposing the opposite: that it were not valid. This would be the case if specific volume ν were not only dependent on temperature and pressure, like in (2.5), but also on *spatial gradients* of temperature and pressure. This would call for an extended non-equilibrium thermodynamics. The 'local equilibrium assumption' amounts to assuming that such spatial gradients are negligible; this allows us to apply, *locally*, the thermodynamic equilibrium relations, such as the equation of state.

In thermodynamics two types of processes are distinguished: according to whether they are reversible or irreversible. In a reversible process all intermediate states are equilibrium states, whereas in an irreversible process they are not. Strictly speaking, the former is not a process (since the only way to change an equilibrium state is by bringing the system *out* of equilibrium), but rather a chain of disconnected equilibrium states. Nevertheless, it is often useful to consider quasi-static processes, as practical approximations to reversible processes, which take place slowly enough for the state to be always very close to thermodynamic equilibrium; one can then assume that the equation of state is valid throughout the process. In the applications, discussed here and in the following chapter, we assume this to be the case.

Besides state variables, which characterize the state of a system independently of how it came into that state, there are also quantities – heat and work – which are the exact reverse in that *they do not refer to a state*, but to the *way* in which one state transforms into another. The First Law of thermodynamics connects the two; in it, the state variable (specific) *internal energy* ϵ is postulated, which can change either by heat (dQ) or by work done on the system (dW):

$$d\epsilon = dQ + dW\,.$$

All quantities are here taken per unit of mass, hence the dimension J kg^{-1}. If, after some process, a system returns to its initial state (a cyclic process), then the First Law guarantees that ϵ takes again its original value; but in the course of the process heat may have been partly converted into work – the principle of a heat engine.

For quasi-static processes, the work done on the system can be expressed as

$$dW \;\; = \;\; -pd\nu\,.$$

In the Second Law another state variable, (specific) entropy η, is postulated,

[1] 1 mol contains 6.02×10^{23} molecules.

[2] 78% N_2, 21% O_2, 1% Ar, with respective molmasses (in grams) of 28, 32 and 40.

which has the property that

$$d\eta \geq \frac{dQ}{T}\,,$$

where the equality sign holds for reversible processes. Here η denotes the entropy per unit of mass, hence the adjective 'specific'; its dimension is J K^{-1} kg^{-1}. For *reversible* processes, we can combine both laws to obtain

$$d\epsilon = Td\eta - pd\nu\,. \tag{2.6}$$

This important equation is often referred to as the *thermodynamic identity*.

2.3.2 Open systems

We speak of an *open* system if the material substance in question is subject to change; this happens for instance if, in addition to the main substance, a second constituent is present whose concentration can change by gain, loss, or redistribution. In the geophysical context, this role is played by salinity in the ocean and humidity in the atmosphere, assuming that we may conceive seawater, or air, simply as a two-component fluid. This is a valid approach, for the following reasons. In the case of seawater, the primary substance (water) is pure; the secondary substance has many constituents, but throughout the ocean they contribute in a nearly fixed proportion,[3] and this is why we can simply think of it as one substance ('salt'). In the troposphere it is the other way round: the primary substance is a composite, but in fixed proportions (see note on p. 27), and hence can be thought of as one substance ('dry air'); the secondary substance, water vapour, is pure. The concentration of the secondary substance is expressed by the state variable S, a dimensionless quantity, which stands for 'salinity'.[4] In what follows, one may as well read S as 'specific humidity' if one has in mind the troposphere instead of the ocean.[5] Notice that there are now *three* (instead of two) independent state variables; specific volume ν, for example, is now a function of the three state variables p, T and S.

For open systems, an extra term has to be added to the thermodynamic identity (2.6), representing the effect on the energetics of the system of any changes in the concentration of the second constituent:

$$\boxed{d\epsilon = Td\eta - pd\nu + \mu dS\,.} \tag{2.7}$$

[3]The main constituents are: chloride, 55%; sodium, 30%; sulfate, 8%; magnesium, 4%; potassium, and calcium, 1%.

[4]It is expressed in 'PSU', Practical Salinity Unit, or in g/kg, or in promille.

[5]Here and elsewhere, we ignore the thermodynamic complexities of phase changes such as condensation, although they have some relevance to the study of internal waves: in the crests of large-amplitude waves in the atmosphere, condensation may take place in the rising parcels (due to adiabatic cooling), producing nice patterns of clouds. In Australia, this phenomenon is known as 'Morning Glory', see Figure 8.2.

Here μ is the *chemical potential* (dimension: J kg^{-1}).[6] It is important to note that μ involves two arbitrary constants, A and B, say: there is no empirical way to distinguish μ from $\mu' = \mu + A + BT$. Apart from these arbitrary additive terms, μ can be determined empirically by indirect means [18]. At first sight, it may seem as if the indeterminacy of μ renders (2.7) utterly meaningless; upon closer examination, however, this problem evaporates (see Section 2.3.4).

A two-component system in a gravity field is in a state of *thermodynamic equilibrium* when it satisfies the following three conditions, which were first formulated by Gibbs, in 1876 [31, pp. 144-147]:

i) hydrostatic equilibrium;

ii) a uniform temperature: $T = const$;

iii) a uniform chemical potential: $\mu = const$.

The third condition implies a remarkably strong vertical salinity gradient, of about 4 PSU per km! This is illustrated in the fictitious example of Figure 3.4, upper panels. In reality, salinity shows no such gradient (see Figure 2.2b), indicating that the ocean is far removed from thermodynamic equilibrium.

The state of thermodynamic equilibrium would be approached if the system were isolated for a sufficiently long time; molecular diffusion would then act to make temperature and the chemical potential (*not* salinity!) uniform.

2.3.3 Some definitions

The specific heat (heat capacity per unit of mass) at constant pressure is defined as

$$c_p = \frac{dQ}{dT}\Big|_{pS} . \tag{2.8}$$

The indices denote which state variables are held fixed. An alternative expression for c_p can be derived as follows. Since η is a state variable, we may write

$$d\eta \;=\; \Big(\frac{\partial\eta}{\partial p}\Big)_{TS} dp + \Big(\frac{\partial\eta}{\partial T}\Big)_{pS} dT + \Big(\frac{\partial\eta}{\partial S}\Big)_{Tp} dS\,. \tag{2.9}$$

In particular, for an isobaric and isohaline process ($dp = dS = 0$), we obtain

$$c_p \;=\; T\Big(\frac{\partial\eta}{\partial T}\Big)_{pS}\,, \tag{2.10}$$

where we also used the Second Law ($dQ = Td\eta$).

The thermal expansion coefficient α, the haline expansion coefficient β, and the speed of sound c_s (see also Appendix B) are defined by

$$\alpha = \frac{1}{\nu}\Big(\frac{\partial\nu}{\partial T}\Big)_{pS}; \qquad \beta = -\frac{1}{\nu}\Big(\frac{\partial\nu}{\partial S}\Big)_{pT}; \qquad c_s^{-2} = \Big(\frac{\partial\rho}{\partial p}\Big)_{\eta S}\,. \tag{2.11}$$

[6] More precisely, μ is the difference between the chemical potentials for salt and pure water [55, §15].

2.3.4 The Gibbs potential and the equation of state

The Gibbs potential, also called free enthalpy, is defined as $G = \epsilon - T\eta + p\nu$. Together with the thermodynamic identity (2.7), this implies

$$dG \quad = \quad \nu dp - \eta dT + \mu dS\,. \tag{2.12}$$

So, changes in G are expressed in terms of changes in pressure, temperature and salinity (or specific humidity). Observationally, these are the variables most directly accessible in the ocean or atmosphere; this places the Gibbs potential at the center of our thermodynamic considerations. We shall assume that the Gibbs potential – as a function of pressure, temperature and salinity – is known:

$$\boxed{G = G(p, T, S)\,.} \tag{2.13}$$

This is the fundamental *equation of state*. Unlike the thermodynamic identity (2.7), or (2.12), whose form remains the same whatever the medium, the equation of state depends in its form on the type of medium under consideration. For dry air, regarded as an ideal gas, we would have

$$G(p, T) = R_d T \log p + \tfrac{7}{2} R_d T (1 - \log T)$$

(apart from two arbitrary constants, see below). For seawater, any reasonably accurate form involves lengthy polynomial expressions. The task of such a construction, based on measurements, was accomplished by Feistel & Hagen [17], who were the first to fully exploit the usefulness of the Gibbs potential in the thermodynamics of seawater; this section is indeed largely based on their work. For the present discussion, the main point is that we can regard $G(p, T, S)$ as empirically established; its possibly unappealing functional form need not concern us here.

We should add the proviso that G involves four arbitrary constants, A to D:

$$G' = G + (A + BT)S + (C + DT)$$

is empirically indistinguishable from G. One can choose these constants arbitrarily. Crucially, they leave the thermodynamic identity, (2.7) or (2.12), unaffected ('gauge-invariance').[7]

For known $G(p, T, S)$, all other relevant thermodynamic variables are obtained simply by taking derivatives; we summarize the main expressions in Table 2.1 (without proof; they follow, directly or indirectly, from the thermodynamic identity (2.12)). Partial derivatives of G are indicated by indices; it is understood that the two state variables to be kept fixed are the complementary ones, so, e.g.,

$$G_p = \Big(\frac{\partial G}{\partial p}\Big)_{TS}\,.$$

[7]Specifically, one finds $\epsilon' = \epsilon + AS + C$, $\eta' = \eta - BS - D$, $\mu' = \mu + A + BT$, implying the required invariance.

specific volume	$\nu = G_p$
specific heat	$c_p = -TG_{TT}$
thermal expansion	$\alpha = G_{pT}/G_p$
haline expansion	$\beta = -G_{pS}/G_p$
entropy	$\eta = -G_T$
internal energy	$\epsilon = G - TG_T - pG_p$
chemical potential	$\mu = G_S$
speed of sound c_s	$c_s^2 = G_p^2 G_{TT}/(G_{pT}^2 - G_{pp}G_{TT})$

Table 2.1: Expressions of various state variables in terms of the Gibbs potential and its derivatives.

This procedure provides us with a collection of equations, each of which can be regarded as an 'equation of state' in its own right. The myriad of equations of state thus obtained is perhaps bewildering; it is therefore important to realize that there is, in essence, only *one* equation of state, that for the Gibbs potential, from which all the others derive. Another strength of employing the Gibbs potential lies in the fact that the ensuing equations of state are, *by construction*, mutually consistent. To illuminate this point, let us pretend (falsely) that they are independent of one another; we consider a system whose specific volume is given by $\nu = c_0 + c_1 p + C_2 T + c_3 T^2$, while we also assume that $c_p = const.$ Now, the first expression implies that G depends quadratically on T, which, in turn, implies that c_p must depend linearly on T. Hence our choice of constant c_p was inconsistent with that of ν. Such an inconsistency could *never* have arisen if we had derived ν and c_p from one and the same Gibbs potential G.

Figure 2.2 shows vertical profiles of state variables that are important in ocean physics. A few comments are in order. From Figure 2.2a,b we see that temperature and salinity vary little below 3 km depth. So, in the lowest 7 km, the steady decrease with depth of μ and c_p, and the steady increase of ρ, c_s and α, can be ascribed to increasing pressure. In the upper 1 km, on the other hand, c_s and α *de*crease with depth; here the rapid decrease of temperature dominates the effect of increasing pressure. Notice, finally, the near constancy of entropy in the lowest 7 km, as well as the small increase of temperature with depth. We return to these aspects in the following chapter, where we also discuss the concepts of potential density and potential temperature.

2.4 Complete set of governing equations

Section 2.2, the mechanics part, left us with four equations – (2.3) and (2.4) – for the five unknowns $\vec{u}$, p and ρ. We can now add the equation of state in terms of the Gibbs potential, (2.13), which also provides us with an expression for density ρ, via $\nu = G_p(p, T, S)$. By introducing the additional variables T and S, however, we are now left with five equations and seven unknowns!

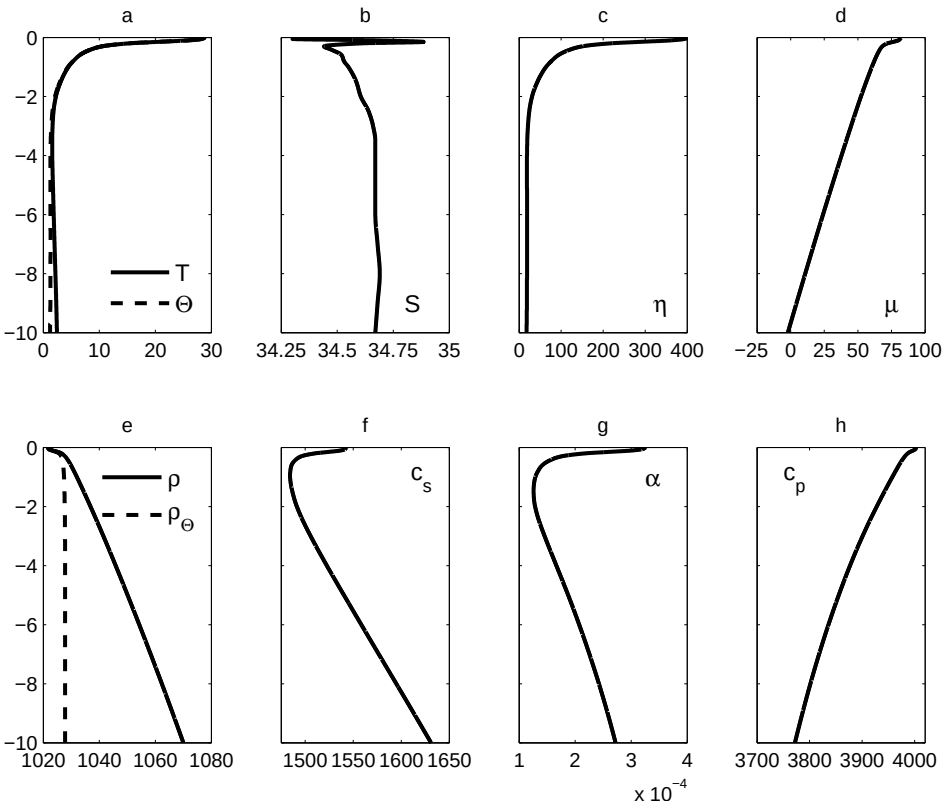

Fig. 2.2: Several thermodynamic state variables as a function of depth (in km), in the Mindanao Trench. The measured profiles of temperature T (in $^\circ$C) and salinity S (in PSU), shown in panels **a** and **b**, are taken from [39, Table 6]. From these profiles, other thermodynamic variables are derived, using the equation of state for the Gibbs potential (2.13), as determined by [17], along with the relations from Table 2.1. Thus the following profiles are obtained: **a**) potential temperature Θ; **c**) entropy η (J kg^{-1} K^{-1}); **d**) chemical potential μ (J kg^{-1} PSU^{-1}); **e**) in-situ density ρ and potential density ρ_Θ (kg m^{-3}); **f**) the speed of sound c_s (m s^{-1}); **g**) the thermal expansion coefficient α (K^{-1}); **h**) specific heat c_p (J K^{-1} kg^{-1}). We note that the profiles for entropy and the chemical potential (panels **c** and **d**) are not unique in the sense that they depend on the arbitrary constants in the Gibbs potential, see the footnote on p. 30; the profiles shown here follow from the choices made in [17].

Clearly, what is still missing is information about sources (or sinks) of salt and heat. Prescribing them provides two equations. However, the set is then still not closed because the sources of heat are prescribed via the Second Law, which involves entropy, so that an eighth variable is added to the list of unknowns. But this is resolved by using $\eta = -G_T$, which follows from (2.12) and thus implicitly hinges on the thermodynamic identity (2.7), and hence on the First Law. With this, the set is closed.

To work out these ideas in some detail, we prescribe possible sources and sinks of salinity, as well as any irreversible processes which may distribute it, by a term Q_1; thus

$$\boxed{\frac{DS}{Dt} = Q_1 \,.} \tag{2.14}$$

For the present purposes, Q_1 need not be specified, but, if needed, it could be expressed in terms of the state variables that have already been introduced. Irreversible changes, for example, can be expressed by a so-called constitutive relation; in its simplest form it can be taken as a diffusive term.[8]

Similarly, we express sources and sinks of heat, and irreversible processes enhancing the entropy, by a term Q_2, which features in the Second Law:

$$\boxed{T\frac{D\eta}{Dt} = Q_2 + T\eta_{_S}\, Q_1 \,.} \tag{2.15}$$

Here we presumed that, like Q_1, Q_2 is itself invariant with respect to the four arbitrary constants in the Gibbs potential (e.g., Q_2 consists of a diffusive term in temperature). Since η, on the left-hand side, is not invariant, this necessitates the inclusion of the second term on the right-hand side.[9]

Finally, we use the implicit form of the First Law:

$$\boxed{\eta = -G_{_T}(p, T, S) \,.} \tag{2.16}$$

All in all, we have gathered eight equations – (2.3) and (2.4) from fluid mechanics; (2.13) and (2.16) from equilibrium thermodynamics; (2.14) and (2.15) from non-equilibrium thermodynamics – for the eight unknowns $\vec{u}$, p, ρ, T, S and η, so that the set is now formally closed.

As a preparation for the next section, we derive a useful expression connecting the thermodynamic state variables η, S, p and ν. This is done as follows.

[8] As explained above, molecular diffusion strives to make the chemical potential uniform, suggesting a term like $Q_1 \sim \nabla^2\mu$. However, this would violate the requirement of invariance for the constant B, which stems from the Gibbs potential (see Section 2.3.4). To resolve this, we must introduce an extra term, such that $Q_1 \sim \nabla^2\mu - \mu_{_T}\nabla^2 T$ (the subscript denotes the partial derivative with respect to T), in which case invariance is ensured. Interestingly, this naturally couples the diffusion of the chemical potential to that of temperature, reflecting their coupling in the conditions for thermodynamic equilibrium, see p. 29.

[9] We note that in convective equilibrium, as opposed to thermodynamic equilibrium, the entropy is near-uniform, and it is generally assumed that turbulent mixing produces such a state. Now, interestingly, diffusion of entropy cannot stand by itself, because $Q_2 \sim \nabla^2\eta$ violates the requirement of invariance. This necessitates the form $Q_2 \sim \nabla^2\eta - \eta_{_S}\nabla^2 S$, ensuring invariance. This couples the diffusion of entropy to that of salinity. Note that in the deep ocean *both* are indeed nearly uniform, see Figure 2.2b,c.

Since ν is a state variable, we can write

$$\frac{D\nu}{Dt} = G_{pp}\frac{Dp}{Dt} + G_{pT}\frac{DT}{Dt} + G_{pS}\frac{DS}{Dt}\,,$$

where we expressed all coefficients in terms of derivatives of the Gibbs potential. Combining this with (2.9), cast in terms of material derivatives ($d \to D/Dt$), we can eliminate DT/Dt to obtain, after some rewriting,

$$-\frac{1}{\nu^2}\frac{D\nu}{Dt} \;=\; \frac{1}{c_s^2}\frac{Dp}{Dt} - \frac{\alpha T}{\nu c_p}\Big[\frac{D\eta}{Dt} - \eta_{\scriptscriptstyle S}\frac{DS}{Dt}\Big] + \frac{\beta}{\nu}\frac{DS}{Dt}\,,$$

where we used the expressions from Table 2.1. With (2.14) and (2.15), this becomes

$$-\frac{1}{\nu^2}\frac{D\nu}{Dt} \;=\; \frac{1}{c_s^2}\frac{Dp}{Dt} - \frac{\alpha}{\nu c_p}Q_2 + \frac{\beta}{\nu}Q_1\,. \tag{2.17}$$

2.5 Internal-wave dynamics

The only state which – from a fundamental point of view – requires no explanation is the state of thermodynamic equilibrium (characterized by a uniform temperature and chemical potential, and by hydrostatic equilibrium). Obviously the oceans and atmosphere are far removed from such a state; there are diabatic processes at work,[10] which find their origin in the differential heating by the sun, and which directly or indirectly drive the large-scale circulations in the oceans and atmosphere. All in all, these processes are responsible for the creation and maintenance of the distribution of temperature (and salinity or humidity) that we find in the oceans and atmosphere.

This distribution, in turn, determines the stratification in density. The dynamics of internal waves, and indeed their very existence, depends crucially on this stratification. While recognizing the all-important role of the stratification for our subject, we shall however not pursue the question of what gives the stratification the form it has. In other words, we shall regard the stratification as something which needs no further explanation, as something *given*. This allows us to ignore henceforth the diabatic processes, so that we can take $Q_1 = Q_2 = 0$ (i.e. processes are isohaline and isentropic). The assumption underlying this approach is that the time scale characteristic of internal waves (the wave period), is much smaller than that characteristic of diabatic processes. This means that at the time-scale of internal waves, the state of the medium would not change significantly if diabatic processes were momentarily 'switched off'. Furthermore, the motions of the fluid parcels, insofar they are due to the presence of an internal wave, can be regarded as an adiabatic process.[11]

[10] In this section we use the term 'diabatic process' in the loose sense of referring to any changes in heat or salinity.

[11] While this approach is sensible as a first-order approximation, it clearly cannot be more than that, given the role internal waves play in ocean mixing! (See the qualitative discussion in Section 1.4.)

With $Q_1 = Q_2 = 0$, then, (2.17) reduces to

$$\frac{D\rho}{Dt} \;=\; \frac{1}{c_s^2}\,\frac{Dp}{Dt}\,, \tag{2.18}$$

where we replaced ν with density, $\rho = 1/\nu$.

It is now convenient to formally eliminate temperature from the problem by using $\rho = G_p^{-1}(p, T, S)$ and assuming that we can express T in terms of ρ, p and S. This would then allow us to obtain an expression of the form $c_s = \Gamma(\rho, p, S)$, for some function Γ.

The complete set of equations now consists of the momentum equations (2.3) (for the moment, we ignore the mechanical forcing/friction term $\vec{F}$, which we re-introduce in later chapters if needed), the equation for conservation of mass (2.4), the 'energy' equation (2.18), the advection equation for salinity (2.14), and the equation of state for c_s (derived from the Gibbs potential). The starting point for the study of internal waves thus becomes:

$$\frac{D\vec{u}}{Dt} + 2\vec{\Omega}\times\vec{u} = -\frac{1}{\rho}\nabla p - g\hat{z} \tag{2.19a}$$

$$\frac{D\rho}{Dt} + \rho\nabla\cdot\vec{u} = 0 \tag{2.19b}$$

$$\frac{D\rho}{Dt} = \frac{1}{c_s^2}\,\frac{Dp}{Dt} \tag{2.19c}$$

$$\frac{DS}{Dt} = 0 \tag{2.19d}$$

$$c_s = \Gamma(\rho, p, S)\,. \tag{2.19e}$$

The set consists essentially of the five equations (2.19a,b,c) for the five unknowns $\vec{u}$, ρ and p, while (2.19e) serves as an auxiliary identity for calculating c_s, in which S is needed from (2.19d).

In conclusion, it is worthwhile to emphasize that (2.19b) and (2.19c), while both featuring $D\rho/Dt$, stem from very different physical principles: the former expresses conservation of mass, the latter originates from the First and Second Laws of thermodynamics, as is clear from its earlier form (2.17), and is thus associated with an energy equation.

Further reading

Many subtleties of the fundamentals of fluid mechanics are discussed in the encyclopedic overview by Serrin [76]. Rewarding is also the book by Lin & Segel [52] and its companion volume by Segel [75].

There are many good books on thermodynamics; especially lucid is Sommerfeld's textbook [78]; it provides both an axiomatic approach to thermodynamics (first part) and a treatment based on statistical mechanics, in particular the kinetic theory of gases (second part).

Thermodynamics is often treated somewhat negligibly in oceanographic textbooks, but in meteorology this is different; a good example is the book by Dutton [13].

Appendix A: a closer look at the Coriolis Force

In the momentum equation (2.3) we have the Coriolis force $2\vec{\Omega} \times \vec{u}$, which gives rise to four terms, with coefficients $\tilde{f} = 2\Omega \cos\phi$ and $f = 2\Omega \sin\phi$. These terms represent a deflecting force, which produces an acceleration perpendicular to velocity, as summarized in Table 2.2. Each of the terms can be derived from elementary mechanical principles, in the following way.

Initial velocity:	Induced Coriolis acceleration (in NH):
eastward (u)	southward ($-fv$) & vertically upward ($\tilde{f}w$)
northward (v)	eastward (fu)
vertically upward (w)	westward ($-\tilde{f}u$)

Table 2.2: The effect of the components of the Coriolis force in the Northern Hemisphere (NH). In the Southern Hemisphere (SH), f is negative, so in the column on the right, 'southward' is then replaced by 'northward', and 'eastward' by 'westward'.

From the terrestrial perspective, a fluid parcel, at rest on the Earth, is subject to a balance of forces. From the perspective of the 'fixed stars', the parcel traverses a latitudinal circle of radius $R_\phi = R\cos\phi$, at an eastward velocity $U = \Omega R_\phi$. The centripetal force (U^2/R_ϕ) required for this circular motion is provided by gravity and pressure gradients. We now consider three cases in which the parcel, instead of being at rest, has a velocity of its own. (I) If the parcel has an eastward velocity u, its total velocity will be $U + u$, enhancing the required centripetal force by an amount $2Uu/R_\phi$ (to first order, assuming $u \ll U$), or, using the definition of U, $2\Omega u$. From the terrestrial perspective, this increment acts as a centrifugal force, tending to sweep the parcel outward in the *latitudinal* plane. This outward acceleration can be decomposed into a radial component $2\Omega u\cos\phi = \tilde{f}u$ and a southward one, $-2\Omega u\sin\phi = -fu$, in the Northern Hemisphere (northward in the SH). This is precisely the result stated in the first row of Table 2.2. (II) If, in the NH, the parcel moves northward to latitude $\phi' = \phi + \delta\phi$, its latitudinal circle becomes smaller by an amount $-R\sin\phi\,\delta\phi$ (to first order). By conservation of angular momentum ($UR_\phi = (U+u)R_{\phi'}$), it will obtain an eastward velocity of its own: $u = \Omega\sin\phi R\delta\phi$. Parcels at rest at latitude ϕ', meanwhile, rotate at $U' = \Omega R_{\phi'} \approx U - \Omega\sin\phi R\delta\phi$. With respect to those parcels, the initially northward moving parcel will thus get an excess of eastward velocity of $2\Omega\sin\phi R\delta\phi = fR\delta\phi$, equivalent to an eastward acceleration fv, as stated in the second row of Table 2.2. (III) Finally, we consider a parcel that moves initially upward (i.e. radially outward), to $R' = R + \delta R$. Again from

the perspective of the 'fixed stars' and by conservation of angular momentum ($UR_\phi = (U+u)R'_\phi$), the parcel will get a smaller eastward velocity by an amount $u = -\Omega\cos\phi\delta R$. Ambient parcels at rest at this higher altitude have an eastward velocity $U' = \Omega R'_\phi$, an excess compared to U, that is, of $\Omega\cos\phi\delta R$. With respect to those parcels, the initially upward moving parcel will thus get an excess of westward velocity of $-2\Omega\cos\phi\delta R = -\tilde{f}\delta R$, equivalent to a westward acceleration $-\tilde{f}w$, as stated in the third row of Table 2.2.

Appendix B: sound waves

In Eq. (2.11) we have introduced the "speed of sound" c_s as yet another thermodynamic state variable. We have yet to show that it indeed *acts* as the speed of sound. In the set (2.19) there are a number of restoring forces at work; the two that are essential to internal waves are gravity (i.e. buoyancy) and the Coriolis force. For sound waves, on the other hand, pressure gradients act as the restoring force. In order to isolate pure sound waves from the problem, we should therefore abandon the other restoring forces; hence we take $g = 0$ and $\vec{\Omega} = 0$. We moreover ignore salinity, for the sake of simplicity. We consider a motionless state of thermodynamic equilibrium; both ρ and p are uniform in such a state, and we denote their values by ρ_c and p_c, respectively. Furthermore $\vec{u} = 0$. We now consider small perturbations with respect to this equilibrium state: $\rho = \rho_c + \rho'$, $p = p_c + p'$ and a velocity $\vec{u}'$. The 'smallness' means that we may neglect products of perturbation (i.e. primed) terms; hence we obtain from (2.19):

$$\rho_c \frac{\partial \vec{u}'}{\partial t} = -\nabla p' \tag{2.20a}$$

$$\frac{\partial \rho'}{\partial t} + \rho_c \nabla\cdot\vec{u}' = 0 \tag{2.20b}$$

$$\frac{\partial \rho'}{\partial t} = \frac{1}{c_s^2}\frac{\partial p'}{\partial t} \tag{2.20c}$$

$$c_s = \Gamma(\rho_c, p_c)\,. \tag{2.20d}$$

The first three of these equations are easily combined into one:

$$\frac{\partial^2 p'}{\partial t^2} - c_s^2\nabla^2 p' = 0\,, \tag{2.21}$$

which is the *wave-equation*. The simplest case occurs if we restrict the problem to one spatial dimension, x say, in which case the general solution can be written

$$p' = F(x + c_s t) + G(x - c_s t)\,,$$

for arbitrary functions F and G, describing left- and rightward propagating waves, respectively. This bornes out the role of c_s as the speed of sound.

A few inferences are in order. First, sound waves are *dispersionless*, i.e. their phase speed does not depend on wavelength. This is, in fact, obvious from

everyday experience: if sound waves were dispersive, speech, not to speak of music, would result in an unintelligible cacophony. (Some would argue that modern music occasionally attains this quality even with dispersionless sound waves.) The color of a voice – that what allows us tell one voice from another – stems from the intensity of "overtones", i.e. the tones whose frequencies are a multiple of the basic tone. If they would travel at different speeds, the signal would entirely lose its coherence.

Second, the speed of sound c_s is the "Laplacian" one, defined by (2.11):

$$c_s^{-2} = \left(\frac{\partial \rho}{\partial p}\right)\Big|_{\eta S},$$

i.e. (de)compression of fluid parcels is assumed to take place under constant entropy η (and salinity). This expression is another way of writing (2.18), which was, after all, based on the assumption that entropy and salinity are constant. Without proof, we note that for an ideal diatomic gas (dry air), the expression reduces to $c_s^2 = \frac{7}{5} R_d T$.

Historically, this expression was preceded by the "Newtonian" speed of sound, in which temperature instead of entropy was assumed to be constant during (de)compression. For dry air, this assumption leads to a smaller speed, given by $\tilde{c}_s^2 = R_d T$. This would have been correct if parcels move so slowly that their excess in temperature during compression is annulled by the exchange of heat with their surroundings. This is however not the case [70, §246]; the observed speed of sound in dry air agrees with the "Laplacian" one.

Chapter 3

Local static stability

3.1 The buoyancy frequency

In later chapters we will think of internal waves as oscillatory motions in an ocean or atmosphere that is otherwise at rest, in other words as perturbations to a static background state. This is why we must first consider in detail the static state itself (this chapter), before turning to internal waves. Central in the theory of internal waves stands the so-called Brunt-Väisälä or buoyancy frequency (N), which is a measure of the strength of the vertical stratification in density; it is an indicator for the local gravitational stability of the stratification. A typical distribution of N in the ocean is shown in Figure 1.6.

Generally, the concept of 'stability' tells us how a system responds to a perturbation. To use the concept in a meaningful way, one has to specify precisely *what sort* of perturbations one is looking at. Here, in dealing with the question of local static stability, we consider perturbations in the form of infinitesimal vertical displacements of fluid parcels (we will see in a moment that this specification is not sufficient). We call a fluid stably stratified if the displaced parcel tends to return to its original position, unstably stratified if it tends to move further away from its original position, and neutrally stratified if it tends to stay where it is. A naive look at Figure 2.2e (solid line) would perhaps suggest that the density distribution is quite stable since the parcels are denser at deeper positions. The real question, however, is whether they would still be denser if they were displaced to a higher position; we have to take into account the effect that quantities like the parcel's temperature and density may change by (de)compression as the parcel is vertically displaced. To make the argument explicit, let us assume that the vertical density distribution, belonging to the static state, is known: $\rho = \rho_0(z)$. Let a parcel be moved from its initial position $z - \delta z$, where its density was $\rho_0(z - \delta z)$, upwards to a new position z; we denote its change in density by $\delta\rho$ (Figure 3.1). The parcel will move back towards its initial position when it is heavier than the surrounding fluid; hence the criterion

for stability is:

$$\rho_0(z-\delta z)+\delta\rho \quad > \quad \rho_0(z) \tag{3.1}$$

or, in a Taylor expansion about z,

$$\rho_0(z)-\frac{d\rho_0}{dz}\,\delta z+\cdots+\delta\rho \quad > \quad \rho_0(z)\,.$$

For infinitesimal displacements ($\delta z \to 0$) this becomes

$$-\frac{d\rho_0}{dz}+\frac{\delta\rho}{\delta z} \quad > \quad 0\,. \tag{3.2}$$

This expression contains two types of density gradients: the first term simply gives the rate at which the static density varies with height; the second, the rate at which the parcel's own density changes during its vertical displacement. It is the difference between the two that determines whether the stratification is stable or not.

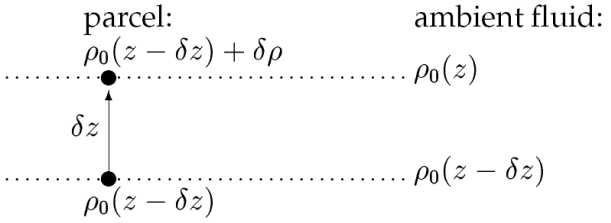

Fig. 3.1: A fluid parcel (black blob) is moved upward from position $z-\delta z$ to z.

At this stage we can formally introduce the buoyancy frequency N as

$$\boxed{N^2=-\frac{g}{\rho_0}\left(\frac{d\rho_0}{dz}-\frac{\delta\rho}{\delta z}\right)} \tag{3.3}$$

where g is the acceleration due to gravity. Hence, in terms of N, the criterion for local static stability becomes

$$\boxed{N^2>0\,.} \tag{3.4}$$

Similarly, the fluid is unstably stratified if $N^2<0$, and neutrally stratified if $N^2=0$.

The next problem is to find an expression for $\delta\rho/\delta z$. It is here that thermodynamic principles come into play: the problem cannot be resolved unless we specify the perturbation also from a *thermodynamic* point of view. We will assume that the displacement of the parcel is such that its entropy and salinity

(or humidity) are conserved. The conservation of entropy is guaranteed if the parcel exchanges no heat and salinity with its surroundings and moves sufficiently slowly to remain (nearly) in thermodynamic equilibrium (see eq. (2.15)). This does not contradict the fact that the fluid *as a whole* may well be far removed from thermodynamic equilibrium: the 'thermodynamic system' that we consider is the *parcel*, and the surrounding fluid plays a role merely in providing the pressure that the parcel experiences.

Before we proceed to derive expressions for N in terms of observable quantities (temperature, salinity, density), we furnish N with a simple interpretation. We consider again the vertical displacement of the parcel (δz), and denote its vertical acceleration by $\ddot{\delta z}$ (i.e. the second time derivative of δz). After its displacement from $z-\delta z$ to z (see Figure 3.1), the parcel experiences a buoyancy (or Archimedean) force, which is gravity g times the difference in density between the ambient fluid and the parcel. To a first approximation, this force is given by $g(\delta z\, d\rho_0/dz - \delta\rho)$. By Newton's second law, the force equals the acceleration of the parcel (times its density); the latter is, again to first approximation, $\rho_0(z)\,\ddot{\delta z}$. Hence

$$\ddot{\delta z} + N^2\,\delta z = 0\,.$$

This equation describes a harmonic oscillator, similarly as for a spring in classical mechanics: for $N^2 > 0$, the parcel oscillates vertically at frequency N about its equilibrium position, the role of N being analogous to the stiffness of a spring.

3.2 N in terms of density

The criterion for stability (3.4) attains its meaning from the definition of N in (3.3). We now have to find expressions for N in terms of observable quantities. Using the results from thermodynamics that were discussed in Chapter 2, we derive an expression for N in terms of the static temperature and salinity profiles (see next section), and one in terms of the static density profile $\rho = \rho_0(z)$ alone.[1] In practice, density is not measured directly but obtained via the equation of state $\nu = G_p(p, T, S)$ (see Section 2.3.4), using known vertical profiles of pressure, temperature and salinity.

In the static state, pressure and density satisfy the hydrostatic balance (see (2.19a) with $\vec{u} = 0$):

$$\frac{dp_0}{dz} \;=\; -\rho_0 g\,. \tag{3.5}$$

By assumption, the parcel's entropy and salinity are conserved during its displacement, hence we can use (2.18) to express its change in density; this implies

$$\delta\rho \;=\; c_s^{-2}\,\delta p\,. \tag{3.6}$$

[1] Here and in the rest of this chapter, the index '0' refers to *static* profiles, while δ refers to changes concerning the displaced parcel.

Here c_s is the speed of sound in the *parcel*, not that in the fluid surrounding it. At each instant during the displacement, the pressure that the parcel experiences is given by the local static value p_0; hence, by the hydrostatic balance (3.5), the change in pressure δp becomes $\delta p = -\rho_0 g \delta z$. Collecting these results in (3.3), we obtain

$$\boxed{N^2 = -\frac{g}{\rho_0}\left(\frac{d\rho_0}{dz} + \frac{\rho_0 g}{c_s^2}\right).} \tag{3.7}$$

The terms on the right-hand side have *opposite* signs, the first term on the right-hand side having a stabilizing effect, while the second term has a destabilizing effect. In the larger part of the ocean (i.e. outside the thermocline), as well as in the troposphere, the two terms are of the same order of magnitude. This is easily verified for the ocean, using Figure 2.2e,f. Density ρ increases from 1022 at the surface to 1070 $\mathrm{kg\,m^{-3}}$ at 10 km depth (testifying to the compressibility of seawater!), giving a gradient $d\rho_0/dz = -0.0048$ $\mathrm{kg\,m^{-2}}$. The mean speed of sound is $c_s = 1545$ $\mathrm{m\,s^{-1}}$, so that $\rho_0 g/c_s^2$ becomes 0.0043 $\mathrm{kg\,m^{-2}}$, being of about the same magnitude, but of opposite sign, as the density gradient. In other words, the second term cannot be neglected; in fact, it acts as an essential 'correction' to the first term, to annul the effects of compressibility.

3.3 N in terms of temperature and salinity

Here we derive an alternative expression for N^2, in terms of the (more directly observable) distributions of static temperature and salinity. Choosing pressure, temperature and salinity as the independent state variables, we can express the parcel's change in density $\delta\rho$ (see Figure 3.1) as

$$\delta\rho = \left(\frac{\partial\rho}{\partial p}\right)_{TS}\delta p + \left(\frac{\partial\rho}{\partial T}\right)_{pS}\delta T + \left(\frac{\partial\rho}{\partial S}\right)_{pT}\delta S\,. \tag{3.8}$$

By assumption, the parcel's salinity is conserved ($\delta S = 0$), so the last term vanishes. Because of the conservation of entropy and salinity, (2.9) and (2.16) together imply that the parcel's change in temperature can be expressed as $\delta T = -\delta p\, G_{pT}/G_{TT}$. Using Table 2.1 gives

$$\delta T = \frac{\alpha T}{\rho c_p}\,\delta p\,. \tag{3.9}$$

The quantities in the coefficient of δp refer to the *parcel's* state, and thus change as the parcel is moved. The change in pressure δp follows from the hydrostatic balance (3.5); using this and (3.9), we obtain from (3.8):

$$\delta\rho \;=\; -\left(\left(\frac{\partial\rho}{\partial p}\right)_{TS} + \left(\frac{\partial\rho}{\partial T}\right)_{pS}\frac{\alpha T}{\rho c_p}\right)\rho_0 g\delta z\,. \tag{3.10}$$

We assume the static temperature profiles $T_0(z)$ and $S_0(z)$ to be known; the static density gradient can then be written as

$$\frac{d\rho_0}{dz} = \left(\frac{\partial\rho}{\partial p}\right)_{TS} \frac{dp_0}{dz} + \left(\frac{\partial\rho}{\partial T}\right)_{pS} \frac{dT_0}{dz} + \left(\frac{\partial\rho}{\partial S}\right)_{pT} \frac{dS_0}{dz}\,. \tag{3.11}$$

Substitution of (3.10) and (3.11) in (3.3) gives

$$\boxed{N^2 = g\left[\alpha\left(\frac{dT_0}{dz} + \frac{\alpha T_0 g}{c_p}\right) - \beta\,\frac{dS_0}{dz}\right]} \tag{3.12}$$

where we used the definitions of α and β in (2.11), and the hydrostatic balance (3.5). The thermodynamic parameters α, β and c_p are functions of pressure, temperature and salinity, for which we can use the local static values p_0, T_0 and S_0, because the displacement of the parcel is assumed to be infinitesimal.

3.4 A practical example

In the ocean, vertical distributions of temperature and salinity are obtained from CTD profiling, a device that measures conductivity (hence salinity), temperature, and depth (or pressure, rather). For example, Figure 1.6 is based on 185 such profiles, made at a range of latitudes. In this section we look further into the practical usage of (3.7); as it turns out, obtaining respectable profiles of N is a little less straightforward than it may seem. (The usage of (3.12) is similar, and is not further discussed.)

We consider the profile at latitude 40°N in Figure 1.6. From the measured profiles of temperature and salinity (Figure 3.2a,b) we calculate density ρ and the speed of sound c_s using the equation of state for the Gibbs potential, see Section 2.3.4. This yields the profiles shown in Figure 3.2c,d.[2] Next we calculate the gradient of density, using a 2nd-order central difference scheme:

$$\left.\frac{\partial\rho}{\partial p}\right|_k = \frac{\rho_{k+1} - \rho_{k-1}}{2\Delta p}\,,$$

i.e. the gradient at pressure level k is calculated by taking the difference of density at levels $k+1$ and $k-1$ divided by the pressure difference between the levels, $2\Delta p$; here $\Delta p = 2$ dbar $= 2 \times 10^4$ Pa.

We can now use (1.1), which is the equivalent of (3.7); this produces the noisy line in Figure 3.3a. The noisiness is caused by small-scale fluctuations in the density gradient. They are of some interest in themselves; they are sometimes due to small steps in temperature and salinity caused by double diffusion. It is

[2] Plots like these commonly have depth as the vertical coordinate rather than pressure, even though the latter is used in the equation of state; a convenient and accurate formula for the conversion of pressure (in dbar) to depth (in m) is $z = (1 - c_1)p - c_2 p^2$ with $c_1 = (5.92 + 5.25\sin^2\phi) \times 10^{-3}$ and $c_2 = 2.21 \times 10^{-6}$, where ϕ is latitude [74].

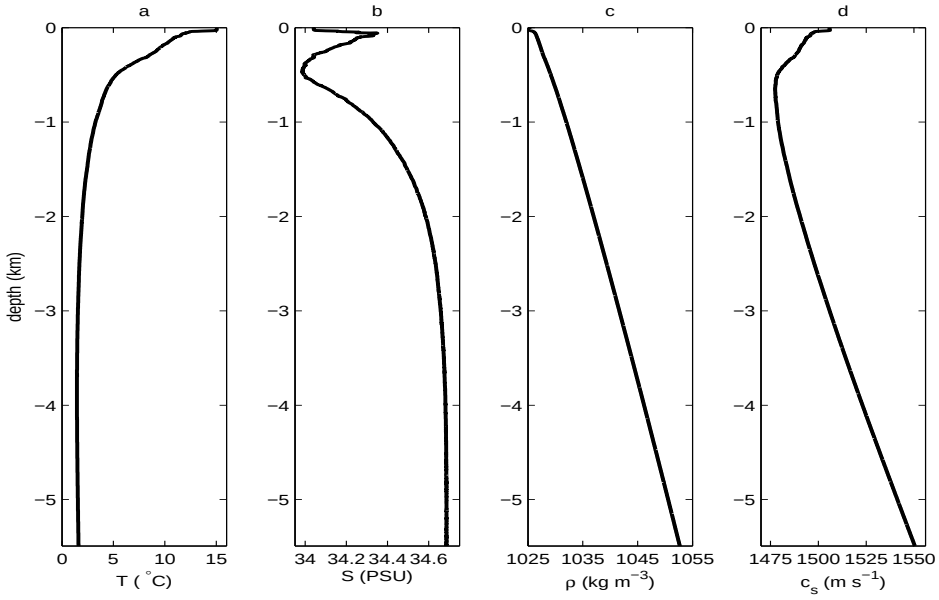

Fig. 3.2: Profiles of temperature, salinity, density and the speed of sound, at 40°N in the Pacific Ocean, from WOCE section P14, see Figure 1.6.

important to realize that Figure 3.3a is based on just one profile of temperature and salinity; repeated profiling at the same location over longer periods of time would partly remove the wiggles. In the absence of repeated profiles, it makes sense to smooth out the wiggles by taking a running mean. Here the running mean at a point is calculated by taking the average of its value and those of its 7 upper and 7 lower neighbours; the smoothing results in the thick blue line in Figure 3.3a. Finally, incidental negative values are set to zero, which allows us to calculate N, shown in Figure 3.3b.

Notice that there are two clearly identifiable peaks in N in the upper layer. The largest peak occurs in the upper 100 m, and has a seasonal character; it largely disappears during winter. Hence its name, seasonal thermocline. (In the tropics, however, it has a permanent character.) Beneath it, near 400 m, lies a smaller peak that is not affected by seasonal influences and hence is called a permanent pycnocline. Its depth depends on the geographical location; in the Bay of Biscay, for example, it lies at about 1000 m depth.

3.5 Special types of stratification

Below we discuss what the stratification is like for two cases of fundamental importance.

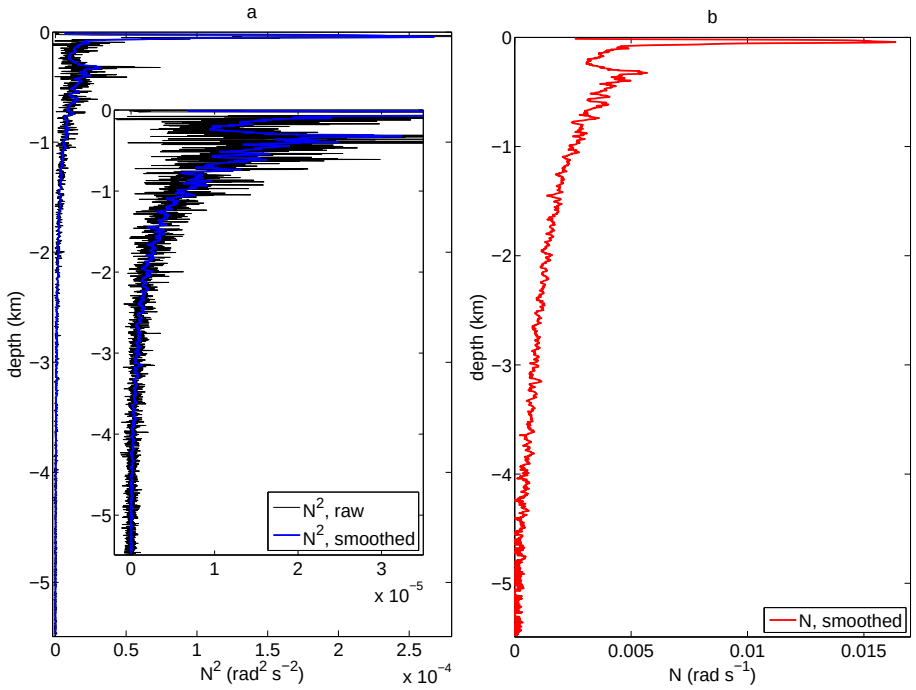

Fig. 3.3: Profiles of N^2 (panel **a**) and N (panel **b**) at 40°N in the Pacific Ocean (WOCE section P14, see Figure 1.6). In **a**, the noisy thin black line shows N^2 obtained from (1.1), using the original data at pressure intervals of 2 dbar; the inserted figure shows a zoom. The blue thick line shows the result of smoothing by taking the running mean over 15 points (i.e. stretches of 30 dbar); moreover, incidental negative values are set equal to zero. Panel **b** shows the square root of the smoothed profile of panel **a**.

3.5.1 A fluid in thermodynamic equilibrium

The state of thermodynamic equilibrium is special from a fundamental point of view, because of all the states with the same energy (i.e. internal plus potential energy) and the same mass, the equilibrium state has the largest entropy. Any isolated body of fluid will evolve toward thermodynamic equilibrium, due to molecular diffusion; the eventual state is one in which temperature and chemical potential are uniform (Figure 3.4, upper panels).

In thermodynamic equilibrium, then, temperature is uniform, $T_0(z) = T_c$. If no salt is present ($S_c = 0$), then the stratification follows from (3.12) as

$$N^2 \quad = \quad \frac{\alpha^2 T_c g^2}{c_p} \,. \tag{3.13}$$

The right-hand side is always positive, so *a fluid in thermodynamic equilibrium is always stably stratified*. This conclusion holds a fortiori if salinity is present, since in thermodynamic equilibrium salinity increases strongly with depth (see the end of Section 2.3.2, and Figure 3.4, upper panels), the chemical potential

being uniform. The last term in (3.12) then makes the stratification even more stable.

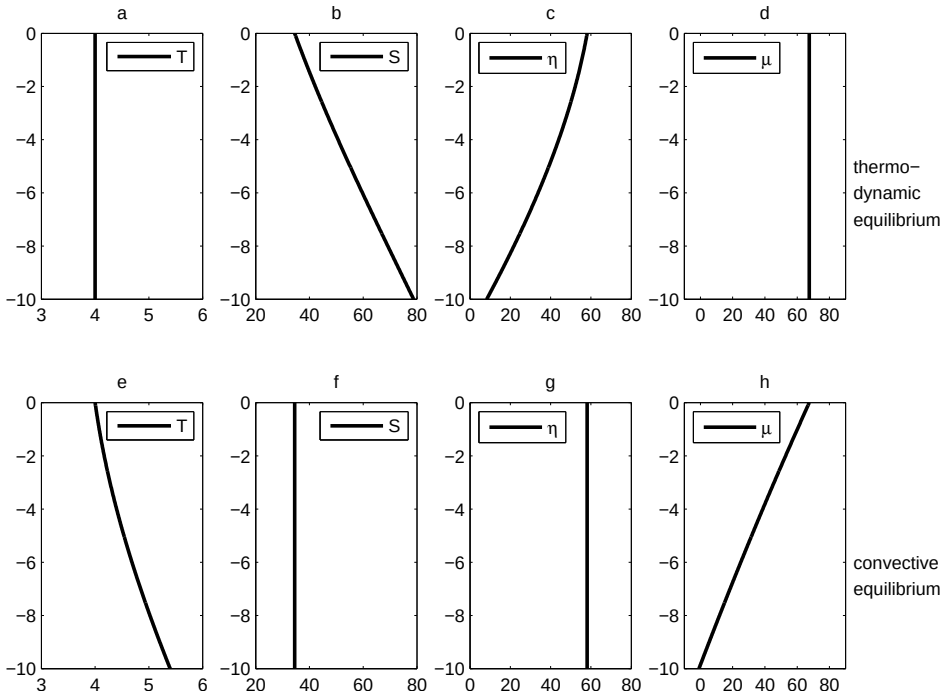

Fig. 3.4: Two fictitious ocean states. Upper panels: thermodynamic equilibrium; lower panels: convective equilibrium. In the upper panels temperature is chosen uniform (**a**), while salinity (**b**) is chosen such that the chemical potential μ (**d**) is uniform; hence a state of thermodynamic equilibrium. In the lower panels, salinity is chosen uniform (**f**), while temperature (**e**) is chosen such that entropy η (**g**) is uniform; this describes a state in which entropy and salinity are mixed, as in "convective equilibrium". Units as in Figure 2.2.

3.5.2 A turbulently mixed fluid

Molecular diffusion is a rather slow process. In geophysical fluid dynamics, it is turbulent (rather than molecular) mixing that often stamps the state of the fluid; it presumably tends to mix entropy and salinity, instead of temperature and chemical potential (see also footnotes on p. 33).

Let the static entropy profile be given by $\eta_0(z)$. We can then express the static density gradient as

$$\frac{d\rho_0}{dz} = \left(\frac{\partial\rho}{\partial p}\right)_{\eta S}\frac{dp_0}{dz} + \left(\frac{\partial\rho}{\partial\eta}\right)_{pS}\frac{d\eta_0}{dz} + \left(\frac{\partial\rho}{\partial S}\right)_{p\eta}\frac{dS_0}{dz} \tag{3.14}$$

$$= -\frac{\rho_0 g}{c_s^2} + \left(\frac{\partial\rho}{\partial\eta}\right)_{pS}\frac{d\eta_0}{dz} + \left(\frac{\partial\rho}{\partial S}\right)_{p\eta}\frac{dS_0}{dz} \tag{3.15}$$

where we used (3.5) and (3.6). Thus (3.7) can be rewritten as

$$N^2 \quad = \quad -\frac{g}{\rho_0}\Big[\Big(\frac{\partial\rho}{\partial\eta}\Big)_{pS}\frac{d\eta_0}{dz} + \Big(\frac{\partial\rho}{\partial S}\Big)_{p\eta}\frac{dS_0}{dz}\Big]\,. \tag{3.16}$$

Now, if entropy and salinity are uniform – a state called "convective equilibrium" (Figure 3.4, lower panels) – we obtain $N^2 = 0$, i.e. the stratification is neutrally stable.

Such a state is sometimes referred to as 'adiabatic' or 'isentropic'.[3] This terminology is misleading, for contrary to what the terms may suggest, *diabatic* processes must be involved to maintain such a state; otherwise the system would evolve towards thermodynamic equilibrium – an isothermal state. Put differently, for the state to stay away from thermodynamic equilibrium, it is necessary that the entropy that is inevitably being continually produced in the interior of the fluid, be expulsed to its surroundings; this is accomplished by a flow of heat through the system, with low-entropic incoming and high-entropic outgoing energy. The Earth's atmosphere/ocean system is a case in point; low-entropic short-wave radiation is received from the sun, while the Earth emits high-entropic long-wave radiation. The flow of energy is energetically neutral, but allows entropy to be expulsed. Similarly, living organisms do not eat to gain energy, but to keep down entropy (and thus to postpone as long as possible the sorry state of thermodynamic equilibrium).[4]

3.6 Potential density

The relation between potential density and static stability is a complicated one. It is instructive to introduce first the concept of a 'generalized' potential density, which does provide a simple connection to static stability. From it, the potential density, in its ordinary sense, can be deduced, but then the simple connection to static stability is lost.

3.6.1 Generalized potential density

Imagine that a fluid parcel were moved from its initial vertical position z to a certain reference level z_r, under conservation of its entropy and salinity. Then we define the density that the parcel would attain at its new position (i.e. at z_r) as its potential density $\rho_r(z_r, z)$. Usually one takes z_r fixed (see next section), but here we shall regard z_r as a *variable* (hence the adjective 'generalized').

The potential density can be obtained by integrating (3.6); this gives

$$\rho_r(z_r, z) \quad = \quad \rho_0(z) + g\int_{z_r}^{z} dz'\,\frac{\rho_0(z')}{c_s^2(z, z')}\,.$$

[3] The terms 'adiabatic' and 'isentropic' are not equivalent and should therefore not be used interchangeably. For example, the free expansion of a gas in a cylinder, being an irreversible process, is adiabatic (i.e. there is no heat exchange with its surroundings), but not isentropic.

[4] See also the thought-provoking discussion of Emden's "Warum heizen wir im Winter?" by Sommerfeld [78, p. 38].

The left-hand side thus denotes the density a parcel, originally at z, would attain when brought adiabatically to z_r. The speed of sound in the parcel, c_s, depends on its original height z, as well as on the height z', which it passes during the displacement. This is because the speed of sound depends on the state variables pressure, entropy and salinity, the first of which takes the ambient value $p_0(z')$, whereas the last two retain their original values $\eta_0(z)$ and $s_0(z)$. This is an important point, because differentiation to z now still yields an integral-term:

$$\frac{\partial \rho_r}{\partial z}(z_r, z) \quad = \quad \frac{d\rho_0}{dz} + \frac{\rho_0 g}{c_s^2} + g \int_{z_r}^{z} dz' \, \rho_0(z') \frac{\partial c_s^{-2}}{\partial z}(z, z') \, .$$

Comparison with (3.7) shows that

$$N^2(z) \quad = \quad -\frac{g}{\rho_0} \frac{\partial \rho_r}{\partial z}(z_r, z) + \frac{g^2}{\rho_0} \int_{z_r}^{z} dz' \, \rho_0(z') \frac{\partial c_s^{-2}}{\partial z}(z, z') \, . \tag{3.17}$$

The quantity ρ_r depends on two independent variables: z_r and z. By traversing the z_r, z-plane along a certain line, we obtain a transect of the surface $\rho_r(z_r, z)$. Of special importance is the transect along the line $z_r = z$; this yields the static density profile, since $\rho_r(z, z) = \rho_0(z)$. Furthermore, for $z_r = z$, all that remains of (3.17) is the simple expression

$$N^2(z) \quad = \quad -\frac{g}{\rho_0} \frac{\partial \rho_r}{\partial z}(z, z) \, . \tag{3.18}$$

Notice that the derivative on the right-hand side is *not* the gradient of the static density ρ_0. We follow the line $z_r = z$, along which (as we saw above) the values of ρ_r are indeed equal to ρ_0; the partial derivative in (3.18), however, does not produce the gradient of this curve, but rather the derivative in an oblique direction (i.e. in a direction parallel to the z-axis).[5]

3.6.2 The common concept of potential density

Potential density, in its ordinary sense, is defined in the same way as was ρ_r above, except that z_r is taken to be *fixed*. To be specific, we shall choose $z_r = 0$ (sea surface, level of standard atmospheric pressure), but the following statements are equally valid for any other fixed value. To make this change in the meaning of z_r explicit, we introduce the symbol

$$\rho_\Theta(z) = \rho_r(0, z) \, .$$

Notice the difference from the previous section (where we followed the line $z_r = z$): we now travers the z_r, z-plane along a line $z_r = const.$

The difference between (in-situ) density ρ and potential density ρ_Θ is shown in Figure 2.2e. In an incompressible fluid the two profiles would be identical;

[5] One may introduce a function $\rho_l(z)$, the 'local' potential density [87], defined by equating $d\rho_l/dz$ with the partial derivative on the right-hand side of (3.18); this provides us with a function whose gradient is proportional to N^2.

thus, the fact that the profiles (and their gradients, in particular) are quite distinct testifies to the importance of compressibility.

It now follows immediately from (3.17) that

$$N^2(z) \;=\; -\frac{g}{\rho_0}\frac{d\rho_\Theta}{dz}(z) + \frac{g^2}{\rho_0}\int_0^z dz'\,\rho_0(z')\frac{\partial c_s^{-2}}{\partial z}(z,z')\,. \tag{3.19}$$

Due to the presence of the second term on the right-hand side, the signs of N^2 and $-d\rho_\Theta/dz$ need not be the same; in other words, the sign of the gradient of ρ_Θ is not, in principle, indicative of static (in)stability. In practice, however, the second term on the right-hand side usually turns out to be negligible unless the stratification comes close to neutral stability, $N \approx 0$. In other words, it is a fact of experience that the gradient of potential density is indicative of (in)stability under most, *but not all* oceanographic conditions. The deeper parts of the South-Atlantic Ocean provide a negative example; here the exact criterion proves that the stratification is stable, whereas the sign of $d\rho_\Theta/dz$ is positive (see [68, §7.352]). One usually deals with this problem by dividing the water column in a certain number of layers, and choosing a separate reference level z_r for each layer. This is certainly more accurate, but it still does not produce the exact result, for this requires the introduction of infinitely many reference levels, as in the previous section. Hence it is advisable to use always the exact criterion (3.4), together with either (3.7) or (3.12).

It is only for the special case of uniform salinity that one finds a simple expression in terms of potential density. From its definition it follows that potential density ρ_Θ is a function only of the parcel's entropy and salinity. Thus we can write

$$\frac{d\rho_\Theta}{dz} \;=\; \Big(\frac{\partial\rho_\Theta}{\partial\eta}\Big)_S\frac{d\eta_0}{dz} + \Big(\frac{\partial\rho_\Theta}{\partial S}\Big)_\eta\frac{dS_0}{dz}\,.$$

Recalling (3.16), we see that if salinity is uniform ($S_0(z) = S_c$), the following identity holds:

$$N^2 \;=\; -\frac{g}{\rho_0}\Big(\frac{\partial\rho}{\partial\eta}\Big)_{pS}\Big(\frac{\partial\rho_\Theta}{\partial\eta}\Big)_S^{-1}\frac{d\rho_\Theta}{dz}\,. \tag{3.20}$$

If the partial derivatives on the right-hand side have constant signs throughout the fluid, then one can conclude that the sign of $d\rho_\Theta/dz$ is indicative of (in)stability.

3.7 Limitations of the concept of local stability

So far we have assumed that vertical displacements are *infinitesimal*, hence we speak of *local* stability. This concept works fine if the entire column is stably stratified. However, layers that are locally *un*stable will collapse and be subject to turbulent mixing. This means that we have to take into account the *finite* displacement of parcels; the criterion of local stability is then no longer

necessarily meaningful. In fact, convective mixing may extend into locally stable layers.

To find the vertical extent of the region affected by local instability, a simple rule can be established. Parcels at the top of the locally unstable layer will descend, parcels at its base will rise. These descending and rising movements are not necessarily limited to the locally unstable layer itself; they will continue until the parcels arrive in a surrounding that matches their instanteneous density. This is illustrated in Figure 3.5. In this atmospheric example, effects of humidity are neglected. Local stability can then be defined by the gradient of potential density (see (3.20)), or of the similarly defined potential temperature. The criterion of local stability identifies a layer of unstable stratification (Figure 3.5a). The descending parcel from the top of this layer, and the rising parcel from its base (denoted by circles), will continue their vertical motion until they arrive at a level at which the potential temperature equals their own (Figure 3.5b). So, the effectively unstable layer is much thicker than the criterion of local stability would suggest.

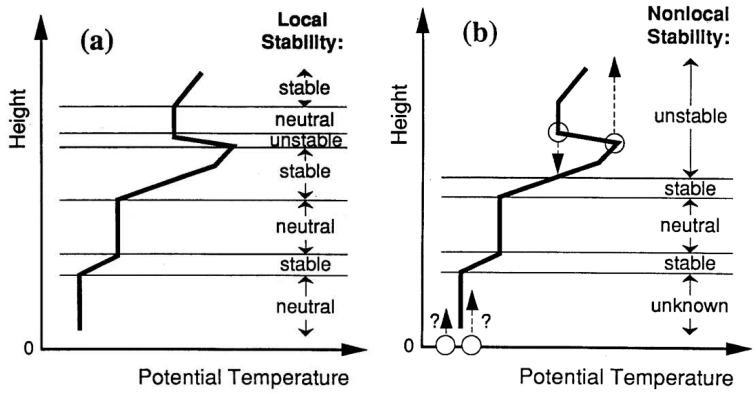

Fig. 3.5: A fictitious atmospheric profile of potential temperature. Negative (positive) gradients are indicative of locally unstable (stable) layers, as shown in **a**. However, descending and rising parcels from the locally unstable layer will continue their movement until they arrive at a level that matches their own potential temperature (**b**). As a result, neighbouring *locally stable* layers will be subject to convective mixing as well. (The question marks in **b** refer to the possibility that surface heating may render the locally neutrally stratified bottom layer unstable.) From [81].

Chapter 4

Approximations

4.1 Introductory remarks

The fundamental equations of geophysical fluid dynamics defy direct analytical treatment. Even in a numerical approach one faces formidable complications. Hence it is desirable to introduce approximations and thereby to simplify the problem. The procedure by which this is done usually consists of a series of steps. First each variable in the equation is replaced by its typical scale times a dimensionless variable. Then, by collecting scales in each term, certain dimensionless parameters are identified, some of which may be considered small on empirical grounds, or just by assumption. Finally, the terms involving a small parameter are neglected.

Implicit in this procedure is the idea 'the smaller a term, the less important'. An utterly trivial example may serve to make this look less obvious. Consider the algebraic equation $a = b + c$, and let us assume that all terms are of order one; hence none of them is negligible. Next we rewrite the equation by defining $\tilde{b} = b + 1000$ and $\tilde{c} = c - 1000$, giving $a = \tilde{b} + \tilde{c}$. The term on the left-hand side is now three orders of magnitude smaller than either of the terms on the right, and (according to a naive way of reasoning) can therefore be considered negligible. At the same time, obviously, the term cannot have become negligible merely by rewriting the equation.

The problem lies in the mistaken notion that terms can be negligible in an equation as such. Upon closer examination, what happens in an approximating procedure is this: we approximate, i.e. *replace*, the exact fields (e.g. current velocity, density etc.) by slightly different ones. Not surprisingly, these approximated fields satisfy an equation that *differs* from the original one. This may seem a rather scholastic point, but the following example will illustrate its specific meaning.

In a later section we will encounter an equation that provides the boundary

condition at the upper surface; after having been scaled, the equation reads:

$$\epsilon \frac{\partial \hat{p}'}{\partial \hat{t}} \quad = \quad \hat{w}\,, \tag{4.1}$$

where a 'hat' indicates that the variable is scaled and dimensionless ($\hat{p}'$ denotes the departure of pressure from its hydrostatic value). Empirically it is known that the dimensionless parameter ϵ is much smaller than one. What can we infer from that? Obviously, one cannot say that "the left-hand side is small compared with the right" (a lapsus in a famous textbook); the equality sign, after all, tells us that the two sides are *equal*. What we can do, however, is to expand the variables $\hat{p}'$ and $\hat{w}$ in a series, in which ϵ serves as the small parameter:

$$\hat{p}' = p^{(0)} + \epsilon p^{(1)} + \epsilon^2 p^{(2)} + \cdots \qquad\qquad \hat{w} = w^{(0)} + \epsilon w^{(1)} + \epsilon^2 w^{(2)} + \cdots\,,$$

where all the $p^{(i)}$ and $w^{(i)}$ are assumed to be of order one. Subsitution in (4.1) then gives, to first order,

$$w^{(0)} \quad = \quad 0\,, \tag{4.2}$$

showing that $w^{(0)}$ vanishes at the surface. This is a simpler equation than (4.1), but notice that (4.1) and (4.2) are about *different* variables; the former deals with the original vertical velocity $\hat{w}$, the latter with its first-order approximation $w^{(0)}$.

However, this precise way of writing becomes unwieldy when several approximations are made, involving different small parameters; in the end, we would be working with things like $w^{(0,0,0,0,0,0)}$. This is why the explicit expansion is usually left out altogether, so that we jump directly from (4.1) to (4.2), the latter moreover being written as $\hat{w} = 0$ (incorrectly suggesting that we are still dealing with the original variable; and opening the trap of substituting approximated expressions back into the original equations...). Furthermore, we describe the process somewhat fuzzily by saying that we have approximated (4.1) by (4.2); in the next sections we often follow this parlance, for the sake of brevity.

4.2 Overview

Our starting point is the set (2.19). For later convenience, we rewrite the continuity equation (2.19b) using the energy equation (2.19c):

$$\rho\Big(\frac{D\vec{u}}{Dt} + 2\vec{\Omega}\times\vec{u}\Big) = -\nabla p - \rho g\hat{z} \tag{4.3a}$$

$$\rho\nabla\cdot\vec{u} = -\frac{1}{c_s^2}\frac{Dp}{Dt} \tag{4.3b}$$

$$\frac{D\rho}{Dt} = \frac{1}{c_s^2}\frac{Dp}{Dt} \tag{4.3c}$$

$$\frac{DS}{Dt} = 0 \tag{4.3d}$$

$$c_s = \Gamma(\rho, p, S)\,, \tag{4.3e}$$

Approximation:	*meaning:*	*removes:*
Quasi-incompressibility	divergence-free velocity, a.o.	sound waves
Linearization (*)	advective terms removed	wave interactions
f-plane	fixed latitude	Rossby/planetary waves
'Traditional Approx.' (*)	$\tilde{f}$-terms neglected	short low-freq. int. IGW
Hydrostaticity (*)	vertical acceleration neglected	gyroscopic waves
Geostrophy (**)	Coriolis balances pressure grad.	(internal) IGW

Table 4.1: Various approximations, what they mean, and what phenomenon is removed by making them. 'IGW' stands for inertio-gravity waves.

from which the unknowns $\vec{u}$, ρ, p and S are to be solved.

This set appears intractable; in an attempt to derive analytical solutions, we would encounter a number of hurdles. Mathematically, the hurdles appear in the form of nonlinear terms, as well as terms with non-constant coefficients. Below we discuss under what circumstances these terms can be neglected. Physically, the hurdles just mean that the equations in their present form are 'too rich'; they contain phenomena in which we are not (or not primarily) interested, such as sound waves, large-amplitude waves etc. By making the approximations discussed below, we discard these phenomena one by one.

Common approximations and their consequences to wave phenomena are listed in Table 4.1; approximations marked by an asterisk will not (**), or not always (*), be adopted in this book.

4.3 Quasi-incompressibility

All approximations discussed in this section rest on the assumption that density varies relatively little in the domain under consideration.[1] In other words, we assume that the (vertical) scale over which density varies by one order of magnitude is much larger than the scale of the phenomena we are looking at. This requirement is very well satisfied in the ocean, and reasonably well in the atmosphere if we restrict ourselves to vertical scales of the order of 100 m.

4.3.1 Momentum equations

In classical mechanics, two concepts of mass are sometimes distinguished: inertial and gravitational mass.[2] Inertial mass is a measure of how strongly a body resists a change in velocity when subjected to a force; gravitational mass determines the force by which a body is attracted in a field of gravity. The former has a passive connotation, the latter an active one.

[1] In the literature, these approximations, or a subset of them, are often referred to as "the" Boussinesq approximation. Since opinions differ on exactly what is to be understood by this term, its usage tends to confuse rather than to clarify the issue; for this reason, we avoid the term altogether.

[2] It was shown observationally by Eötvös that the two can be identified numerically; later, conceptual equality was implied by Einstein's theory of general relativity.

It is instructive to consider first a simple mechanical problem: Atwood's machine, see Figure 4.1. The net force working on the connected bodies is $g\Delta m$; this force pulls at a total mass of $2m + \Delta m$. Hence the acceleration is

$$a = \frac{g\Delta m}{2m + \Delta m}\,.$$

Now, if $\Delta m \ll m$, then we can approximate this expression by

$$a \approx \frac{g\Delta m}{2m}\,.$$

This amounts to neglecting the difference in mass, but only so far as the *inertial* mass is concerned; in the gravitational mass the difference remains essential since otherwise their would be no net force.

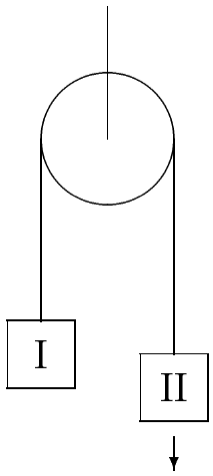

Fig. 4.1: Atwood's machine: two bodies are connected by a string passing over a pulley. The mass of body **II**, $m + \Delta m$, slightly exceeds that of body **I**, m.

In the ocean, density varies by at most a few percentage points (see Figure 2.2e, solid line). This suggests that we may replace density by a constant reference value ρ_* in the momentum equations (4.3a), but of course only where density plays its "inertial" role. Hence

$$\rho_*\Big(\frac{D\vec{u}}{Dt} + 2\vec{\Omega} \times \vec{u}\Big) = -\nabla p - \rho g\hat{z}\,. \tag{4.4}$$

This natural and seemingly innocent approximation deserves further comment. One important quantity in fluid dynamics is vorticity, $\nabla \times \vec{u}$. Now, how can vorticity be created? According to (4.3a) we have

$$\frac{\partial}{\partial t}(\nabla \times \vec{u}) + \cdots = -\nabla \times \Big(\frac{\nabla p}{\rho}\Big)\,, \tag{4.5}$$

stating that pressure gradients can create vorticity, while gravity cannot. (The advective and Coriolis terms play no role in this argument and are rendered by dots.)

Repeating the same procedure in (4.4), we find

$$\frac{\partial}{\partial t}(\nabla \times \vec{u}) + \cdots = -\nabla \times \left(\frac{\rho g \hat{z}}{\rho_*}\right), \tag{4.6}$$

stating that gravity can create vorticity, while pressure gradients cannot – the exact *reverse* of what we found in the original equation (4.5)!

Lest we conclude that this consequence wholly discredits the validity of the approximation, we should realize that gravity and pressure gradients are closely connected in geophysical flows. For example, horizontal density gradients result in pressure gradients due to gravity. In practice, the reversal of roles is thus fairly harmless.

4.3.2 Mass and energy equations

As in the previous chapter, we will think of internal waves as a perturbation of a (known) static background state that has only vertical dependences; thus

$$p = p_0(z) + p'(t, \vec{x}) \tag{4.7}$$
$$\rho = \rho_0(z) + \rho'(t, \vec{x}) \tag{4.8}$$

(and similarly for salinity); the primed fields are associated with the internal-wave motion. The static fields obey the hydrostatic balance:

$$\frac{dp_0}{dz} = -\rho_0 g\,. \tag{4.9}$$

At this stage, these steps neither presume nor imply anything, for the fields p and ρ can always be written like this. The right-hand side of (4.4) can now be written as

$$-\nabla p' - \rho' g \hat{z}$$

and (4.3b,c) become

$$\rho \nabla \cdot \vec{u} = -\frac{1}{c_s^2}\frac{Dp'}{Dt} + w\frac{\rho_0 g}{c_s^2} \tag{4.10}$$
$$\frac{D\rho'}{Dt} + w\frac{d\rho_0}{dz} = \frac{1}{c_s^2}\frac{Dp'}{Dt} - w\frac{\rho_0 g}{c_s^2}\,. \tag{4.11}$$

To proceed, we must assume something about the time-scale of the motion in question.[3] In geophysical fluid dynamics, one may (roughly) distinguish two

[3] In scaling procedures like this, there is always an element of reasoning in a circle: the resulting approximated equations we "prove" to be valid simply reflect the primary balance we assumed in the first place; in other words, 'what you want is what you get'. Indeed, nothing is strictly proven; we only delineate the conditions under which the approximated equations can be expected to be valid; and, in the end, only a comparison with empirical observation can bear out their validity.

types of motion. One is the large-scale flows, with time scales (much) larger than f^{-1}; here, the primary balance in the horizontal momentum equations is between the Coriolis force and the pressure gradient (geostrophy). The other one is inertio-gravity waves, with time scales smaller than f^{-1}; here the primary balance is typically between acceleration and the pressure gradient. For our purposes, the second assumption is the most appropriate one. Denote the scales of t, x, u and p' by T, L, U and P; then, by assumption:

$$\rho_* \frac{U}{T} \sim \frac{P}{L} \,.$$

Since we are dealing with waves, the length and time scales are related by $L/T = C$, the phase speed, which for internal waves in the oceans typically is of the order of 1 $\mathrm{m\,s^{-1}}$. We may thus write

$$P \sim \rho_* C U \,. \tag{4.12}$$

At this stage, it is useful to introduce one more type of velocity, namely the phase speed of long surface waves: $c_{sf} = (gD)^{1/2}$, where D is water depth. We summarize the typical orders of magnitude of the four speeds we have introduced so far, all in meters per second:

$$U \sim O(10^{-1})\,; \quad C \sim O(1)\,; \quad c_{sf} \sim O(10^1\,, 10^2)\,; \quad c_s \sim O(10^3)\,\mathrm{m\,s^{-1}}\,.$$

They denote the particle velocity and phase speed of internal waves, the phase speed of surface waves, and the speed of sound in seawater, respectively. We thus have

$$\boxed{U \ll C \ll c_{sf} \ll c_s \,.} \tag{4.13}$$

These estimates form an important guide to the approximations made below.

We first consider (4.10), and denote the typical scales above each term:

$$\overbrace{\rho \frac{\partial u}{\partial x}}^{\rho_* U/L} + \overbrace{\rho \frac{\partial v}{\partial y}}^{\rho_* U/L} + \overbrace{\rho \frac{\partial w}{\partial z}}^{\rho_* W/H} = - \overbrace{\frac{1}{c_s^2} \frac{\partial p'}{\partial t}}^{P/(Tc_s^2)} - \overbrace{\frac{u}{c_s^2} \frac{\partial p'}{\partial x}}^{UP/(Lc_s^2)} - \overbrace{\frac{v}{c_s^2} \frac{\partial p'}{\partial y}}^{UP/(Lc_s^2)} - \overbrace{\frac{w}{c_s^2} \frac{\partial p'}{\partial z}}^{WP/(Hc_s^2)} + \overbrace{w \frac{\rho_0 g}{c_s^2}}^{\rho_* W g/c_s^2} \,.$$

Here we introduced the vertical internal-wave scales W and H,[4] as yet unrelated to the other scales. By comparing the first term on the right with the first term on the left, we find from (4.12) and (4.13) that their ratio is $C^2/c_s^2 \ll 1$, so the latter dominates. We find the same for the first two advective terms on the right, even if we would take $U \sim C$ (as is the case for strongly nonlinear waves, see next section). Similarly, we find that the third term on the left dominates the fourth term on the right. It also dominates the fifth term on the right, since

[4]For clarity, we emphasize that H does not represent the scale of the amplitude of the internal wave, but rather the vertical scale over which the internal-wave field varies by one order of magnitude; this may often be simply water depth itself.

$gH < gD = c_{sf}^2 \ll c_s^2$. [5] In other words, each of the terms on the right is much smaller than at least one of the terms on the left, so we can conclude that the velocity field is to a close approximation given by a field $\vec{u}^{(0)}$ that obeys (cf. Section 4.1)

$$\nabla \cdot \vec{u}^{(0)} = 0\,. \tag{4.14}$$

An immediate consequence is that *volumes* of water parcels are now conserved, which combines awkwardly with the variation in density that is still allowed. As a result, *mass is no longer conserved*. For wave-phenomena, this problem is not too serious, because the oscillatory nature of the motion implies that mass gains are followed by, and indeed annulled by mass losses, and vice versa.

For later usage, we note that the scale W can be expressed via (4.14) as

$$W \sim UH/L\,. \tag{4.15}$$

We now consider the energy equation (4.11), and focus our attention on the right-hand side:

$$\cdots = + \overbrace{\frac{1}{c_s^2}\frac{\partial p'}{\partial t}}^{P/(Tc_s^2)} + \overbrace{\frac{u}{c_s^2}\frac{\partial p'}{\partial x}}^{UP/(Lc_s^2)} + \overbrace{\frac{v}{c_s^2}\frac{\partial p'}{\partial y}}^{UP/(Lc_s^2)} + \overbrace{\frac{w}{c_s^2}\frac{\partial p'}{\partial z}}^{WP/(Hc_s^2)} - \overbrace{w\frac{\rho_0 g}{c_s^2}}^{W\rho_* g/c_s^2}$$

Using (4.12) and (4.15), we find that the ratio of the first and last term scales as $C^2/(gH)$. Now, the vertical scale of internal waves is often nearly waterdepth itself, in which case $C^2/(gD) = C^2/c_{sf}^2 \ll 1$ by (4.13). Even if we have much smaller scales (for example when an internal-wave is trapped in the thermocline), then we would still have $H \sim 10^1$ to 10^2 m, so that the condition is still met. The first term on the right is thus much smaller than the last one; the advective terms too are much smaller, as one sees by a similar argument. Hence, the fields are to a good approximation given by

$$\frac{D\rho'^{(0)}}{Dt} + w^{(0)}\Big(\frac{d\rho_0}{dz} + \frac{\rho_0 g}{c_s^2}\Big) = 0\,, \tag{4.16}$$

where we brought the remaining term from the right-hand side to the left.

4.3.3 Resulting equations

In summary, we have reduced (4.3a-c) – via (4.4), (4.14) and (4.16) – to

$$\frac{D\vec{u}}{Dt} + 2\vec{\Omega} \times \vec{u} = -\frac{1}{\rho_*}\nabla p' + b\hat{z} \tag{4.17a}$$

$$\nabla \cdot \vec{u} = 0 \tag{4.17b}$$

$$\frac{Db}{Dt} + N^2 w = 0\,, \tag{4.17c}$$

[5] Notice that this condition may pose a problem in the atmosphere, where the speed of sound is much smaller than in the ocean: $c_s \sim 300\,\mathrm{m\,s^{-1}}$. In that case one must assume that H, the vertical scale of the internal-wave, is sufficiently small (i.e. much smaller than the thickness of the troposphere); otherwise the approximation (4.14) will be invalid.

where we dropped the scripts $\{\cdot\}^{(0)}$, and introduced a convenient quantity called *buoyancy*:

$$b = -g\rho'/\rho_* \,. \tag{4.18}$$

Buoyancy can be interpreted as (minus) the effective gravity, associated with the Archimedean force. We also used the definition of the buoyancy frequency from (3.7):

$$N^2 = -\frac{g}{\rho_*}\left(\frac{d\rho_0}{dz} + \frac{\rho_0 g}{c_s^2}\right), \tag{4.19}$$

but with ρ_0 replaced by ρ_* in the denominator on the right-hand side, in line with the approximation made in the previous section. Since N refers to the static state, we can evaluate the speed of sound c_s as a function of the prescribed static fields, i.e. $c_s = \Gamma(\rho_0, p_0, S_0)$; c_s can thus be regarded as a known function of z.

The upshot is that effects of compressibility can be neglected *except* in the definition of the background stratification (4.19). The importance of the second term on the right-hand side of (4.19) was earlier demonstrated in Section 3.2.

The simplifications introduced in this section come at a price, in that they distort a number of fundamental balances; we discussed this already for the vorticity balance and mass conservation. It is beyond the scope of this text to discuss in detail that the energy balance, too, can get distorted. If diabatic processes were included, one would find that a supply (or loss) of heat is represented by an extraction (or supply) of mass; in other words, in the approximated equations heat is represented as a form of mass, not energy (an ironic reversal of 19th-century progress in thermodynamics). In thermodynamic cycles, e.g. thermal convection, one then finds that gravity is doing net work in the model equations, as opposed to pressure in reality [89].

Notwithstanding these problems, the set (4.17) has the merit that sound waves have been filtered out from the equations; this is due to the approximation of (4.3b) by (4.17b).

4.4 The 'Traditional Approximation'

The so-called 'Traditional Approximation' consists in neglecting the terms with $\tilde{f}$, i.e. the Coriolis terms involving the cosine of latitude. In the momentum equations (4.17a), written out in its components, there are two such terms:

$$\frac{Du}{Dt} - fv + \tilde{f}w = -\frac{1}{\rho_*}\frac{\partial p'}{\partial x} \tag{4.20a}$$

$$\frac{Dv}{Dt} + fu = -\frac{1}{\rho_*}\frac{\partial p'}{\partial y} \tag{4.20b}$$

$$\frac{Dw}{Dt} - \tilde{f}u = -\frac{1}{\rho_*}\frac{\partial p'}{\partial z} + b\,. \tag{4.20c}$$

At mid-latitudes the parameter $\tilde{f}$ is of course equally large as f; it seems odd, then, to neglect the terms with $\tilde{f}$ while retaining those with f. However, the

difference between the two is that $\tilde{f}$ is always associated with *vertical* motions (either a vertical velocity or a vertical acceleration), whereas f is associated with horizontal motions alone. Low-frequency motions are usually predominantly horizontal, so that the effects of $\tilde{f}$ are generally slight. Moreover, strong stratification, in the sense that $N \gg \Omega$, tends to suppress vertical motions, thus diminishing the role of $\tilde{f}$. Still, weakly stratified regions occur in the abyssal ocean (see Figure 1.6); in these regions, 'non-traditional' effects can be quite significant, especially for near-inertial internal waves (i.e. internal waves at frequencies close to $|f|$). More generally, other kinds of motions can be affected as well, such as deep convection, Ekman layers, and equatorial flows, see [30].

The mathematical structure of the problem depends on whether one makes the Traditional Approximation or not. With $\tilde{f}$ included, solutions can no longer be obtained by separation of horizontal and vertical variables. For the standard internal-wave problems, however, this obstacle proves illusory. In the following chapters, we therefore include sections on 'non-traditional' results, which are obtained without any difficulty.

4.5 Linearization

Nonlinear terms are those that contain products of unknowns. So in (4.17a), we identify the nonlinear term $(\vec{u} \cdot \nabla)\vec{u}$; (4.17c) similarly contains advective terms. These terms are responsible for interactions among waves. Omitting them greatly simplifies the problem, for this allows wave solutions to be superposed, the superposition being a solution, too. Generally speaking, nonlinear terms are negligible if the wave amplitude is sufficiently small. In the remainder of this section, we specify what is here meant by 'small'.

In (4.17a), we consider the material derivative Du/Dt:

$$\overbrace{\frac{\partial u}{\partial t}}^{U/T} + \overbrace{u\frac{\partial u}{\partial x}}^{U^2/L} + \overbrace{v\frac{\partial u}{\partial y}}^{U^2/L} + \overbrace{w\frac{\partial u}{\partial z}}^{UW/H} + \cdots$$

where the dots stand for the Coriolis terms and pressure gradient. Using the scales of Section 4.3.2, we find that the ratio of each of the nonlinear terms and the time-derivative is U/C, where $C = L/T$ is the phase speed of the internal wave. Thus, nonlinear terms can be neglected if $U \ll C$, i.e. if the horizontal velocity of the water parcels is much smaller than the phase speed. Typically, internal-wave phase speeds are of the order of 1 meter per second, so horizontal currents should be less than, say, 10 cm per second, for nonlinear terms to be negligible. The other momentum equations yield the same criterion, and so does (4.17c).

4.5.1 Resulting equations

Under the assumption, then, that all nonlinear terms can be neglected, (4.17) reduces to

$$\frac{\partial u}{\partial t} - fv + \tilde{f}w = -\frac{1}{\rho_*}\frac{\partial p'}{\partial x} \tag{4.21a}$$

$$\frac{\partial v}{\partial t} + fu = -\frac{1}{\rho_*}\frac{\partial p'}{\partial y} \tag{4.21b}$$

$$\frac{\partial w}{\partial t} - \tilde{f}u = -\frac{1}{\rho_*}\frac{\partial p'}{\partial z} + b \tag{4.21c}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{4.21d}$$

$$\frac{\partial b}{\partial t} + N^2 w = 0\,. \tag{4.21e}$$

From these five equations, it is easy to derive a single equation for one of the unknowns. Taking $\partial/\partial z$ of (4.21b), and $\partial/\partial y$ of (4.21c), and subtracting the results, gives

$$\frac{\partial}{\partial t}\Big(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\Big) \;=\; \Big(\tilde{f}\,\frac{\partial}{\partial y} + f\,\frac{\partial}{\partial z}\Big)u + \frac{\partial b}{\partial y}\,. \tag{4.22}$$

Similarly, taking $\partial/\partial z$ of (4.21a), and $\partial/\partial x$ of (4.21c), gives

$$\frac{\partial}{\partial t}\Big(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\Big) \;=\; \Big(\tilde{f}\,\frac{\partial}{\partial y} + f\,\frac{\partial}{\partial z}\Big)v - \frac{\partial b}{\partial x}\,, \tag{4.23}$$

where we used the continuity equation, (4.21d). Finally, combining $\partial/\partial y$ of (4.21a) and $\partial/\partial x$ of (4.21b):

$$\frac{\partial}{\partial t}\Big(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\Big) \;=\; \Big(\tilde{f}\,\frac{\partial}{\partial y} + f\,\frac{\partial}{\partial z}\Big)w\,, \tag{4.24}$$

where we used, again, (4.21d). On the left-hand sides of (4.22), (4.23) and (4.24), we recognize the vorticity, $\nabla \times \vec{u}$, discussed in Section 4.3.1. Vorticity is a measure of the rotation of water parcels, i.e. their change in orientation.[6] We noted earlier that the creation of vorticity by gravity (here represented by buoyancy b), rather than by pressure gradients, is an artifact of assuming quasi-incompressibility.

Subtracting $\partial^2/\partial y \partial t$ of (4.22) and $\partial^2/\partial x \partial t$ of (4.23), gives

$$\frac{\partial^2}{\partial t^2}\Big(\nabla_h^2 w - \frac{\partial}{\partial z}\Big[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\Big]\Big) + \Big(\tilde{f}\,\frac{\partial}{\partial y} + f\,\frac{\partial}{\partial z}\Big)\frac{\partial}{\partial t}\Big(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\Big) - \nabla_h^2\frac{\partial b}{\partial t} \;=\; 0\,,$$

where we defined

$$\nabla_h^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\,.$$

[6] This type of movement should not be confused with those of parcels traversing for example a circle in a *translational* movement, in which case the orientation of the parcels stays the same.

The term in square brackets can be rewritten with (4.21d); the Coriolis terms, with (4.24); and the last term, with (4.21e). The end result is an equation for w alone:

$$\boxed{\frac{\partial^2}{\partial t^2}\nabla^2 w + (\vec{f}\cdot\nabla)^2 w + N^2\nabla_h^2 w = 0\,,} \tag{4.25}$$

where $\vec{f} = (0, \tilde{f}, f)$ and

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\,.$$

It can be shown that the other variables (u, v, p' and b) satisfy the same equation. The reason for selecting the vertical velocity component lies in the fact that the boundary conditions, to be introduced later, are most easily posed in terms of this variable.

Eq. (4.25) will be the starting point for further analysis in following chapters. We note that the buoyancy frequency N is allowed to vary with z, in which case (4.25) has a non-constant coefficient.

4.5.2 Energy equation

A simple but instructive equation can be derived from (4.17a,b,c,e); multiplying them by u, v, w and b/N^2, respectively, and adding up the resulting equations, one obtains

$$\frac{1}{2}\,\rho_*\,\frac{\partial}{\partial t}\Big[u^2 + v^2 + w^2 + b^2/N^2\Big] + \vec{u}\cdot\nabla p' = 0\,. \tag{4.26}$$

This is an energy equation: the energy density E, being the sum of kinetic and potential energy, is given by[7]

$$E = \frac{1}{2}\,\rho_*\,\Big[u^2 + v^2 + w^2 + b^2/N^2\Big]\,. \tag{4.27}$$

The other term in (4.26), $\vec{u}\cdot\nabla p'$, can also be written as $\nabla\cdot(\vec{u}p')$ (because of (4.21d)) and thus represents the divergence of a *flux*. If we integrate (4.26) over a volume, we find that temporal changes in the volume-integrated energy density must be equal to the flux through the boundaries of the volume. In other words, (4.26) expresses conservation of energy. Note that the Coriolis force plays no role in this equation; it acts as a deflecting force, perpendicular to the motion itself, and hence does no work.

[7] It is important to note that (4.27) is only valid if one of the following conditions is fulfilled: a) waves are linear, or b) the buoyancy frequency N is constant, or both. For nonlinear waves in a medium with variable N, the expression becomes more complicated, and involves derivatives in N: $E = \frac{1}{2}\,\rho_*\,[u^2 + v^2 + w^2 + b^2/N^2 - \frac{1}{3}(b^3/N^6)\,d(N^2)/dz + \cdots]$, see [42] and [58, p. 211].

4.6 Rigid-lid approximation

The ocean supports not only internal waves, but of course also surface waves. We wish to concentrate on the former. Yet, we cannot simply assume that the surface is still, because internal-wave propagation is accompanied by small elevations and depressions of the surface; this merely reflects the pressure gradients inherent to the passage of an internal wave. These vertical excursions of the surface are very small compared to those in the ocean's interior. The reason for this is simple: the same force that is able to give isopycnals in the interior large vertical excursions, can only produce a very small excursion of the surface, because the latter has a much larger vertical density gradient (that between air and water) than has the ocean's interior. This fact can be exploited to simplify the boundary condition at the surface, leading to the so-called 'rigid-lid approximation'.

Let the free surface be described by

$$z = \eta(t, x, y)\,,$$

whence follows

$$\begin{aligned} w(t, x, y, \eta) &= \frac{D\eta}{Dt} \\ &= \frac{\partial \eta}{\partial t} + u\,\frac{\partial \eta}{\partial x} + v\,\frac{\partial \eta}{\partial y}\,. \end{aligned}$$

Moreover, we suppose that the atmospheric pressure at the free surface, p_a, is constant:

$$p(t, x, y, \eta) = p_0(\eta) + p'(t, x, y, \eta) = p_a\,.$$

where we wrote pressure p as the sum of the static (p_0) and dynamic (p') parts, as in (4.7). A Taylor expansion about $z = 0$ gives

$$\begin{aligned} w(t, x, y, 0) + \eta\,\frac{\partial w}{\partial z} + \cdots &= \frac{\partial \eta}{\partial t} + u(t, x, y, 0)\,\frac{\partial \eta}{\partial x} + \cdots \\ &\quad + v(t, x, y, 0)\,\frac{\partial \eta}{\partial y} + \cdots \\ p_0(0) + p'(t, x, y, 0) + \eta\left(\frac{dp_0}{dz} + \frac{\partial p'}{\partial z}\right) + \cdots &= p_a\,. \end{aligned}$$

Assuming amplitudes to be small, we can, like in Section 4.5, neglect product of perturbation terms (i.e. involving u, v, w, η and p'), so that

$$\begin{aligned} w &= \frac{\partial \eta}{\partial t} \quad \text{at } z = 0 \\ p_0 + p' + \eta\frac{dp_0}{dz} &= p_a \quad \text{at } z = 0\,. \end{aligned}$$

Using the hydrostatic balance, (4.9), we can combine the two equations into

$$\frac{\partial p'}{\partial t} = w\rho_0 g \quad \text{at } z = 0\,. \tag{4.28}$$

We cast this equation in a nondimensional form, by using the scale for pressure p' from (4.12), and that for w from (4.15). Moreover, we take for the constant reference density, ρ_*, the static surface value: $\rho_0(0) = \rho_*$. By writing $w = W\hat{w}$, $p' = P\hat{p}'$ and $t = T\hat{t}$, we then obtain

$$\frac{C^2}{c_{sf}^2}\frac{\partial \hat{p}'}{\partial \hat{t}} \quad = \quad \hat{w} \quad \text{at } z = 0\,. \tag{4.29}$$

Recall that $C = L/T$ is a measure of the phase speed of internal waves, while c_{sf} is that of surface gravity waves ($c_{sf}^2 = gH$, H water depth). Now, according to (4.13), the ratio of the two is small. The underlying reason is that these phase speeds are proportional to the square root of effective gravity; for surface waves, this is simply g, but for internal waves, it is the much smaller g', being proportional to relative differences in density in the ocean's interior. Hence $C \ll c_{sf}$, or expressed in terms of a small parameter ϵ,

$$\epsilon = \frac{C^2}{c_{sf}^2} \ll 1\,.$$

We thus arrive at (4.1). As explained in Section 4.1, we can now expand $\hat{p}$ and $\hat{w}$ in a series in which ϵ serves as the small parameter, and to lowest order one finds

$$w^{(0)} \quad = \quad 0 \quad \text{at } z = 0\,. \tag{4.30}$$

This is the so-called 'rigid-lid approximation'. It means that surface elevations are neglected, as if the surface were covered by a rigid plate. At such a plate, vertical velocities would be zero, as in (4.30). The pressure, however, is *not* zero at the plate: pressure variations that would otherwise have been associated with elevations of the free surface, are now exerted by the rigid plate, in response to the internal-wave motions. (An example of the pressure field is shown in Figure 5.4).

At the bottom, the boundary condition is more straightforward, namely that of no normal flow. We describe the bottom by

$$z = -h(x, y)\,.$$

The boundary condition can then be written as

$$w \quad = \quad -u\frac{\partial h}{\partial x} - v\frac{\partial h}{\partial y} \quad \text{at } z = -h\,. \tag{4.31}$$

Slopes of the ocean floor are predominantly less than 0.02, see Figure 4.2.

In many cases we will assume, for the sake of simplicity, that the bottom is horizontal, with water depth H, in which case (4.31) becomes

$$w \quad = \quad 0 \quad \text{at } z = -H\,. \tag{4.32}$$

The idealized character of this assumption is evident from a look at real ocean topography, see for example Figure 1.6!

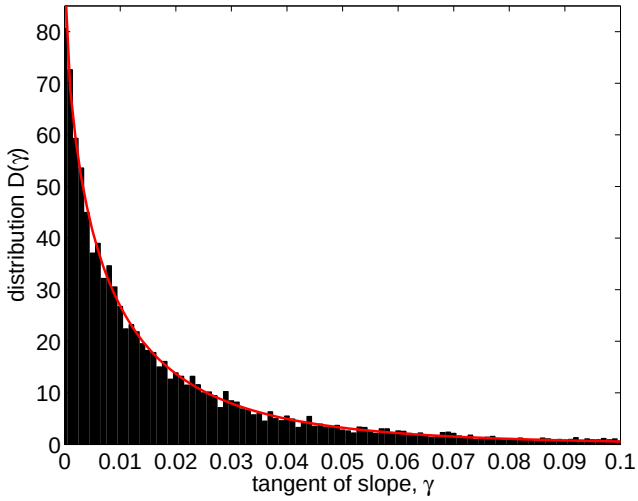

Fig. 4.2: Slope distribution based on several profiles from a topographic database. Slopes are mostly gentle; 67% of the slopes is less than 0.02. The red curve shows a fit with a Burr distribution $a(1+\gamma^b)^{-c}$, with $a = 91$, $b = 0.65$ and $c = 25$. From [29].

4.7 Preview: methods of solution

As a preparation for the following two chapters, in which we discuss linear internal-wave solutions, we first take a closer look at (4.25):

$$\frac{\partial^2}{\partial t^2}\nabla^2 w + (\vec{f}\cdot\nabla)^2 w + N^2\nabla_h^2 w = 0\,. \tag{4.33}$$

Supposing waves to be sinusoidal in time, at frequency ω, i.e.

$$w \sim \exp(-i\omega t)\,,$$

we find from (4.25), [8]

$$(N^2-\omega^2+\tilde{f}^2)w_{yy} + 2f\tilde{f}w_{yz} + (f^2-\omega^2)w_{zz} + (N^2-\omega^2)w_{xx} = 0\,. \tag{4.34}$$

Recall that x is the west-east coordinate, and y, south-north. These coordinates play different roles in (4.34). In other words, there is an anisotropy in the horizontal plane: northward propagating waves, say, behave differently from eastward propagating ones.

However, under the Traditional Approximation, where $\tilde{f}$ is neglected, the anisotropy disappears, since (4.34) then becomes

$$(N^2-\omega^2)\nabla_h^2 w - (\omega^2-f^2)w_{zz} = 0\,. \tag{4.35}$$

[8] Strictly, another symbol than w should be used in (4.34); from the context it will however be always clear whether w is meant to indicate the time-dependent w from (4.33), or the purely spatially dependent one from (4.34).

Here x and y play identical roles. This means that if we consider propagating in the x-direction, taking $\partial/\partial y = 0$, we may apply the results thus obtained also to any other direction of wave propagation.

Taking $\partial/\partial y = 0$, then, (4.35) becomes

$$(N^2 - \omega^2)w_{xx} - (\omega^2 - f^2)w_{zz} = 0\,. \tag{4.36}$$

There are now two main roads to solution, which we briefly outline here; a detailed discussion follows in Chapters 5 and 6.

For *uniform stratification*, i.e. $N = const$, all coefficients of (4.36) are constant, and the general solution can be written

$$w = F(\mu_+ x - z) + G(\mu_- x - z)\,, \tag{4.37}$$

where F and G are arbitrary functions. The coefficients $\mu_\pm$ are given by

$$\mu_\pm = \pm\Big(\frac{\omega^2 - f^2}{N^2 - \omega^2}\Big)^{1/2}. \tag{4.38}$$

As discussed below, the physical significance of the coordinates $\mu_\pm x - z$, the so-called *characteristic coordinates*, lies in the fact that internal-wave energy propagates along lines $\mu_\pm x - z = const.$, which are diagonals in the x, z-plane (cf. Figure 1.9).

This *method of characteristics* forms the subject of Chapter 6. It rests on the assumption of constant N, but involves no assumption concerning the boundaries of the system. It can therefore be applied to, for example, the problem of internal-wave reflection from a sloping bottom.

If, on the other hand, the buoyancy frequency varies with the vertical – as it does particularly strongly in the seasonal thermocline, see Figures 1.6 and 3.3b – then another method may be used that involves no assumption on $N(z)$, but does require boundaries (bottom, surface) to be *horizontal*. This is the *method of vertical modes*, in which we write

$$w = W(z)\exp ikx\,, \tag{4.39}$$

substitution of which into (4.36) gives

$$W'' + k^2\,\frac{N^2(z) - \omega^2}{\omega^2 - f^2}\,W = 0\,. \tag{4.40}$$

(Primes denotes derivatives to z.) Together with the boundary conditions $W = 0$ at $z = -H, 0$ (bottom, surface), this constitutes a Sturm-Liouville problem. Its solution is formed by a set of eigenvalues k_n and corresponding eigenfunctions W_n. For some special choices of $N(z)$, W_n can be obtained in analytical form; otherwise (4.40) has to be solved numerically. Either way, the general solution consists of an arbitrary superposition of modes. This method forms the subject of the next chapter.

We thus have two methods at our disposal, involving different assumptions. The case of overlapping validity occurs when stratification is constant and

boundaries are horizontal. Despite their different appearances, (4.37) on the one hand, and the solution to (4.39,4.40) on the other, are then entirely equivalent (see Section 6.3). There is however also a case in which neither method works: when both stratification and water depth vary. This occurs, for example, in the upper part of the water column, near the seasonal thermocline, over the continental slope. This is a very important region in the study of internal waves, since internal tides are generated here. To treat this problem satisfactorily, one thus has to resort to numerical methods. Nonetheless, the two analytical methods discussed in the following chapters already provide important clues to how internal waves propagate. We also discuss how the two methods of solution can be applied in the more general 'non-traditional' case.

Chapter 5

Internal-wave propagation I: method of vertical modes

The ocean's boundaries in the vertical, bottom and surface, naturally confine the propagation of internal waves; the boundaries act as a *waveguide*. A more subtle form of waveguide is a region of strong stratification in the water column, which acts as a waveguide for high-frequency internal waves. In either case, the internal wave has the character of a standing wave in the vertical, while it propagates in the horizontal. These properties are exploited in the method of *vertical modes*. It has the advantage that it works for any buoyancy profile $N(z)$, but it requires boundaries to be horizontal, since the solution is obtained by a separation of the horizontal and vertical variables.

5.1 General formulation

The starting point is (4.25),

$$\frac{\partial^2}{\partial t^2}\nabla^2 w + (\vec{f}\cdot\nabla)^2 w + N^2\nabla_h^2 w = 0\,. \tag{5.1}$$

Recall that $\vec{f} = (0, \tilde{f}, f)$. We assume first that $\tilde{f} = 0$ (Traditional Approximation), and defer a discussion of non-traditional effects to Section 5.6. This assumption renders the horizontal plane isotropic, in the sense that the direction of wave propagation becomes immaterial. Withous loss of generality, then, we can consider wave propagation in the west-east direction, i.e. along x; the results will similarly apply to any other direction. We seek solutions of the form

$$w = W(z)\exp i(kx - \omega t)\,. \tag{5.2}$$

(We take ω to be positive.) Substitution in (5.1), with $\tilde{f} = 0$, results in an ordinary differential equation for W:

$$\boxed{W'' + k^2 \, \frac{N^2(z) - \omega^2}{\omega^2 - f^2} \, W = 0 \, .} \tag{5.3}$$

Primes denote derivatives to z. Since k here occurs as a square, k^2, solutions with positive k (describing rightward propagating waves), which we will denote by k^+, imply that another solution exists with $k^- = -k^+$, describing leftward propagating waves. So, without loss of generality, we can assume k to be positive, adding afterward solutions with $-k$, if needed.

In addition to (5.3), we pose the boundary conditions at the surface (regarded as a rigid-lid, see Section 4.6) and the horizontal bottom:

$$W = 0 \quad \text{at} \;\; z = 0, \, -H \, . \tag{5.4}$$

Together, (5.3) and (5.4) form a Sturm-Liouville problem, which for fixed ω has an infinite number of solutions W_n (eigenfunctions, vertical modes) with corresponding eigenvalues k_n.

The other variables u, v, p and b can be expressed similarly as (5.2), with

$$U = \frac{i}{k} \, W' \, ; \quad V = \frac{f}{\omega k} \, W' \, ; \quad P = i\rho_* \frac{\omega^2 - f^2}{\omega k^2} \, W' \, ; \quad B = -\frac{iN^2}{\omega} \, W \, . \tag{5.5}$$

These expressions follow from (4.21). Notice that (5.4) and (5.5) imply that the *vertically integrated horizontal velocities are zero*, i.e.,

$$\int_{-H}^{0} dz \, u = 0 \, ; \quad \int_{-H}^{0} dz \, v = 0 \, .$$

This property distinguishes internal waves from surface waves.

The general solution of w consists of the superposition

$$w = \sum_n W_n(z) \Big[a_n^{\pm} \exp i(k_n^{\pm} x - \omega t) \Big] \, , \tag{5.6}$$

with arbitrary complex constants $a_n^{\pm}$. The terms with index 'plus' describe rightward propagating waves, those with 'minus', leftward propagating ones. It is understood that the real part is being taken.

5.1.1 Oscillatory versus exponential behaviour

For convenience, we introduce

$$m^2(z) = k^2 \, \frac{N^2(z) - \omega^2}{\omega^2 - f^2} \, . \tag{5.7}$$

Solutions to (5.3) may exhibit two kinds of behaviour, depending on the sign of m^2. Firstly, *oscillatory* in those parts of the water column where m is real, i.e.

$m^2(z) \geq 0$. For this to be possible, one of the two following inequalities must hold throughout this part of the water column:

$$\boxed{\text{(I)} \quad N(z) \leq \omega \leq |f| \quad \text{or} \quad \text{(II)} \quad |f| \leq \omega \leq N(z)\,.} \tag{5.8}$$

The prevailing situation in the ocean and atmosphere is $N > |f|$, but $|f|$ may exceptionally exceed N in extremely weakly stratified regions, such as convective layers. In either case, the range of allowable wave frequencies is delineated by a lower and upper bound.

For internal waves, of a given frequency ω, to exist at all, one of the inequalities in (5.8) should be satisfied in at least part of the water column. Elsewhere, the opposite may hold, $m^2(z) < 0$, which gives rise to the second type of behaviour: *exponential*-like decay, describing the rapid decrease of the wave-amplitude outside the waveguide. Here neither of the inequalities (5.8) is fulfilled.

5.1.2 Orthogonality

Suppose W_n satisfies (5.3), with eigenvalue k_n:

$$W_n'' + k_n^2 \, \frac{N^2(z) - \omega^2}{\omega^2 - f^2} \, W_n = 0\,. \tag{5.9}$$

Let W_l be another eigenfunction , with $k_l \neq k_n$. Multiplying W_n'' by W_l, and applying partial integration twice, gives

$$\begin{aligned}
\int_{-H}^{0} dz \, W_l W_n'' &= W_l W_n'|_{-H}^{0} - \int_{-H}^{0} dz \, W_l' W_n' \\
&= -W_l' W_n|_{-H}^{0} + \int_{-H}^{0} dz \, W_l'' W_n \\
&= \int_{-H}^{0} dz \, W_l'' W_n \,,
\end{aligned}$$

where we used the boundary conditions (5.4), stating that W vanishes at the surface and bottom. Hence, multiplying (5.9) by W_l gives

$$\int_{-H}^{0} dz \, W_l'' W_n + k_n^2 \int_{-H}^{0} dz \, \frac{N^2(z) - \omega^2}{\omega^2 - f^2} \, W_l W_n = 0\,. \tag{5.10}$$

Since W_l with associated k_l, too, satisfies (5.9), we have

$$(k_n^2 - k_l^2) \int_{-H}^{0} dz \, \frac{N^2(z) - \omega^2}{\omega^2 - f^2} \, W_l W_n = 0\,.$$

Since, by assumption, $k_l \neq k_n$, it follows that

$$\int_{-H}^{0} dz \, \frac{N^2(z) - \omega^2}{\omega^2 - f^2} \, W_l W_n = 0\,. \tag{5.11}$$

i.e. W_n and W_l are orthogonal.

This is, in fact, a general property of Sturm-Liouville problems, a more general formulation of which is

$$(pW')' - qW + \lambda rW = 0\,,$$

with $W(z_1) = W(z_2) = 0$ (or, more generally still, boundary conditions involving its derivative); λ is the eigenvalue. Here q and r are supposed to be continuous functions, and p continuously differentiable. Moreover, a common restriction is $p > 0$, $q \geq 0$ and $r > 0$. Eq. (5.3) is a special case, with $p = 1$ and $q = 0$, $\lambda = k^2$ being the eigenvalue. However, as hinted in Section 5.1.1, the coefficient $r = (N^2 - \omega^2)/(\omega^2 - f^2)$ may change its sign in the vertical, contrary to the common assumption $r > 0$. This more complicated case is discussed in [44, p. 237]. The main outcome is that there exists both an infinite set of negative eigenvalues λ_n^- and of positive eigenvalues λ_n^+, tending to $-\infty$ and ∞, respectively, for $n \to \infty$. Only the positive ones are physically meaningful in our case. The upshot is that we can expect a set of eigenvalues extending to infinitely short waves, $k_n \to \infty$, for large n.

5.1.3 Hydrostatic approximation

For low-frequency internal waves, such as internal tides, the hydrostatic approximation is often appropriate, i.e. the vertical acceleration $\partial w/\partial t$ in (4.17c) can be neglected. It can be readily verified that this amounts to assuming $N \gg \omega$ in (5.3), so that we can write

$$W'' + \bar{k}^2 N^2(z)\, W = 0\,. \tag{5.12}$$

with $\bar{k} = k/(\omega^2 - f^2)^{1/2}$. The vertical structure of the modes W_n now no longer depends on the wave frequency ω. The hydrostatic approximation thus brings about a considerable simplification.

5.2 Uniform stratification

We start with the simple case: that of uniform stratification, i.e. $N = const.$ There are two things to be resolved from (5.3): 1) the eigenvalues k_n and their relation to wave frequency ω, which constitutes the *dispersion relation*; and 2) the vertical structure of the modes W_n.

In this case the solution of the eigenfunctions W_n is obvious and could be written down immediately, but we derive them here in a systematic way to illustrate the general procedure that works also for non-constant N:

- Derive the general solution of (5.3), which involves two arbitrary (real) constants;
- Pose the boundary conditions in matrix form;

- To have non-trivial solutions, the matrix has to be singular; this requirement yields the *dispersion relation*;
- Finally, one constant can be chosen arbitrarily, fixing the eigenfunction W_n.

We can write (5.3) as

$$W'' + m^2 W = 0\,,$$

where m, defined in (5.7), is now independent of z. Wave-like solutions are found if m is real. We thus restrict ourselves to wave frequencies satisfying (5.8). Unless stated otherwise, we shall assume $N > |f|$, a condition that is usually satisfied in the ocean and lower atmosphere. Hence

$$|f| \le \omega \le N\,.$$

The general solution now reads

$$W = C_1 \sin mz + C_2 \cos mz\,.$$

The boundary conditions (5.4) can be gathered in a matrix:

$$\begin{pmatrix} 0 & 1 \\ -\sin mH & \cos mH \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\,. \tag{5.13}$$

5.2.1 Dispersion relation

To have non-trivial solutions for the pair (C_1, C_2), the determinant of the matrix in (5.13) must be zero, implying $\sin mH = 0$; hence

$$m_n = \pm\frac{n\pi}{H}\,, \qquad \text{for} \quad n = 1, 2, 3, \cdots\,.$$

Using the definition of m, (5.7), we obtain the *dispersion relation*:

$$\boxed{k_n = \pm\frac{n\pi}{H}\left(\frac{\omega^2 - f^2}{N^2 - \omega^2}\right)^{1/2}\,, \qquad n = 1, 2, 3, \cdots\,.} \tag{5.14}$$

For a given frequency ω, one thus finds an infinite number of eigenvalues k_n, which serve as horizontal wavenumbers. Waves become shorter, i.e. $|k_n|$ increases, with increasing modenumber n. Alternatively, taking the wavenumber k and modenumber n as the independent variables, we can rewrite (5.14) to express the wave frequency as a function of k and n:

$$\omega^2 = \frac{N^2 k^2 + f^2(\frac{n\pi}{H})^2}{k^2 + (\frac{n\pi}{H})^2}\,. \tag{5.15}$$

The lower bound of the frequency domain, $|f|$, is attained in the long-wave limit $|k| \to 0$; the upper bound, N, in the short-wave limit $|k| \to \infty$, see Figure 5.1

(upper panel). For positive k, the function $\omega(k)$ is monotonically increasing, as is clear from the fact that its derivative is positive. This derivative, the horizontal group velocity $c_g = d\omega/dk$, is obtained from differentiating (5.15):

$$c_g = \frac{k(\frac{n\pi}{H})^2(N^2-f^2)}{[N^2k^2+f^2(\frac{n\pi}{H})^2]^{1/2}\,[k^2+(\frac{n\pi}{H})^2]^{3/2}} \tag{5.16}$$

$$= \pm\Big(\frac{H}{n\pi}\Big)\,\frac{(\omega^2-f^2)^{1/2}(N^2-\omega^2)^{3/2}}{\omega(N^2-f^2)}\,. \tag{5.17}$$

In (5.17), the plus-sign applies if k is positive, the minus-sign if k is negative. We thus see that k and c_g have the same horizontal direction if $|f| < N$, and the opposite direction if $|f| > N$. The horizontal phase speed $c = \omega/k$ is given by

$$c = \frac{[N^2k^2+f^2(\frac{n\pi}{H})^2]^{1/2}}{k[k^2+(\frac{n\pi}{H})^2]^{1/2}} \tag{5.18}$$

$$= \pm\Big(\frac{H\omega}{n\pi}\Big)\,\Big(\frac{N^2-\omega^2}{\omega^2-f^2}\Big)^{1/2}\,. \tag{5.19}$$

The group velocity c_g indicates how fast the wave energy travels; the phase speed c, how fast the wave's crests and troughs travel. Clearly, the former is the most relevant quantity because it determines where the internal waves manifest themselves (there are no waves if there is no wave-energy).

For fixed ω, we see from (5.17) and (5.19) that the group velocity and phase speed both are inversely proportional to modenumber n; *higher modes propagate more slowly*. Eq. (5.17) moreover implies that the group velocity vanishes at the extremes of the frequency domain, i.e. for $\omega = |f|, N$. The phase speed, on the other hand, tends to infinity at the lower bound $|f|$. This is illustrated in Figure 5.2.

Next we consider how c_g and c vary if we take modenumber n fixed ($n = 1$, say) and choose k as the independent variable; here we use (5.16) and (5.18), see Figure 5.1 (middle and lower panels). The phase speed c varies with k, i.e. internal waves are *dispersive*. Differentiating $c = \omega/k$ to k, we obtain the generally valid identity

$$k\frac{dc}{dk} = c_g - c\,.$$

So, for dispersive waves the group and phase velocity are unequal. The dispersiveness means that a superposition of first-mode waves (or any other fixed mode-number) involving different wavenumbers, and hence different frequencies, cannot propagate as a coherent entity, but will disperse, because different components have different phase and group speeds.

Finally, we consider the low-frequency limit, or equivalently, long-wave limit. Assuming $kH \ll 1$, we can expand (5.15) as

$$\omega^2 = f^2 + \Big(\frac{NH}{n\pi}\Big)^2 k^2 + \cdots\,, \tag{5.20}$$

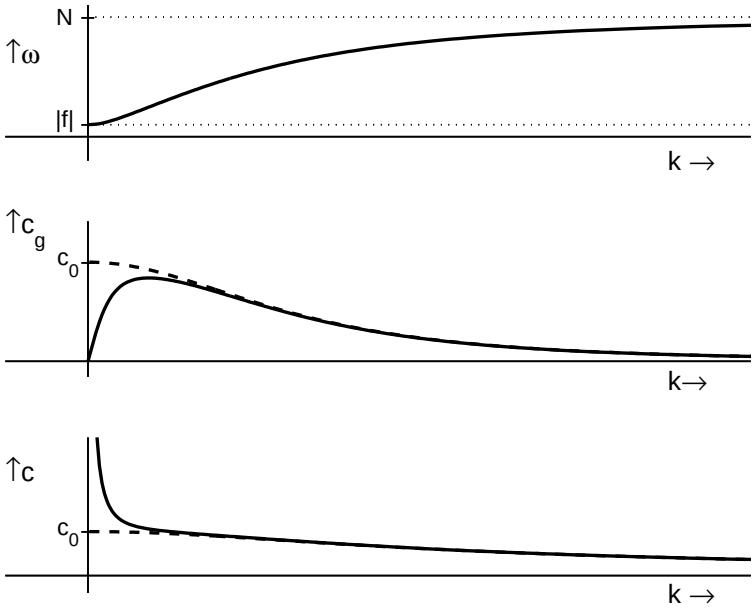

Fig. 5.1: Profiles of wave frequency ω, group velocity c_g, and phase speed c, each as a function of wavenumber k (chosen positive), for arbitrary but fixed modenumber n; these plots illustrate (5.15), (5.16) and (5.18), respectively. In the middle and lower panels, the behaviour for $f = 0$ is also shown (dashed line). The value $c_0 = NH/n\pi$ denotes the linear long-wave phase and group speed for $f = 0$.

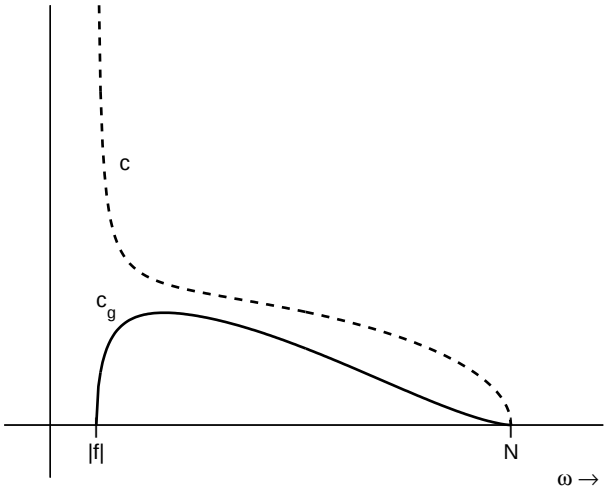

Fig. 5.2: Group and phase speeds as a function of wave frequency ω, illustrating (5.17) and (5.19) for arbitrary but fixed modenumber n, taking $N > |f|$ and k positive.

where we retained only the first two terms of the expansion. This expression follows also directly from (5.14), if we suppose $\omega \ll N$. Still another way of obtaining (5.20) would have been to make the hydrostatic approximation (see Section 5.1.3); these are all different renderings of the same statement. In (5.14), the waves are still dispersive (i.e. $c = \omega/k$ depends on k), but this is now entirely due to Coriolis effects. If $f = 0$, long waves become dispersionless, having a group and phase speed $c_0 \equiv NH/n\pi$ (see dashed lines in Figure 5.1, middle and lower panels). Typical values for c_0 in the ocean are of the order of $1\,\mathrm{m\,s^{-1}}$ (take, for example, $n = 1$, $N = 1 \times 10^{-3}\,\mathrm{rad\,s^{-1}}$, $H = 4000\,\mathrm{m}$). A useful measure of the importance of Coriolis effects is the ratio $r_d = c_0/|f|$, the so-called internal Rossby radius of deformation. Internal waves having a wavelength of the order of r_d, or larger, are significantly affected by Coriolis effects. For waves much shorter than r_d, Coriolis effects can generally be neglected (but see Section 5.6.2!). Notice that short waves are dispersive due to gravity, an effect neglected in (5.20), but present in the exact dispersion relation (5.15).

5.2.2 Modal structure

Returning to (5.13), we see that C_2 must be zero, while C_1 may take any (real) value. Without loss of generality, we can choose $C_1 = 1$ for all n (thus normalizing the amplitude at 1), since we already included an arbitrary (complex) constant a_n in the series (5.6). With this, the vertical modes become

$$\boxed{W_n = \sin\left(\frac{n\pi z}{H}\right), \qquad n = 1, 2, 3, \cdots .} \tag{5.21}$$

They have the remarkable property of being independent of the wave frequency ω. This is a peculiarity of the case $N = const$; for general profiles $N(z)$, the modes take different structures for different frequencies (see following sections). The first three verical modes of (5.21) are shown in Figure 5.3.

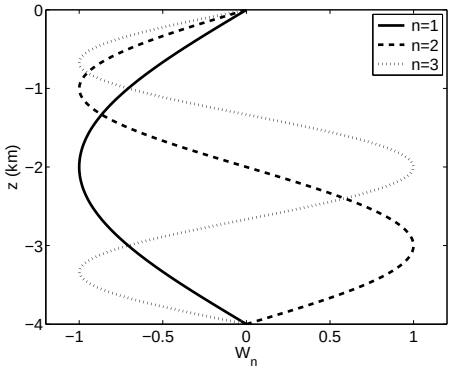

Fig. 5.3: The first three vertical modes W_n, from (5.21), for $H = 4000$ m.

With (5.21), the general solution for rightward propagating waves is obtained from (5.6), by selecting positive k (i.e. k^+; hereafter, we drop the 'plus'). Assuming a_n to be real, and taking the real part of (5.6), we find

$$w = \sum_n a_n \sin\left(\frac{n\pi z}{H}\right) \cos(k_n x - \omega t)\,. \tag{5.22}$$

Using (5.5), we obtain for the horizontal velocity component u, the tranverse velocity v, pressure p and buoyancy b,

$$u = -\sum_n a_n \frac{n\pi}{k_n H} \cos\left(\frac{n\pi z}{H}\right) \sin(k_n x - \omega t) \tag{5.23}$$

$$v = \frac{f}{\omega}\sum_n a_n \frac{n\pi}{k_n H} \cos\left(\frac{n\pi z}{H}\right) \cos(k_n x - \omega t) \tag{5.24}$$

$$p = -\rho_* \frac{\omega^2 - f^2}{\omega}\sum_n a_n \frac{n\pi}{k_n^2 H} \cos\left(\frac{n\pi z}{H}\right) \sin(k_n x - \omega t) \tag{5.25}$$

$$b = \frac{N^2}{\omega}\sum_n a_n \sin\left(\frac{n\pi z}{H}\right) \sin(k_n x - \omega t)\,. \tag{5.26}$$

Another important quantity is the displacement of the isopycnals (levels of constant potential density, strictly speaking). Let the isopycnal that lies at depth z_0 in the state of rest be described by

$$z = z_0 + \zeta(t, x, z_0)\,.$$

Then

$$w(t, x, z) = \frac{\partial \zeta}{\partial t}(t, x, z_0) + u(t, x, z)\,\frac{\partial \zeta}{\partial x}\,.$$

Making a Taylor expansion about $z = z_0$, and neglecting nonlinear terms, gives

$$w(t, x, z_0) = \frac{\partial \zeta}{\partial t}(t, x, z_0)\,.$$

Hence the solution for ζ,

$$\zeta = -\frac{1}{\omega}\sum_n a_n \sin\left(\frac{n\pi z}{H}\right) \sin(k_n x - \omega t)\,, \tag{5.27}$$

where we substituted z_0 by z. Comparing this with (5.26), we see that $\zeta = -b/N^2$.

We now show examples of u, v, w, b and ζ for the first three modes individually, using (5.22)–(5.27) at $t = 0$, with $a_n = 1$. From (5.23) and (5.25) we see that u and p differ only by a positive coefficient; they have identical structures and are therefore represented in one and the same plot. Similarly, b and ζ differ in essence only by a minus sign. The results are shown in Figure 5.4. Notice that v owes its presence to the Coriolis force, and would be zero for $f = 0$. For the n-th mode, u changes its sign n times in the vertical (as does v); w and ζ

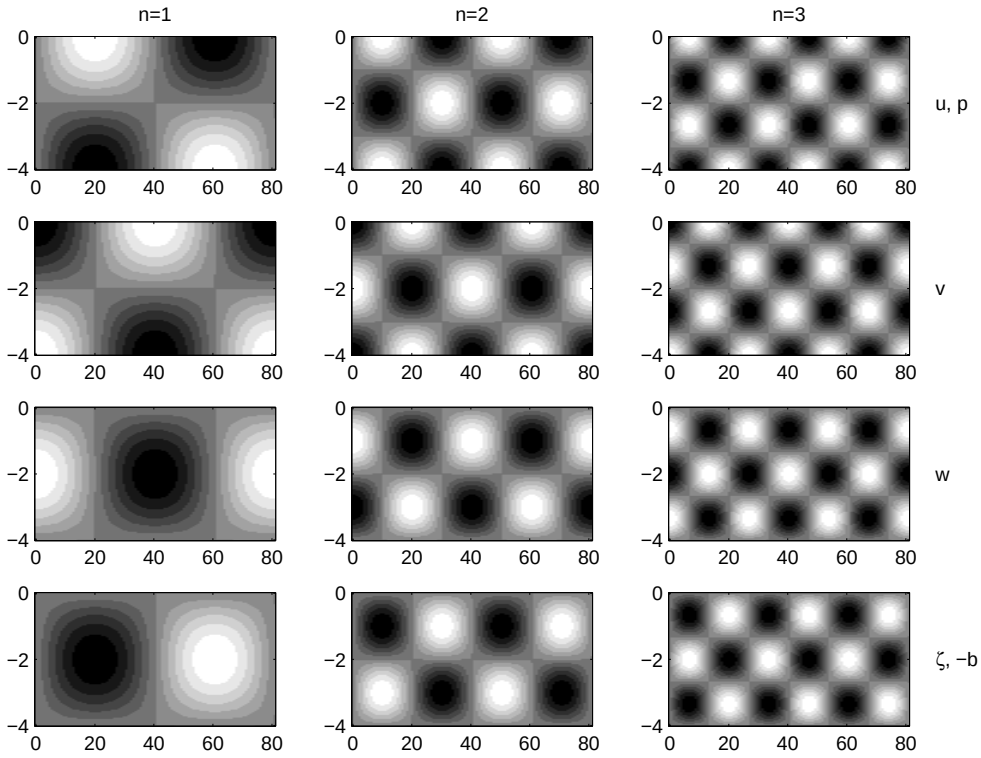

Fig. 5.4: Spatial structure of the first three modes, for constant stratification, at $t = 0$. Upper panels show the horizontal velocity component u; the second row, the tranverse velocity v; the third row, the vertical velocity w; and the lowest panels the isopycnal elevation ζ. The first and fourth rows also represent pressure p and minus buoyancy b, respectively. White denotes negative values; black, positive ones. Parameters are: $N = 1 \times 10^{-3}$, $f = 1 \times 10^{-4}$, and $\omega = 1.405 \times 10^{-4}$ (the semi-diurnal lunar tidal frequency, M_2), all in $\mathrm{rad\,s^{-1}}$. Along the vertical is water depth, with $H = 4$ km; horizontal distances are also in km.

each $n-1$ times. So, for example, at a certain horizontal position the isopycnals corresponding to the second mode are elevated in one part of the water column, and depressed in the other part.

Notice that the pressure is not zero at the surface (rigid-lid); on the contrary, it has a local extremum there. If the surface were allowed to move freely, these regions of high (low) pressure at $z = 0$ would be associated with an elevation (depression) of the surface. By comparing the pattern of p with that of ζ for the first mode, we see that depressions in the interior are accompanied by an elevated pressure at $z = 0$, and vice versa. In other words, surface displacements would be in anti-phase with the displacements in the interior.

In time, the patterns all move steadily to the right (not shown), but at lower speeds for higher modenumbers, because the phase speed c in (5.19) is inversely

proportional to n.

5.2.3 Superposition of modes

We also see from Fig. 5.21 that the second mode is twice as short as the first one; the third three times as short, etc. Indeed, (5.14) implies that the wavenumbers are commensurable:

$$k_n/k_l = n/l\,.$$

(We note that this property does *not* hold for general $N(z)$; it is a peculiarity of the case of constant N.) As a consequence, a superposition of modes must be horizontally periodic. An example of a superposition of 10 modes, representing u, is shown in Figure 5.5 at several stages during half a period (for clarity, we have enlarged the horizontal domain by a factor of two). At first sight, Figure 5.5 is very surprising: it shows that the modes – each of which travels to the right at its *own* phase speed c_n –, when superposed give rise to 'beams' that stay in place! The water parcels oscillate predominantly within these beams, in a parallel direction, as is shown by the red arrows depicting the (u, w) vectorial velocity field. (Due to Coriolis effects, there is also a component v in the transverse direction, which is not shown.)

Upon closer examination, we see that phase propagation is in a direction *perpendicular* to the beam. For example, in the interval $x = 0 - 40\,\mathrm{km}$, the phases move up- and rightward, perpendicularly to the beam itself. The energy, meanwhile, must flow *along* the beam, which is, in this interval, down- and rightward. Vertical opposition of direction in phase and energy propagation is found in the other intervals as well. As shown in Chapter 6, this rule holds generally when $N > |f|$.

The regular pattern of diagonals of Figure 5.5 invites the question as to what underlying principle imposes this orderliness. For this, we have to reconsider the nature of the modal solution. With (5.2), we consider rightward propagating waves, whose vertical structure is described by $W(z)$. This structure turns out to be sinusoidal, as shown in (5.21) and Figure 5.3. We can thus interpret each W_n as a standing wave in the vertical, i.e. a combination of up- and downward propagating waves; here, $m_n = n\pi/H$ serves as the vertical wavenumber. Now, from (5.7), we find that the ratio of the vertical and horizontal wavenumbers, m_n and k_n, is given by

$$\frac{m_n}{k_n} = \pm\frac{N^2 - \omega^2}{\omega^2 - f^2}\,, \tag{5.28}$$

which is *independent of modenumber* n. In other words, one and the same angle, in the x, z-plane, pervades *all* modes. So it is, after all, not surprising to find a well-defined pattern of diagonals in Figure 5.5. More specifically, (5.28) denotes the tangent of the angle that the wavevector $\vec{k} = (k, m)$ makes with the horizontal, the wavevector pointing in the direction of phase propagation. We inferred earlier that energy propagates perpendicularly to this; hence its slope must be given by the inverse of (5.28), namely $\pm(\omega^2 - f^2)/(N^2 - \omega^2)$. These properties will be established in a much more direct way in Chapter 6.

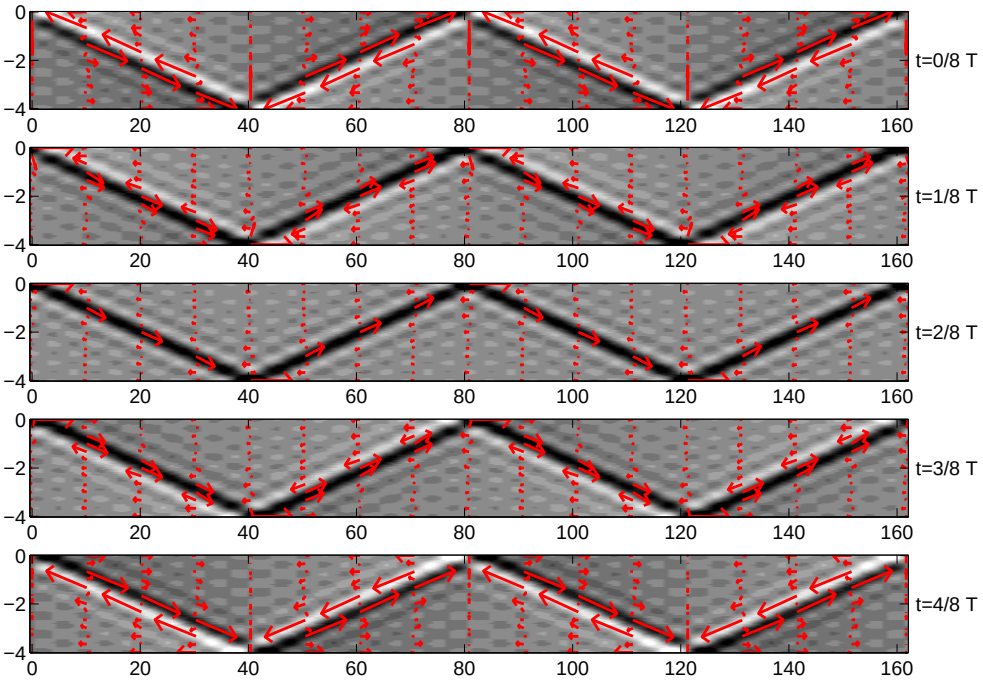

Fig. 5.5: Snapshots of a superposition of 10 modes, showing the horizontal current velocity u at intervals of one-eighth of wave period T. In total, the panels cover half a period. Time progresses downward. White denotes negative values; black, positive ones. For all modes, $a_n = 1$. The arrows, in red, depict the u, w-field. Parameters as in Figure 5.4.

Finally, we consider how the pattern changes if we add more modes. This is illustrated in Figure 5.6: the more modes are involved, the clearer the beam becomes. In this case it becomes also much finer with more modes, but the extent of this effect depends on the way modes are superposed. Here we have $a_n = 1$, but for $a_n = 1/n$, say, the high modes would have less influence on the structure of the beam. We note that decreasing a_n with n are generally found in generation problems, such as for internal tides (Chapter 7).

In conclusion, we emphasize that the case of constant N is simple and instructive, but in some ways atypical. Two key properties – independence of the modal structure on wave frequency, and commensurability of wavenumbers –, which lend much simplicity to this case, do not in general hold for profiles $N(z)$.

5.3 Varying N: two layers

In the previous section, it was shown that a superposition of modes gives rise to internal-wave beams, directed along diagonals in the x, z-plane. As an aside, we note that the frequency ω chosen in Figures 5.4-5.6, corresponds to the

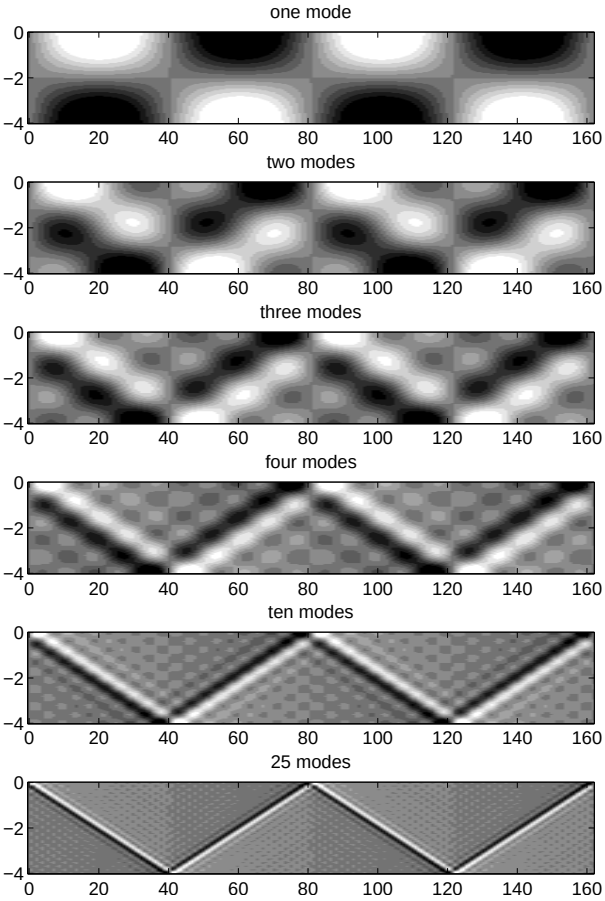

Fig. 5.6: Superpositions with an increasing number of modes, each at $t = 0$, showing the horizontal current velocity u. White denotes negative values; black, positive ones. Parameters as in Figure 5.4.

semi-diurnal lunar frequency M_2. It is therefore interesting to compare the pattern found in Figure 5.5 with that of an observed internal-tide beam, shown in Figure 1.9. We see that the angles are roughly similar: for every kilometer the beam traverses in the vertical, it traverses about 10 km in the horizontal. A noticeable difference between the figures is, however, that the beams in Figure 5.5 are straight, while the observed beam is bended; it becomes steeper in the deep ocean. A clue to this behaviour is found in Figure 1.6, which shows that stratification N generally decreases with depth in the abyssal ocean.

To take into account this effect, we have to abandon the assumption of constant N. A simple way of doing this is to take two layers of different constants

N_1 and N_2:

$$N(z) = \begin{cases} N_1 & -d < z < 0 \quad \text{(upper layer)} \\ N_2 & -H < z < -d \quad \text{(lower layer)}\,. \end{cases} \tag{5.29}$$

The governing equation is still (5.3), which we now apply to each layer separately. Accordingly, we define, as in (5.7),

$$m_{1,2}^2 = k^2\,\frac{N_{1,2}^2 - \omega^2}{\omega^2 - f^2}\,.$$

The solution to (5.3) can now be written, with arbitrary constants C_1 and C_2,

$$W(z) = \begin{cases} C_1 \sin m_1 z & -d < z < 0 \quad \text{(upper layer)} \\ C_2 \sin m_2(z+H) & -H < z < -d \quad \text{(lower layer)}\,, \end{cases} \tag{5.30}$$

where the boundary conditions at surface and bottom, $W(0) = W(-H) = 0$, are already satisfied. However, continuity of W and W' at the transition between the layers has to be imposed as well. These conditions can together be written in matrix form as

$$\begin{pmatrix} \sin m_1 d & \sin m_2(H-d) \\ m_1 \cos m_1 d & -m_2 \cos m_2(H-d) \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.31}$$

Requiring the determinant to be zero yields the *dispersion relation*

$$m_2 \sin m_1 d \cos m_2(H-d) + m_1 \cos m_1 d \sin m_2(H-d) = 0\,. \tag{5.32}$$

For given wave frequency ω (and constants f, N_1, N_2, d and H), we find the wavenumbers k as zeros of this equation, whence we obtain $m_{1,2}$. However, the equation is transcendent (i.e. cannot be solved by analytical means), so we have to resort to numerical methods to find the eigenvalues k. In fact, one can simply plot the left-hand side and pick out the zeros, see, e.g., Figure 5.8. Finally, using (5.31), we can express C_2 in terms of C_1:

$$C_2 = -\frac{\sin m_1 d}{\sin m_2(H-d)}\,C_1\,.$$

The remaining coefficient C_1, which fixes the amplitude, can be chosen arbitrarily.

The physics contained in this simple solution is already rich; in the next sections, we consider three cases to illustrate this.

5.3.1 Refraction and internal reflection

We choose the lower layer to be less strongly stratified than the upper one, see Figure 5.7a. Furthermore, we take $f = 1 \times 10^{-4}$ and $\omega = 1.4 \times 10^{-4}$ rad s^{-1}. As result the following inequalities are satisfied:

$$f < \omega < N_{1,2}\,,$$

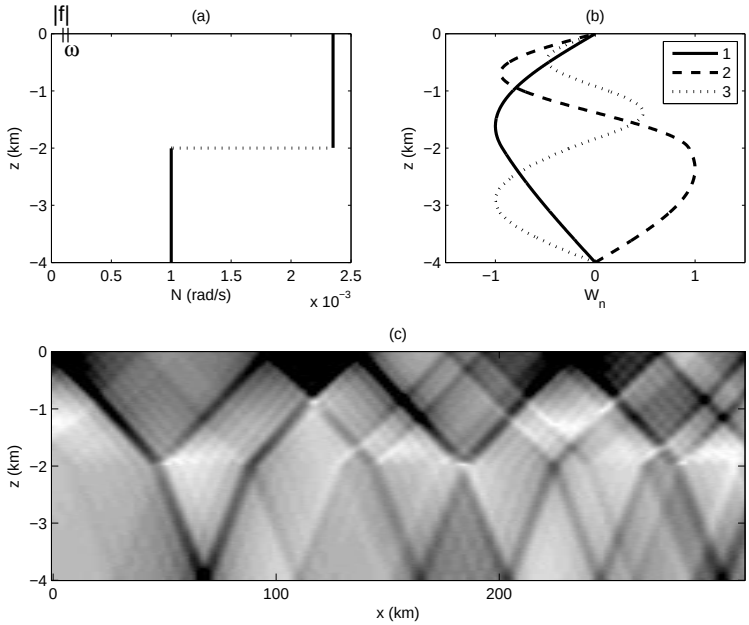

Fig. 5.7: Stratification with two layers of constant N, with $|f| < \omega < N_2 < N_1$ (**a**). Panel **b** shows the first three eigenmodes (5.30), with $C_{1,n}$ chosen such that their amplitudes are one. Modal coefficients are $a_n = 1/n$. The resulting superposition of 20 modes, representing the amplitude of u, is shown in **c**. White denotes zero; black, maximum values.

which implies that internal waves of the chosen frequency can propagate in *both* layers.

Eigenvalues are easily found as the zeros of (5.32), as shown in Figure 5.8; the first three corresponding vertical modes are shown in Figure 5.7b. From (5.2) and (5.5), we obtain u as

$$u = -\sum_n \frac{1}{k_n} W_n' \sin(k_n x - \omega t)\,,$$

with W_n given by (5.30), and normalized to one. A superposition of 20 modes, representing the amplitude of u, is shown in Figure 5.7c. It demonstrates that the beam is slightly steeper in the lower layer (in qualitative agreement with Figure 1.9), which must be due to the weaker stratification in that layer. So, non-uniformity in N causes *refraction*.

Moreover, we see that *internal reflections* occur at the transition between the two layers, yielding a complex pattern. Clearly, the pattern is no longer periodic in x (as it was for constant N, Section 5.2). This reflects the fact that the wavenumbers are now *in*commensurable: $k_n/k_l \neq n/l$. Indeed, the k_n's obtained from Figure 5.8 yield: $k_2/k_1 = 2.34$, $k_3/k_1 = 3.52$, $k_4/k_1 = 4.40$ etc.

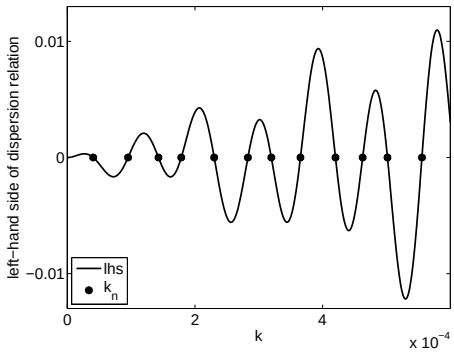

Fig. 5.8: The left-hand side of (5.32) plotted against k; zeros denote eigenvalues k_n, with n increasing from 1 to 12, from left to right.

5.3.2 Trapping of high-frequency waves ($\omega > |f|$)

We now choose a higher wave frequency such that $|f| < N_2 < \omega < N_1$. This means that m_2 becomes imaginary: $m_2 = i\Im(m_2)$. Cosines and sines containing m_2 can be rewritten using the identities

$$\sin(ix) = i\sinh(x)\,; \quad \cos(ix) = \cosh(x)\,. \tag{5.33}$$

As a consequence, the modal solution becomes exponential in z in the lower layer. (Notice that C_2 too becomes imaginary, hence the factor i drops out from (5.30), which thus remains real.)

The dispersion relation (5.32) remains the same, but we can rewrite it as

$$\Im(m_2)\sin m_1 d\cosh\Im(m_2)(H-d) + m_1\cos m_1 d\sinh\Im(m_2)(H-d) = 0\,.$$

We choose the same parameters as in Figure 5.7, except that the wave frequency ω is now higher, see Figure 5.9a. The first three modes are shown in panel Figure 5.9b. Their structure clearly is very different from that in Figure 5.7b. This demonstrates that *the modal structure now depends on the wave frequency*, in contrast to the absence of any such dependence in the case of uniform N (Section 5.2.2). Specifically, the modes now oscillate only in the upper layer, and decay exponentially in the lower one. Since higher modes have larger k_n, and hence larger $|m_2|$, they decay more rapidly than lower modes. Figure 5.9c shows a superposition of 15 modes; they add up to form a beam that is *trapped* in the upper layer. In other words, the upper layer now acts as a waveguide.

5.3.3 Trapping of low-frequency waves ($\omega < |f|$)

In contrast with the previous case, internal waves can also be trapped in the layer of *weakest* stratification. For this to happen, wave frequencies need to be sub-inertial, i.e. $\omega < |f|$. As an extreme example, we consider a neutrally stratified lower layer, $N_2 = 0$. Furthermore, we take $f = 1 \times 10^{-4}$ and $N_1 =$

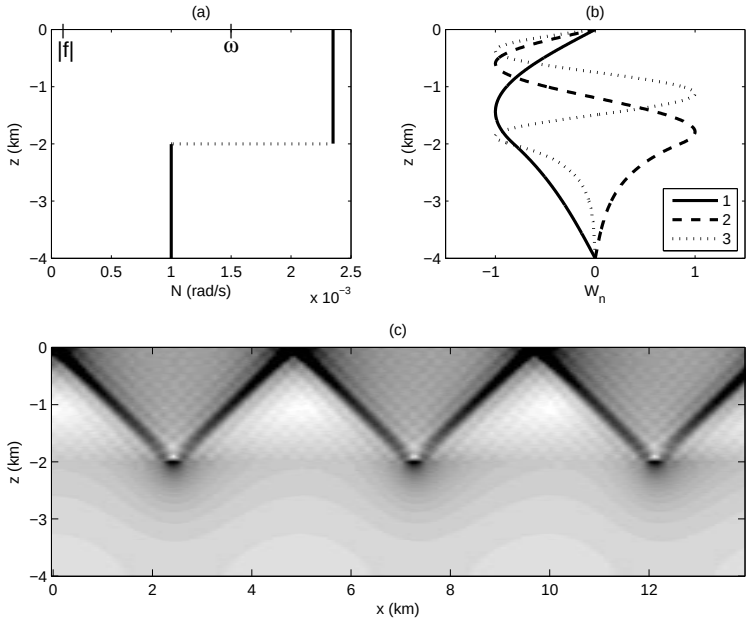

Fig. 5.9: Stratification with two layers of constant N; parameters are as in Figure 5.7, except for the wave frequency, which is now such that $|f| < N_2 < \omega < N_1$ (**a**). Panel **b** shows the first three eigenmodes (5.30), normalized to one. The resulting superposition of 15 modes, representing the amplitude of u, is shown in **c**. White denotes zero; black, maximum values.

5×10^{-4} rad s^{-1}. For the wave frequency ω we choose the diurnal tidal frequency $\mathrm{K}_1(=\Omega)$, 7.292×10^{-5} rad s^{-1}. So we have

$$N_2 < \omega < |f| < N_1\,.$$

According to (5.8), wave propagation can now only occur in the lower layer. Moreover, since $N_2 = 0$, the Coriolis force acts as the sole restoring force, in which case internal waves are called *gyroscopic* (Section 1.2.2).

Figure 5.10c confirms that it is now the lower layer that acts as a waveguide; hardly any signal penetrates into the upper layer. This is confirmed by the fact that the eigenvalues are very nearly given by

$$k_n \approx \frac{n\pi}{H/2}\left(\frac{\omega^2-f^2}{-\omega^2}\right)^{1/2},$$

as if the total water depth were just that of the lower layer, $H/2$.

From (5.16) we see that group velocity is now negative, implying that energy propagates to the left, while phase propagation is to the right (k being positive). This horizontal opposition is typical of gyroscopic waves.

A final comment concerns the effect of the hydrostatic approximation, briefly discussed in Section 5.1.3. This approximation amounts to assuming $N \gg \omega$.

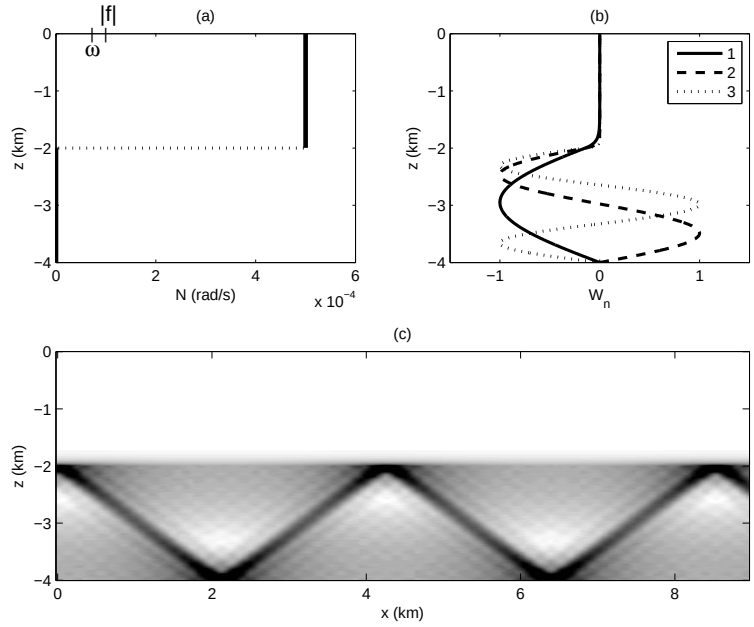

Fig. 5.10: Stratification with two layers of constant N, the lower layer being neutrally stable, $N_2 = 0$, with $N_2 < \omega < |f| < N_1$ (**a**). Panel **b** shows the first three eigenmodes (5.30), normalized to one. A superposition of 15 modes, representing the amplitude of u, is shown in **c**. White denotes zero; black, maximum values.

Obviously, one cannot assume this and later take $N = 0$. This indicates that the hydrostatic approximation is inappropriate in the presence of very weakly stratified layers; in fact, it would remove the entire class of sub-inertial waves.

5.4 A simple model for the ocean's stratification

There is not, in principle, a limitation to extending the procedure of the previous section to any number of layers. There is a good reason, indeed, to consider a *three*-layer system. The models of stratification discussed in Sections 5.2 and 5.3 miss an important feature of the ocean's stratification as depicted in, for example, Figure 3.3b, namely the strong peak in N, representing the seasonal thermocline. Schematically the structure, from surface to bottom, is as follows: a very weakly stratified upper mixed layer, a seasonal thermocline, and a fairly weakly stratified abyssal ocean, in which N decreases slowly with depth (apart from the presence of a second, less pronounced maximum, the permanent pycnocline, which we will ignore here). These features can be captured, though in an idealized way, by a model of stratification consisting of three layers of constant

N:

$$N^2(z) = \begin{cases} 0 & -d < z < 0 \quad \text{(mixed layer)} \\ g'/\epsilon & -d-\epsilon < z < -d \quad \text{(thermocline)} \\ N_c^2 & -H < z < -d-\epsilon \quad \text{(abyss)}\,. \end{cases} \tag{5.34}$$

see Figure 5.11. In the upper mixed layer, we take $N = 0$. The thermocline is represented by a thin layer of thickness ϵ and amplitude $N^2 = g'/\epsilon$. Notice that the area of N^2 enclosed by the thermocline equals g', irrespective of the value of ϵ; this property makes it easy to estimate the key parameter g' from empirical profiles of N^2. The whole 'abyssal' lower layer, finally, is simply represented by constant $N = N_c$.

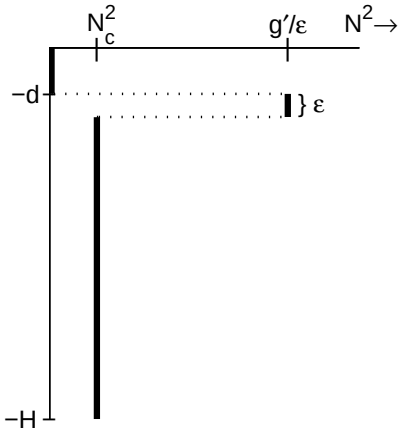

Fig. 5.11: Three-layer representation of the ocean's stratification: an upper mixed layer of thickness d, a thermocline of thickness ϵ, and a weakly stratified deep layer. Notice that the area enclosed by the thermocline is g', which serves as a measure of its strength.

We take the idealization one step further by taking $\epsilon \to 0$; this reduces the thermocline to an *interface*. In this limit, g' has a simple interpretation: it is g times the relative jump in density across the interface (this will be demonstrated in Section 8.3.2). We can now write (5.34) succinctly as

$$N^2(z) = g'\delta(z+d) + N_c^2\Theta(-z-d)\,, \tag{5.35}$$

where δ and Θ are the delta-distribution and Heaviside stepfunction, respectively (see Appendix).

5.4.1 Dispersion relation and vertical modes

We assume wave frequencies to lie in the interval

$$|f| < \omega < N_c\,.$$

We solve (5.3) separately for the mixed and abyssal layer. In the mixed layer, where $N = 0$, the solution to (5.3) is exponential and can be written

$$W_u = C_1 \sinh q_u z\,, \qquad \text{for } -d < z < 0\,, \tag{5.36}$$

with

$$q_u = k\left(\frac{\omega^2}{\omega^2 - f^2}\right)^{1/2}.$$

The other solution, $\cosh(\cdots)$ can be ignored a priori in view of the boundary condition at the surface, (5.4). In the abyssal layer, solutions to (5.3) are sinusoidal:

$$W_l = C_2 \sin q_l(z + H)\,, \qquad \text{for } -H < z < -d\,, \tag{5.37}$$

with

$$q_l = k\left(\frac{N_c^2 - \omega^2}{\omega^2 - f^2}\right)^{1/2}.$$

The cosine solution $\cos(\cdots)$ can be ignored in view of the boundary condition at the bottom.

We now have two pieces of the solution, W_u and W_l; they need to be matched at the thermocline. One obvious matching condition is continuity of W, i.e.

$$W_u(-d) = W_l(-d)\,. \tag{5.38}$$

We have yet to take into account the presence of the thermocline. This can be done by returning to the original profile (5.34), and taking the integral of (5.3) over the thermocline:

$$\begin{aligned}&\int_{-d-\epsilon}^{-d} dz \left\{W'' + k^2\,\frac{N^2(z) - \omega^2}{\omega^2 - f^2}\,W\right\}\\ &= W'(-d) - W'(-d-\epsilon) + k^2\,\frac{g'/\epsilon - \omega^2}{\omega^2 - f^2}\int_{-d-\epsilon}^{-d} dz\,W = 0\,.\end{aligned}$$

For small ϵ, the last integral reduces to $\epsilon W(-d)$; hence in the limit $\epsilon \to 0$,

$$W_u'(-d) - W_l'(-d) + \frac{g' k^2}{\omega^2 - f^2}\,W(-d) = 0\,. \tag{5.39}$$

For the evaluation of the last term, it is immaterial which of the two, W_u or W_l, is taken, because of (5.38). Without a thermocline ($g' = 0$), the first derivative of W would be continuous. We see from (5.39) that the presence of the thermocline creates a discontinuity in W', and hence, by (5.5), a jump in the horizontal velocities u and v across the interface. Thus, the thermocline is accompanied by a strong vertical shear.

We gather (5.38) and (5.39) in a matrix

$$\begin{pmatrix} \sinh q_u d & \sin q_l(H-d) \\ q_u \cosh q_u d - \frac{g' q_u^2}{\omega^2}\sinh q_u d & -q_l \cos q_l(H-d) \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.40}$$

The dispersion relation follows from the requirement that the determinant be zero:

$$q_l \cos q_l(H-d) \sinh q_u d + q_u \sin q_l(H-d) \Big[\cosh q_u d - \frac{g' q_u}{\omega^2} \sinh q_u d \Big] = 0\,. \quad (5.41)$$

One can obtain numerically the zeros k_n of this transcendental equation. We choose $C_2 = 1$; using (5.36), (5.37) and (5.51), we obtain the vertical modes

$$W_n(z) = \begin{cases} -\frac{\sin q_l(H-d)}{\sinh q_u d} \sinh q_u z & -d < z < 0 \\ \sin q_l(z+H) & -H < z < -d\,. \end{cases} \quad (5.42)$$

Notice that the coefficients q_u and q_l depend on modenumber n, via k_n.

5.4.2 Scattering at the thermocline

An example of the first five vertical modes is shown in Figure 5.12a, and a superposition of 25 modes in 5.12b. The value used here for g' is $0.005\,\mathrm{m\,s^{-2}}$.

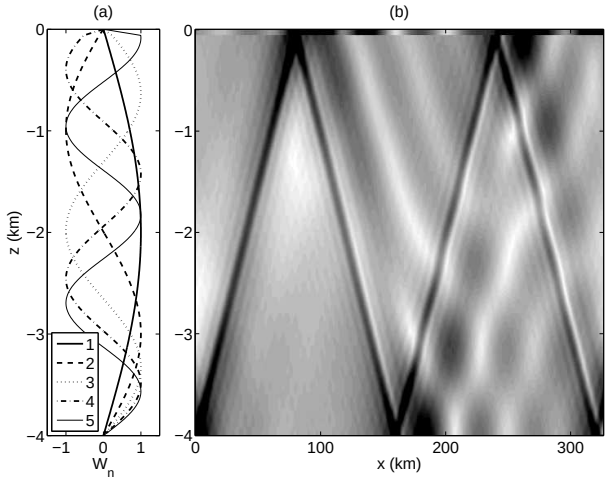

Fig. 5.12: The first five modes for the three-layer system (**a**), here with $g' = 0.005\,\mathrm{m\,s^{-2}}$. Other parameters are: $d = 60\,\mathrm{m}$ (mixed-layer depth), $N_c = 2 \times 10^{-3}$, $f = 1 \times 10^{-4}$ and $\omega = 1.4 \times 10^{-4}$, all in $\mathrm{rad\,s^{-1}}$; modal coefficients are $a_n = 1/n$. In **b**, a superposition of 25 modes, representing the amplitude of u. White denotes zero; black, maximum values.

In the lower layer internal-wave beams are clearly visible, as in previous sections, but where they impinge on the thermocline, strong currents appear in the mixed layer; the beam, meanwhile, becomes slightly less intense. Further rightward, we see broad, rather weak beams radiating downward from the thermocline. There is not really a transfer of energy, of course; this is a linear solution, and there is no interaction among modes. It is just the superposition of modes that at some places yields strong currents in the mixed layer. A different

way to look at this is as the occurrence of internal reflections, discussed in Section 5.3.1. The thermocline forms a strong inhomogeneity in the stratification, and causes an impinging beam to scatter, due to internal reflections. Observational evidence from the Bay of Biscay, discussed in Chapter 8, demonstrates that this process may lead to the generation of so-called solitary waves.

In is instructive to consider also the case of an extremely strong thermocline, taking, for example, $g' = 0.5\,\mathrm{m\,s^{-2}}$ (this is an admittedly unrealistic value!). The result is shown in Figure 5.13. The thermocline is now so strong that beams in the lower layer reflect from it as if it were a rigid surface, leaving almost no trace in the mixed layer (Figure 5.13c). The first mode, on the other hand, is accompanied by strong currents in the upper mixed layer, which show a periodic pattern (Figure 5.13b). The interpretation of these results is further explored in the next section.

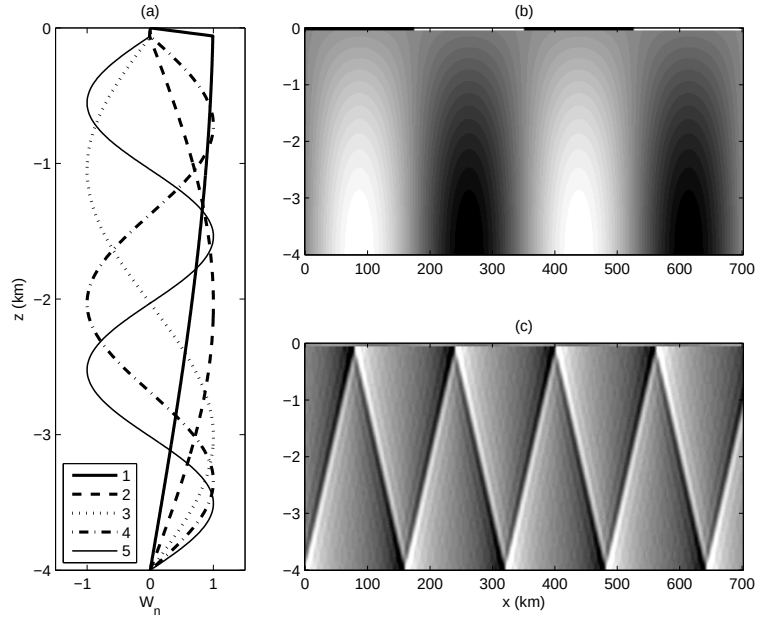

Fig. 5.13: The first five modes for the three-layer system (**a**), now with $g' = 0.5\,\mathrm{m\,s^{-2}}$. Other parameters as in previous figure. In **b**, the contribution of the first mode to u, at $t = 0$, is shown; in **c**, a superposition of modes 2 to 14. In **b** and **c**, white denotes negative values; black, positive.

5.4.3 Interfacial waves

A special case of interest is $N_c = 0$. Obviously, we then have to drop the restriction $\omega < N_c$. The other restriction, $\omega > |f|$, is however retained. With $N_c = 0$, the stratification of (5.35) becomes that of a so-called *two-layer model*,

i.e. there are two neutrally stratified layers ($N = 0$) separated by an interface.[1] This model can sometimes be applied to internal waves in shallow regions such as the continental shelf, or to high-frequency internal waves trapped at the thermocline, such as solitons (see Chapter 8). A two-layer model is however wholly unsuitable for the description of internal waves in the deep ocean.

With $N_c = 0$, q_l becomes imaginary, and we can write $q_l = iq_u$. Using the identities (5.33), we can derive the dispersion relation for the two-layer system from (5.41):

$$\omega^2 = \frac{g' q_u}{\coth q_u d + \coth q_u (H-d)} \tag{5.43}$$

$$= f^2 + \frac{g' k^2}{q_u(\coth q_u d + \coth q_u (H-d))}\,. \tag{5.44}$$

The connection between ω and k is implicitly contained in (5.43),[2] which yields ω as a function of q_u; then k follows from

$$k = q_u \left(\frac{\omega^2 - f^2}{\omega^2}\right)^{1/2}.$$

The result is shown in Fig 5.14. For any choice of ω, there is now only *one* k, and hence only one mode, the interfacial mode,

$$W(z) = \begin{cases} -\frac{\sinh q_u(H-d)}{\sinh q_u d} \sinh q_u z & -d < z < 0 \\ \sinh q_u (z+H) & -H < z < -d\,. \end{cases} \tag{5.45}$$

This expression follows from (5.42), leaving out the factor i. The maximum of W occurs at the interface (thermocline); from there, it decreases steadily towards the surface and bottom. This means that W' scales as $[W]/d$ in the upper layer ($[W]$ being the scale of W), and as $[W]/(H-d)$ in the lower layer. The continuity equation (4.21d) implies that u must scale with d^{-1} in the upper layer, and with $(H-d)^{-1}$ in the lower layer. Typically d is much smaller than H, and this explains the strong currents in the mixed layer in Figure 5.13b.

Without rotation ($f = 0$), we can replace q_u with k in (5.44). For long

[1] This is not be confused with the stratification considered in Section 5.3, which consists of two layers with constant *buoyancy frequency* N; there N is discontinuous, but density itself is continuous. In the present section, the latter is no longer the case: a jump in density occurs across the interface, at which N is infinite, i.e. a delta-peak.

[2] Notice that an explicit dependence is found for high-frequency waves ($\omega \gg |f|$), in which case $q_u \approx k$. In fact, replacing q_u by k in (5.44) yields a relation that turns out to be very accurate even for *low*-frequency waves.

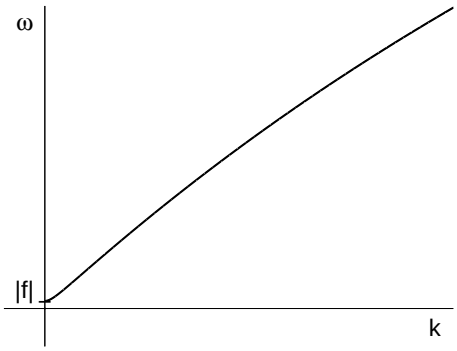

Fig. 5.14: The dispersion relation (5.44), plotted as ω vs. k. Here for the parameter values $g' = 0.01\,\mathrm{m\,s^{-2}}$, $d = 100\,\mathrm{m}$, $H = 4000\,\mathrm{m}$, and $f = 1\times10^{-4}\,\mathrm{rad\,s^{-1}}$. The function is monotonically increasing; hence for every $\omega > |f|$ there is only *one* wavenumber k.

waves, $kH \ll 1$, we then find[3]

$$\begin{aligned}
\omega^2 &= \frac{g'k^2d(H-d)}{(H-d)[kd\coth kd] + d[k(H-d)\coth k(H-d)]} \\
&= \frac{g'k^2d(H-d)}{(H-d)[1+\frac{1}{3}(kd)^2+\cdots] + d[1+\frac{1}{3}(k(H-d))^2+\cdots]} \\
&= g'k^2\,\frac{d(H-d)}{H}\,\frac{1}{1+\frac{1}{3}k^2d(H-d)+\cdots} \\
&= c_0^2k^2\,[1-\tfrac{1}{3}k^2d(H-d)+\cdots]\,, \qquad (5.46)
\end{aligned}$$

with

$$c_0^2 = g'\,\frac{d(H-d)}{H}\,,$$

the linear long-wave phase speed for interfacial waves. Interfacial waves propagate horizontally, just like surface waves. The main difference between the two is the strongly reduced effective gravity, g' instead of g. For *surface* waves, the long-wave phase speed is given by gH; so g is here replaced with g', and total water depth H with an 'equivalent' depth $d(H-d)/H$, i.e. the product of the thickness of each layer, divided by the total water depth. In practice, we often have $d \ll H$ so that $c_0^2 \approx g'd$. Typical values are $g' = 0.01\,\mathrm{m\,s^{-2}}$ and $d = 100\,\mathrm{m}$, giving a phase speed of one meter per second.

Returning now to the interpretation of Figure 5.13, we first notice that modes 2 to 4 strongly resemble modes 1 to 3 for a layer of constant stratification, cf. Figure 5.3; in particular, the second mode in Figure 5.13a has one maximum halfway down the water column, and vanishes at the thermocline, as if it were a rigid surface. Thus, a superposition of modes $2, 3, 4\cdots$ creates beams in the lower layer, as illustrated in Figure 5.13c. The first mode in Figure 5.13a stands

[3]Using the Taylor expansion, for $x \ll 1$, $x\coth x = 1 + \frac{1}{3}x^2 - \frac{1}{45}x^4 + \cdots$.

apart from the higher modes in that it has its maximum at the thermocline; in this, we recognize the interfacial mode (5.45). Thus, the wave propagation in Figure 5.13b,c consists of two distinct, independent elements: 1) an interfacial mode, and 2) the higher modes which form beams in the lower, constantly stratified layer. The two kinds of motion are here uncoupled, because the thermocline is extremely strong. For a weaker, and indeed more realistic thermocline, like in Figure 5.12, the two are coupled, as is clear from the fact that the third and fifth modes also have a local extremum at the thermocline. So, in that case, one cannot really speak of pure interfacial waves or pure beams; they appear in mixed form.

To conclude the discussion on interfacial waves, we mention that there is a caveat to the notion that the two-layer system supports only one mode. If we relax the restriction $\omega > |f|$, we find that there is an infinite number of subinertial modes, $\omega < |f|$, both in the upper mixed layer and in the lower neutrally stratified layer ($N_c = 0$). These are the gyroscopic waves already discussed in Section 5.3.3.

5.5 Linearly varying N^2: Airy functions

There are various choices of profiles $N(z)$ that allow for a solution in terms of special functions. Several of them are conveniently listed in [72, §3.4]. We will not repeat them here; the physics contained in these solutions is not essentially different from what has been discussed in previous sections. There is however one case that deserves our attention, because it provides a generic description of the behaviour near a 'turning point', i.e. the depth where internal waves turn from oscillatory to exponentially decaying.

We consider continuous profiles $N(z)$, and assume that $N(z) > |f|$ throughout the water column. Suppose that at some depth z_*, a transition occurs from $N(z) > \omega$ to $N(z) < \omega$. A Taylor expansion around this depth gives

$$N^2(z) = N_0^2 + \lambda z + \cdots , \tag{5.47}$$

with $\lambda = N^{2'}(z_*)$ and $N_0^2 = N^2(z_*) - \lambda z_*$. In other words, in the vicinity of the depth where internal waves turn from oscillatory to exponentially decaying, we may regard the profile of N^2 as linear in z.

This provides the motivation for examining the case (5.47) more closely. We do this by applying (5.47) to the whole water column. The equation for vertical modes, (5.3), can then be written

$$W'' + \hat{k}^2 \, (z + A_0) \, W = 0 \, . \tag{5.48}$$

with $\hat{k}^2 = k^2\lambda/(\omega^2 - f^2)$ and $A_0 = (N_0^2 - \omega^2)/\lambda$. With the coordinate transformation $\hat{z} = -\hat{k}^{2/3}[z + A_0]$, (5.48) reduces to the *Airy equation*:

$$\frac{d^2W}{d\hat{z}^2} - \hat{z}W = 0 \, . \tag{5.49}$$

Its general solution can be written in terms of the Airy functions Ai and Bi:

$$W = C_1 \mathrm{Ai}(\hat{z}) + C_2 \mathrm{Bi}(\hat{z})\,.$$

As illustrated in Figure 5.15, Ai and Bi are oscillatory for $\hat{z} < 0$; for $\hat{z} > 0$, Ai

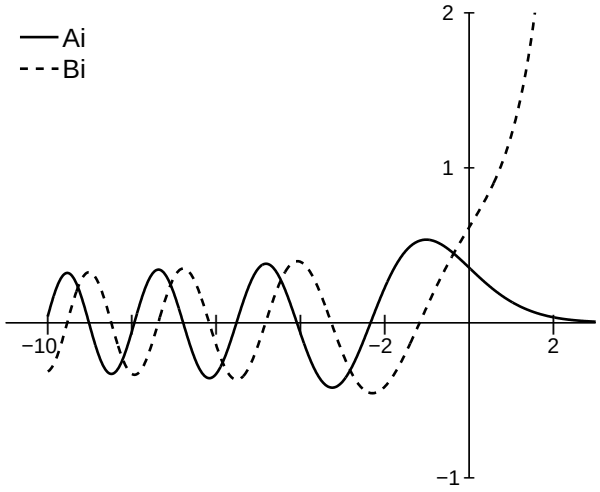

Fig. 5.15: The Airy functions Ai and Bi, which are two independent solutions of (5.49).

decays exponentially, while Bi grows exponentially.[4] Transformed back to the original coordinate z, the solution reads

$$W = C_1 \mathrm{Ai}(-\hat{k}^{2/3}[z + A_0]) + C_2 \mathrm{Bi}(-\hat{k}^{2/3}[z + A_0])\,. \tag{5.50}$$

Application of the boundary conditions (5.4) gives

$$\begin{pmatrix} \mathrm{Ai}(-\hat{k}^{2/3}A_0) & \mathrm{Bi}(-\hat{k}^{2/3}A_0) \\ \mathrm{Ai}(-\hat{k}^{2/3}[A_0 - H]) & \mathrm{Bi}(-\hat{k}^{2/3}[A_0 - H]) \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.51}$$

The determinant has to be zero to have non-trivial solutions for (C_1, C_2); this requirement yields the dispersion relation

$$\mathrm{Ai}(-\hat{k}^{2/3}A_0)\mathrm{Bi}(-\hat{k}^{2/3}[A_0 - H]) - \mathrm{Bi}(-\hat{k}^{2/3}A_0)\mathrm{Ai}(-\hat{k}^{2/3}[A_0 - H]) = 0\,. \tag{5.52}$$

This transcendent equation can be solved numerically (for example by simply plottoing the left-hand side and picking out the zeros), and yields the eigenvalues k_n, $n = 1, 2, 3 \ldots$.

Using (5.51), we can express C_2 in terms of C_1 (which we give an additional index n, because different values may be chosen for different modes); the vertical

[4] We do not here dwell on the analytical properties of Airy functions; for integral representations, Taylor series, asymptotic expansions etc., see [4].

modes (5.50) can then be written

$$W_n = C_{1,n}\Big(\mathrm{Ai}(-\hat{k}_n^{2/3}[z+A_0]) - \frac{\mathrm{Ai}(-\hat{k}_n^{2/3}A_0)}{\mathrm{Bi}(-\hat{k}_n^{2/3}A_0)}\,\mathrm{Bi}(-\hat{k}_n^{2/3}[z+A_0])\Big)\,. \qquad (5.53)$$

In the examples discussed below, we choose $C_{1,n}$ such that $\max\{W_n\} = 1$.

We consider two examples, for different wave frequencies; the first showing refraction, the second, trapping. In both cases, we choose $H = 4000\,\mathrm{m}$, $N_0 = 1.5 \times 10^{-3}\,\mathrm{rad\,s^{-1}}$, $\lambda = 5 \times 10^{-10}\,\mathrm{rad^2\,m^{-1}\,s^{-2}}$, and $f = 10^{-4}\,\mathrm{rad\,s^{-1}}$. The buoyancy frequency N thus decreases from N_0 at the surface to $5 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$ at the bottom; this covers more or less the range of values found in the deeper layers of the ocean, see Figure 1.6.

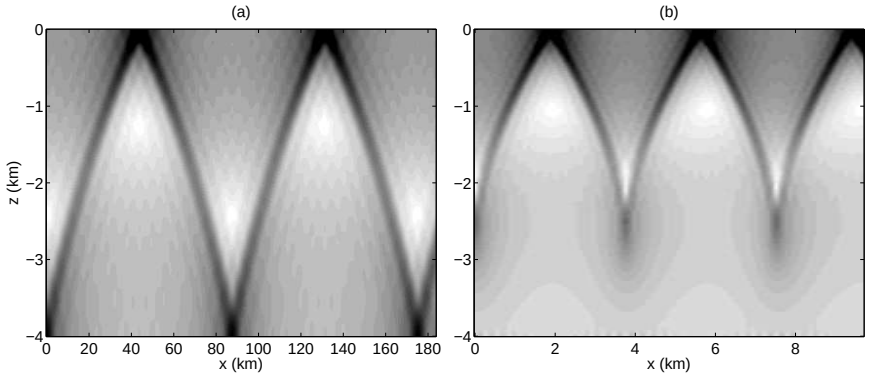

Fig. 5.16: Solution for a linearly varying $N(z)$; the amplitude of u is shown for a superposition 15 modes, with modal coefficients $a(n) = 1/n$. White denotes zero; black, maximum values. In **a**, internal-wave beams can propagate at any depth, but are refracted due to the decrease of N with depth. In **b**, a higher wave frequency is chosen, such that $|f| < N < \omega$ in the deeper part of the water column; hence, internal waves are trapped in the upper layer.

First we choose $\omega = 1.4 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$ (semi-diurnal internal tides). For this value, we have $|f| < \omega < N(z)$ for all z; the waves can thus propagate throughout the water column. A superposition of 15 modes, depicting the amplitude of u, is shown in Figure 5.16a. The beams are slightly bended due to the decreasing stratification N with depth.

Taking now a higher frequency, $\omega = 1.0 \times 10^{-3}\,\mathrm{rad\,s^{-1}}$, for which the inequality $N(z) > \omega$ is satisfied only in the upper 2500 m of the water column, so that the waves are trapped there; they reflect at the base of that layer, as is illustrated in Figure 5.16b.

5.6 Non-traditional effects

To conclude this chapter, we briefly discuss how the results obtained so far would change if one abandons the Traditional Approximation. We return to

(5.1):

$$\frac{\partial^2}{\partial t^2}\nabla^2 w + (\vec{f}\cdot\nabla)^2 w + N^2\nabla_h^2 w = 0\,, \tag{5.54}$$

where $\vec{f} = (0, \tilde{f}, f)$. It is convenient to rotate the coordinate system in the horizontal plane, at an arbitrary angle α in anti-clockwise direction:

$$x' = x\cos\alpha + y\sin\alpha\,; \quad y' = -x\sin\alpha + y\cos\alpha\,,$$

as illustrated in Figure 5.17. Hence

$$\begin{aligned}
\frac{\partial}{\partial x} &= \frac{\partial x'}{\partial x}\frac{\partial}{\partial x'} + \frac{\partial y'}{\partial x}\frac{\partial}{\partial y'} = \cos\alpha\frac{\partial}{\partial x'} - \sin\alpha\frac{\partial}{\partial y'} \\
\frac{\partial}{\partial y} &= \frac{\partial x'}{\partial y}\frac{\partial}{\partial x'} + \frac{\partial y'}{\partial y}\frac{\partial}{\partial y'} = \sin\alpha\frac{\partial}{\partial x'} + \cos\alpha\frac{\partial}{\partial y'}\,.
\end{aligned}$$

With this transformation, (5.54) becomes

$$\frac{\partial^2}{\partial t^2}\nabla'^2 w + (\vec{f}'\cdot\nabla')^2 w + N^2{\nabla'_h}^2 w = 0\,, \tag{5.55}$$

with

$$\nabla'^2 = \frac{\partial^2}{\partial x'^2} + \frac{\partial^2}{\partial y'^2} + \frac{\partial^2}{\partial z^2}\,; \quad {\nabla'_h}^2 = \frac{\partial^2}{\partial x'^2} + \frac{\partial^2}{\partial y'^2}\,; \quad \vec{f}' \equiv (\tilde{f}\sin\alpha, \tilde{f}\cos\alpha, f)\,.$$

Notice that the angle α is exclusively connected to the non-traditional $\tilde{f}$, confirming the fact that the angle becomes immaterial under the Traditional Approximation (isotropy).

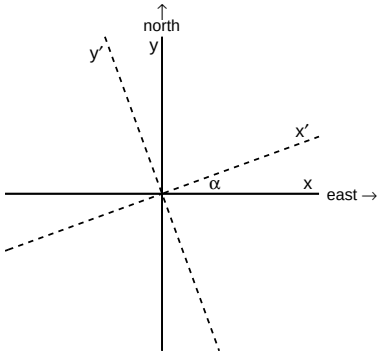

Fig. 5.17: Coordinate transformation to the primed system, at an angle α with respect to the original east-north system.

We consider waves travelling in the x'-direction, with $\partial/\partial y' = 0$; this leaves the geographical direction of propagation still arbitrary, via α. Moreover, we assume

$$w \sim \exp(-i\omega t)\,.$$

With this, (5.55) reduces to

$$A\frac{\partial^2 w}{\partial x'^2} + 2B\frac{\partial^2 w}{\partial x' \partial z} + C\frac{\partial^2 w}{\partial z^2} = 0\,. \tag{5.56}$$

The coefficients A, B and C are defined by

$$A = N^2 - \sigma^2 + f_s^2\,; \quad B = f f_s\,; \quad C = f^2 - \sigma^2\,; \quad \text{with } f_s = \tilde{f}\sin\alpha\,.$$

Non-traditional effects are now represented by f_s. Of particular significance is the *mixed derivative* in (5.56), which would disappear under the Traditional Approximation ($f_s = 0$, hence $B = 0$). This term precludes solutions obtained by separation of variables, i.e. real solutions of the form $F(x)G(z)$. Still, it is straightfoward to solve (5.56) if we use the transformation

$$w = W(z)\exp ik(x' - Bz/C)\,, \tag{5.57}$$

by which (5.56) becomes

$$\frac{d^2W}{dz^2} + k^2\,\frac{B^2 - AC}{C^2}\,W = 0\,. \tag{5.58}$$

This expression replaces the 'traditional' (4.40), and solutions are obtained similarly as in previous sections. We consider here the case of constant N, so that A is constant.

5.6.1 Frequency range

For sinusoidal solutions to be possible in (5.58), $B^2 - AC$ must be positive. This condition can be rewritten as

$$\omega_{\min} < \omega < \omega_{\max}\,,$$

with

$$\omega_{\min,\max} = \frac{1}{\sqrt{2}}\Big([N^2 + f^2 + f_s^2] \mp \Big\{[N^2 + f^2 + f_s^2]^2 - (2fN)^2\Big\}^{1/2}\Big)^{1/2}\,. \tag{5.59}$$

A principal difference with the 'traditional' range (i.e. $\omega_{\min} = \min(|f|, N)$ and $\omega_{\max} = \max(|f|, N)$) is that both bounds now depend on both latitude (via f and f_s) and stratification N. Moreover, it can be proven that the range is, in general, enhanced:

$$\omega_{\min} \le \min(|f|, N) < \omega < \max(|f|, N) \le \omega_{\max}\,.$$

The only exception is for waves travelling in the west-east direction, so that $\alpha = 0$, in which case (5.59) reduces to the traditional bounds.

The width of the extension of the range depends on stratification N; in a strongly stratified fluid, $N \gg \Omega$, (5.59) can be approximated by

$$\omega_{\min} = |f|\,\Big[1 - \frac{f_s^2}{2N^2} + O\Big(\frac{\Omega^4}{N^4}\Big)\Big]\,; \quad \omega_{\max} = N\Big[1 + \frac{f_s^2}{2N^2} + O\Big(\frac{\Omega^4}{N^4}\Big)\Big]\,. \tag{5.60}$$

Thus, in this limit, the lower and upper bounds approach the traditional values. However, it can be seen from Figure 1.6 that the condition $N \gg 2\Omega = 1.5 \times 10^{-4}$ (rad s^{-1}) is *not* satisfied in the deeper layers of the ocean, and here non-traditional effects can be expected to be important.

The contrast between traditional and non-traditional ranges is most clearly seen for south-north propagation ($\alpha = \pi/2$) in regions where $N \approx 0 \ll |f|$ (in convective layers, for example); then we have, taking $N = 0$,

$$0 < \omega < |f| \quad \text{(Traditional Approximation)}\,; \quad 0 < \omega < 2\Omega \quad \text{(non-traditional)}\,.$$

The difference is stark for an important class of internal waves: the semi-diurnal internal tides, whose frequency is slightly less than 2Ω. The latter range indicates that they can exist at *all* latitudes in neutrally stratified waters, whereas the former range resticts their habitat to polar seas.

More generally, under the Traditional Approximation the ranges $N > |f|$ and $N < |f|$ are mutually exclusive; internal waves existing in one range are precluded from the other. Without the Traditional Approximation, however, there is always an overlap, especially around near-inertial frequencies; this overlap includes a band of sub-inertial ($\omega < |f|$) and super-inertial ($\omega > |f|$) frequencies. In stratified fluids, the former range may be small (as indicated by (5.60)), but as will be demonstrated in the following sections, the behaviour of these sub-inertial waves is altogether different from that of 'traditional' low-frequency internal waves.

5.6.2 Dispersion relation for constant N

For constant stratification, solutions of (5.58), satisfying the boundary conditions (5.4), are given by

$$W_n = \sin\left(\frac{n\pi z}{H}\right). \tag{5.61}$$

This is identical to the traditional expression (5.21), but notice that it no longer represents the sole dependence on z; there is now an *extra* vertical dependence via (5.57). Substitution of (5.61) in (5.58) yields the dispersion relation

$$k_n = \mp\frac{n\pi}{H}\,\frac{C}{(B^2 - AC)^{1/2}} = \pm\frac{n\pi}{H}\,\frac{\omega^2 - f^2}{[(\omega^2 - \omega_{\min}^2)(\omega_{\max}^2 - \omega^2)]^{1/2}}\,, \tag{5.62}$$

where $\omega_{\min,\max}$ are given by (5.59). (The second equality is readily verified by noticing that the zeros of $B^2 - AC$ occur, by definition, at $\omega = \omega_{\min}, \omega_{\max}$.) One important consequence of (5.62) is immediately clear: waves become short not only in the high-frequency limit $\omega \to \omega_{\max}$ (as under the Traditional Approximation), but now also in the *low*-frequency limit $\omega \to \omega_{\min}$. This affects, moreover, the vertical scale as well, via (5.57). The sub-inertial short-wave limit is illustrated in Figure 5.18.

It is also clear that the thick solid and dashed lines in Figure 5.18 cross the frequency $|f|$ at an angle, instead of horizontally as under the Traditional

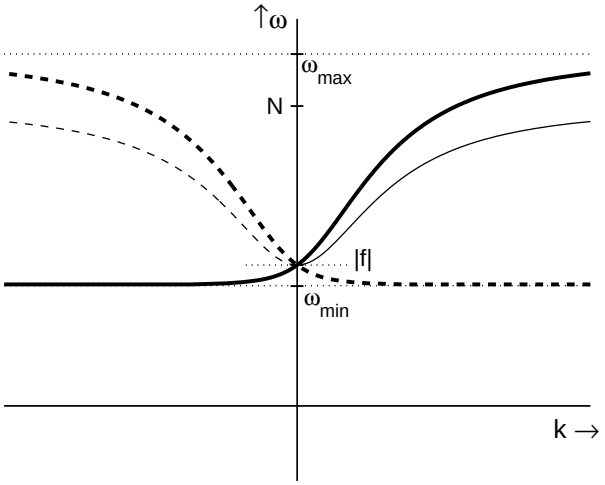

Fig. 5.18: The dispersion relation (5.62), showing both the plus-branch (solid thick line), and the minus-branch (dashed thick line). For comparison, the traditional result (5.14) is also shown (thin solid and dashed lines). After [28].

Approximation. This indicates the group velocity does not vanish at $\omega = |f|$ if non-traditional terms are included; indeed, it vanishes only at the upper and lower bounds, $\omega_{\max}$ and $\omega_{\min}$. (This can be verified by actually deriving c_g, which we leave as an exercise for the reader.)

5.6.3 Expressions of other fields

The expressions listed in (5.5) undergo a modification if non-traditional terms are included. To derive them, we need to return to (4.21), now in terms of the rotated coordinate system (Figure 5.17):

$$\frac{\partial u}{\partial t} - fv + f_c w = -\frac{1}{\rho_*}\frac{\partial p'}{\partial x'} \tag{5.63a}$$

$$\frac{\partial v}{\partial t} + fu - f_s w = -\frac{1}{\rho_*}\frac{\partial p'}{\partial y'} \tag{5.63b}$$

$$\frac{\partial w}{\partial t} + f_s v - f_c u = -\frac{1}{\rho_*}\frac{\partial p'}{\partial z} + b \tag{5.63c}$$

$$\frac{\partial u}{\partial x'} + \frac{\partial v}{\partial y'} + \frac{\partial w}{\partial z} = 0 \tag{5.63d}$$

$$\frac{\partial b}{\partial t} + N^2 w = 0\,. \tag{5.63e}$$

Here u and v now denote the velocity components in the x' and y' direction, respectively. In line with (5.57), we take w of the form

$$w = \sum_n a_n W_n(z) \exp[ik_n(x' - Bz/C) - i\omega t]\,. \tag{5.64}$$

From this, incidentally, the non-separable nature of the solution is evident.

Using (5.64), we obtain from (5.63):

$$U_n = \frac{i}{k_n} W_n' - \frac{f f_s}{\omega^2 - f^2} W_n \tag{5.65a}$$

$$V_n = \frac{f}{\omega k_n} W_n' + \frac{i\omega f_s}{\omega^2 - f^2} W_n \tag{5.65b}$$

$$P_n = i\rho_* \frac{\omega^2 - f^2}{\omega k_n^2} W_n' + i\rho_* \frac{f_c}{k_n} W_n \tag{5.65c}$$

$$B_n = -\frac{iN^2}{\omega} W_n \,. \tag{5.65d}$$

Notice that all expressions except the one for B_n have an extra term with respect to the 'traditional' (5.5). The full expression for u is now formed by replacing W_n with U_n in (5.64), and similarly for the other variables.

5.6.4 Superposition of modes

We show two examples of internal-wave beams as they manifest themselves without the Traditional Approximation. Starting from the general solution for rightward propagating waves (5.64), with k_n positive, we obtain the expression for u using (5.65a). We assume N constant with $N > |f|$, and consider a sub-inertial and super-inertial case.

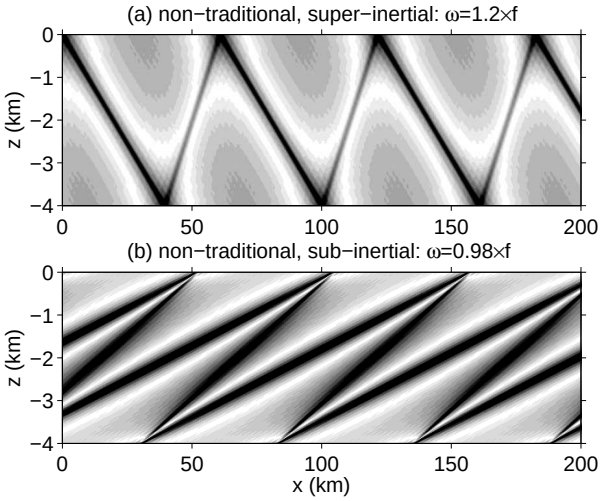

Fig. 5.19: A superposition of 25 modes, showing $|u|$ at $t = 0$. In **a** the wave frequency is super-inertial, in **b**, sub-inertial. In both panels, parameter values are: $N = 5 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$, $\phi = 45°\mathrm{N}$ (latitude), and $\alpha = \pi/2$ (i.e. the x-axis is directed south-north, positive northward). The modal coefficients are $a_n = 1/n$. White denotes zero; black, maximum values.

The sub-inertial case shown in Figure 5.19b would not exist under the Traditional Approximation. It is marked by a curious retrograde kind of propagation. Phase propagation is of course to the right, since we have chosen k_n positive. The direction of energy propagation can be inferred from Figure 5.18. We have here a sub-inertial wave, $\omega < |f|$, with k positive, so the dashed thick line on the right applies. It descends, hence $d\omega/dk$ is negative, implying that energy propagates to the *left*, opposite to phase propagation.

In the super-inertial case (Figure 5.19a), the thick solid line applies, so energy propagates to the right. An obvious difference from earlier 'traditional' examples (see, e.g., Figure 5.6) is that there is now an asymmetry in the zigzag pattern of beams; the up- and downward beams are not equally steep. This fact will be derived in a straightfoward way in the following chapter, Section 6.6.

Appendix: the delta-distribution

Consider the family of functions f_ϵ, with $\epsilon > 0$, defined by

$$f_\epsilon(x) = \begin{cases} \frac{1}{2\epsilon} & |x| \le \epsilon \\ 0 & |x| > 0\,. \end{cases}$$

This family forms a δ-approximation, i.e. $f_\epsilon(x) \to \delta(x)$. Because of its singular character, δ is not a function in the proper sense of the word (unlike f_ϵ), but a so-called distribution (an object that maps a function onto a number). The mathematical details need not concern us here. One property of practical importance can be derived us follows. For an arbirary continuous function g, we find

$$\int_{-\infty}^{\infty} dx\, f_\epsilon(x) g(x) = \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} dx\, g(x) = g(c)\,,$$

for a certain $-\epsilon < c < \epsilon$ (mean-value theorem for integrals). In the limit $\epsilon \to 0$ we thus find

$$\int_{-\infty}^{\infty} dx\, \delta(x) g(x) = g(0)\,. \tag{5.66}$$

A second useful property of the δ-distribution follows from the the primitive of the family f_ϵ:

$$F_\epsilon(x) = \begin{cases} 0 & x < -\epsilon \\ \frac{x+\epsilon}{2\epsilon} & |x| \le \epsilon \\ 1 & x > \epsilon\,, \end{cases}$$

so $F'_\epsilon = f_\epsilon$. In the limit $\epsilon \to 0$, F_ϵ becomes the Heaviside stepfunction:

$$\Theta(x) = \begin{cases} 0 & x < 0 \\ 1 & x \ge 0\,. \end{cases}$$

Thus, the δ-distribution can be interpreted as the derivative of stepfunction:

$$\frac{d\Theta}{dx}(x) = \delta(x)\,. \tag{5.67}$$

Chapter 6

Internal wave-propagation II: method of characteristics

In the previous chapter, we found that a superposition of modes produces internal-wave beams, which propagate diagonally in the x, z-plane. In this chapter we derive this fact in a much more direct way. Also, we treat problems that are beyond the scope of the method of vertical modes, such as reflection from sloping boundaries, which is a particularly relevant topic, given the distinctly bumpy nature of ocean bottom topography (as illustrated in Figure 1.6).

6.1 Basic properties of internal waves

Our starting point is (4.35), which governs the spatial structure of internal waves:

$$(N^2 - \omega^2)\nabla_h^2 w - (\omega^2 - f^2)w_{zz} = 0\,. \tag{6.1}$$

This expression is based on the Traditional Approximation; we defer a discussion of non-traditional effects to Section 6.6. We first consider plane waves, i.e. $\partial/\partial y = 0$. Since the roles of x and y are entirely similar in (6.1), the results obtained for propagation in the x direction will equally apply to any other horizontal direction. We note that a fully three-dimensional problem is discussed in Section 6.4. Hereafter we assume N to be constant; effects of vertically varying N are discussed in Section 6.5.

6.1.1 Dispersion relation and corollaries

With $\partial/\partial y = 0$, then, (6.1) reduces to

$$(N^2 - \omega^2)\frac{\partial^2 w}{\partial x^2} - (\omega^2 - f^2)\frac{\partial^2 w}{\partial z^2} = 0\,. \tag{6.2}$$

The dispersion relation, which provides the connection between wave frequency and wavenumbers, can be found by substituting $w = \exp i(kx + mz)$ into (6.2):

$$\omega^2 = \frac{N^2k^2 + f^2m^2}{k^2 + m^2}\,. \tag{6.3}$$

We can simplify this expression by writing the wavevector in polar coordinates,

$$\vec{k} = (k, m) = \kappa(\cos\theta, \sin\theta)\,; \quad \kappa = (k^2 + m^2)^{1/2}\,, \tag{6.4}$$

where κ is the length of the wavevector, and θ the angle between the wavevector and the horizontal. With this, (6.3) becomes

$$\boxed{\omega^2 = N^2\cos^2\theta + f^2\sin^2\theta\,.} \tag{6.5}$$

This relation lies at the heart of internal-wave theory. It contains the two key parameters N and f, which represent the two restoring forces at work in internal waves: buoyancy and the Coriolis force. Remarkably, (6.5) shows that the wave frequency depends on the direction θ of the wavevector, but *not* on its length κ. This has as an immediate and important consequence that the group velocity vector

$$\vec{c}_g = \Big(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial m}\Big)$$

must be perpendicular to the wavevector. In other words, the propagation of internal-wave energy is perpendicular to that of lines of constant phase, see Figure 6.1. To prove this, we write, in a general way,

$$\omega(k(\kappa,\theta), m(\kappa,\theta)) = \bar{\omega}(\kappa,\theta)\,.$$

(Here we introduce a new symbol $\bar{\omega}$, to be identified with (6.5), because its functional dependence on κ and θ is formally different from that of $\omega(k, m)$.) Hence

$$\frac{\partial\bar{\omega}}{\partial\kappa} = \frac{\partial\omega}{\partial k}\frac{\partial k}{\partial\kappa} + \frac{\partial\omega}{\partial m}\frac{\partial m}{\partial\kappa} = \vec{c}_g\cdot\vec{k}/\kappa\,.$$

Now, since ω in (6.5) does not depend on κ, it follows that $\vec{c}_g\cdot\vec{k} = 0$, i.e.

$$\boxed{\vec{c}_g \perp \vec{k}\,.} \tag{6.6}$$

As a result, the group velocity makes an angle θ with the *vertical*.

Another way to arrive at this property is of course by calculating the components of the group velocity explicitly from (6.3); cast in polar coordinates, the result is

$$\vec{c}_g = \frac{(N^2 - f^2)\cos\theta\sin\theta}{\kappa\omega}\,(\sin\theta, -\cos\theta)\,. \tag{6.7}$$

This confirms (6.6). It does, however, provide some extra information. The sign of the horizontal component of the group velocity vector is determined by $(N^2 -$

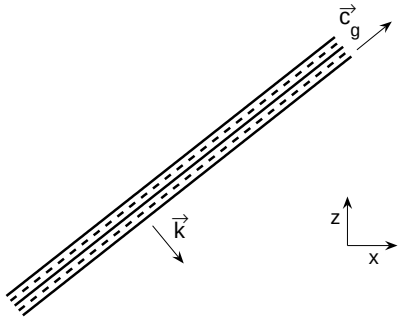

Fig. 6.1: Properties of an internal-wave beam, depicted schematically. The dashed and solid diagonals denote lines of constant phase; they propagate in the direction of the wavevector, $\vec{k}$, which in this example points right- and downward. The energy, on the other hand, propagates in the direction of the group velocity vector, $\vec{c}_g$, which is *perpendicular* to $\vec{k}$. In this example, we assume $N > |f|$; the vectors $\vec{k}$ and $\vec{c}_g$ then point in the same horizontal direction, but are vertically opposite. (If $N < |f|$, it is the other way round.)

$f^2) \cos\theta$ (the remaining factor being positive), implying that it has the same sign as k if $N > |f|$, and the opposite sign if $N < |f|$. The sign of the vertical component of the group velocity vector is determined by $-(N^2 - f^2) \sin\theta$, which is opposed to the sign of m if $N > |f|$, and the same if $N < |f|$.

Writing $\omega^2 = \omega^2(\cos^2\theta + \sin^2\theta)$ in (6.5), we obtain an alternative form of the dispersion relation,

$$\cot^2\theta = \frac{\omega^2 - f^2}{N^2 - \omega^2}\,. \tag{6.8}$$

This implies that the right-hand side must be non-negative, hence either

$$\boxed{\text{(I)} \quad N \le \omega \le |f| \quad \text{or} \quad \text{(II)} \quad |f| \le \omega \le N} \tag{6.9}$$

must hold, depending on which of the two, $|f|$ or N, is the largest. These expressions delineate the range of allowable internal-wave frequencies, and are in agreement with the earlier result (5.8). Two extreme cases can occur: $\omega \to |f|$ if $\theta \to \pi/2$, see (6.5); then the wavevector $\vec{k}$ becomes vertical, and the group velocity vector $\vec{c}_g$, horizontal. The other extreme is $\omega \to N$ (if $\theta \to 0$), then the wavevector becomes horizontal, and the group velocity vector, vertical.

Wave motions in fluids are of course associated with oscillating parcels. Their movements follow directly from the continuity equation (4.21d), by substituting $\vec{u} = \exp i(kx + ly + mz - \omega t)$, which implies

$$\boxed{\vec{u} \perp \vec{k}\,.} \tag{6.10}$$

In this section we assumed $l = 0$, and so the parcels oscillate parallel to the lines of constant phase (the dashed and solid lines in Figure 6.1), so far as the x, z-plane is concerned. Due to Coriolis effects, there is also a component v perpendicular to that plane.

6.1.2 General solution

The steepness of the group velocity $\vec{c}_g$ follows from (6.7) as $-\cos\theta/\sin\theta = -\cot\theta$. We denote this steepness by $\mu_\pm$, with two possible signs, according to (6.8),

$$\mu_\pm = \pm\Big(\frac{\omega^2 - f^2}{N^2 - \omega^2}\Big)^{1/2} . \tag{6.11}$$

Given the physical significance of $\mu_\pm$ as the steepness at which energy propagates in the x, z-plane, it makes sense to introduce new coordinates that are in line with this propagation:

$$\xi_\pm = \mu_\pm x - z\,, \tag{6.12}$$

see Figure 6.2. Energy thus propagates along lines $\xi_\pm = const.$

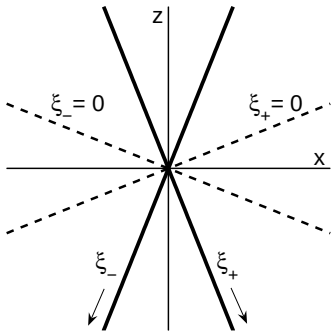

Fig. 6.2: Characteristic coordinates $\xi_+ = \mu_+ x - z$ and $\xi_- = \mu_- x - z$ (solid thick lines), here for $|\mu_\pm| = 0.4$. Internal-wave energy propagates along lines $\xi_+ = const$ or $\xi_- = const$, two examples of which are shown as dashed lines; they have steepness $\mu_\pm$. Notice that lines with $\xi_\pm = const > 0$ lie *below* the ones sketched here, implying that both ξ_+ and ξ_- increase in the downward direction, as indicated by the arrows.

Adopting ξ_+ and ξ_- as the new independent variables, instead of x and z, in fact translates into a mathematical simplification of (6.2). Writing

$$w(x,z) = \bar{w}(\xi_+(x,z), \xi_-(x,z))\,,$$

we have

$$\frac{\partial w}{\partial x} = \frac{\partial \bar{w}}{\partial \xi_+}\,\frac{\partial \xi_+}{\partial x} + \frac{\partial \bar{w}}{\partial \xi_-}\,\frac{\partial \xi_-}{\partial x} = \mu_+\,\frac{\partial \bar{w}}{\partial \xi_+} + \mu_-\,\frac{\partial \bar{w}}{\partial \xi_-}\,,$$

and similarly for $\partial w/\partial z$ and second derivatives. As a result, (6.2) reduces to the simple form

$$\frac{\partial^2 \bar{w}}{\partial \xi_+ \partial \xi_-} = 0\,. \tag{6.13}$$

The general solution of this equation can be written (dropping the bar)

$$\boxed{w = F(\xi_+) + G(\xi_-)\,,} \tag{6.14}$$

for arbitrary functions F and G.

To show an example, we take

$$F = \exp(-\xi_+^2) \exp ik\xi_+\,, \tag{6.15}$$

and the same functional dependence for $G(\xi_-)$. Together with the time factor $\exp -i\omega t$, the real part becomes

$$w = \exp(-\xi_+^2)\cos(k\xi_+ - \omega t) + \exp(-\xi_-^2)\cos(k\xi_- - \omega t)\,, \tag{6.16}$$

and is illustrated in Figure 6.3 for $t = 0$. (Notice that positive k here means *down*ward phase propagation, corresponding to the direction of increasing $\xi_\pm$, see Figure 6.2.) The Gaussian factor gives the beams their confined form, while the cosine describes the phase propagation within the beams, in a direction perpendicular to the beams themselves (as sketched in Figure 6.1).

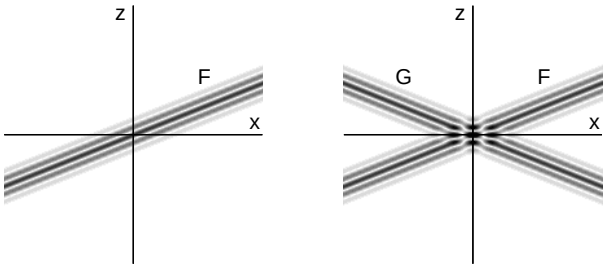

Fig. 6.3: On the right, the magnitude of the solution (6.16), $|w|$, shown at $t = 0$ for $k = 5$ and $\mu_+ = -\mu_- = 0.4$ (cf. Figure 6.2). On the left, only F is shown.

6.1.3 Kinetic and potential energy

From (4.27) we have the expressions for kinetic and potential energy,

$$E_k = \frac{1}{2}\,\rho_*\left[u^2 + v^2 + w^2\right]; \quad E_p = \frac{1}{2}\,\rho_*\, b^2/N^2\,.$$

From the general solution for w, (6.14), we can derive the expressions for u, v and b, using (4.21b,d,e), with $\partial/\partial y = 0$ and $\tilde{f} = 0$; this gives

$$u = \mu_+^{-1} F(\xi_+) + \mu_-^{-1} G(\xi_-) \tag{6.17}$$

$$v = -i \frac{f}{\omega} \left[\mu_+^{-1} F(\xi_+) + \mu_-^{-1} G(\xi_-) \right] \tag{6.18}$$

$$b = -i \frac{N^2}{\omega} \left[F(\xi_+) + G(\xi_-) \right] . \tag{6.19}$$

Selecting now either F or G, *but not both*, we find for the ratio of potential and kinetic energy,[1]

$$\frac{E_p}{E_k} = \frac{N^2(\omega^2 - f^2)}{\omega^2(N^2 - f^2) + f^2(N^2 - \omega^2)} . \tag{6.20}$$

This expression is remarkably simple; in particular, it is spatially uniform. However, this would no longer be the case if one includes both F and G. The easiest way to see this is by considering a basin of finite depth (such as the ocean!), in which case it does not suffice to include just F, because wave reflections from the boundaries (bottom or surface) will result in propagation along the other characteristic, and hence require the inclusion of G. In particular, near a horizontal bottom the potential energy must vanish (hence $E_p/E_k \to 0$) because no vertical excursions can occur there; in that case, (6.20) is clearly not applicable. Still, (6.20) would remain valid in a *depth-averaged* sense, as can be checked using the modal expressions (5.5) with constant N.

6.2 Reflection from a sloping bottom

In the previous section we considered plane waves extending into infinity, spatially. Physical boundaries like a (sloping) bottom of course render these solutions invalid. However, they can be easily modified to accomodate for the presence of such boundaries. We consider here the effect of a uniformly sloping bottom whose normal lies in the same vertical plane as the wavevector $\vec{k} = (k, 0, m)$; a more general situation is examined in Section 6.4.

We have seen from the dispersion relation (6.5) that the wave frequency is determined by the angle of propagation (θ). Vice versa, for a given frequency ω, the angle θ is fixed (modulo π). This has an important implication for internal-wave reflection from boundaries, such as a sloping bottom: after reflection, energy must again propagate at an angle θ *with the vertical*, since the wave frequency has not changed. This property is quite unlike that of for example reflection of light from a mirror, where the angle of incidence also equals that of reflection but with both angles being defined *with respect to the normal of the reflecting surface.*

[1] In view of the complex nature of the expressions, we calculate E_p as $E_p = \frac{1}{2}\rho_* bb^*/N^2$ etc.

The precise way in which internal waves reflect can be readily established from the general solution (6.14). Without loss of generality, we take F to represent the incident wave. We consider reflection from a uniform slope $z = \gamma x$, with $\gamma < 0$. Applying the boundary condition at the slope enables us to express the reflected wave G in terms of F. Let u be the velocity component in the direction x; using the continuity equation $u_x + w_z = 0$, we obtain

$$u = \mu_+^{-1} F(\xi_+) + \mu_-^{-1} G(\xi_-)\,.$$

The boundary condition to be posed at the slope is that of zero normal velocity, i.e. $w = \gamma u$ at $z = \gamma x$. Hence

$$(1 - \gamma/\mu_+)F([\mu_+ - \gamma]x) + (1 - \gamma/\mu_-)G([\mu_- - \gamma]x) = 0 \quad \text{for all } x\,.$$

Since this must hold for all x, we can express G in terms of F as

$$G(\xi_-) = \lambda F(\lambda \xi_-) \quad \text{with} \quad \lambda = \frac{\mu_+ - \gamma}{\mu_- - \gamma}\,. \tag{6.21}$$

(Here we used $\mu_-/\mu_+ = -1$, implied by (6.11).) This expression has no meaning if $\lambda \to \infty$, or

$$\mu_- = \gamma\,,$$

i.e. the direction of energy propagation of the reflected wave coincides with the bottom slope; this is called *critical reflection*. The linear solution then breaks down because G becomes infinite; to treat this case properly, nonlinear and viscous effects need to be included. We restrict ourselves to examining the case of non-critical reflection, $\mu_- \neq \gamma$.

Substituting (6.21) into (6.14) provides the general solution of the linear internal-wave reflection problem:

$$w = F(\xi_+) + \lambda F(\lambda \xi_-)\,. \tag{6.22}$$

Figure 6.4 shows two examples, with the incident beam F as defined in (6.15). The upper panel shows a case of *sub-critical reflection*, i.e. the slope is less steep than the reflected beam (G); as a result, the reflected beam continues in the same horizontal direction as the incident beam. In the lower panel, reflection is *super-critical*, the slope being steeper than the reflected beam, which now bounces back horizontally. In both cases, the reflected beam is less wide, and more intense than the incident beam; that this must be so is readily understood from geometric arguments (trace the lines within which the incident beam is confined, and follow their reflection from the sloping wall). Conversely, reflection from a slope with positive γ would result in defocussing. This is also evident from the way λ modifies the argument in the reflected $F(\lambda\xi_-)$ in (6.22); writing $\mu \equiv \mu_+ = -\mu_-$, we have $|\lambda| = |(\mu - \gamma)/(\mu + \gamma)|$, which is smaller than 1 for positive γ, and larger than 1 for negative γ.

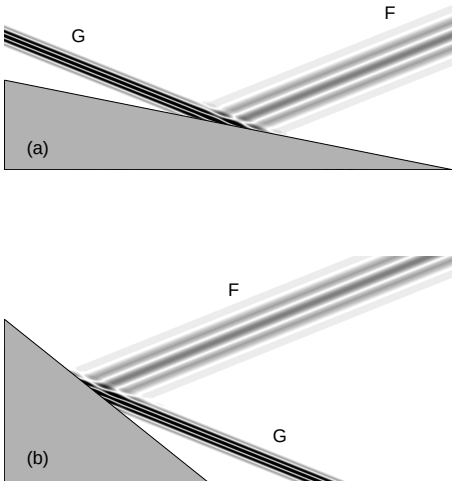

Fig. 6.4: Reflection from a uniform slope, showing $|w|$, for $\mu_+ = 0.4$. The incident beam $F(\xi_+)$, entering from the upper right corner, is prescribed by (6.15), with $k = -5$ (i.e. upward phase propagation); the total solution then follows from the general expression (6.22). In **a**, the slope is $\gamma = -0.2$, so $|\gamma| < |\mu_-|$, in which case we speak of *sub*-critical reflection; here $\lambda < 0$, so the reflected beam has downward phase propagation. In **b**, we have $\gamma = -0.8$, so $|\gamma| > |\mu_-|$, giving *super*-critical reflection; here $\lambda > 0$, so the reflected beam has upward phase propagation, like the incident beam. This is consistent with the assumption that $F(\xi_+)$ is the incident beam, provided that $N > |f|$; for $N < |f|$, consistency requires a reversal of roles.

6.3 Propagation between two horizontal boundaries

A special case of reflection occurs for a horizontal boundary, $\gamma = 0$. Then $\lambda = \mu_+/\mu_- = -1$, and the solution (6.22) becomes

$$w = F(\xi_+) - F(-\xi_-)\,. \tag{6.23}$$

We interpret this as a reflection from the surface (rigid-lid), at $z = 0$. If we moreover include a horizontal bottom, placed at $z = -H$ where $w = 0$, then (6.23) implies

$$F(\mu_+ x + H) = F(-\mu_- x - H) \quad \text{for all } x\,.$$

With $-\mu_- = \mu_+ \equiv \mu$, this condition means that F must be periodic in x, with period $2H/\mu$. This is the distance between subsequent reflections at the surface (or bottom).

Since F is periodic, we can write it as a (complex) Fourier series of period

$2H$:

$$F(\xi_+) = \sum_{n=0}^{\infty} a_n \exp(in\pi\xi_+/H)\,,$$

with complex coefficients a_n. Using this series in (6.23) gives

$$\begin{aligned} w &= \sum_{n=0}^{\infty} a_n \Big[\exp(in\pi\xi_+/H) - \exp(-in\pi\xi_-/H) \Big] \\ &= \sum_{n=0}^{\infty} a_n \Big[\exp(in\pi(\mu x - z)/H) - \exp(in\pi(\mu x + z)/H) \Big] \\ &= -2i \sum_{n=0}^{\infty} a_n \sin(n\pi z/H) \exp(in\pi x\mu/H)\,. \end{aligned}$$

Adding the time-factor finally results in

$$w = -2i \sum_{n=0}^{\infty} a_n \sin\left(\frac{n\pi z}{H}\right) \exp(i[k_n x - \omega t])\,. \tag{6.24}$$

where we introduced horizontal wavenumbers $k_n = n\pi\mu/H$. In this expression, we recognize the solution for rightward propagating waves in terms of *vertical modes*, (5.22). This proves the equivalence between the two methods of Chapters 5 and 6; the present configuration, viz. horizontal boundaries and constant N, allows both methods to be applied. At the same time, each has its own advantage. The usage of the coordinates x and z, in the method of vertical modes, is natural in view of the geometry of the problem, but makes it less easy to grasp how the waves actually propagate. In this respect, the solution is easier to interpret when expressed in terms of characteristic coordinates, since they reflect the direction of energy propagation.

6.4 Three-dimensional reflection

So far we have considered plane waves propagating in the x, z-plane, and their reflection from a sloping boundary whose normal lies in the same vertical plane. We now examine the more general case of a given incident wave

$$w_i = \exp i(kx + ly + mz - \omega t)$$

reflecting from an arbitrarily oriented uniform slope

$$z = \gamma x + \delta y\,.$$

The incident wave w satisfies (6.1), as does the reflected wave

$$w_r = Q \exp i(Kx + Ly + Mz - \omega t)\,,$$

where Q, K, L and M are to be found. The total vertical velocity is $w = w_i + w_r$. In view of the boundary condition, we need the other velocity components as

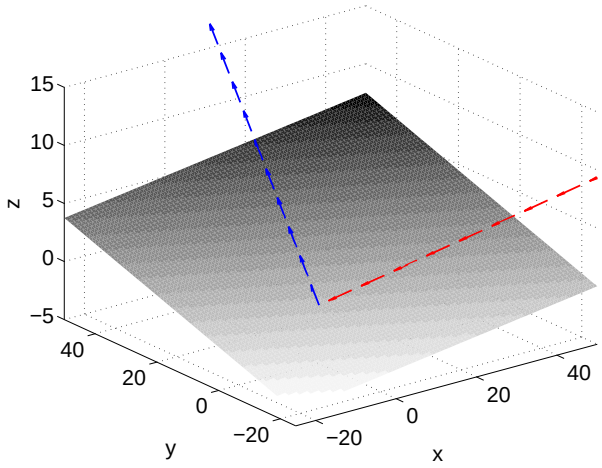

Fig. 6.5: A schematic view of the three-dimensional reflection problem: an incident wave (in red), with wavevector $\vec{k} = (k, l, m)$ reflects from a uniform slope whose normal does not lie in the same vertical plane as the wavevector $\vec{k}$. The problem is to find the reflected wave (in blue), i.e. its wavevector $\vec{K}$ as well as its amplitude Q.

well. They can be obtained from (4.21d) and (4.24), with $\tilde{f} = 0$ (Traditional Approximation):

$$u_i = \frac{m(fl - i\omega k)}{i\omega(k^2 + l^2)}\, w_i\,; \quad v_i = -\frac{m(fk + i\omega l)}{i\omega(k^2 + l^2)}\, w_i\,, \tag{6.25}$$

and analogous expressions for the reflected wave (replace index i by r, and k, l, m by K, L, M). These expressions, incidentally, determine also the polarization of the horizontal velocity field.

The boundary condition to be applied is that of zero normal velocity at the slope, i.e.

$$w = \gamma u + \delta v \quad \text{at} \quad z = \gamma x + \delta y\,. \tag{6.26}$$

This condition takes the form

$$(\cdots)\exp i\Big([k + \gamma m]x + [l + \delta m]y\Big) + (\cdots)\exp i\Big([K + \gamma M]x + [L + \delta M]y\Big) = 0\,,$$

for all x and y. For this to hold, we must have

$$k + \gamma m = K + \gamma M\,; \quad l + \delta m = L + \delta M\,.$$

Writing $M = qm$ (q to be found) gives

$$K = k + \gamma(1 - q)m\,; \quad L = l + \delta(1 - q)m\,. \tag{6.27}$$

The problem has thus been reduced to finding just two coefficients: q and Q. Both the incident and reflected wave have to satisfy the dispersion relation that follows from (6.1),

$$(N^2 - \omega^2)(k^2 + l^2) - (\omega^2 - f^2)m^2 = 0\,, \tag{6.28}$$

or, in short-hand notation,

$$Cm^2 + D(k^2 + l^2) = 0\,, \tag{6.29}$$

with $C = f^2 - \omega^2$ and $D = N^2 - \omega^2$. The wavevector (K, L, M) too satisfies (6.29); hence, using (6.27),

$$C[qm]^2 + D([k + \gamma(1-q)m]^2 + [l + \delta(1-q)m]^2) = 0\,.$$

Here the terms $D(k^2+l^2)$ can be replaced using (6.29); this produces a common factor $(1-q)$. Since $q = 1$ represents the trivial but unphysical solution of an incident wave passing through the slope, which we ignore, we can divide by this factor to obtain

$$q = \frac{D((k+\gamma m)^2 + (l+\delta m)^2)}{m^2(C + D(\gamma^2 + \delta^2))}\,. \tag{6.30}$$

Applying the boundary condition (6.26), finally, yields the complex amplitude factor of the reflected wave; after some rewriting

$$Q = -q\,\frac{\omega[Cm - D(\gamma k + \delta l)] + iDf(\delta k - \gamma l)}{\omega[CM - D(\gamma K + \delta L)] + iDf(\delta K - \gamma L)}\,. \tag{6.31}$$

With this, the reflected wave has been fully determined.

Critical reflection occurs when the factor q, and hence Q, becomes infinite, i.e.

$$\gamma^2 + \delta^2 = -\frac{C}{D} = \frac{\omega^2 - f^2}{N^2 - \omega^2}\,, \tag{6.32}$$

stating that the bottom slope has the same angle with the horizontal plane as has energy propagation. This can be seen as follows. The direction of energy propagation is found by differentiating (6.28) to k, l and m, respectively,

$$\vec{c}_g = \Big(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial l}, \frac{\partial\omega}{\partial m}\Big) = \frac{1}{\omega\kappa^2}\left((N^2-\omega^2)k,\, (N^2-\omega^2)l,\, (f^2-\omega^2)m\right), \tag{6.33}$$

with κ the length of the wavevector, $\kappa^2 = k^2+l^2+m^2$. As in the two-dimensional case, we have $\vec{c}_g \perp \vec{k}$; this can be seen by taking the inner product of (6.33) with $\vec{k}$ and using (6.28). Now, the angle θ of the wavevector $\vec{k}$ with the horizontal plane follows from (6.29): $\tan^2\theta = -D/C$. Since this angle depends only on the wave frequency, it must be the same for the reflected wavevector (K, L, M). Hence the energy propagation of both the incident and reflected waves has an angle with the horizontal plane whose tangent is given by the square root of the right-hand side of (6.32).

Finally, we note that taking $l = 0$ and using (6.5) leads us from (6.33) to the earlier expression (6.7).

6.5 Non-uniform stratification

So far we have assumed N to be constant. We now return to (6.2) to see what happens when N is allowed to vary vertically:

$$(N^2(z) - \omega^2)\frac{\partial^2 w}{\partial x^2} - (\omega^2 - f^2)\frac{\partial^2 w}{\partial z^2} = 0\,. \tag{6.34}$$

6.5.1 WKB approximation

We will first assume that N varies weakly with z (the meaning of 'weak' is specified below), and proceed in a heuristic way. Trying a solution of the form

$$w = \exp i[kx + \varphi(z)]\,, \tag{6.35}$$

we find from substitution in (6.34):

$$-k^2(N^2(z) - \omega^2) - (\omega^2 - f^2)[-(\varphi')^2 + i\varphi''] = 0\,,$$

(primes denote derivatives to z). Or, in short-hand notation,

$$-(\varphi')^2 + i\varphi'' + q^2(z) = 0\,, \quad \text{with} \quad q^2 = k^2\frac{N^2 - \omega^2}{\omega^2 - f^2}\,. \tag{6.36}$$

If N were constant, we would have $\varphi(z) = mz$, and hence $\varphi'' = 0$. So, for weakly varying N, it makes sense to assume that this second derivative is a relatively small term in (6.36). We thus find by successive approximation:

$$\begin{aligned}
(\varphi_1')^2 &= q^2 \\
(\varphi_2')^2 &= q^2 + i\varphi_1'' \\
(\varphi_n')^2 &= q^2 + i\varphi_{n-1}'' \qquad (n \geq 3)\,.
\end{aligned}$$

Here φ_1, φ_2 etc are increasingly accurate representations of φ. The lowest-order approximation satisfies $\varphi_1' = \pm q$, hence

$$\varphi_1 = \pm \int dz\; q\,.$$

(More precisely, the integrals are to be read as $\int_0^z d\bar{z}\, q(\bar{z})$.) The second-order approximation φ_2 now satisfies

$$\begin{aligned}
\varphi_2' &= \pm\Big(q^2 \pm iq'\Big)^{1/2} \\
&= \pm q\Big(1 \pm i\tfrac{q'}{q^2}\Big)^{1/2} \\
&\approx \pm q\Big(1 \pm \tfrac{iq'}{2q^2}\Big) \\
&= \pm q + \tfrac{iq'}{2q}\,.
\end{aligned} \tag{6.37}$$

(In the first three expressions, the two $\pm$'s must have the *same* sign, so either both $+$, or both $-$, because of the relation between the φ's: $\varphi_2 = \varphi_1 + \cdots$.) Hence

$$\varphi_2 = \frac{i}{2}\ln q \pm \int dz\, q\,.$$

We thus find for the z-dependent part of w in (6.35),

$$w \sim C_-\,\frac{1}{\sqrt{q}}\,e^{-i\int dz\, q} + C_+\,\frac{1}{\sqrt{q}}\,e^{+i\int dz\, q}\,, \tag{6.38}$$

with arbitrary constants $C_\pm$. This describes vertically propagating waves. We may interpret q as a kind of vertical wavenumber, since for constant N we would have found $\varphi = mz$. The approximation (6.37) hinges on the assumption $q' \ll q^2$. Now, q' is related to the vertical variation of N, so the inequality can be interpreted as requiring that N varies slowly compared to the vertical scale of the wave, q^{-1}. In other words, waves have to be sufficiently short for the approximation to be valid.

The assumption breaks down, of course, near vertical positions where $N(z)$ approaches ω, in which case $q \to 0$. These are the so-called turning points, at which the local behaviour can be described by Airy functions (see also Section 5.5, and, in particular, Figure 5.16b). The solution then consists of various parts, close to and away from the turning point; they can matched by techniques commonly used in quantum mechanics (see, e.g., [56]). We do not pursue this topic here.

If we assume $q^2 > 0$, then no turning points occur. Still, as pointed out above already, the approximation (6.38) loses its validity if N varies too strongly. Physically, this can be interpreted as follows. In (6.38), the terms describe up- and downward propagating waves, which are entirely *independent* of one another; this is expressed by the fact that the constants C_+ and C_- can be prescribed independently. Now, if N varies strongly, this will no longer be the case because *internal reflections* occur, as sketched in Figure 6.6.

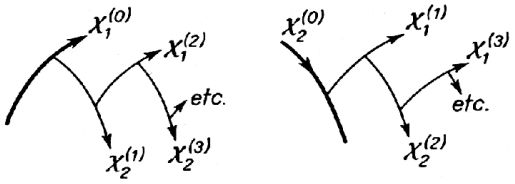

Fig. 6.6: A schematic view of how an upward propagating beam (on the left), in a vertically varying medium, undergoes not only refraction but also internal reflections; these reflected beams, in turn, also undergo reflections, giving rise to an increasingly complicated pattern of beams. On the right, the same for a downward main beam. From [6].

Having outlined the restrictions of the approximation, we now turn to the case in which all the assumptions are satisfied. The vertical wavenumber $m(z)$ can then be defined by

$$m(z) = \frac{d\varphi}{dz} \approx q(z) = \pm k \Big(\frac{N^2(z) - \omega^2}{\omega^2 - f^2} \Big)^{1/2} .$$

So, whereas the horizontal wavenumber remains constant, the vertical wavenumber is affected by the vertically varying medium.[2]

[2]This is a special case of the more general 'ray theory', see [48, §6].

6.5.2 Characteristic coordinates

We start with a short mathematical excursion. We can still define characteristic coordinates $\xi_\pm$; they are constant along the curves

$$\frac{dz}{dx} = \mu_\pm(z) \equiv \pm\Big(\frac{\omega^2 - f^2}{N^2(z) - \omega^2}\Big)^{1/2}. \tag{6.39}$$

This equation cannot be solved analytically except for very special choices of $N(z)$; thus, one cannot, in general, obtain the characteristic coordinates in explicit form. Whether explicitly obtainable or not, using these coordinates as the independent variables does *not* reduce (6.34) to the previous simple form (6.13); its right-hand side is no longer zero. Hence (6.14) no longer provides a solution.[3] This is, again, due to the occurrence of internal reflections, which connects the flow of energy on one characteristic $\xi_+ = const$ with that on the other $\xi_- = const$; thus the two can no longer be independently prescribed, as was the case in (6.14).

However, if the stratification varies sufficiently weakly, we may ignore internal reflections, and find the path of energy propagation by numerically integrating (6.39). This is a sensible approach in the deep ocean, but is bound to be invalid in the seasonal thermocline (see Figure 1.6 for a typical distribution of N). For strongly varying N, the method of vertical modes (Chapter 5) is the more suitable one, at least in the absence of strongly varying depth.

6.6 Non-traditional effects

To conclude this chapter, we briefly discuss how the results obtained so far would change if one abandons the Traditional Approximation. We return to (5.56),

$$A\frac{\partial^2 w}{\partial x'^2} + 2B\frac{\partial^2 w}{\partial x' \partial z} + C\frac{\partial^2 w}{\partial z^2} = 0\,, \tag{6.40}$$

with $A = N^2 - \omega^2 + f_s^2$, $B = f f_s$, $C = f^2 - \omega^2$, and $f_s = \tilde{f}\sin\alpha$. We recall that α is the angle of the axis x' with the west-east direction (Figure 5.17). We assume $N = const$.

Requiring hyperbolicity of (6.40), $B^2 - AC > 0$, translates into the bounds of the frequency domain as stated in (5.59).

We can again introduce characteristics coordinates as in (6.12) and Figure 6.2 (but now respect to the x', z-system rather than x, z),

$$\xi_\pm = \mu_\pm x' - z\,. \tag{6.41}$$

However, $\mu_\pm$ are now defined differently:

$$\mu_\pm = \frac{-B \pm (B^2 - AC)^{1/2}}{A}\,. \tag{6.42}$$

[3] For a special class of profiles with $(\omega^2 - f^2)/(N^2(z) - \omega^2) = (c_1 z + c_2)^2$, with arbitrary $c_{1,2}$, a simple solution can still be obtained, see [47, §151].

An important departure from traditional theory is that they are not equally large: $|\mu_+| \neq |\mu_-|$. This means that internal-wave energy propagates at two different slopes, depending on which of the two, $\xi_+ = const$ or $\xi_- = const$, applies. This is sketched in Figure 6.7.

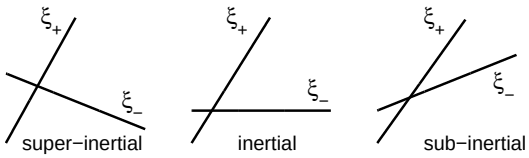

Fig. 6.7: A schematic view of the direction of the lines of constant ξ_+ and ξ_-, assuming $|f| < (N^2 + f_s^2)^{1/2}$ (a condition usually satisfied in the ocean). Three important cases are distinguished, from left to right: $(N^2 + f_s^2)^{1/2} > \omega > |f|$, $\omega = |f|$, and $\omega < |f|$. Notice that the first and third cases are in qualitative agreement with the direction of the beams shown in Figure 5.19a and 5.19b, respectively.

Using the characteristic coordinates instead of x' and z, (6.40) transforms into $w_{\xi_+\xi_-} = 0$; hence the general solution has the same form as the traditional expression (6.14):

$$w = F(\xi_+) + G(\xi_-)\,. \tag{6.43}$$

The problem of reflection from a uniform slope can thus be easily solved, as in Section 6.3. However, the expressions derived there are not all carried over to the non-traditional case, because we no longer have $\mu_-/\mu_+ = -1$.

Assuming solutions of the form $w = \exp i(kx + mz)$, we find that the traditional dispersion relation (6.5) is now to be replaced by

$$\omega^2 = N^2 \cos^2\theta + (f_s \cos\theta + f \sin\theta)^2\,. \tag{6.44}$$

But, again, ω depends only on the angle θ of the wavector, not on its length κ, so $\vec{c}_g \perp \vec{k}$ still holds. From this, the group velocity can be easily derived; the qualitative behaviour of the components is shown in Figure 6.8.

Segments where the wave frequency increases (decreases) as a function of θ correspond to energy propagation along a μ_+ (μ_-) characteristic. The wavevector and group velocity vector are either vertically or horizontally opposed, but not both (Figure 6.8b,c). In Figure 6.8a, dotted lines are drawn at two frequency levels: $\omega = (N^2 + f_s^2)^{1/2}$ and $\omega = |f|$. These are precisely the values at which the coefficients A and C in (6.40) change sign, respectively. Their special significance is furthermore seen from the simple rule that one can distill from Figure 6.8: for the μ_+ characteristic, the wavevector and group velocity are vertically opposed if $\omega_{\min} < \omega < (N^2 + f_s^2)^{1/2}$, and horizontally opposed if $(N^2 + f_s^2)^{1/2} < \omega < \omega_{\max}$; for the μ_- characteristic, they are vertically opposed if $|f| < \omega < \omega_{\max}$, and horizontally opposed if $\omega_{\min} < \omega < |f|$. This rule is modified for weak stratification, when $(N^2 + f_s^2)^{1/2} < |f|$. All the possible situations are summarized in Table 6.1.

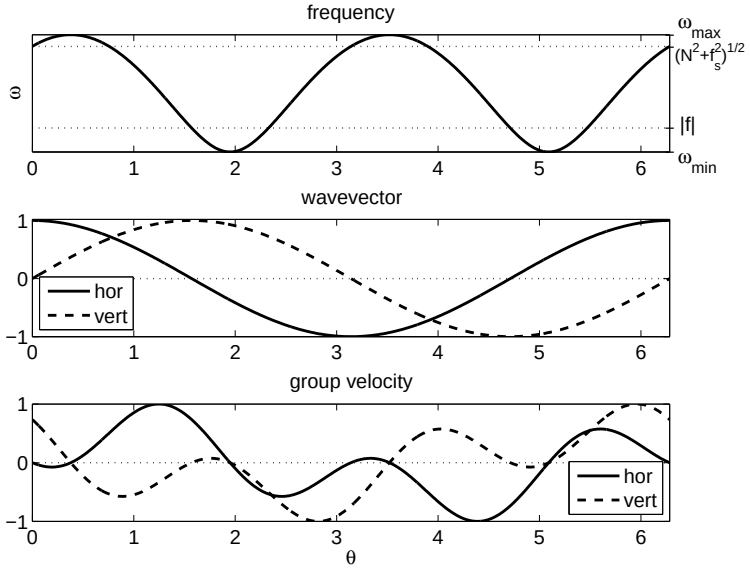

Fig. 6.8: The dispersion relation (6.44) for internal inertio-gravity, and related quantities, for $|f| < (N^2+f_s^2)^{1/2}$ (regime I of Table 6.1). In **a**, wave frequency ω versus θ, the angle of the wavevector with the horizontal. Segments where ω increases (decreases) correspond to the characteristic μ_+ (μ_-). In **b**, the horizontal and vertical components of the wavevector are shown (normalized), also as a function of θ. Panel **c** shows the components of the group velocity factor (multiplied by ω and normalized). Parameter values are $\phi = 45°$N and $\alpha = \pi/2$ (i.e. the x-axis is directed south-north, positive northward); for optimal clarity, N has been chosen only marginally larger than 2Ω ($N = 1.5 \times 10^{-4}$rad s^{-1}).

Parameter regime:	*Frequency intervals:*	*Characteristics:*
regime I:	$\omega_{\min} < \omega < \lvert f\rvert$	$\mu_+ > 0$ (V) & $\mu_- > 0$ (H)
$\lvert f\rvert < (N^2 + f_s^2)^{1/2}$	$\lvert f\rvert < \omega < (N^2 + f_s^2)^{1/2}$	$\mu_+ > 0$ (V) & $\mu_- < 0$ (V)
	$(N^2 + f_s^2)^{1/2} < \omega < \omega_{\max}$	$\mu_+ < 0$ (H) & $\mu_- < 0$ (V)
regime II:	$\omega_{\min} < \omega < (N^2 + f_s^2)^{1/2}$	$\mu_+ > 0$ (V) & $\mu_- > 0$ (H)
$(N^2 + f_s^2)^{1/2} < \lvert f\rvert$	$(N^2 + f_s^2)^{1/2} < \omega < \lvert f\rvert$	$\mu_+ < 0$ (H) & $\mu_- > 0$ (H)
	$\lvert f\rvert < \omega < \omega_{\max}$	$\mu_+ < 0$ (H) & $\mu_- < 0$ (V)

Table 6.1: The two possible regimes, depending on which of the two, $|f|$ or $(N^2 + f_s^2)^{1/2}$, is the largest. Each regime gives rise to three frequency intervals, for which the signs of μ_+ and μ_- are listed, as well as the corresponding behaviour of the wavevector and group velocity vector: 'H' indicates that they are horizontally opposed; 'V', that they are vertically opposed.

This variety of behaviour is to be contrasted with the traditional limit ($f_s = 0$), where one finds that they are always vertically opposed in the 'strongly'

stratified regime (i.e. $N > |f|$), and always horizontally opposed in the 'weakly' stratified regime ($N < |f|$).

Finally, we emphasize that properties such as the steepness of characteristics, the dispersion relation, and the lower and upper bounds of the frequency window, all depend on the angle α in $f_s = \tilde{f} \sin \alpha$, i.e. on the orientation in the horizontal geographical plane. There is a *horizontal anistropy*, unlike under the Traditional Approximation, where internal waves behave the same way for all horizontal directions of wave propagation. Non-traditional effects are weakest for zonal propagation ($\alpha = 0, \pi$) and strongest for meridional propagation ($\alpha = \pm\pi/2$).

Chapter 7

Internal tides

At the origin of internal waves in the ocean lie two major sources: the wind, which generates near-inertial internal waves, and the barotropic tide, which generates internal tides. In this chapter we discuss the latter. We start with a simple model of the barotropic tide, propagating along a continental slope, which serves as a basis for developing the theory of internal-tide generation.

7.1 Barotropic tides

There are many tidal components, but the two main classes are those of diurnal and semi-diurnal periods, and of the latter, the lunar component M_2 is the most important one. Although surface tides have been measured for centuries in coastal regions, it has only recently become clear exactly how they propagate in the deep ocean. The advancements in satellite altimetry now make it possible to measure ocean surface displacements as small as a centimeter. An example of how the surface tides propagate is shown in Figure 7.1.

As explained qualitatively in Section 1.3, the surface tides are responsible for the generation of internal tides. To put this idea into a more quantitative but tractable form, we need first of all a simple model for the propagation of barotropic tides near topographic features. We shall consider a barotropic tide propagating along a continental slope, described by $z = -h(x)$. The equations to be satisfied are the linear shallow-water equations,

$$U_t - fV = -g\eta_x\,; \quad V_t + fU = -g\eta_y\,; \quad \eta_t + (hU)_x + hV_y = 0\,. \tag{7.1}$$

Here U is the cross-slope component, V the along-slope one, and η the surface displacement, all associated with the barotropic tide.

Suppose first that this slope were a vertical wall (at $x = 0$), covering the entire water column. Clearly, we must require $U = 0$ at the wall. If we furthermore assume that there is no cross-slope component away from the slope either, then we arrive at the well-known Kelvin-wave solution:

$$U = 0\,; \quad V = \frac{ag}{c_0} e^{fx/c_0} \exp i(ly - \omega t)\,; \quad \eta = a e^{fx/c_0} \exp i(ly - \omega t)\,,$$

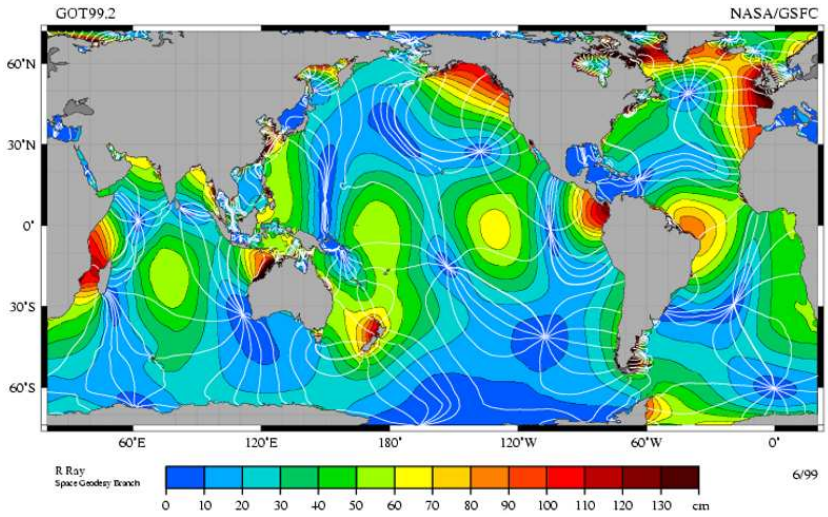

Fig. 7.1: The pattern of the semi-diurnal lunar tidal component M_2, established using Topex/Poseidon altimetry data. Colours indicate the amplitude of the surface elevation (in cm); white lines represent lines of constant phase. They are organized around so-called amphidromic points: points where the amplitude vanishes and at which phase lines come together. Phases move counterclockwise around amphidromic points in the Northern Hemisphere, and clockwise in the Southern Hemisphere. Figure from http://svs.gsfc.nasa.gov/stories/topex/tides.html.

with $c_0 = (gH)^{1/2}$ (with H the water depth in the deep ocean, typically 4 km), $l = \omega/c_0$, ω the tidal frequency, and a an arbitrary amplitude factor. The wave propagates as a long wave in the along-slope direction, with the wall on its right-hand side (Northern Hemisphere). Both the surface excursion η and the along-slope velocity V decay exponentially off the wall, obeying the geostrophic balance. These features are illustrated in Figure 7.2.

A real continental slope, even if conceived as purely vertical, differs crucially from the previous model in that it does not cover the whole water column, but leaves a narrow opening in the upper part of it (typically of a height of the order of 200 m), connecting the deep ocean to a *continental shelf*. This opening, small as it may seem in comparison with H, profoundly changes the solution.

To see this, consider the configuration shown in Figure 7.3, where the shelf has width L. The equations to be satisfied are again (7.1); here $h(x)$ equals H in the deep ocean ($x < 0$) and H_s on the shelf ($x > 0$). We anticipate that, not unlike the Kelvin wave, the tide propagates along the continental slope, i.e. each of the fields is proportional to $\exp i(ly - \omega t)$. This implies that the x dependent parts (denoted by hats) must satisfy

$$\hat{U} = \frac{ig}{\omega^2 - f^2}(fl\hat{\eta} - \omega\hat{\eta}')\,; \quad \hat{V} = \frac{g}{\omega^2 - f^2}(\omega l\hat{\eta} - f\hat{\eta}')\,,$$

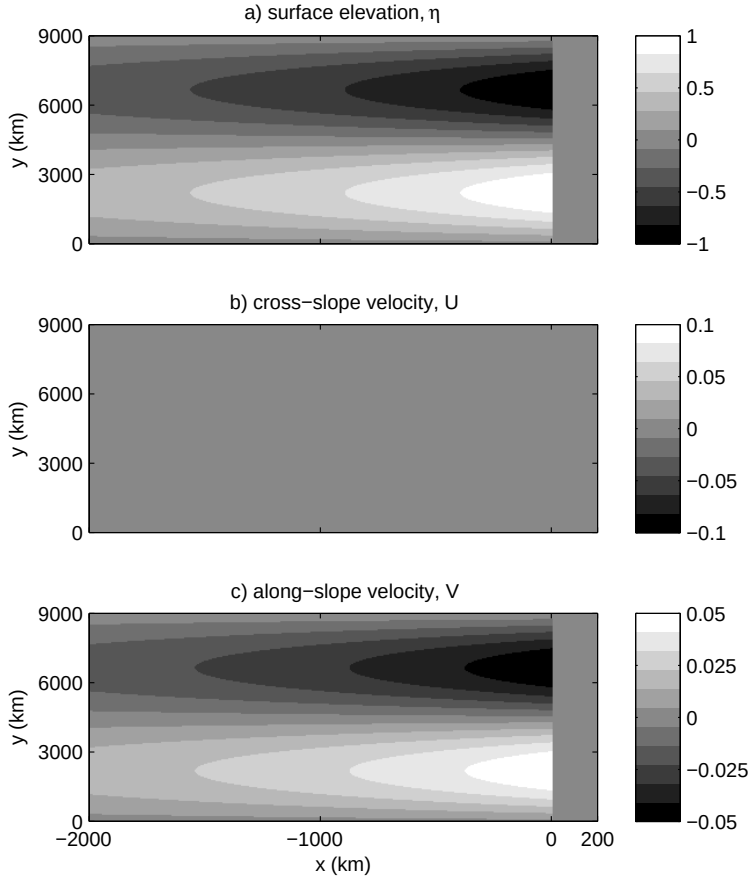

Fig. 7.2: The classical Kelvin wave (top view), travelling over uniform depth along a vertical wall at $x = 0$. Propagation is in the positive y-direction. The fields η and V decay exponentially off the wall. Note the total absence of a cross-slope velocity component. Units are: η in m, U and V in $\mathrm{m\,s^{-1}}$. Parameters are $\omega = 1.4 \times 10^{-4}$ $\mathrm{rad\,s^{-1}}$, $f = 1 \times 10^{-4}$ $\mathrm{rad\,s^{-1}}$, $H = 4$ km. The along-slope wavenumber is $l = 7.071 \times 10^{-7}$ $\mathrm{rad\,m^{-1}}$.

and

$$\hat{\eta}'' - \Big[l^2 - \frac{\omega^2 - f^2}{gh}\Big]\hat{\eta} = 0\,,$$

where primes denote derivatives to x. We assume the factor in brackets to be positive in the deep ocean (and denote it by m^2), and negative on the shelf ($-m_s^2$).[1] We thus impose an exponential solution in the deep ocean, and a

[1]This assumption presupposes that we are considering super-inertial tides, i.e. $\omega > |f|$, which applies to diurnal tides at latitudes lower than approximately 30°N/S, and to semi-diurnal tides everywhere except in polar regions. For sub-inertial tides, on the other hand, a different type of solution applies, the so-called double Kelvin-wave, see [48].

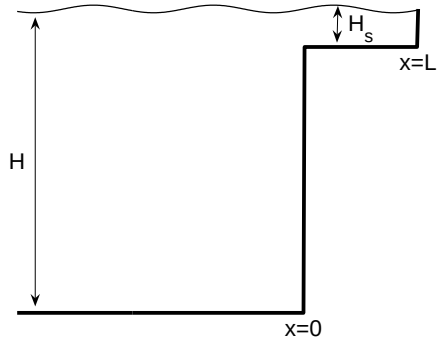

Fig. 7.3: A schematic representation of the transition between the deep ocean and the continental shelf.

sinusoidal one on the shelf:

$$\hat{\eta} = \begin{cases} a \exp mx & x < 0 \\ a(\cos m_s x + \gamma \sin m_s x) & x > 0\,. \end{cases}$$

Here a is an arbitrary coefficient, fixing the amplitude of the signal; γ is determined below. The choice on m and m_s implies a restriction on l; it is now bounded by

$$\frac{\omega^2 - f^2}{gH} < l^2 < \frac{\omega^2 - f^2}{gH_s}\,.$$

The two conditions to be imposed at $x = 0$ are continuity of η (already satisfied) and continuity of cross-slope transport: $HU|_{0^-} = H_s U_{0^+}$; the latter implies

$$H(fl - \omega m) = H_s(fl - \omega\gamma m_s)\,,$$

so that

$$\gamma = \frac{H\omega m - (H - H_s)fl}{H_s \omega m_s}\,.$$

Finally, we impose the condition $U|_{x=L} = 0$ (zero normal velocity at the end of the shelf), implying

$$(fl - \omega m_s \gamma)\cos m_s L + (fl\gamma + \omega m_s)\sin m_s L = 0\,. \tag{7.2}$$

This is the dispersion relation; it yields wavenumber l for given tidal frequency ω (and fixed constants f, H, H_s and L). The equation is transcendental and cannot be solved by analytical means, but it can easily be solved numerically (e.g. using the Newton algorithm, or simply by inspection, plotting the left-hand side as a function of l).

An example is shown in Figure 7.4. The presence of the shelf has two principal consequences: the wave is shortened (i.e. l is larger) and hence its phase speed is decreased; and, more importantly, there is now a *cross*-slope velocity.

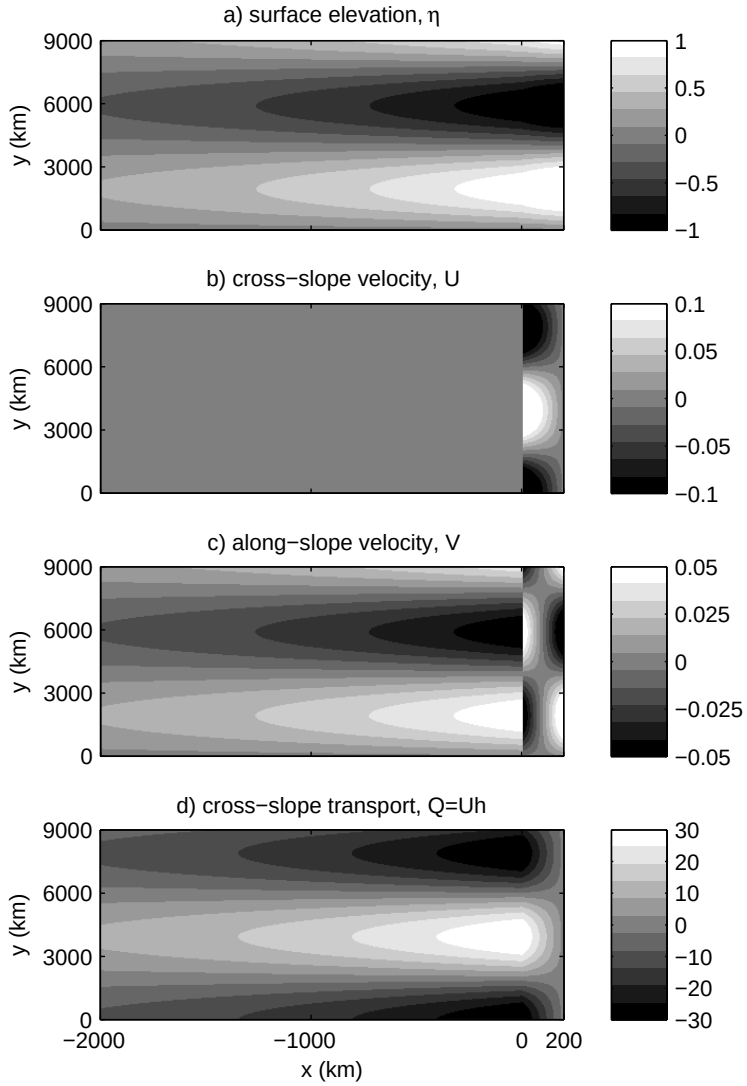

Fig. 7.4: The presence of a continental shelf (at $x > 0$) leads to a *modified* Kelvin wave. Note the presence of a cross-slope velocity component in **b**, and the associated *cross-slope transport* ($\mathrm{m^2 s^{-1}}$) in **d**. Here $H_s = 200$ m and $L = 200$ km; remaining parameters as in Figure 7.2. The dispersion relation (7.2) gives $l = 7.968 \times 10^{-7}$ rad $\mathrm{m^{-1}}$.

From Figure 7.4b this is obvious for the shelf-region, although in the deep ocean the component is too weak to be visible. Yet, it is there, too, as becomes clear when we look at the cross-slope *transport*, Q (Figure 7.4d). In accordance with the boundary conditions we imposed, the transport is continuous over the shelf break ($x = 0$), where it takes its largest values, and vanishes at the end of the

shelf ($x = L$).

Despite the utter simplicity of the model, the value for the cross-slope transport in Figure 7.4d is realistic (given that we have chosen a realistic value for the surface amplitude, η, cf. Figure 7.1). In the Bay of Biscay, for example, the cross-slope transport near the shelf break typically lies in the range 30-40 $\mathrm{m}^2\,\mathrm{s}^{-1}$ [49], although at some locations values as high as $100\,\mathrm{m}^2\,\mathrm{s}^{-1}$ may be attained [27].

The presence of a cross-slope velocity component, in combination with variation in depth, implies the presence of a *vertical* velocity, because of the boundary condition $W = Uh_x$ at the bottom. In the present case, the vertical velocity has a singular character (a δ-peak), being concentrated at $x = 0$. For a smoother topography, this singularity disappears. In either case, the presence of this barotropically induced vertical velocity in a *stratified* fluid means that the isopycnal surfaces are periodically lifted up and pulled down. Like in a stretched rope or string, an imposed oscillation at one point engenders waves travelling away from that point. Here these waves are the *internal tides*, which propagate away from their source according to the laws established in previous chapters. In the remainder of this chapter, we work out some specific examples.

First, we will simplify the problem a bit further. As can be seen from Figure 7.4d, the cross-slope transport Q varies spatially, both in the along- and cross-slope direction. However, so far as the generation of internal tides is concerned, the transport matters only in the generation region itself, i.e. over the slope (which in the above example is concentrated at $x = 0$). The transport typically varies over scales of order thousand km in the along-slope direction, and of order hundred km in the cross-slope direction. So, if we zoom in into a region around the shelf break, within a distance of the order of a few tens of km, the transport will appear nearly spatially uniform. This leads us to a somewhat crude but very convenient idealization: we will henceforth ignore the spatial variability altogether, and prescribe the cross-slope transport simply as $Q = Q_0 \sin \omega t$, with amplitude Q_0, a constant.

7.2 Boundary vs. body forcing

Our starting point is, again, (4.25),

$$\frac{\partial^2}{\partial t^2}\nabla^2 w + (\vec{f}\cdot\nabla)^2 w + N^2\nabla_h^2 w = 0\,. \tag{7.3}$$

To simplify matters, we adopt the Traditional Approximation ($\tilde{f} = 0$); moreover, we assume uniformity in one of the horizontal directions, $\partial/\partial y = 0$. Notice that a velocity component v will still be present in the y direction, due to Coriolis effects.

The assumption $\partial/\partial y = 0$ allows us to introduce a streamfunction Ψ, defined via $u = \Psi_z$ and $w = -\Psi_x$, by which the continuity equation (4.21d) is

automatically satisfied. In terms of the streamfunction, (7.3) becomes

$$\nabla^2\Psi_{tt} + f^2\Psi_{zz} + N^2\Psi_{xx} = 0\,. \tag{7.4}$$

with $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial z^2$.

In previous chapters we have interpreted w in (7.3) as an internal-wave field, but it need not be restricted to that, as is clear from the derivation in Section 4.5.1. In this chapter, w, or Ψ, will stand for the internal-tide field *plus a prescribed barotropic tide field*. We can effectuate this by imposing the following boundary condition for Ψ:

$$\Psi = 0 \quad \text{at } z = 0\,; \qquad \Psi = Q_0 \exp(-i\omega t) \quad \text{at } z = -h(x)\,, \tag{7.5}$$

where $-h(x)$ describes the bottom topography; Q_0 is a constant, and ω, the tidal frequency. Together, these conditions impose a spatially uniform, time-oscillating transport, as envisaged at the end of the previous section. This is immediately clear from vertically integrating $u = \Psi_z$, which gives

$$\int_{-h(x)}^{0} dz\, u = \Psi|_{-h(x)}^{0} = -Q_0 \exp(-i\omega t)\,.$$

Moreover, (7.5) imply that the boundary conditions (4.30) and (4.31) are fulfilled. For the latter, this can be seen by writing the second condition in (7.5) as

$$\Psi(t, x, z)|_{z=-h(x)} = \Psi(t, x, -h(x)) = Q_0 \exp(-i\omega t)\,,$$

whose total derivative to x is zero, hence

$$\Psi_x - h_x\Psi_z = 0 \quad \text{at } z = -h(x)\,,$$

which is equivalent to (4.31).

The problem is to solve, for given $N(z)$ and $h(x)$, eq. (7.4) along with the boundary conditions (7.5). The solution Ψ will contain the prescribed barotropic tide (here simplified to an oscillating transport), which acts as the forcing mechanism, and a baroclinic field, which is the response to the forcing.

In this formulation of the problem, the forcing is imposed via the *boundary conditions*. An alternative, equivalent way of posing the problem is to include the forcing in the equation itself, as a *body-force term*. Following [22], we write

$$\Psi = \Psi_0 + \psi\,, \quad \text{with } \Psi_0 = -\frac{zQ_0}{h(x)} \exp(-i\omega t)\,, \tag{7.6}$$

substitution of which into (7.4) gives

$$\nabla^2\psi_{tt} + f^2\psi_{zz} + N^2\psi_{xx} = zQ_0(N^2 - \omega^2)\Big(\frac{1}{h}\Big)_{xx} \exp(-i\omega t)\,. \tag{7.7}$$

In terms of ψ, the boundary conditions (7.5) become

$$\psi = 0 \quad \text{at } z = 0\,; \qquad \psi = 0 \quad \text{at } z = -h(x)\,. \tag{7.8}$$

The forcing is no longer present in the boundary conditions; instead, it is now imposed via the right-hand side of (7.7).

From (7.8) it is immediately clear that the vertically integrated horizontal velocity associated with ψ must be zero. Still, we cannot interpret ψ as a purely baroclinic field; over the slope, it also contains the nonhydrostatic part of the barotropic field [22]. However, under the hydrostatic approximation (i.e. assuming $\omega \ll N$), the interpretation of ψ as the baroclinic field is correct.

7.3 Generation over a step-topography

We return to the configuration shown in Figure 7.3, with one important modification: we remove the vertical boundary on the continental shelf, so that the shelf is now open and extends to $x \to \infty$. Assuming constant N, we reproduce, in slightly modified form, the solution previously derived by [79].

The problem is to solve (7.4), subject to the boundary conditions (7.5). Without loss of generality, we may write Ψ as in (7.6). Since the deep ocean ($x < 0$) and the continental shelf ($x > 0$) both have a horizontal bottom, we can, for each separately, write ψ in terms of the modal expressions of Section 5.2. Thus,

$$\Psi = \begin{cases} -\frac{zQ_0}{H}\exp(-i\omega t) + \sum_n a_n \sin(\frac{n\pi z}{H})\exp i(-k_n x - \omega t) & \text{for } x < 0\,; \\ -\frac{zQ_0}{H_s}\exp(-i\omega t) + \sum_n a_{s,n} \sin(\frac{n\pi z}{H_s})\exp i(k_{s,n} x - \omega t) & \text{for } x > 0\,, \end{cases} \tag{7.9}$$

where a_n and $a_{s,n}$ are arbitrary complex coefficients; k_n and $k_{s,n}$ are positive wavenumbers defined by (5.14):

$$k_{(s,)n} = \frac{n\pi}{H_{(s)}}\left(\frac{\omega^2 - f^2}{N^2 - \omega^2}\right)^{1/2}, \qquad n = 1, 2, 3, \cdots . \tag{7.10}$$

In (7.9) we anticipate that waves will propagate *away* from the source, i.e. to the left in the deep ocean, and to the right over the shelf. Also, by using sine series in (7.9), the boundary conditions (7.5) are automatically satisfied.

The problem has thus been reduced to finding the coefficients a_n and $a_{s,n}$. They are obtained from imposing appropriate matching conditions at $x = 0$. Moreover, the second boundary condition in (7.5) is to be taken into account at $x = 0$; specifically, this means that the 'deep-sea' solution Ψ_- (i.e. the upper expression in (7.9)) should satisfy, at the vertical slope,

$$\Psi_-|_{x=0} = Q_0 \exp(-i\omega t) \quad \text{for } -H < z < -H_s\,.$$

Hence

$$-\frac{zQ_0}{H} + \sum_n a_n \sin\left(\frac{n\pi z}{H}\right) = Q_0 \quad \text{for } -H < z < -H_s\,. \tag{7.11}$$

Moreover, above the slope, we impose continuity of the streamfunction Ψ as well as of its horizontal derivative Ψ_x (the latter condition implies continuity of

the vertical velocity, w). So, for $-H_s < z < 0$,

$$-\frac{zQ_0}{H} + \sum_n a_n \sin\left(\frac{n\pi z}{H}\right) = -\frac{zQ_0}{H_s} + \sum_n a_{s,n} \sin\left(\frac{n\pi z}{H_s}\right) \quad (7.12)$$

$$-\sum_n a_n k_n \sin\left(\frac{n\pi z}{H}\right) = \sum_n a_{s,n} k_{s,n} \sin\left(\frac{n\pi z}{H_s}\right). \quad (7.13)$$

Since (7.11)–(7.13) contain no complex parts, the coefficients a_n and $a_{s,n}$ will be real.

The conditions (7.11) and (7.12) have identical left-hand sides, and can be combined to

$$-\frac{zQ_0}{H}+\sum_n a_n \sin\left(\frac{n\pi z}{H}\right) = \begin{cases} Q_0 & \text{for } -H < z < -H_s \\ -\frac{zQ_0}{H_s} + \sum_n a_{s,n} \sin(\frac{n\pi z}{H_s}) & \text{for } -H_s < z < 0\,. \end{cases}$$

We multiply this expression by $\sin(m\pi z/H)$ and integrate over the vertical interval $(-H, 0)$ to obtain, after some rewriting,

$$a_m = -\frac{2Q_0}{\alpha(m\pi)^2}\sin(m\pi\alpha) + \frac{2\alpha}{\pi}\sin(m\pi\alpha)\sum_n \frac{n(-1)^n}{(m\alpha)^2 - n^2}\, a_{s,n}\,. \quad (7.14)$$

with $\alpha = H_s/H$. Here we used the integral identities listed in Appendix A, at the end of this chapter.

We multiply the remaining condition (7.13) by $\sin(m\pi z/H_s)$ and integrate over $(-H_s, 0)$; this yields

$$a_{s,m} = -\frac{2\alpha(-1)^m}{\pi}\sum_n \frac{n\sin(n\pi\alpha)}{(n\alpha)^2 - m^2}\, a_n\,, \quad (7.15)$$

where we used (7.10) and the last two expressions of Appendix A.

If we truncate the series at a certain modenumber M, we can write (7.14) and (7.15) in matrix form as $a = F + Aa_s$ and $a_s = Ba$, respectively, where A and B are $M \times M$ matrices, and F a vector. Hence a can be obtained by matrix inversion: $a = (I - AB)^{-1}F$, where I is the identity matrix. For given parameters, this is easily done using numerical tools.

A solution thus obtained, involving 25 modes, is shown in Figure 7.5. Here the horizontal baroclinic velocity is shown,

$$u' = \begin{cases} \sum_n a_n \frac{n\pi}{H}\cos(\frac{n\pi z}{H})\exp i(-k_n x - \omega t) & \text{for } x < 0\,; \\ \sum_n a_{s,n} \frac{n\pi}{H_s}\cos(\frac{n\pi z}{H_s})\exp i(k_{s,n} x - \omega t) & \text{for } x > 0\,. \end{cases} \quad (7.16)$$

Figure 7.5 demonstrates that there are two beams, one propagating into the deep ocean, the other onto the shelf. The extent (i.e. width) of the former is in this case determined by the two oceanward characteristics that can be drawn from the shelf break: one left- and downward, the other, left- and upward. The latter leads, after reflection from the surface, to a characteristic parallel to the former; the beam is confined precisely between these two characteristics. We

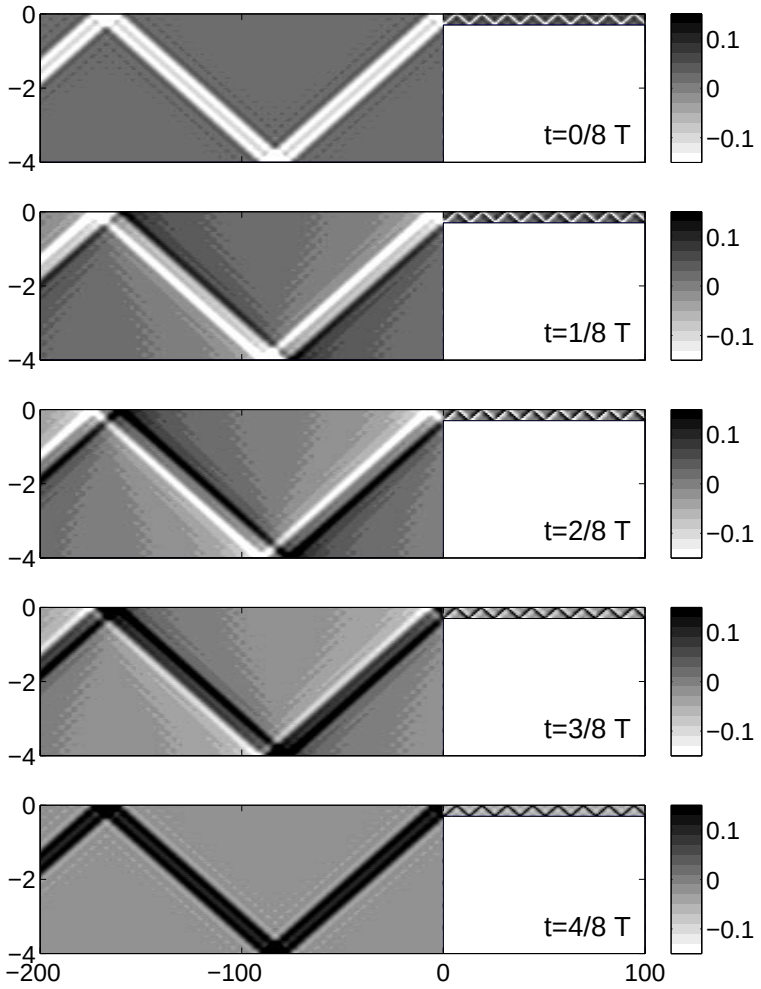

Fig. 7.5: Internal-tide generation over a steep continental slope: the horizontal baroclinic velocity (in $\mathrm{m\,s^{-1}}$) at five instances during half a tidal period. Parameter values are: $N = 2 \times 10^{-3}$, $f = 1.0 \times 10^{-4}$ (latitude $\phi = 45°\mathrm{N}$), $\omega = 1.4 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$; $H = 4000\,\mathrm{m}$, $H_s = 300\,\mathrm{m}$, and $Q_0 = 100\,\mathrm{m^2\,s^{-1}}$; 25 modes are included.

note that this is rather due to the presence of a sharp corner; for a more realistic, smoother topography, no such surface reflection occurs (see Section 7.5).

From the development in time in Figure 7.5, we see that in the beam descending into the deep ocean, phase propagation is upward, implying that energy must propagate downward (see Section 6.1.1); after bottom reflection, this is reversed.

As noted before (Section 7.1), the present value of $Q_0 = 100\,\mathrm{m^2\,s^{-1}}$ is representative for some locations the Bay of Biscay. Apart from a qualitiative similarity between the descending beams in Figures 1.9 and 7.5, the horizontal

velocities found here, of a few tens of $\mathrm{cm\,s^{-1}}$, indeed correspond to observed values [69]. This similarity even extends to the so-called *conversion rate*, the amount of energy transferred from barotropic to baroclinic tides, per unit of time and unit of lateral extent. This amount must be equal to the vertically-integrated energy flux in the deep ocean and over the shelf together. As noted in Section 4.5.2, the energy flux is given by $u'p'$, the primes here denoting baroclinic fields. Without entering into the technical details here, we mention the result from [79], which states the vertically-integrated flux in terms of the modal coefficients,

$$F = \int dz\, \langle u'p' \rangle = \frac{\rho_* \pi}{4\omega} [(N^2 - \omega^2)(\omega^2 - f^2)]^{1/2} \sum_n n(a_n^2 + a_{s,n}^2)\,.$$

Here $\langle \cdot \rangle$ denotes the mean over a tidal period. The outcome for the parameters of Figure 7.5 is 9.7×10^3 $\mathrm{kW\,m^{-1}}$. By far the largest part goes into the deep ocean; the energy flux onto the shelf is only 2% of the total amount.

Finally, we look at the convergence of the solution. There are two aspects to this. First, how quickly the coefficient a_n, for any fixed n, converges as one increases the number of modes (M) involved in the matrix inversion. This convergence is rapid; for example, for the first mode a_1, the difference between using $M = 1$ or $M = 25$ amounts to an increase of only 0.2%. Second, and physically more interesting, the behaviour of a_n with n. This is illustrated in Figure 7.6, where we plot $|na_n|$ (the reason for including a factor n is that the horizontal velocity u' is proportional to na_n rather than to a_n). Clearly, the lowest modes are the most important ones. We note, however, that this result hinges on the assumption of constant N. In the Bay of Biscay, for example, it was found that the third mode is the dominant one [69]; this can be explained by the fact that this mode has a local maximum in the seasonal thermocline (cf. Figure 5.12), and is more strongly forced than other modes. In this respect, it is worthwhile to note that the forcing of internal tides tends to be stronger in regions of high N, as is seen from the fact that the body-force term on the right-hand side of (7.7) increases with increasing N.

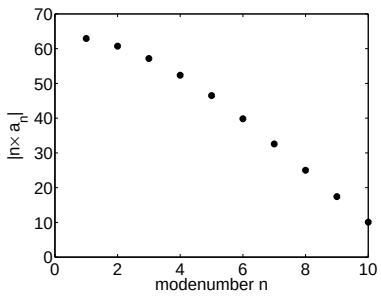

Fig. 7.6: The contribution of the lowest modes to u', for the parameters of Figure 7.5.

7.4 Solutions for infinitesimal topography

Our starting point is again the equation for internal tides, with the barotropic forcing term on its right-hand side, (7.7),

$$\nabla^2\psi_{tt} + f^2\psi_{zz} + N^2\psi_{xx} = zQ_0(N^2-\omega^2)\Big(\frac{1}{h}\Big)_{xx}\exp(-i\omega t)\,, \tag{7.17}$$

which is to be solved subject to the boundary conditions (7.8),

$$\psi = 0 \quad \text{at } z = 0\,; \qquad \psi = 0 \quad \text{at } z = -h(x)\,. \tag{7.18}$$

In the previous section, the problem was solvable due to the fact that the deep-ocean basin and continental shelf were both assumed to be of constant depth. In this section, we make an even more radical assumption: we assume the topography to be *infinitesimal*, so that we can solve the equation with the same set of vertical modes for the entire domain. However, the topographic form is left arbitrary (provided that it has a small amplitude); moreover, we pose the problem in terms of an arbitrary profile of stratification, $N(z)$.

For constant h the forcing term in (7.17) would of course vanish altogether. However, if we write

$$h(x) = H - r(x)\,, \qquad \text{with } |r| \ll H\,,$$

and constant H, we can make the following approximation in the forcing term, using the Taylor expansion $(1+y)^\alpha = 1 + \alpha y + \cdots$,

$$\frac{1}{h} = \frac{1}{H-r(x)} = \frac{1/H}{1-r(x)/H} \approx \frac{1}{H}\left(1 + \frac{r(x)}{H}\right).$$

Eq. (7.17) then becomes

$$\nabla^2\psi_{tt} + f^2\psi_{zz} + N^2\psi_{xx} = zQ_0(N^2-\omega^2)H^{-2}r_{xx}\exp(-i\omega t)\,. \tag{7.19}$$

Furthermore, we may apply the second boundary condition in (7.18) simply at $z = -H$, so

$$\psi = 0 \quad \text{at } z = 0\,; \qquad \psi = 0 \quad \text{at } z = -H\,. \tag{7.20}$$

The problem to be solved is now (7.19), subject to (7.20), for arbitrary $r(x)$ and $N(z)$. We try a solution of the form

$$\psi = \sum_n a_n(x)\phi_n(z)\exp(-i\omega t)\,, \tag{7.21}$$

where ϕ_n are solutions of (5.3),

$$\phi_n'' + k_n^2\,\frac{N^2(z)-\omega^2}{\omega^2-f^2}\,\phi_n = 0\,, \tag{7.22}$$

and hence are orthogonal (see Section 5.1.2). Substitution of (7.21) in (7.19), then multiplication by ϕ_m and vertical integration, gives,

$$a_n''(x) + k_n^2 a_n(x) = \frac{d_n Q_0}{H^2}\,r_{xx}\,, \tag{7.23}$$

with

$$d_n = \frac{\int_{-H}^{0} dz\, z\, \phi_n''}{\int_{-H}^{0} dz\, \phi_n'' \phi_n}\,. \tag{7.24}$$

The problem has thus been reduced to solving (7.22) and (7.23). In Chapter 5 several examples were given of profiles $N(z)$ for which (7.22) can be solved analytically; we will consider some of them in the following sections.

The other equation, (7.23), is easy to solve; its general solution reads

$$a_n = C_{1,n}\exp(ik_n x)+C_{2,n}\exp(-ik_n x)+\frac{d_n Q_0}{H^2}\int_0^x dy\, \cos k_n(x-y) r_y(y)\,, \tag{7.25}$$

in which the first two terms (with arbitrary complex coefficients $C_{1,n}$ and $C_{2,n}$) describe free left- and rightward propagating waves; the third term is a particular solution to (7.23), as can be verified by substitution.

However, this is not by itself an appropriate solution to our problem; in order that the solution contain only waves emanating from the forcing region, we need to exclude unforced waves entering from $\pm\infty$. This is done by imposing *radiation conditions*; for $k_n > 0$, they read[2]

$$a_n \sim \begin{cases} \exp(-ik_n x) & \text{as } x \to -\infty \\ \exp(ik_n x) & \text{as } x \to +\infty\,. \end{cases}$$

These requirements allow us to determine the constants $C_{1,n}$ and $C_{2,n}$. Writing, in (7.25), the cosine in terms of complex exponential functions, we have

$$\begin{aligned}\int_0^x dy\, \cos k_n(x-y) r_y(y) &= \tfrac{1}{2}\exp(ik_n x)\int_0^x dy\, \exp(-ik_n y)\, r_y(y) \\ &\quad +\tfrac{1}{2}\exp(-ik_n x)\int_0^x dy\, \exp(ik_n y)\, r_y(y)\,,\end{aligned}$$

so that a_n becomes

$$\begin{aligned}a_n &= \exp(ik_n x)\Big[C_{1,n}+\frac{d_n Q_0}{2H^2}\int_0^x dy\, \exp(-ik_n y)\, r_y(y)\Big] \\ &\quad +\exp(-ik_n x)\Big[C_{2,n}+\frac{d_n Q_0}{2H^2}\int_0^x dy\, \exp(ik_n y)\, r_y(y)\Big]\,.\end{aligned}$$

To satisfy the radiation conditions, we must choose

$$\begin{aligned}C_{1,n} &= \frac{d_n Q_0}{2H^2}\int_{-\infty}^{0} dy\, \exp(-ik_n y)\, r_y(y) \\ C_{2,n} &= -\frac{d_n Q_0}{2H^2}\int_0^{\infty} dy\, \exp(ik_n y)\, r_y(y)\,.\end{aligned}$$

[2]NB: The radiation conditions concern the direction of *energy* propagation. We implicitly assume here that $|f| < N$, in which case phases and energy have the same *horizontal* direction of propagation (see Sections 5.2.1 and 6.1.1), so that we may pose the conditions by requiring that phases leave the domain at infinity.

Hence

$$\boxed{a_n = \frac{d_n Q_0}{2H^2}\Big[A_n(x)\exp(ik_n x) - B_n(x)\exp(-ik_n x)\Big]\,,} \tag{7.26}$$

with

$$\begin{aligned} A_n(x) &= \int_{-\infty}^{x} dy\, \exp(-ik_n y)\, r_y(y) \\ B_n(x) &= \int_{x}^{\infty} dy\, \exp(ik_n y)\, r_y(y)\,. \end{aligned}$$

With this, a closed solution satisfying the radiation conditions has been obtained for general (small) topography $r(x)$.

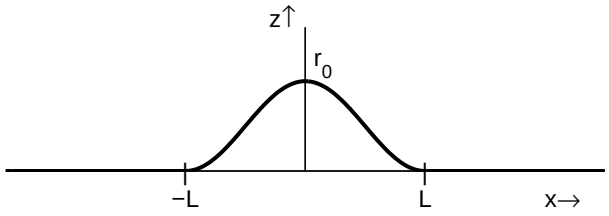

Fig. 7.7: A seamount, defined by (7.27).

As an example, we consider a symmetric seamount, described by a continuously differentiable function:

$$r(x) = \begin{cases} 0 & \text{for } x < -L \\ \frac{1}{2} r_0[1 + \cos(\pi x/L)] & \text{for } -L < x < L \\ 0 & \text{for } x > L\,, \end{cases} \tag{7.27}$$

with amplitude r_0 and width $2L$ (Figure 7.7). For this r, the functions $A_n(x)$ and $B_n(x)$ can be evaluated (Appendix B). One important feature deserves mention: it turns out that the modes travelling away from the topography are proportional to

$$R(k_n L) = \frac{\sin(k_n L)}{\pi^2 - (k_n L)^2}\,,$$

which is one of the two factors determining the strength of each mode (the other one being d_n in (7.26)). The dependence of R on the product $k_n L$ is shown in Figure 7.8; the response is strong for a certain range of values (2 to 5, say), and very weak as $k_n L$ becomes large. Still, even for certain low values the response may vanish altogether.

7.4.1 Uniform stratification

Here and in later sections, we consider various types of stratification. The simplest choice is, of course, a constant N, as in Section 5.2.2. We then have

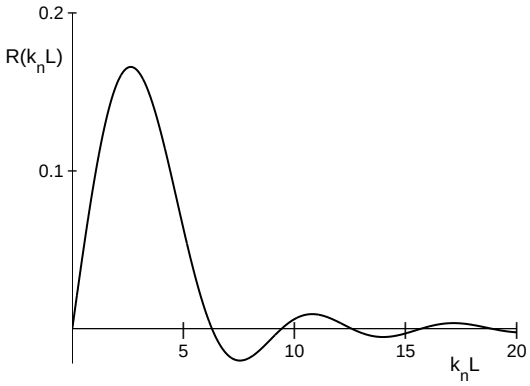

Fig. 7.8: The response function R, one of the factors determing how strongly each mode contributes to the internal tidal signal.

from (7.22),

$$\phi_n = \sin\left(\frac{n\pi z}{H}\right), \qquad n = 1, 2, 3, \cdots . \tag{7.28}$$

and, from the dispersion relation (5.14),

$$k_n = \frac{n\pi}{H}\left(\frac{\omega^2 - f^2}{N^2 - \omega^2}\right)^{1/2}, \qquad n = 1, 2, 3, \cdots . \tag{7.29}$$

where we selected positive k_n, in accordance with the way the solution was constructed in the previous section.

Notice that we need not include a constant coefficient in (7.28), since it would automatically disappear from the problem through the combination $d_n\phi_n$. So, the solution is fully determined by (7.21),

$$\psi = \sum_n a_n(x)\phi_n(z)\exp(-i\omega t)\,, \tag{7.30}$$

with a_n from (7.26), and ϕ_n from (7.28). The real part of (7.30) is implied. The coefficient d_n is here given by

$$d_n = -\frac{2H}{n\pi}(-1)^n\,.$$

It follows from the expressions in Appendix B that $A_n, B_n \sim k_n^{-2}$ for large n. Together with $d_n \sim n^{-1}$ and $\phi_n' \sim n$, this means that, for mode n, $u_n = \psi_{n,z} \sim n^{-2}$, implying convergence of the series $u = \sum_n u_n$. (This, however, depends on the choice of the topography, $r(x)$; if it were not continuously differentiable, i.e. if it were having sharp corners, the series would yield logarithmic singularities on the characteristics emanating from these corners.)

The baroclinic horizontal cross-slope velocity,

$$u = \sum_n a_n(x)\phi_n'(z)\exp(-i\omega t)\,. \tag{7.31}$$

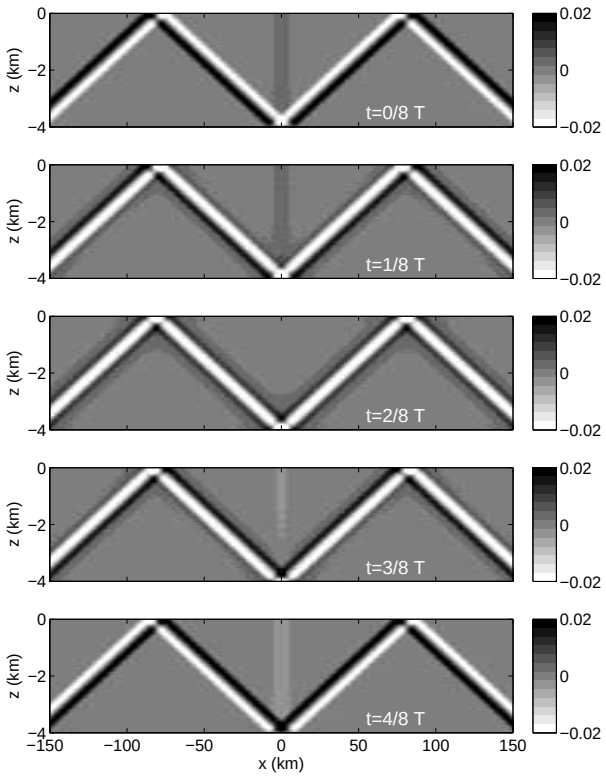

Fig. 7.9: Internal-tide generation over a small seamount (depicted in Figure 7.7): the horizontal baroclinic velocity u (in $\mathrm{m\,s^{-1}}$) at five instances during half a tidal period. Parameter values are: $N = 2 \times 10^{-3}$, $f = 1.0 \times 10^{-4}$ (latitude $\phi = 45°\mathrm{N}$), $\omega = 1.4052 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$ ($\mathrm{M_2}$ tidal frequency); $H = 4000\,\mathrm{m}$, $r_0 = 500\,\mathrm{m}$, $L = 10\,\mathrm{km}$, and $Q_0 = 100\,\mathrm{m^2\,s^{-1}}$; 25 modes are included.

is shown in Figure 7.9, at five different moments during half a tidal cycle. We see that two beams emanate from the seamount, which is centered around $x = 0$: one leftward propagating, the other, rightward. This fact is easily deduced from the direction of phase propagation and the rules for energy and phase propagation established in Section 5.2.1, here for $N > |f|$.

7.4.2 Three-layer model

The assumption of constant N, made in the previous section, provides a rather inadequate description of the ocean's stratification; notably, it lacks the thermocline. To capture this important feature, we return to the 3-layer model discussed in Section 5.4:

$$N^2(z) = \begin{cases} 0 & -d < z < 0 & \text{(mixed layer)} \\ g'/\epsilon & -d-\epsilon < z < -d & \text{(thermocline)} \\ N_c^2 & -H < z < -d-\epsilon & \text{(abyss)}\,, \end{cases}$$

see also Figure 5.11. Again, we will take the middle, representing the thermocline, to be infinitely thin ($\epsilon \to 0$), thus reducing it to an interface.

We recapitulate the earlier results that we need here. The wavenumbers k_n are solved (numerically) from the dispersion relation (5.41),

$$q_l \cos q_l(H-d) \sinh q_u d + q_u \sin q_l(H-d) \Big[\cosh q_u d - \frac{g' q_u}{\omega^2} \sinh q_u d \Big] = 0\,, \quad (7.32)$$

where H is water depth, d the thickness of the upper mixed layer, g' a measure of the strength of the thermocline. The abyssal stratification N_c (a constant) features in

$$q_l = k_n \Big(\frac{N_c^2 - \omega^2}{\omega^2 - f^2} \Big)^{1/2}\,.$$

The vertical wavenumber associated with the mixed layer is

$$q_u = k_n \Big(\frac{\omega^2}{\omega^2 - f^2} \Big)^{1/2}\,.$$

The vertical modes are given by (5.42),

$$\phi_n(z) = \begin{cases} -\frac{\sin q_l(H-d)}{\sinh q_u d} \sinh q_u z & -d < z < 0 \\ \sin q_l(z+H) & -H < z < -d\,. \end{cases} \quad (7.33)$$

The coefficients d_n are defined by (7.24); the integral expressions needed here are listed in Appendix B. With k_n and d_n thus obtained, (7.21) is fully determined, with a_n given by (7.26). From this, we can derive u as in (7.31). Its real part is shown in Figure 7.10.

The presence of the thermocline clearly has a marked influence on the wave pattern, as is evident from a comparison with Figure 7.9, which lacked a thermocline. As in the previous example, the beams emanate from the seamount, but when they impinge on the thermocline (near $x = \pm 80\,\mathrm{km}$), they are strongly 'scattered'. As a result, the energy is no longer concentrated in one beam, but is spread out widely. The physical mechanism behind this is the occurrence of *internal reflections*, which are due to the strong inhomogeneity of the medium, i.e. the strongly vertically varying $N(z)$. Successive internal reflections then cause the spreading, as sketched in Figure 6.6.

The effect of the beams, as they hit the thermocline, can also be seen from the interfacial displacement itself. They follow from $w = \eta_t$, where w can expressed in terms of the streamfunction. To find the *total* displacement, we have to return to the original streamfunction $\Psi = \Psi_0 + \psi$, as introduced in (7.6); we then have $\eta_t = -\Psi_x$. Since we assume the topography to be small ($h = H - r$, $|r| \ll H$), we can approximate the derivative $\Psi_{0,x}$ as

$$\Psi_{0,x} = -\Big(\frac{zQ_0}{h(x)} \Big)_x \exp(-i\omega t) \approx -\frac{zQ_0}{H^2}\, r_x \exp(-i\omega t)\,.$$

Hence

$$\eta = i\, \frac{zQ_0}{\omega H^2}\, r_x \exp(-i\omega t) - \frac{i}{\omega} \sum_n a_{n,x}(x) \phi_n(z) \exp(-i\omega t)\,. \quad (7.34)$$

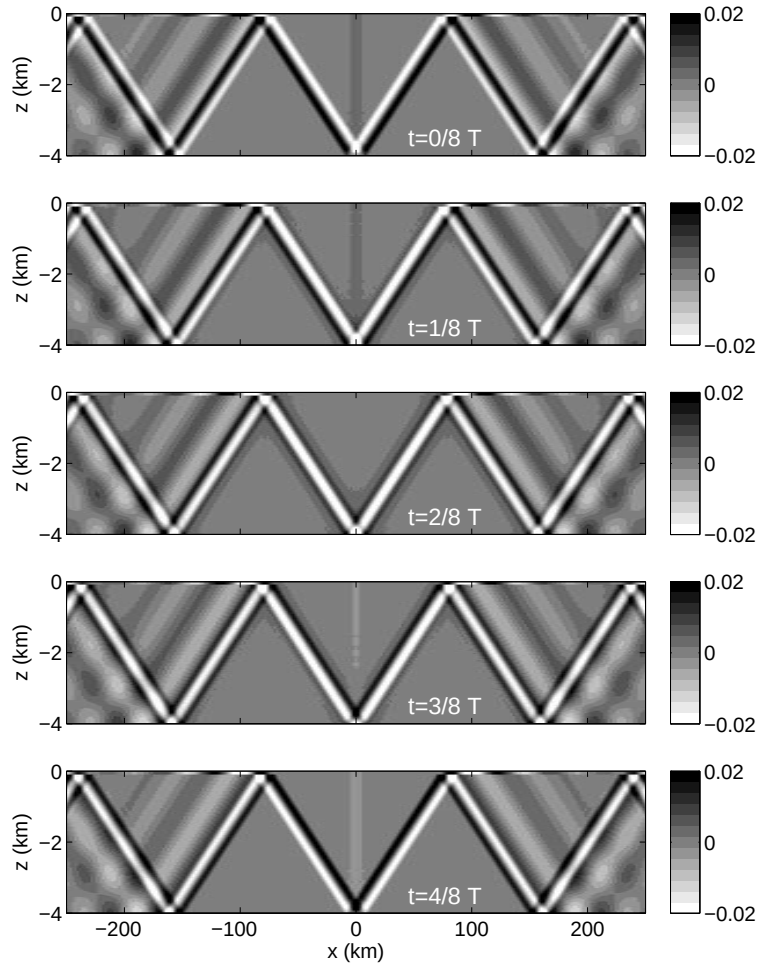

Fig. 7.10: Internal-tide generation over a small seamount (depicted in Figure 7.7): the horizontal baroclinic velocity u (in $\mathrm{m\,s^{-1}}$) at five instances during half a tidal period. Here the stratification of Figure 5.11 is used (for $\epsilon \to 0$), with $d = 100\,\mathrm{m}$ (mixed-layer thickness), $g' = 0.005\,\mathrm{m\,s^{-2}}$ (strength of the thermocline), and $N_c = 2 \times 10^{-3}\,\mathrm{rad\,s^{-1}}$ (abyssal stratification). The remaining parameters (H, f, ω, r_0, L and Q_0) are as in the previous figure; here, too, 25 modes are used.

The derivative of a_n follows from (7.26):

$$a_{n,x} = \frac{d_n Q_0}{2H^2}\left[2r_x(x) + ik_n\Big(A_n(x)\exp(ik_n x) + B_n(x)\exp(-ik_n x)\Big)\right],$$

where the term $2r_x$ stems from taking the derivatives of A_n and B_n; the effect of this term is of course restricted to the layer over the seamount, like the term $\Psi_{0,x}$. The result of (7.34), evaluated at $z = -d$, i.e. at the thermocline, is shown in Figure 7.11, for different stages during a full tidal cycle (time progresses

downward). Clearly, the interface is virtually at rest over the seamount, and at some distance away from it. However, at the location where the beam impinges on the thermocline ($x = \pm 80\,\mathrm{km}$), the interface is brought into oscillation; waves propagate away from this point. They become gradually smaller, and the cause of this attenuation is seen in Figure 7.10: they leak into the deep ocean as *beams*. The upshot of this simple example is that internal tides are neither purely interfacial nor pure beams for this kind of stratification; they appear in mixed form. This is an important feature of some internal-tide observations, as is further discussed in Section 8.5.2.

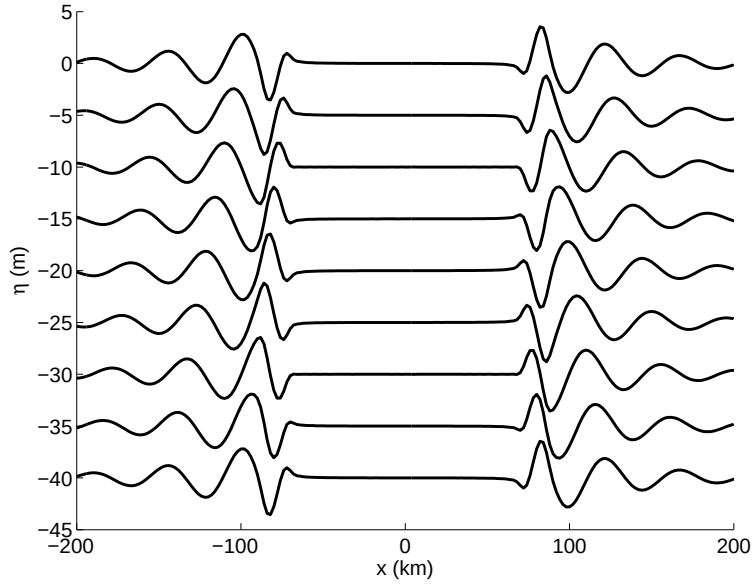

Fig. 7.11: The evolution of the interfacial (i.e. thermocline) displacement, at intervals of 1/8 tidal period; time progresses downward; subsequent profiles are given a shift of -5, for better visibility. The same parameters as in the previous figure.

7.4.3 Interfacial tides

A special case of the stratification used in the previous section is that with $N_c = 0$; the stratification is then confined to the interface (Section 5.4.3). As a consequence, there is only *one* mode (as long as $\omega > |f|$): the interfacial mode. The barotropic tidal forcing will in this case produce purely interfacial tides. Its wavenumber follows from (5.44), and the modal structure is given by (5.45):

$$\phi(z) = \begin{cases} -\frac{\sinh q_u(H-d)}{\sinh q_u d} \sinh q_u z & -d < z < 0 \\ \sinh q_u(z+H) & -H < z < -d\,. \end{cases} \tag{7.35}$$

The expression for d_n is given in Appendix B.

The two-layer system, which involves only one mode, is rather inadequate for a description of internal tides in the deep ocean, but can sometimes be applied to shallower regions such as the continental shelf. We therefore choose parameters suitable for this region. In the modal sum of the velocity field u, (7.31), there is now of course only one mode involved; the result is shown in Figure 7.12. The propagation is purely horizontal, with oppositely directed currents in upper and lower layer. The currents are approximately inversely proportional to the thickness of the respective layers; hence the stronger currents in the thin upper mixed layer.

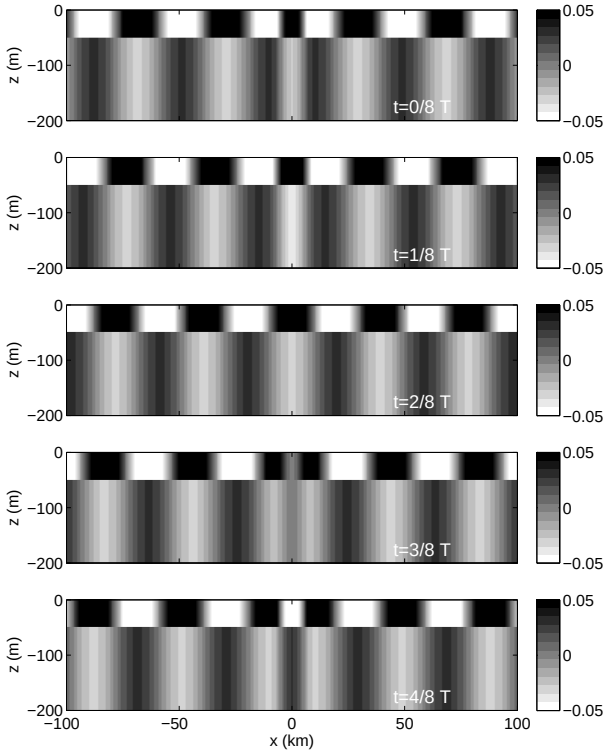

Fig. 7.12: Internal-tide generation over a small ridge (depicted in Figure 7.7) on the continental shelf: the horizontal baroclinic velocity u (in $\mathrm{m\,s^{-1}}$) at five instances during half a tidal period. Here the stratification of Figure 5.11 is used (for $\epsilon \to 0$), with $d = 50\,\mathrm{m}$ (mixed-layer thickness), $g' = 0.01\,\mathrm{m\,s^{-2}}$ (strength of the thermocline), and $N_c = 0$. The remaining parameters are $H = 200\,\mathrm{m}$ and $r_0 = 50\,\mathrm{m}$; L, f, ω, and Q_0 are as in previous figures.

The total interfacial displacement η is given by (7.34), involving now only one mode. The result and is shown in Figure 7.13, for 8 different phases of the tide.

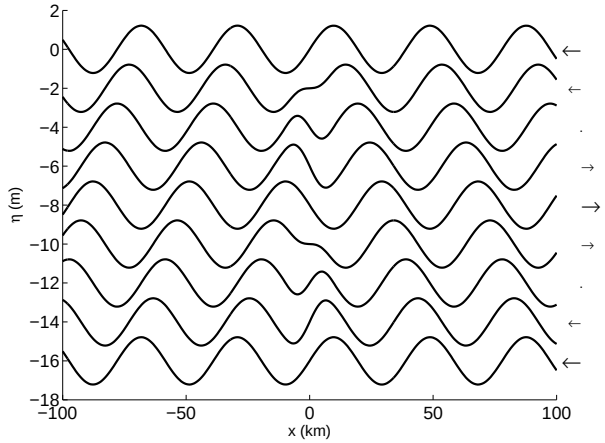

Fig. 7.13: Internal-tide generation over a small ridge, showing the total interfacial displacement $\eta_0 + \eta$ at $t = 0$ and at later stages separated by intervals of 1/8 tidal period. Time progresses downward; subsequent profiles are given a shift of -2, for better visibility. On the right, arrows indicate the direction of the barotropic tidal flow at each stage.

7.5 Energetics and conversion rates

One can roughly distinguish two main regions of internal-tide generation: 1) the continental slope, where generation is concentrated near the shelf break, and 2) deep ocean ridges and seamounts. For the latter, the geographical distribution of internal-tide sources is obtained from estimates of barotropic tidal energy dissipation, based on satellite altimetry; an example is shown in Figure 1.8.

This pattern is reproduced fairly well by results from numerical modelling, see Figure 7.14. Here the topography was assumed to be infinitesimal, as in preceding sections, but two-dimensionality was taken into account. The three essential ingredients of any internal-tide generation model are stratification, topography, and the barotropic flow. The three were here based on empirical data sets; for the stratification, a constant value, representative of the deepest layer, was used for each position. Despite all the caveats and simplications, the resulting pattern looks sensible, and the total amount internal-tide energy flux was found to be about 1 TW, in agreement with estimates based on satellite altimetry. The major influence of topographic features such as the Mid-Atlantic Ridge is obvious from Figures 1.8 and 7.14.

As briefly discussed in Section 1.4, internal tides (and, more generally, internal waves) are thought to play an important role in ocean mixing. This happens through a transfer of energy to shorter scales, the first stage of which may be, for example, near-critical reflection from slopes, or nonlinear evolution to shorter waves such as solitons (see the next chapter). This goes down to the scales at which waves break and turbulent mixing occurs (Figure 1.10). The very uneven distribution of internal-tide sources, as demonstrated in Figures 1.8 and

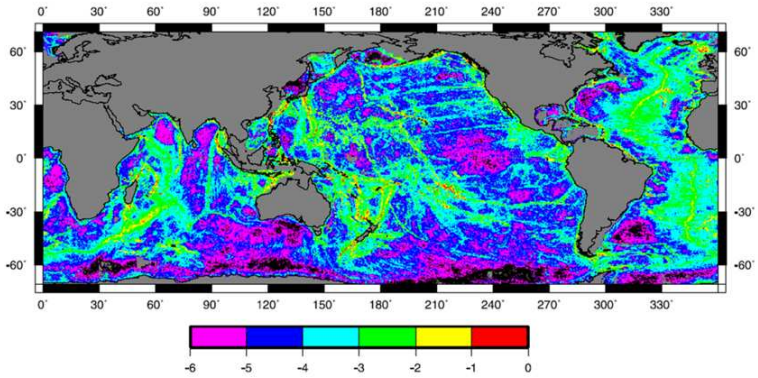

Fig. 7.14: Global distribution of M_2 internal tidal energy fluxes, in W/m^2, calculated numerically from an internal-tide generation model. The color scale is logarithmic, e.g., -3 means 10^{-3} W/m^2. From [63].

7.14, means that deep ocean mixing, too, must be distributed unevenly. This is an important point in the modelling of ocean circulation; it means that one should not prescribe a simple uniform eddy diffusivity throughout the ocean, but rather take into account the geographical distribution of the mixing intensity. The ocean circulation, in turn, contributes to the meridional heat transport (the lion share of which, though, is due to atmospheric circulation). So the heat transport, too, is influenced by the way mixing is distributed through the geographical spreading of internal-wave sources. This is illustrated in Figure 7.15.

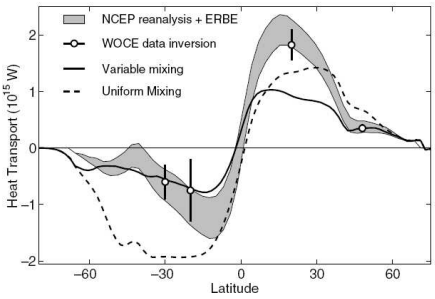

Fig. 7.15: The meridional heat transport for the global ocean. The gray band and circles are indirectly based on empirical data. The solid and dashed lines are results from a global circulation model, one with uniform mixing (dashed), in the other (solid), the geographical distribution of internal tide generation sources was taken into account. The latter seems to correspond better to the data; at any rate, the difference between the solid and dashed lines is conspicuous. From [77].

The continental slope, too, forms an important source region of internal tides. A very simplified analysis of this problem was made in Section 7.3. Attempts to

model analytically a more realistic setting are thwarted by the complications of variable topography and stratification, which imply that neither the method of vertical modes (Chapter 5) nor that of characteristics (Chapter 6) is applicable. However, the equations can be solved numerically, and this provides valuable insights. For one thing, the steepness of the slope matters; in Section 6.2 we discussed the distinction between sub- an supercritical slopes. Continental slopes are, generally, supercritical (apart from, of course, the upper and lower parts). An example of internal tide generation at a supercritical slope is shown in Figure 7.16 (left). We see a beam emanating from the shelf break; by comparison, very little goes onto the continental shelf. It is now interesting to see how the oceanward energy flux (F_1) depends on the steepness of the slope. This is illustrated in Figure 7.16 (right); note that the vertical scale is logarithmic. The angle α is a measure of the steepness of the slope; $\alpha > 1$ means supercritical slopes; $\alpha < 1$, subcritical slopes. We see that the flux varies only weakly with the angle of the slope as long as we are in the supercritical regime. The most extreme case of a supercritical slope would be the vertical slope considered in Section 7.3, and, as noticed there, the energy flux derived from this model is fairly close to results from more sophisticated modelling, involving realistically shaped slopes. This fits in with the result from Figure 7.16 for supercritical slopes. For subcritical slopes, on the other hand, we see that the oceanward flux is much smaller, and now depends strongly on the angle of the slope.

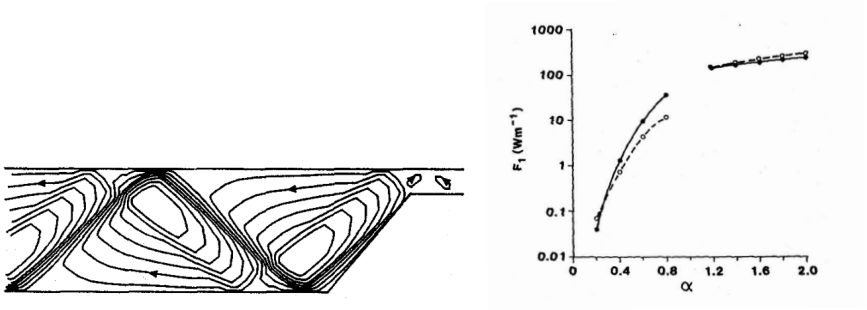

Fig. 7.16: Results from a numerical model. *Left:* The streamfunction showing internal-tide generation over a supercritical slope. *Right:* Dependence of the oceanward energy flux (F_1) on the steepness of the slope, α ($\alpha < 1$, subcritical; $\alpha > 1$, supercritical). The dashed lines are from numerical model calculations; the solid lines are from a representation in terms of powers of α: $F_1 \sim \alpha^5$ for the subcritical domain, $F_1 \sim \alpha$ for the supercritical domain. From [8].

Turning, finally, to a more realistic setting, we show model results based on topography and stratification in the Bay of Biscay (Figure 7.17). In the left panel we see the conversion rate, in W/m^3. It shows where the energy is converted from barotropic into baroclinic tides. The conversion rate forms a source term in the energy equation, which can be written,

$$E_t + \nabla(p'\vec{u}) = -\rho_* bW \,,$$

where E is the energy density, p' baroclinic pressure, $\vec{u}$ the baroclinic velocity, ρ_* a constant reference value for density, b the buoyancy, and W the vertical velocity component of the barotropic tide. We derived earlier an energy equation but without source term (Section 4.5.2); now the barotropic tide is responsible for an input of energy, represented by the term on the right-hand side. Since internal tides are periodic in time, time-averaging over a tidal period will make the first term disappear (in other words, at every position, the mean energy density stays the same), resulting in a balance between the flux gradient and the source term. Stated otherwise, we may take a surface (or volume) integral, implying that the nett flux through the boundaries of the chosen area must equal the integral input of energy by the source. In this sense we can take the surface integral in 7.17 (left panel). This gives the value of the integrated conversion rate, as stated at the bottom of the figure, in W/m. The 'per meter' refers to the transverse direction; so if the slope would stretch over one thousand kilometers (say) in the Bay of Biscay, we would have to multiply by this number to get the total conversion for this region.

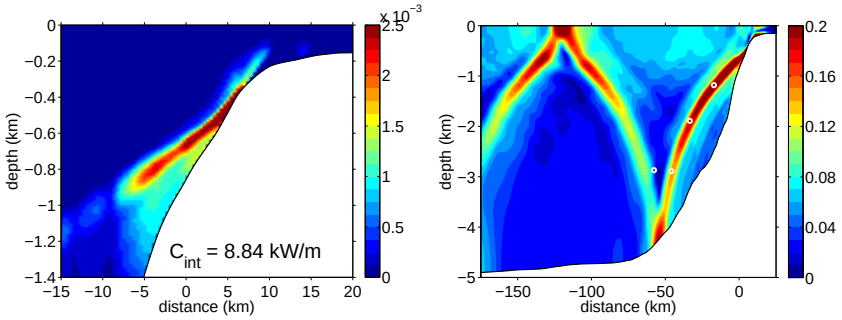

Fig. 7.17: Results from a numerical model, here applied to the Bay of Biscay. *Left:* The spatial distribution tidally-averaged conversion rate, $C = -\rho_* \langle bW \rangle$, in W/m^3 (the brackets stand for tide averaging); the integrated value over the entire area is also stated. *Right:* The internal tide emanating from the continental slope, here depicted in terms of the amplitude of the cross-slope velocity u, in m/s. The white dots denote the position of the observed beam, derived from Figure 1.9. From [27].

An important aspect of the conversion rate is its dependence on buoyancy b, which contains the *baroclinic* field itself. In other words, we cannot derive the conversion rate from the barotropic field (W) alone, one needs to know the internal tide as well. This means that the conversion rate at any position is not only determined by the local barotropic forcing, but also by the internal tide generated there as well as by internal tides passing that position but generated elsewhere.

The resulting field is shown in Figure 7.17 (right panel), representing the cross-slope baroclinic velocity. A clear beam is present, emanating from the slope, propagating downward, and then reflecting from the bottom. The circles from the observations shown in Figure 1.9 have been inserted for comparison.

We note that these model results were obtained for winter stratification, lacking the seasonal thermocline. The beam is generated too deeply to be influenced by the thermocline (which occupies a layer from about 50 to 150 meters beneath the surface), but after bottom reflection the beam returns to the surface, and would, in summer, cross the thermocline. This gives rise to an interesting phenomenon discussed in Section 8.5.2.

Appendix A: Integral expressions I

Here we list the identities used in Section 7.3 to solve the internal-tide generation problem over a step-topography. To obtain (7.14), we used

$$\begin{aligned}
\int_{-H}^{0} dz\, z \sin\Big(\frac{m\pi z}{H}\Big) &= -H^2 \frac{(-1)^m}{m\pi} \\
\int_{-H}^{0} dz\, \sin\Big(\frac{n\pi z}{H}\Big) \sin\Big(\frac{m\pi z}{H}\Big) &= \frac{H}{2}\,\delta_{nm} \\
\int_{-H}^{-H_s} dz\, \sin\Big(\frac{m\pi z}{H}\Big) &= \frac{H}{m\pi}\Big[(-1)^m - \cos(m\pi\alpha)\Big] \\
\int_{-H_s}^{0} dz\, z \sin\Big(\frac{m\pi z}{H}\Big) &= \Big(\frac{H}{m\pi}\Big)^2\Big[\sin(m\pi\alpha) - m\pi\alpha\cos(m\pi\alpha)\Big] \\
\int_{-H_s}^{0} dz\, \sin\Big(\frac{n\pi z}{H_s}\Big) \sin\Big(\frac{m\pi z}{H}\Big) &= \frac{H_s n}{\pi}\,\frac{(-1)^n \sin(m\pi\alpha)}{(m\alpha)^2 - n^2}\,.
\end{aligned}$$

and to obtain (7.15),

$$\begin{aligned}
\int_{-H_s}^{0} dz\, \sin\Big(\frac{n\pi z}{H}\Big) \sin\Big(\frac{m\pi z}{H_s}\Big) &= \frac{H_s m}{\pi}\,\frac{(-1)^m \sin(n\pi\alpha)}{(n\alpha)^2 - m^2} \\
\int_{-H_s}^{0} dz\, \sin\Big(\frac{n\pi z}{H_s}\Big) \sin\Big(\frac{m\pi z}{H_s}\Big) &= \frac{H_s}{2}\,\delta_{nm}\,.
\end{aligned}$$

Appendix B: Integral expressions II

The solution (7.26) contains the integral expressions $A_n(x)$ and $B_n(x)$. For the seamount given by (7.27), they read:

$$A_n = r_0 \begin{cases} 0 & \text{for } x < -L \\ \frac{\pi^2 \exp(ik_nL) + [i\pi k_n L \sin(\pi x/L) + \pi^2 \cos(\pi x/L)]\exp(-ik_n x)}{2[\pi^2 - (k_nL)^2]} & \text{for } -L < x < L \\ \frac{i\pi^2 \sin(k_nL)}{\pi^2 - (k_nL)^2} & \text{for } x > L\,, \end{cases}$$

and

$$B_n = r_0 \begin{cases} -\frac{i\pi^2 \sin(k_nL)}{\pi^2 - (k_nL)^2} & \text{for } x < -L \\ \frac{-\pi^2 \exp(ik_nL) + [i\pi k_n L \sin(\pi x/L) - \pi^2 \cos(\pi x/L)]\exp(ik_n x)}{2[\pi^2 - (k_nL)^2]} & \text{for } -L < x < L \\ 0 & \text{for } x > L\,. \end{cases}$$

In Sections 7.4.2 and 7.4.3 we use the three-layer model, in which the middle layer (describing the thermocline) is in fact reduced to an interface. We provide here the technical details of the calculation of d_n for these cases. Recall that d_n is given by (7.24):

$$d_n = \frac{\int_{-H}^{0} dz\, z\, \phi_n''}{\int_{-H}^{0} dz\, \phi_n'' \phi_n} \,. \tag{7.36}$$

The function ϕ_n is given by (7.33); ϕ_n itself is continuous, but its derivative is not: at the thermocline ($z = -d$) it jumps from one value to another. As a consequence, its second derivative will contain a delta-distribution. This is also evident from (7.22), since N^2 can here be written as

$$N^2(z) = g'\delta(z+d) + N_c^2 \Theta(-z-d) \,.$$

So, the integrals that feature in (7.36) can be split up into three parts: a term due to the presence of the thermocline (concentrated at $z = -d$), and integrals over the upper and lower layers ($\int_{-d}^{0}$ and $\int_{-H}^{-d}$, respectively). The latter are given by

$$\begin{aligned}
I_1 &= \int_{-d}^{0} dz\, z\, \phi_n'' = c q_u^2 \int_{-d}^{0} dz\, z\, \sinh(q_u z) \\
&= c[q_u d \cosh(q_u d) - \sinh(q_u d)] \,, \\
I_2 &= \int_{-H}^{-d} dz\, z\, \phi_n'' = -q_l^2 \int_{-H}^{-d} dz\, z\, \sin q_l(z+H) \\
&= q_l H - \sin q_l(H-d) - q_l d \cos q_l(H-d) \,, \\
I_3 &= \int_{-d}^{0} dz\, \phi_n'' \phi_n = c^2 q_u^2 \int_{-d}^{0} dz\, \sinh^2(q_u z) \\
&= \tfrac{1}{4} c^2 q_u [\sinh(2q_u d) - 2q_u d] \,, \\
I_4 &= \int_{-H}^{-d} dz\, \phi_n'' \phi_n = -q_l^2 \int_{-H}^{-d} dz\, \sin^2 q_l(z+H) \\
&= \tfrac{1}{4} q_l [\sin 2q_l(H-d) - 2q_l(H-d)] \,,
\end{aligned}$$

with

$$c = -\frac{\sin q_l(H-d)}{\sinh q_u d} \,.$$

The coefficient d_n is thus given by

$$d_n = \frac{k_n^2 g' d (\omega^2 - f^2)^{-1} \sin q_l(H-d) + I_1 + I_2}{-k_n^2 g' (\omega^2 - f^2)^{-1} \sin^2 q_l(H-d) + I_3 + I_4} \,, \tag{7.37}$$

where the first term in the numerator and denominator represents the effect of the thermocline.

These expressions are to be modified for the case examined in Section 7.4.3; there the stratification in the lower layer is removed ($N_c = 0$), making ϕ_n hyperbolic in the lower layer, as expressed in (7.35). Replacing q_l by iq_u in (7.37), including the integrals $I_{1,2,3,4}$, we obtain a complex d_n; multiplying it by i then yields the proper expression for the interfacial case.

Chapter 8

Internal solitons

In previous chapters we considered the linear equations, which can be assumed to be valid as long as the amplitude of the waves is small (Section 4.5). This assumption is however not always satisfied in observed internal waves. In this chapter, we turn to a particular class of waves in which nonlinearity is essential, the so-called internal solitons. They consist of a single depression (or elevation), and retain their form while propagating, and even when interacting with other solitons. They are often observed in the oceans and atmosphere.

Linear problems are relatively easy to solve because the sum of any two solutions is itself a solution (the superposition of modes in Chapter 5 is a case in point); thus, one solution engenders an infinite number of solutions. In nonlinear problems, this convenient property holds no longer. However, soliton equations are special in that a method has been developed to obtain the evolution of a given initial profile by analytical means. This is discussed in this chapter for the most well-known soliton equation, the Korteweg-de Vries equation, which is widely applied to describe oceanic internal solitons.

8.1 Observations

One of many observations of solitons was made in the Andaman Sea [64]. A single soliton is shown in Figure 8.1 (left panel). Its largest vertical excursions occur in the thermocline, which here lies between 50 and 200 m depth. The depression forms, in fact, part of a train of solitons (right panel); an increase in temperature at a given depth means that isotherms undergo a depression, in agreement with the panel on the left. One noticeable property is the amplitude-ordering of the train: the largest peak arrives first, and each subsequent peak is smaller than its predecessor. A surface manifestation of a group of solitons in the Andaman Sea is shown in Figure 1.1. Further observations (not shown here) reveal that such groups appear every tidal period, suggesting a relation to the internal tide, which is further discussed in Section 8.5.

In this chapter we focus on solitons in the ocean, but spectacular manifesta-

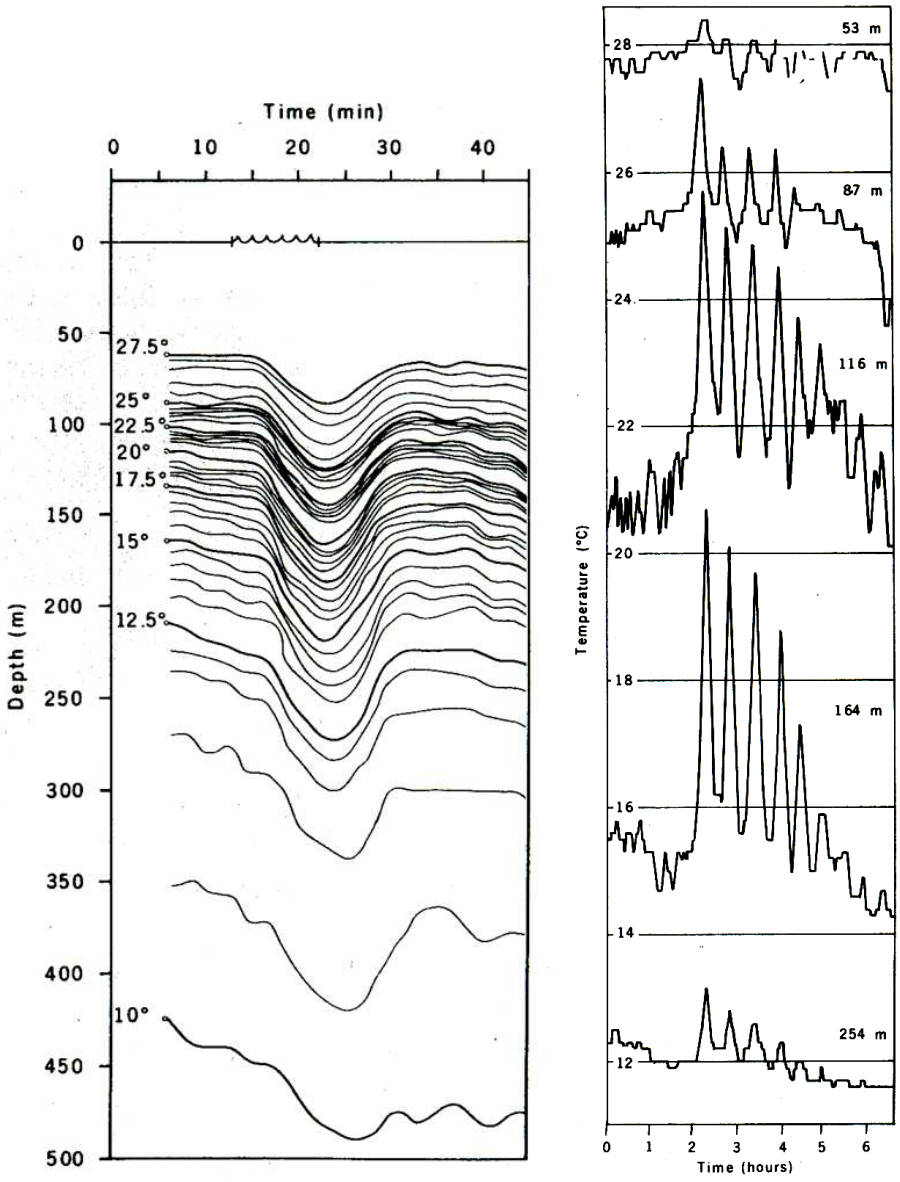

Fig. 8.1: Observations of solitons in the Andaman Sea. *Left:* Isothermal displacements showing a large depression, with vertical excursions of more than 60 m in the course of just ten minutes. Levels of isotherms were here determined by using expendable bathythermographs (XBTs), giving a series of vertical temperature profiles, from which follows the temporal evolution of selected isotherms. *Right:* A train of solitons; temperature was measured continuously at certain fixed depths: **a** 53, **b** 87, **c** 116, **d** 164, and **e** 254 m. The total water depth at this location is about 1100 m. From [64].

tions have also been observed in the atmosphere. In Australia, this phenomenon is known as 'Morning Glory' (Figures 8.2 and 8.3). In the atmosphere, solitons (and internal waves, in general) are rendered visible by adiabatic cooling of rising parcels, leading to condensation of water vapour in the crests of the waves, and hence cloud formation.

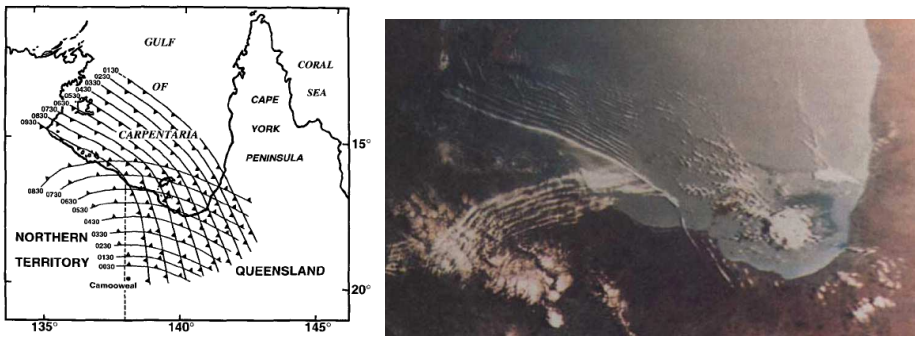

Fig. 8.2: Interacting 'Morning Glories' over Northern Australia. *Left:* Map of the location and sketch of the passage in time of two trains of solitary waves, coming from northeastern and southern directions. *Right:* NOAA-12AVHRR satellite imagery at the early morning of 3 October 1991, showing two interacting trains. From [71].

In the atmosphere, two main types of solitons can be distinguished [73]. The first type is confined to the lower few kilometers of the troposphere, and has horizontal length scales of 100 m to a few kilometers; the 'Morning Glory' solitons belong to this category. They may be generated by, for example, gravity currents or katabatic winds. The second type extends over the entire troposphere and has a much larger horizontal scale, of the order of 100 km; proposed generation mechanisms are shear instability or geostrophic adjustment of large-scale frontal systems. The vertical confinement of solitons, either to the lower few kilometers or to the troposphere as a whole, invites the question as to what causes this confinement. The atmosphere, unlike the ocean, has of course no well-defined upper boundary. Mechanisms that may nevertheless impose a waveguide are a neutrally stratified upper layer, or a very strongly stratified upper layer (causing internal reflections), as discussed in Sections 5.3 and 5.4; shear layers may similarly cause reflections (from so-called critical levels). For a more extensive discussion on the two types of solitons, we refer to [73].

8.2 Korteweg-de Vries (KdV) equation

One of the most well-known equations admitting solitons is the Korteweg-de Vries equation, which was derived at the end of the 19th century, but received little attention until remarkable properties were discovered numerically some sixty years later. It was found that localized solutions exist which regain their form after interactions with similar waves; to emphasize their robust character,

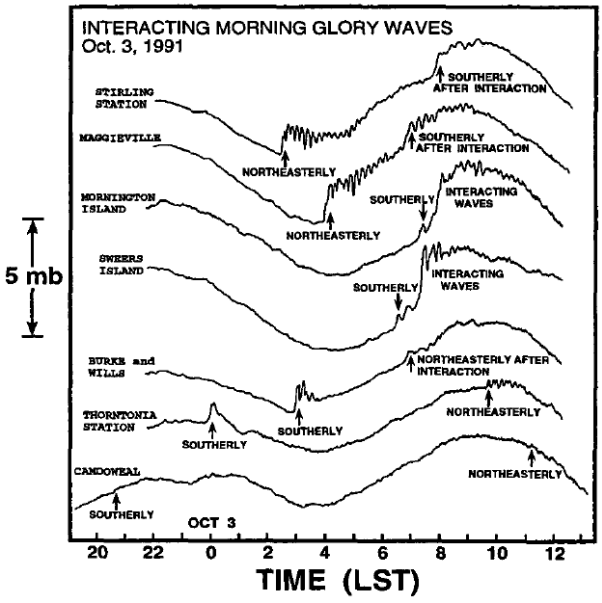

Fig. 8.3: Surface pressure signatures of interacting Morning Glory waves on 3 October 1991 from an array of portable microbarograph stations deployed over the southern Gulf of Carpentaria region of Northern Queensland. The passage of a soliton is marked by a change in pressure of about 1 mb. ('LST' stands for Australian Eastern Standard time, i.e. UTC plus 10 h.) From [71].

the name 'soliton' was coined, the suffix "-on" giving it the flavour of an elementary particle [91]. A second discovery, made soon after, concerned the evolution of a given initial profile; it turned out that this problem can be reduced to solving two *linear* equations [59], as is further discussed in Section 8.4.[1] The equation surfaces in many branches of physics, such as fluid mechanics, plasma physics, and quantum gravity [9].

The KdV equation contains two effects that together allow for the existence of a soliton solution: nonlinearity and dispersion. We first discuss each of them separately.

8.2.1 Effect of nonlinearity

The simplest equation describing the propagation of nonlinear waves is

$$\frac{\partial \eta}{\partial t} + \frac{\partial \eta}{\partial x} + \epsilon\, \eta \frac{\partial \eta}{\partial x} \quad = \quad 0\,. \tag{8.1}$$

[1] For more on the history of the KdV equation, see [86, 57]. In fact, the history of the KdV equation (and the question how the paper by Korteweg and De Vries relates to previous works, notably by Boussinesq) seems to get more intricate as more documents, among them notes by De Vries, are unearthed; for a recent overview, see [11].

Here we have some dimensionless parameter ϵ (the precise form of which is immaterial in the present discussion) serving as a coefficient of the nonlinear term. The field η may describe, for example, an isopycnal excursion.

In implicit form, the general solution of (8.1) can be expressed in terms of an arbitrary function F:

$$\eta(x,t) \quad = \quad F(\xi)\,, \qquad \text{with} \quad \xi = x - (1+\epsilon\eta(x,t))\,t\,. \tag{8.2}$$

It is easily verified that this is indeed a solution; differentiation of (8.2) gives

$$\frac{\partial\eta}{\partial t} = \frac{\partial\xi}{\partial t}\,F' = -\Big(1+\epsilon\eta+\epsilon t\frac{\partial\eta}{\partial t}\Big)\,F'\,, \quad \frac{\partial\eta}{\partial x} = \frac{\partial\xi}{\partial x}\,F' = \Big(1-\epsilon t\frac{\partial\eta}{\partial x}\Big)\,F'\,,$$

with $F' = dF/d\xi$. Hence

$$\begin{aligned} \frac{\partial\eta}{\partial t} &= -\frac{(1+\epsilon\eta)F'}{1+\epsilon tF'} \\ \frac{\partial\eta}{\partial x} &= \frac{F'}{1+\epsilon tF'}\,. \end{aligned} \tag{8.3}$$

Substitution now demonstrates that (8.1) is satisfied.

We have here introduced a 'characteristic' coordinate ξ, which in a mathematical sense is analogous to the characteristics of Chapter 6. Physically, the meaning is of course very different, for in Chapter 6 only spatial coordinates were involved in the transformation, whereas $\xi = const$ here represents *a path in x,t-space, on which η is constant.* For constant $\xi = x_0$, we have the characteristic

$$x \quad = \quad x_0 + (1+\epsilon\eta(x,t))\,t\,. \tag{8.4}$$

On it, we must have, according to (8.2),

$$\eta(x,t) = F(x_0) = \text{constant}\,.$$

Hence (8.4) describes a straight line,

$$x \quad = \quad x_0 + (1+\epsilon F(x_0))\,t\,. \tag{8.5}$$

This expression indicates where, at a certain moment t, the value $\eta = F(x_0)$ is to be found. It moreover shows that higher elevations (i.e. larger $F(x_0)$) propagate faster. An example is shown in Figure 8.4. On the left we have the line

$$x \quad = \quad x_A + (1+\epsilon F(x_A))\,t\,,$$

and the other lines are similarly defined. (Notice that time is shown on the vertical axis in Figure 8.4.) The line in the middle, starting at $x = 0$ shows the location of the wave crest, $\eta = F(x_B)$, as time passes. The crest propagates faster than the neighbouring elevations $F(x_A)$ and $F(x_C)$; in other words, the

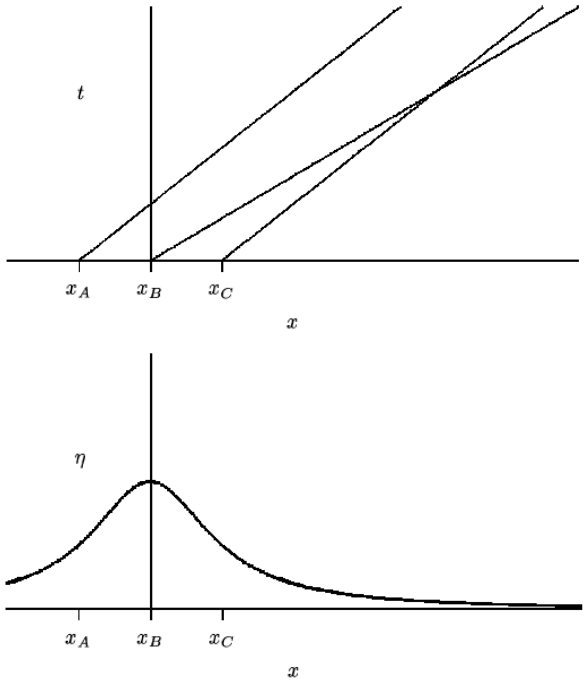

Fig. 8.4: *Lower panel:* An initial profile of η. *Upper panel:* Characteristics indicating where, at time t, the elevation F is located that was initially at x_A, x_B, or x_C. The intersection of lines is indicative of a multi-valued η at that point in t, x-space, and signals wave breaking.

phase speed depends on the elevation F. At a certain moment the lines B and C cross; this is the moment of *wave breaking*.

This wave breaking is clearly a result of the nonlinear term, since the characteristics would all be equally steep if $\epsilon = 0$. The moment of breaking can also derived directly from (8.3). If the wave breaks, $\partial\eta/\partial x$ becomes infinite, and this happens at

$$t = -\frac{1}{\epsilon F'} = \mathcal{O}(\epsilon^{-1})\,.$$

(The minus sign indicates that a rightward propagating wave of elevation will break on its descending side, i.e. where F' is negative.) This expression reveals also that a *time scale*, of order $1/\epsilon$, is associated with wave breaking. So, no matter how small ϵ, the nonlinear term will become important after a sufficiently long time. In other words, for the linear approximation to be valid, smallness of ϵ is not sufficient; a limit is posed as well on the stretch of time that one may consider.

8.2.2 Effect of dispersion

In (5.46), we found that long interfacial waves in a rotationless system ($f = 0$) satisfy the dispersion relation

$$\omega^2 \quad = \quad c_0^2 k^2 \left[1 - \tfrac{1}{3} d(H-d)k^2 + \cdots\right] \tag{8.6}$$

with

$$c_0^2 = g' \, \frac{d(H-d)}{H} \, . \tag{8.7}$$

Here H is the total water depth, d the thickness of the upper layer, and g' reduced gravity, measuring the strength of the thermocline, here represented by an interface. If we restrict ourselves to rightward propagating waves, we can approximate (8.6) by

$$\omega \quad = \quad c_0 k \left[1 - \tfrac{1}{6} d(H-d)k^2 + \cdots\right] . \tag{8.8}$$

From this expression we see that interfacial are dispersive, i.e. the phase speed

$$c(k) = \frac{\omega}{k} = c_0 \left[1 - \tfrac{1}{6} d(H-d)k^2 + \cdots\right] \tag{8.9}$$

varies with wavenumber k. This type of dispersion may be termed *non-hydrostatic* dispersion, since it disappears in the hydrostatic (or long-wave limit, or shallow-water) limit, i.e. if kd, $k(H-d) \to 0$.

A certain initial profile, such as in Figure 8.4 (upper panel), can be described in terms of a Fourier integral,

$$\eta(x,0) \quad = \quad \int_{-\infty}^{\infty} dk \, F(k) \, e^{ikx} \, ,$$

for a certain function F. The profile evolves in time according to

$$\begin{aligned} \eta(x,t) \quad &= \quad \int_{-\infty}^{\infty} dk \, F(k) \, e^{i(kx - \omega(k)t)} \\ &= \quad \int_{-\infty}^{\infty} dk \, F(k) \, e^{ik(x - c(k)t)} \, . \end{aligned}$$

Because of the dispersive character of (8.9), different Fourier components will travel at different speeds. As a result, the the initial profile will lose its form as it falls apart in its Fourier components.

It is easy to verify by substitution of the Fourier integral, that the evolution equation associated with the dispersion relation (8.8) reads

$$\frac{\partial \eta}{\partial t} + c_0 \, \frac{\partial \eta}{\partial x} + \frac{1}{6} \, c_0 d(H-d) \, \frac{\partial^3 \eta}{\partial x^3} \quad = \quad 0 \, . \tag{8.10}$$

The third term represents the effect of non-hydrostatic dispersion.

8.2.3 KdV for interfacial waves

From the previous two sections it is clear that

- non-linearity alone will make a solitary wave profile steepen and break;
- non-hydrostatic dispersion alone will make a solitary wave profile fall apart in its Fourier components.

So, neither of the effects, taken in isolation, can sustain a solitary wave. The crux of the KdV equation is that a *combination* of the effects can. Intuitively this can be understood as follows: nonlinearity, by its steepening effect, tends to drive energy to a short scale; dispersion, on the other hand, tends to spread it out. In a sense, then, the effects are opposed, and the KdV equation demonstrates that they can balance to produce form-preserving waves.

For interfacial waves, described by η, the displacement of the interface from its level of rest, the KdV equation reads

$$\boxed{\frac{\partial \eta}{\partial t} + c_0 \frac{\partial \eta}{\partial x} + \frac{3}{2} \frac{h_1 - h_2}{h_1 h_2} c_0 \eta \frac{\partial \eta}{\partial x} + \frac{1}{6} c_0 h_1 h_2 \frac{\partial^3 \eta}{\partial x^3} = 0 \,.} \tag{8.11}$$

Here we write the thickness of the upper layer as $h_1 = d$, and that of the lower layer as $h_2 = H - d$. The linear long-wave speed for interfacial waves, c_0, is again defined by (8.7), which in terms of $h_{1,2}$ reads

$$c_0^2 = g' \frac{h_1 h_2}{h_1 + h_2} \,. \tag{8.12}$$

The effects of nonlinearity and non-hydrostatic dispersion, discussed in the previous sections, recur in the second and third term of (8.11), respectively.

This form of the KdV equation will be derived in Section 8.3. The essential underlying assumption is that nonlinear and non-hydrostatic effects are weak and of the same order of strength. The derivation then proceeds in two steps; at lowest order, a linear hydrostatic equation is derived, consisting of the first two terms of (8.11). Then, at the next order, nonlinear and non-hydrostatic effects appear, completing (8.11).

From this outline of the derivation it will already be clear that the KdV equation is not exact. It is sometimes indeed necessary to proceed to a higher order. A configuration in which $h_1 = h_2$ (i.e. the upper and lower layer are of equal thickness) is a case in point: the nonlinear term in (8.11) then vanishes altogether. A higher-order extension is then needed (Section 8.6).

8.2.4 A heuristic 'derivation'

If we cast the linear dispersive equation (8.10) in terms of the parameters $h_1 = d$ and $h_2 = H - d$, we obtain

$$\frac{\partial \eta}{\partial t} + c_0 \frac{\partial \eta}{\partial x} + \frac{1}{6} c_0 h_1 h_2 \frac{\partial^3 \eta}{\partial x^3} \quad = \quad 0 \,, \tag{8.13}$$

which is (8.11), except for the nonlinear term. There is a simple heuristic way to obtain the correct form of this term from (8.13), apart from a numerical factor. So, although the equation thus obtained will not be entirely correct, the reasoning leading to it gives some more insight in the nature of the nonlinear term.

The starting point is the term $c_0 \partial\eta/\partial x$, in which c_0 is given by (8.12),

$$c_0^2 = g' \, \frac{h_1 h_2}{h_1 + h_2} \, .$$

Here h_1 and h_2 are the thickness of the upper and lower layer, respectively. The important thing to realize is that if waves attain an appreciable amplitude η, the actual thickness of the upper layer will not be h_1 but $h_1 - \eta$. Similarly, the thickness of the lower layer becomes $h_2 + \eta$. If we substitute this in the expression for c_0, we get

$$\begin{aligned} &\Big[g' \, \frac{(h_1 - \eta)(h_2 + \eta)}{h_1 + h_2} \Big]^{1/2} = \Big[g' \, \frac{h_1 h_2 + \eta(h_1 - h_2) - \eta^2}{h_1 + h_2} \Big]^{1/2} \\ &= c_0 \Big[1 + \eta \frac{h_1 - h_2}{h_1 h_2} - \eta^2 \frac{1}{h_1 h_2} \Big]^{1/2} = c_0 \Big[1 + \eta \frac{h_1 - h_2}{2 h_1 h_2} + \cdots \Big] \, , \end{aligned}$$

where we neglected quadratic terms in η. If replace $c_0 \partial\eta/\partial x$ in (8.13) with the last expression, we obtain the nonlinear term

$$c_0 \, \frac{h_1 - h_2}{2 h_1 h_2} \, \eta \, \frac{\partial \eta}{\partial x} \, ,$$

which is of the same form as the nonlinear term in (8.11), but lacks the factor three.

This 'derivation' makes clear why the nonlinear term cancels if $h_1 = h_2$; the increase of the phase speed by the enlargement of one layer is then precisely compensated by the decrease of the phase speed due to the thinner other layer.

8.2.5 Soliton solution

In short-hand notation, we write (8.11) as

$$\frac{\partial \eta}{\partial t} + c_0 \, \frac{\partial \eta}{\partial x} + A \eta \, \frac{\partial \eta}{\partial x} + B \, \frac{\partial^3 \eta}{\partial x^3} \quad = \quad 0 \, . \tag{8.14}$$

We seek form-preserving solutions,

$$\eta(x,t) = F(\xi) \, , \qquad \text{with} \quad \xi = x - Ct \, .$$

The function F and phase speed C are to be determined. We denote the derivative of F by a prime, $F' = dF/d\xi$. The derivatives of η can be written

$$\frac{\partial \eta}{\partial t} = -C F' \, , \qquad \frac{\partial \eta}{\partial x} = F' \, .$$

Substitution in (8.14) gives

$$-(C-c_0)F' + AFF' + BF''' \;=\; 0\,.$$

The second term can be written as $\frac{1}{2}A(F^2)'$, so that we can integrate the equation to obtain

$$-(C-c_0)F + \tfrac{1}{2}AF^2 + BF'' \;=\; 0\,.$$

Here we have taken the constant of integration equal to zero, as we are looking for solitary-wave solutionsz, for which F and derivatives vanish as $|\xi| \to \infty$. Multiplying by F' now gives

$$-(C-c_0)FF' + \tfrac{1}{2}AF^2F' + BF''F' \;=\; 0\,,$$

or, equivalently,

$$-\tfrac{1}{2}(C-c_0)(F^2)' + \tfrac{1}{6}A(F^3)' + \tfrac{1}{2}B(F'^2)' \;=\; 0\,.$$

After one more integration (and again setting the constant of integration equal to zero), we obtain

$$-(C-c_0)F^2 + \tfrac{1}{3}AF^3 + B(F')^2 \;=\; 0\,.$$

Taking the square root,

$$B^{1/2}\,F' \;=\; [(C-c_0)F^2 - \tfrac{1}{3}AF^3]^{1/2}\,,$$

so that

$$\begin{aligned} d\xi &= \frac{B^{1/2}\,dF}{[(C-c_0)F^2 - \frac{1}{3}AF^3]^{1/2}} \\ &= \frac{B^{1/2}\,dF}{(C-c_0)^{1/2}\,F\,[1-\frac{A}{3(C-c_0)}F]^{1/2}}\,. \end{aligned} \tag{8.15}$$

The second square root can be removed via the transformation

$$F \;=\; \frac{3(C-c_0)}{A}\,\mathrm{sech}^{\,2}q\,, \tag{8.16}$$

and using the identity

$$\tanh^2 q = 1 - \mathrm{sech}^{\,2}q\,.$$

Using moreover the identity

$$\frac{d}{dq}\,\mathrm{sech}\,q = -\mathrm{sech}\,q\tanh q\,,$$

we have

$$dF = -2F\tanh q\;dq\,,$$

so that (8.15) reduces to

$$d\xi \;=\; -2\,dq\,\Big[\frac{B}{C-c_0}\Big]^{1/2},$$

or integrated,

$$q \;=\; -\xi\,\Big[\frac{C-c_0}{4B}\Big]^{1/2},$$

where we left out the constant of integration, which would give a mere phase shift. Finally, substitution in (8.16) gives

$$F \;=\; \frac{3(C-c_0)}{A}\,\mathrm{sech}^{\,2}\left(\xi\Big[\frac{C-c_0}{4B}\Big]^{1/2}\right).$$

This is the soliton solution. It can be more conveniently written in terms of the amplitude and length of the wave,

$$a=\frac{3(C-c_0)}{A} \qquad\qquad l=\Big[\frac{4B}{C-c_0}\Big]^{1/2}.$$

With this, the soliton solution becomes

$$\boxed{\eta = a\,\mathrm{sech}^{\,2}\Big(\frac{x-Ct}{l}\Big),} \tag{8.17}$$

where

$$C=c_0+\frac{aA}{3}, \qquad l=\Big(\frac{12B}{aA}\Big)^{1/2}.$$

Using the explicit form of the coefficients A and B for the two-layer system,

$$\boxed{C=c_0\Big[1+\tfrac{1}{2}\,a\,\frac{h_1-h_2}{h_1h_2}\Big], \qquad l=\frac{2h_1h_2}{[3a(h_1-h_2)]^{1/2}}\,.} \tag{8.18}$$

An example of a soliton profile is shown in Figure 8.5.

Eq. (8.18) demonstrates that the phase speed C (as well as the length l) depend on the wave *amplitude*; this is a manifestation of the nonlinear character of the soliton. It furthermore follows from the second expression in (8.18) that

$$\boxed{a\,(h_1-h_2)>0\,,} \tag{8.19}$$

since the argument of the square root must be positive. From (8.18) and (8.19) follow four important properties of KdV solitons:

1. The phase speed of the soliton, C, *exceeds* the linear long-wave phase speed c_0;

2. Eq. (8.19) implies that the amplitude a must be negative if $h_1 < h_2$ (i.e. if the upper layer is relatively thin, as is normally the case in the ocean); the soliton then manifests itself as a *depression*, as in Figure 8.1 (left panel);

3. Larger solitons (i.e. larger $|a|$) propagate faster than smaller ones. This means that a train of solitons emanating from the same source will be *ordered in amplitude*, the largest leading the group. We recall that this was indeed observed in Figure 8.1 (right panel);

4. Larger solitons are shorter than smaller ones, since l is inversely proportional to $a^{1/2}$; an example is shown in Figure 8.5.

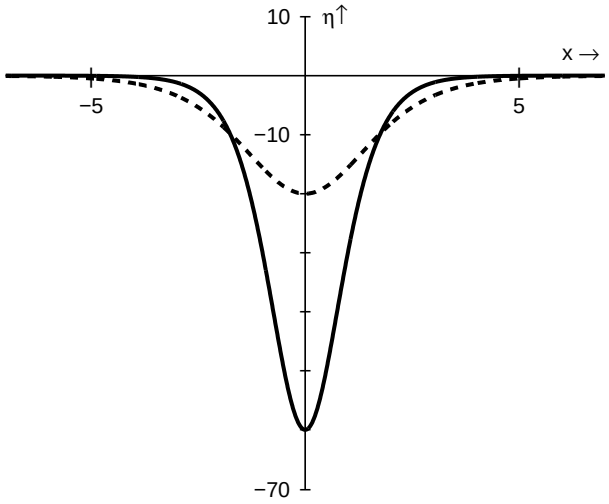

Fig. 8.5: Examples of the soliton solution (8.17), with different amplitudes: $a = -60\,\mathrm{m}$ (solid line) and $a = -20\,\mathrm{m}$ (dashed line). For both, the parameters are $g' = 0.01\,\mathrm{m\,s^{-2}}$, $h_1 = 200\,\mathrm{m}$, $h_2 = 1200\,\mathrm{m}$. The phase speeds C are $1.47\,\mathrm{m\,s^{-1}}$ (solid line) and $1.37\,\mathrm{m\,s^{-1}}$ (dashed line); for comparison, the linear long-wave phase speed is $c_0 = 1.31\,\mathrm{m\,s^{-1}}$. On the horizontal axis, distances are in kilometers; on the vertical, in meters.

Apart from a single-soliton solution, the KdV equation also admits multi-soliton solutions, involving solitons of different amplitudes. We will not dwell on this, but remark that they pass each other (since the phase speed depends on the amplitude), yet come out of the interaction with unchanged shape. During the interaction, they do of course not superpose in a linear way; in fact, the total amplitude during interaction is less than the sum of their amplitudes before (or after) the interaction.

8.3 Derivation of the KdV equation

8.3.1 Basic equations

Our starting point is the set (4.17), where nonlinear terms are retained,

$$\frac{D\vec{u}}{Dt} + 2\vec{\Omega} \times \vec{u} = -\frac{1}{\rho_*}\nabla p' + b\hat{z} \tag{8.20a}$$

$$\nabla \cdot \vec{u} = 0 \tag{8.20b}$$

$$\frac{Db}{Dt} + N^2 w = 0\,. \tag{8.20c}$$

We simplify the problem by neglecting Coriolis effects altogether (a short discussion on Coriolis effects and solitons follows in Section 8.5); moreover, we assume $\partial/\partial y = 0$. In the two-layer system, the upper and lower layer are each neutrally stratified ($N = 0$), the stratification being concentrated as a δ-peak at the interface that separates the layers (see Section 5.4.3). So, if we consider the upper and lower layers separately, we have $N = 0$ in each, and hence $b = 0$, so that (8.20) becomes

$$u_{i,t} + u_i u_{i,x} + w_i u_{i,z} = -\frac{1}{\rho_*}\, p'_{i,x} \tag{8.21}$$

$$w_{i,t} + u_i w_{i,x} + w_i w_{i,z} = -\frac{1}{\rho_*}\, p'_{i,z} \tag{8.22}$$

$$u_{i,x} + w_{i,z} = 0\,. \tag{8.23}$$

Here index i takes the value 1 for the upper layer, and 2 for the lower layer.

We choose the z-axis such that the level of rest of the interface lies at $z = 0$; the upper surface (rigid-lid) lies at $z = h_1$, the bottom at $z = -h_2$. We introduce a new notation for these levels, which proves convenient in later scaled forms: $h_1 = \alpha H$ and $h_2 = (1 - \alpha)H$, where $H = h_1 + h_2$ is the total water depth (Figure 8.6).

The the upper surface (rigid-lid) and bottom the vertical velocity w must vanish,

$$w_1|_{z=\alpha H} = 0 \tag{8.24}$$

$$w_2|_{z=(\alpha-1)H} = 0\,. \tag{8.25}$$

At the interface, $z = \eta(t, x)$, we find by taking D/Dt,

$$w_i|_{z=\eta} = \eta_t + \eta_x u_i|_{z=\eta} \qquad (i = 1, 2)\,. \tag{8.26}$$

We have yet to to take into account the presence of a density jump across the interface, which produces the delta-peak in N^2. This is done by requiring continuity of *total* pressure, for which we have to return to (4.7),

$$p = p_0(z) + p'(t, \vec{x})\,, \tag{8.27}$$

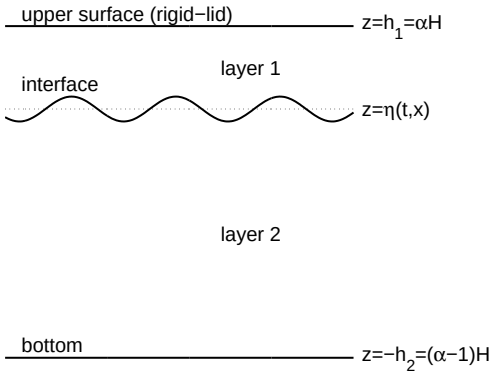

Fig. 8.6: The two-layer system for which the KdV equation is derived. Both upper and lower layer are neutrally stratified, $N = 0$. Across the interface there is a jump in density, giving a delta-peak $N^2 = g'\delta(z)$. The horizontal dotted line indicates the level $z = 0$, the level at which the interface resides at rest.

in which p_0 is given by the hydrostatic balance (4.9),

$$\frac{dp_0}{dz} \quad = \quad -\rho_0 g\,. \tag{8.28}$$

To obtain p_0 we need to know the static density profile ρ_0. In a two-layer system, the densities in each layer are commonly assumed to be constant. This is however in contradiction with the assumption that $N = 0$ in each layer, as is easily seen from (4.19),

$$N^2 = -\frac{g}{\rho_*}\left(\frac{d\rho_0}{dz} + \frac{\rho_0 g}{c_s^2}\right). \tag{8.29}$$

Assuming ρ_0 to be constant produces a layer in which $N^2 < 0$, i.e. an *unstably* stratified layer. So, if we stick to the assumption $N = 0$, we cannot also assume that ρ_0 is constant.

We thus proceed from (8.28), integrating for each layer to obtain the hydrostatic pressure:

$$p_{0,1}(z) = g\int_z^{h_1} d\zeta\,\rho_{0,1}(\zeta)\,, \qquad p_{0,2}(z) = g\int_0^{h_1} d\zeta\,\rho_{0,1}(\zeta) + g\int_z^0 d\zeta\,\rho_{0,2}(\zeta)\,.$$

Requiring now the total pressure, $p_{0,i} + p_i'$ to be continuous at the interface, gives

$$p_2'|_{z=\eta} - p_1'|_{z=\eta} = g\int_0^{\eta} d\zeta\,\rho_{0,2}(\zeta) - g\int_0^{\eta} d\zeta\,\rho_{0,1}(\zeta)\,.$$

For small η, the right-hand side can be expanded as

$$g\int_0^\eta d\zeta\left[\rho_{0,2}(0)+\zeta\frac{d\rho_{0,2}}{dz}(0)+\cdots\right]-g\int_0^\eta d\zeta\left[\rho_{0,1}(0)+\zeta\frac{d\rho_{0,1}}{dz}(0)+\cdots\right]$$
$$g\Big[\eta\rho_{0,2}(0)+\tfrac{1}{2}\eta^2\frac{d\rho_{0,2}}{dz}(0)+\cdots\Big]-g\Big[\eta\rho_{0,1}(0)+\tfrac{1}{2}\eta^2\frac{d\rho_{0,1}}{dz}(0)+\cdots\Big]$$
$$g\Big[\eta\rho_{0,2}(0)-\tfrac{1}{2}\eta^2\frac{\rho_{0,2}g}{c_{s,2}^2}+\cdots\Big]-g\Big[\eta\rho_{0,1}(0)-\tfrac{1}{2}\eta^2\frac{\rho_{0,1}g}{c_{s,1}^2}+\cdots\Big]\,.$$

In the last expression, we used the fact that $N^2=0$ within each layer, and (8.29). Now, the quadratic terms in square brackets scale with the linear terms as $\eta g/c_s^2$, which even for large solitons ($\eta\sim 100\,\mathrm{m}$) is very small, namely of order $O(10^{-3})$. We will therefore neglect them; we can thus write the condition for continuity of pressure as

$$p_2'|_{z=\eta}-p_1'|_{z=\eta}=(\rho_{0,2}(0)-\rho_{0,1}(0))g\eta\,. \tag{8.30}$$

(Hereafter we will drop the argument in $\rho_{0,i}$, it be understood that they are evaluated at the interface.) This condition states that baroclinic pressure is discontinuous across the interface, due to a jump in density. It is clear from (8.21) that this discontinuity will similarly create a discontinuity in the horizontal velocity u_i.

8.3.2 Scaling and small parameters

For u_i we introduce the scale $[u_i]=\epsilon c$, where c is a measure of the phase speed of the wave (to be specified below). Here ϵ acts as the *parameter of nonlinearity*, in accordance with the argument presented in Section 4.5, according to which the ratio of particle velocity and phase speed is a measure of the strength of nonlinearity. Time and length scales of the waves are denoted by $[t]=T$ and $[x]=cT=L$. We scale the vertical with H, the water depth.

As we assume nonlinear terms to be weak, the dominant balance in (8.21) will be between the linear acceleration term and the pressure gradient, giving as a scale for pressure: $[p_i']=\rho_*\epsilon c^2$. The scale for w_i follows from (8.23) as $[w_i]=\epsilon cH/L$, and hence from (8.26), $[\eta]=\epsilon H$. This provides us with an alternative interpretation of the parameter of nonlinearity, namely as the ratio of the wave amplitude and water depth.

Using these scales, we can cast the equations (8.21)–(8.23) in a nondimensional form,

$$u_{i,t}+\epsilon u_iu_{i,x}+\epsilon w_iu_{i,z} = -p_{i,x}' \tag{8.31}$$
$$\delta(w_{i,t}+\epsilon u_iw_{i,x}+\epsilon w_iw_{i,z}) = -p_{i,z}' \tag{8.32}$$
$$u_{i,x}+w_{i,z} = 0\,, \tag{8.33}$$

where we used the same symbols as for the dimensional variables, for convenience. (No confusion will arise from this as we will consider only the nondimen-

sional form in the remainder of this section.) We introduced a new parameter

$$\delta = \left(\frac{H}{L}\right)^2,$$

which is a measure of the strength of *non-hydrostatic dispersion*; the limit $\delta \to 0$ corresponds to the hydrostatic (or long-wave) limit, in which case the waves become dispersionless (Section 8.2.2).

The boundary conditions (8.24)-(8.26) and (8.30) become, in nondimensional form,

$$\begin{aligned} w_1|_{z=\alpha} &= 0 && (8.34)\\ w_2|_{z=\alpha-1} &= 0 && (8.35)\\ w_i|_{z=\epsilon\eta} &= \eta_t + \epsilon\eta_x u_i|_{z=\epsilon\eta} && (8.36)\\ p_2'|_{z=\epsilon\eta} - p_1'|_{z=\epsilon\eta} &= \eta\,. && (8.37) \end{aligned}$$

Here we identified c, the measure of the phase speed, with $c^2 = g'H$, where g' is the reduced gravity $g' = g(\rho_{0,2} - \rho_{0,1})/\rho_*$; the last expression provides an interpretation of g' in terms of density, rather than in terms of N^2, as in Figure 5.11.

The crucial assumption, which leads to the KdV equation, is

$$\boxed{\epsilon = \mathcal{O}(\delta) \ll 1\,.} \qquad (8.38)$$

This means we consider weakly nonlinear, weakly non-hydrostatic effects, which are of similar strength. The last assumption means that we are, in essence, dealing with only *one* small parameter, and we can write

$$\delta = \epsilon\delta_*$$

with $\delta_* = \mathcal{O}(1)$.

The goal is now to reduce the set (8.31)–(8.37) to a single equation for the interfacial displacement η. We thus seek an equation of the form

$$\boxed{D_0(\eta) + \epsilon D_1(\eta) + \epsilon^2 D_2(\eta) + \cdots = 0} \qquad (8.39)$$

where D_i are differential operators. To obtain the KdV equation, it suffices to determine D_0 and D_1.

We develop the variables p_i', u_i, w_i and η in a series, in which ϵ serves as the small parameter; so, for example,

$$p_i' = {p_i}^{(0)} + \epsilon {p_i}^{(1)} + \epsilon^2 {p_i}^{(2)} + \cdots$$

with ${p_i}^{(n)} = \mathcal{O}(1)$.

As a preparation for the following sections, we develop the boundary conditions (8.36) and (8.37) in a Taylor series about $z = 0$; for the left-hand side of (8.36), this gives

$$\begin{aligned} w_i|_{z=\epsilon\eta} &= w_i|_{z=0} + \epsilon\eta w_{i,z}|_{z=0} + \mathcal{O}(\epsilon^2) \\ &= w_i|_{z=0} - \epsilon\eta u_{i,x}|_{z=0} + \mathcal{O}(\epsilon^2) \end{aligned}$$

(in the second equality we used (8.33)), so (8.36) becomes

$$w_i = \eta_t + \epsilon(\eta u_i)_x + \mathcal{O}(\epsilon^2) \qquad \text{at } z = 0\,. \tag{8.40}$$

Similarly, we obtain from (8.37):

$$\eta = (p_2' - p_1')|_{z=0} + \epsilon\eta(p_{2,z}' - p_{1,z}')|_{z=0} + \mathcal{O}(\epsilon^2)\,. \tag{8.41}$$

For later reference, we define the jump in u ('shear') across the interface as $\bar{u} = (u_2 - u_1)|_{z=\epsilon\eta}$, the Taylor expansion of which is

$$\bar{u} = (u_2 - u_1)|_{z=0} + \epsilon\eta(u_{2,z} - u_{1,z})|_{z=0} + \mathcal{O}(\epsilon^2)\,. \tag{8.42}$$

8.3.3 Lowest order

At the lowest order, ϵ^0, (8.31)–(8.33) read

$$\begin{aligned} {u_{i,t}}^{(0)} + {p_{i,x}}^{(0)} &= 0 \\ {p_{i,z}}^{(0)} &= 0 \\ {u_{i,x}}^{(0)} + {w_{i,z}}^{(0)} &= 0\,. \end{aligned}$$

So, at this order, ${p_i}^{(0)}$ and ${u_i}^{(0)}$ are independent of z, which is characteristic of the hydrostatic approximation, which applies at this order because we have neglected the terms with δ_*. The boundary conditions (8.34), (8.35), (8.40) and (8.41) are, at lowest order,

$$\begin{aligned} {w_1}^{(0)}|_{z=\alpha} &= 0 \\ {w_2}^{(0)}|_{z=\alpha-1} &= 0 \\ {w_i}^{(0)}|_{z=0} &= {\eta_t}^{(0)} \\ {p_2}^{(0)} - {p_1}^{(0)} &= \eta^{(0)}\,. \end{aligned}$$

The horizontal momentum equation and boundary condition for pressure can be combined into

$$\bar{u}_t{}^{(0)} + \eta_x{}^{(0)} = 0\,. \tag{8.43}$$

where $\bar{u}^{(0)}$ is the lowest-order shear from (8.42).

Integration of the continuity equation over the upper and lower layers gives, respectively,

$$\begin{aligned} {\eta_t}^{(0)} - \alpha {u_{1,x}}^{(0)} &= 0 \\ {\eta_t}^{(0)} + (1-\alpha){u_{2,x}}^{(0)} &= 0\,. \end{aligned}$$

Multiplying the first by $1-\alpha$, the second by α, and adding up the results gives

$$\eta_t{}^{(0)} + \nu^2 \bar{u}_x{}^{(0)} \quad = \quad 0 \tag{8.44}$$

with $\nu^2 = \alpha(1-\alpha)$.

We combine (8.43) and (8.44) to

$$\eta_{tt}{}^{(0)} - \nu^2 \eta_{xx}{}^{(0)} \quad = \quad 0\,.$$

This equation decribes linear long waves, both left- and rightward propagating. Its general solution can be written as $F(x+\nu t) + G(x-\nu t)$ (cf. Section 6.1.2 for a mathematically similar problem). We restrict ourselves to rightward propagating waves; they are described by

$$\eta_t{}^{(0)} + \nu \eta_x{}^{(0)} \quad = \quad 0\,. \tag{8.45}$$

At lowest order, (8.39) reads

$$D_0(\eta^{(0)}) \quad = \quad 0\,. \tag{8.46}$$

Comparing this with (8.45), we see that the operator D_0 is given by

$$\boxed{D_0(\eta) = \eta_t + \nu \eta_x\,.} \tag{8.47}$$

Finally, we derive some expressions that are useful at the next order. From (8.44) and (8.45), we have

$$\eta^{(0)} = \nu \bar{u}^{(0)}\,. \tag{8.48}$$

The integrated continuity equations imply

$$u_1{}^{(0)} = (\alpha-1)\bar{u}^{(0)} \qquad\qquad u_2{}^{(0)} = \alpha \bar{u}^{(0)} \tag{8.49}$$

$$w_1{}^{(0)} = \frac{\alpha - z}{\alpha}\,\eta_t{}^{(0)} \qquad\qquad w_2{}^{(0)} = \frac{\alpha - 1 - z}{\alpha - 1}\,\eta_t{}^{(0)}\,. \tag{8.50}$$

From (8.49) we see that the horizontal velocities in each layer are inversely proportional to their thickness. Thus, in a shallow upper layer, the currents will be strongest. Moreover, they are directionally opposed, with the total transport being zero.

8.3.4 Next order

At the next order, ϵ^1, (8.31) and (8.33) read

$$\begin{aligned} u_{i,t}{}^{(1)} + u_i{}^{(0)} u_{i,x}{}^{(0)} \quad &= \quad -p_{i,x}{}^{(1)} \\ \delta_* w_{i,t}{}^{(0)} \quad &= \quad -p_{i,z}{}^{(1)} \\ u_{i,x}{}^{(1)} + w_{i,z}{}^{(1)} \quad &= \quad 0\,. \end{aligned}$$

In the first equaion we used ${u_{i,z}}^{(0)} = 0$. The boundary conditions (8.34), (8.35), (8.40) and (8.41) read

$$\begin{aligned} {w_1}^{(1)}|_{z=\alpha} &= 0 \\ {w_2}^{(1)}|_{z=\alpha-1} &= 0 \\ {w_i}^{(1)}|_{z=0} &= {\eta_t}^{(1)} + ({u_i}^{(0)}\eta^{(0)})_x \\ {p_2}^{(1)}|_{z=0} - {p_1}^{(1)}|_{z=0} &= \eta^{(1)} \,. \end{aligned}$$

(In the last equation we used ${p_{i,z}}^{(0)} = 0$.) The vertical shear, (8.42), becomes at this order $\bar{u}^{(1)} = ({u_2}^{(1)} - {u_1}^{(1)})|_{z=0}$.

From the horizontal momentum equation, together with the boundary condition for pressure, we find, using also (8.49):

$$ {\bar{u}_t}^{(1)} + (2\alpha - 1)\bar{u}^{(0)}{\bar{u}_x}^{(0)} + {\eta_x}^{(1)} = 0 \,. \tag{8.51}$$

We now take $\partial/\partial z$ of the horizontal momentum equation (by which the nonlinear terms disappear because ${u_{i,z}}^{(0)} = 0$), and $\partial/\partial x$ of the vertical momentum equation, and substract the results; after one time integration, this gives

$$ {u_{i,z}}^{(1)} = \delta_* {w_{i,x}}^{(0)} \,. $$

Substitution of (8.50) gives, after one intergration to z,

$$\begin{aligned} {u_1}^{(1)} &= \delta_* \frac{\alpha z - z^2/2}{\alpha} {\eta_{xt}}^{(0)} + {u_1}^{(1)}|_{z=0} \\ {u_2}^{(1)} &= \delta_* \frac{(\alpha-1) z - z^2/2}{\alpha - 1} {\eta_{xt}}^{(0)} + {u_2}^{(1)}|_{z=0} \,. \end{aligned}$$

Using these expressions, we integrate the continuity equation over the upper and lower layers, to obtain,

$$\begin{aligned} {\eta_t}^{(1)} + ({u_1}^{(0)}\eta^{(0)})_x - \alpha {u_{1,x}}^{(1)}|_{z=0} - \tfrac{1}{3}\delta_*\alpha^2 {\eta_{xxt}}^{(0)} &= 0 \\ {\eta_t}^{(1)} + ({u_2}^{(0)}\eta^{(0)})_x + (1-\alpha) {u_{2,x}}^{(1)}|_{z=0} - \tfrac{1}{3}\delta_*(1-\alpha)^2 {\eta_{xxt}}^{(0)} &= 0 \,. \end{aligned}$$

where we used the boundary conditions for ${w_i}^{(1)}$. Multiplying these expressions by $1-\alpha$ en α, respectively, and adding up the results, yields

$$ {\eta_t}^{(1)} + (2\alpha - 1)(\bar{u}^{(0)}\eta^{(0)})_x + \nu^2 {\bar{u}_x}^{(1)} - \tfrac{1}{3}\delta_*\nu^2 {\eta_{xxt}}^{(0)} = 0 \tag{8.52}$$

where we simplified the nonlinear term using (8.49).

Combining (8.51) and (8.52) gives

$$ {\eta_{tt}}^{(1)} - \nu^2 {\eta_{xx}}^{(1)} + 3(1-2\alpha)(\eta^{(0)}{\eta_x}^{(0)})_x - \tfrac{1}{3}\delta_*\nu^4 {\eta_{xxxx}}^{(0)} = 0 \,, \tag{8.53}$$

where we used (8.45) and (8.48). Because of (8.45), we may split the spatial derivative at lowest order as

$$ {\eta_x}^{(0)} = -\tfrac{1}{2\nu}({\eta_t}^{(0)} - \nu {\eta_x}^{(0)}) \,. $$

Hence we can extract the operator $\partial/\partial t - \nu\partial/\partial x$ from (8.53):

$$\eta_t{}^{(1)} + \nu\eta_x{}^{(1)} + \tfrac{3}{2\nu}(2\alpha-1)\eta^{(0)}\eta_x{}^{(0)} + \tfrac{1}{6}\delta_*\nu^3\eta_{xxx}{}^{(0)} = 0\,. \tag{8.54}$$

At this order, (8.39) reads

$$D_0(\eta^{(1)}) + D_1(\eta^{(0)}) \quad = \quad 0\,. \tag{8.55}$$

Comparing this with (8.54), we find that the operator D_1 is given by

$$\boxed{D_1(\eta) = \tfrac{3}{2\nu}(2\alpha-1)\eta\eta_x + \tfrac{1}{6}\delta_*\nu^3\eta_{xxx}\,.} \tag{8.56}$$

8.3.5 Final result

Now that we have obtained the operators D_0 and D_1, we can write (8.39) in explicit form as

$$\eta_t + \nu\eta_x + \tfrac{3}{2\nu}\epsilon\,(2\alpha-1)\eta\eta_x + \tfrac{1}{6}\delta\nu^3\eta_{xxx} + \cdots \quad = \quad 0\,. \tag{8.57}$$

The KdV equation now follows if we neglect the higher-order terms (cubic etc.), which are here rendered by dots. When we cast this equation back into dimensional variables, we obtain (8.11). With this, the derivation has been completed.

8.4 Inverse-scattering theory

So far, we have obtained only one solution of the KdV equation, the soliton (8.17), which preserves its form while propagating. Other initial profiles than the soliton will not, in general, be form-preserving. Yet, they may give rise to solitons during their evolution. The question is now, how we can determine whether this is the case, and, more specifically, how many solitons are to be expected. At the end of this section, it will become clear that these questions are surprisingly easy to answer, but to arrive at that simple result, we first need to go through a subject from quantum mechanics, called 'inverse scattering theory'.

In the following sections, we briefly explain this method, which is remarkable because it reduces the initial-value problem for the KdV equation, which is nonlinear, to the problem of solving two *linear* equations, the Schrödinger and Marchenko equations. In practice, the method is applicable in only some cases, because these linear equations are not, in general, easy to solve. However, if we are only interested in the *number* of solitons that will appear from an arbitrary initial profile, then, as it turns out, we need to solve only part of the problem (namely determine the number of discrete eigenvalues of the Schrödinger equation). What is more, it actually turns out that we do not even have to solve *that* problem, because the number of eigenvalues depends merely on the overall integral shape of the initial profile; its specifics are immaterial. So, finally, a

simple expression is obtained for the number of solitons, N, arising from an initial profile $\eta \leq 0$:

$$N \sim \int_{-\infty}^{\infty} dx \sqrt{|\eta(x,0)|}\,. \tag{8.58}$$

In Section 8.5.1 we adapt this expression to conform with the parameters in a two-layer system.

8.4.1 The scattering problem

We start *in medias res* and consider the scattering problem of the Schrödinger equation:

$$-\frac{d^2\psi}{dx^2} + q\psi = \lambda\psi\,. \tag{8.59}$$

The problem is to determine, for a given localized potential $q(x)$,[2] the eigenfunctions ψ, along with its corresponding eigenvalues λ.

As we assume q be localized, (8.59) becomes for large $|x|$

$$-\frac{d^2\psi}{dx^2} \sim \lambda\psi\,. \tag{8.60}$$

There are two cases to be distinguished: $\lambda > 0$ and $\lambda < 0$. In the first case we write $\lambda = k^2$ (k being real). One solution of (8.60) then is e^{-ikx}, which we interpret as a wave coming from $+\infty$. When it arrives in the region where the potential is active, part of the wave will be reflected, giving $R(k)e^{ikx}$ as $x \to \infty$, and the other part will be transmitted, giving $T(k)e^{-ikx}$ as $x \to -\infty$ (Figure 8.7). The reflection coefficient $R(k)$ and transmission coefficient $T(k)$ depend on the potential q. Asymptotically, we thus have

$$\psi(x,k) \sim \begin{cases} e^{-ikx} + R(k)e^{ikx} & x \to \infty \\ T(k)e^{-ikx} & x \to -\infty\,. \end{cases}$$

The second case, $\lambda < 0$, gives rise to a finite number of discrete eigenvalues, together with the corresponding eigenfunctions ψ describing *bound states*; both will be labeled by n. We write $\lambda = -p_n^2$, with p_n real, and choose, without loss of generality, p_n to be positive. For $x \to \infty$, the bound state ψ_n looks like

$$\psi_n(x) \sim c_n e^{-p_n x}\,.$$

We fix the constant c_n by the requirement that $\int_{-\infty}^{\infty} dx\, \psi_n^2(x) = 1$.

The *direct* scattering problem consists in determining $R(k), T(k), c_n$ for a given potential q. The *inverse* scattering problem consists in determining q from given 'scattering data'. For the class of potentials considered here, it suffices to know $\{R(k), c_n, p_n\}$ to uniquely determine the potential (see [1, Theorem 2.5.1]).

[2] i.e. $q \to 0$ for $|x| \to \infty$; some of the later theorems also presume $\int_{-\infty}^{\infty} dx\,(1+|x|)\,|q(x)| < \infty$.

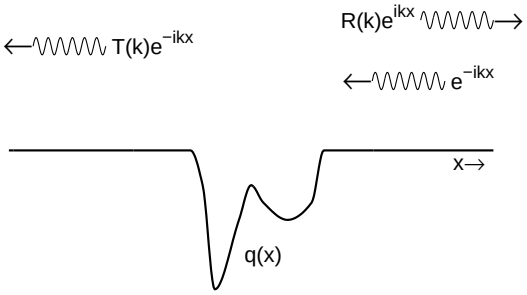

Fig. 8.7: The scattering problem in quantum mechanics: an incoming wave, from the right, is partially reflected at the potential q, and partially transmitted; hence the reflection and transmission coefficients $R(k)$ and $T(k)$.

We now briefly outline. how the inverse scattering problem is solved.[3] For given scattering data $\{R(k), c_n, p_n\}$, we define

$$M(x) = \sum_n c_n^2 e^{-p_n x} + \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, R(k) e^{ikx} \,. \tag{8.61}$$

Then we solve the function $K(x, y)$ from the so-called Marchenko equation,

$$K(x, y) + M(x + y) + \int_x^{\infty} dz\, K(x, z) M(y + z) = 0 \tag{8.62}$$

for $y > x$. Finally, we calculate $K(x, x) \equiv \lim_{y \to x} K(x, y)$, from which the potential q is found as

$$q(x) = -2 \frac{d}{dx} K(x, x) \,. \tag{8.63}$$

We now show that there is a connection, indeed a very useful one, between the KdV equation and this scattering problem. We first slightly generalize the scattering problem, by allowing the potential in (8.59) to depend on a certain parameter, τ say. (This parameter will later be taken to represent time.) The eigenfunction ψ and scattering data $\{R(k), c_n, p_n\}$ will then, in general, also depend on τ. However, the eigenvalue λ (and hence p_n) are *independent* of τ for certain classes of potentials; for example, for q's satisfying

$$\boxed{q_\tau - 6qq_x + q_{xxx} = 0} \tag{8.64}$$

which is, in scaled form, the KdV equation!

Stated more generally, we define operator L by

$$L = -\frac{d^2}{dx^2} + q(x; \tau) \,,$$

[3] Proofs of the correctness of the method are beyond the scope of this text; see, e.g., [1].

and suppose that

$$\psi_\tau = B\psi \tag{8.65}$$

for a certain differential operator B (which may contain derivatives to x only). We then find by taking the derivative of (8.59) to τ,

$$(q_\tau + [L, B])\psi = \lambda_\tau \psi \,, \tag{8.66}$$

where $[L, B] = LB - BL$. The eigenvalues λ are independent of τ if B is such that the left-hand side of (8.66) is zero. This is the case if, for example,

$$B = -4\frac{d^3}{dx^3} + 6q\frac{d}{dx} + 3q_x \tag{8.67}$$

in which case the left-hand side of (8.66) becomes the KdV equation (8.64).

The procedure for solving the KdV equation (8.64) for a given initial profile $q(x;0)$ is now as follows. First, we use $q(x;0)$ as the potential in (8.59), and solve the direct scattering problem, which yields $\{R(k;0), c_n(0), p_n\}$. Then, in a way yet to be specified, we derive from this the time-dependent scattering data, $\{R(k;\tau), c_n(\tau), p_n\}$. This forms the starting point of the inverse problem, in which we solve (8.61), to determine K, in which τ now serves as a parameter; with (8.63), this yields $q(x;\tau)$, the evolution of the initial profile.

It remains to be explained, how the τ-dependence in the scattering data may be obtained. Since k and p_n are, by construction, τ-independent, we need only determine $R(k;\tau)$ and $c_n(\tau)$. Asymptotically, the expression for B, in (8.67), is

$$B \sim -4\frac{d^3}{dx^3} \,.$$

Eq. (8.65) then implies

$$\psi_\tau \sim -4\psi_{xxx} \tag{8.68}$$

for $|x| \to \infty$. For the continuous spectrum, we try an expression of the form

$$\psi(x,k;\tau) \sim h(\tau) \begin{cases} e^{-ikx} + R(k;\tau)e^{ikx} & x \to \infty \\ T(k;\tau)e^{-ikx} & x \to -\infty \,. \end{cases}$$

By substitution in (8.68), we find $h(\tau) = \exp(-4ik^3\tau)$ and

$$R(k;\tau) = R(k;0)e^{8ik^3\tau} \,, \qquad T(k;\tau) = T(k;0) \,. \tag{8.69}$$

For the discrete spectrum, we have

$$\psi_n(x;\tau) \sim c_n(\tau)e^{-p_n x}$$

as $x \to \infty$, so that

$$c_n(\tau) = c_n(0)e^{4p_n^3\tau} \,. \tag{8.70}$$

With this, the τ-dependence of the spectrum has been established.

8.4.2 The meaning of the discrete spectrum

We now show that the presence of a single discrete eigenvalue p in the scattering data implies the emergence of one soliton. Let the scattering data, corresponding to the initial profile, be given by $p > 0$, $c(0)$ and $R(k;0) = 0$, a reflectionless case with one bound state. From (8.69) and (8.70), one finds their evolution in time,

$$\begin{aligned} c(\tau) &= c(0)e^{4p^3\tau} \\ R(k;\tau) &= R(k;0)e^{8ik^3\tau} = 0\,. \end{aligned}$$

Hereafter, we will write $c = c(\tau)$, for the sake of brevity.

We now solve the inverse problem. The definition of M, in (8.61), gives

$$M(x) = c^2 e^{-px}\,.$$

(The τ dependence of M and c is implied.) With this M, we solve the Marchenko equation, (8.62),

$$K(x,y) + c^2 e^{-p(x+y)} + c^2 \int_x^\infty dz\, K(x,z) e^{-p(y+z)} \quad = \quad 0\,,$$

which we rewrite as

$$K(x,y) \quad = \quad -c^2 e^{-py}\Big\{e^{-px} + \int_x^\infty dz\, K(x,z) e^{-pz}\Big\}\,.$$

Substitution of $K(x,y) = e^{-py}w(x)$ gives

$$\begin{aligned} w(x) &= -c^2 e^{-px} - c^2 w(x) \int_x^\infty dz\, e^{-2pz} \\ &= -c^2 e^{-px} - \frac{c^2}{2p} w(x) e^{-2px}\,, \end{aligned}$$

from which we obtain w, and hence

$$K(x,y) \quad = \quad -\frac{c^2 e^{-py}}{e^{px} + \frac{c^2}{2p} e^{-px}}\,.$$

The limit $y \to x$ gives

$$K(x,x) \quad = \quad -\frac{c^2 e^{-px}}{e^{px} + \frac{c^2}{2p} e^{-px}} \quad = \quad -\frac{c^2}{e^{2px} + \frac{c^2}{2p}}\,.$$

Finally, the potential is obtained from (8.63):

$$\begin{aligned} q(x) &= -2\frac{d}{dx} K(x,x) \quad = \quad -\frac{4pc^2}{(e^{px} + \frac{c^2}{2p} e^{-px})^2} \\ &= -\frac{8p^2}{(\frac{\sqrt{2p}}{c} e^{px} + \frac{c}{\sqrt{2p}} e^{-px})^2} \quad = \quad -\frac{8p^2}{(e^{px+\gamma} + e^{-px-\gamma})^2} \\ &= -2p^2 \mathrm{sech}^{\,2}(px + \gamma)\,, \end{aligned}$$

with $\gamma = \ln \frac{\sqrt{2p}}{c}$. Restoring the τ-dependence, due to $c = c(0)e^{4p^3\tau}$, we find,

$$q(x;\tau) \quad = \quad -2p^2 \mathrm{sech}^2(p[x - 4p^2\tau] + \gamma_0)\,,$$

with $\gamma_0 = \ln \frac{\sqrt{2p}}{c(0)}$. This is the soliton. Here p^2 serves as the amplitude, which is proportional to the phase speed, and inversely proportional to the square of the length scale, precisely as in (8.18).

This straightforward derivation shows that a single discrete eigenvalue, with a reflectionless potential, corresponds to one soliton. In fact, this connection can be generalized: irrespective of whether the reflection coefficient is zero or not, the number of solitons always equals the number of discrete eigenvalues.[4] This is a very convenient result, since it means that it suffices to determine the number of discrete eigenvalues for the initial profile if we want to know how many solitons will arise from it. An example of such a calculation is given in the next section.

8.4.3 Calculation of the number of emerging solitons

We now calculate the number of discrete eigenvalues for the initial profile describing a 'aquare well',

$$q(x;0) \quad = \quad \begin{cases} -A_0 & |x| < L \\ 0 & |x| > L\,, \end{cases} \tag{8.71}$$

for constant $A_0 > 0$ (Figure 8.8).

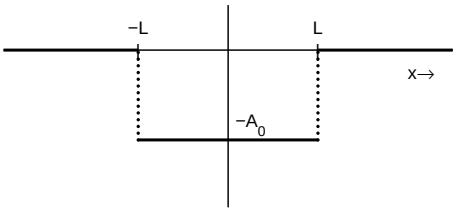

Fig. 8.8: The initial profile for which the number of emerging solitons is calculated.

The discrete eigenvalues of (8.59) follow from

$$-\frac{d^2\psi}{dx^2} + q(x;0)\psi = -p_n^2\psi\,.$$

For $|x| > L$,

$$\frac{d^2\psi}{dx^2} = p_n^2\psi\,,$$

so that the bound solution is there given by

$$\psi \quad = \quad \begin{cases} c_n^+ e^{-p_n x} & x > L \\ c_n^- e^{p_n x} & x < -L\,, \end{cases}$$

[4] The demonstration of this fact is beyond the scope of this text, but see [14, §2.8].

with constants $c_n^{\pm}$. For $|x| < L$,

$$\frac{d^2\psi}{dx^2} + (A_0 - p_n^2)\psi = 0\,.$$

We assume $p_n^2 < A_0$ (otherwise there would be no bound states), and write $r_n^2 = A_0 - p_n^2$. The solution in the interval $|x| < L$ is then

$$\psi = a_n \sin r_n x + b_n \cos r_n x\,,$$

with arbitrary constants a_n and b_n. We now require continuity of ψ and $d\psi/dx$ at $x = \pm L$. This gives four conditions,

$$\begin{aligned}
c_n^+ e^{-p_n L} &= a_n \sin r_n L + b_n \cos r_n L \\
c_n^- e^{-p_n L} &= -a_n \sin r_n L + b_n \cos r_n L \\
-p_n c_n^+ e^{-p_n L} &= a_n r_n \cos r_n L - b_n r_n \sin r_n L \\
p_n c_n^- e^{-p_n L} &= a_n r_n \cos r_n L + b_n r_n \sin r_n L\,.
\end{aligned}$$

The third equation plus p_n times the first one, and the fourth equation minus p_n times the second one, gives, respectively,

$$\begin{aligned}
a_n(p_n \sin r_n L + r_n \cos r_n L) + b_n(p_n \cos r_n L - r_n \sin r_n L) &= 0 \\
a_n(p_n \sin r_n L + r_n \cos r_n L) - b_n(p_n \cos r_n L - r_n \sin r_n L) &= 0\,.
\end{aligned}$$

For non-trivial solutions (i.e. a_n and b_n not both zero), the determinant has to be zero. Hence

$$\frac{p_n}{r_n} = \tan r_n L\,, \quad \text{or} \quad -\frac{r_n}{p_n} = \tan r_n L\,.$$

The first equality implies $a_n = 0$, giving an even solution, $\cos r_n x$; the second one implies $b_n = 0$, giving the odd solution, $\sin r_n x$. Introducing $y = r_n L$, we can write these equalities more conveniently as

$$\frac{[(L\sqrt{A_0})^2 - y^2]^{1/2}}{y} = \tan y\,, \quad \text{or} \quad -\frac{y}{[(L\sqrt{A_0})^2 - y^2]^{1/2}} = \tan y\,. \tag{8.72}$$

It is now easy to see by graphical inspection (see Figure 8.9), that there are N intersections (excluding $y = 0$) for

$$\frac{(N-1)\pi}{2} \le L\sqrt{A_0} < \frac{N\pi}{2}\,.$$

(For convenience, we treat the special case $L\sqrt{A_0} = (2n+1)\pi/2$ as if there were an intersection on the asymptote, at $-\infty$.) The number of bound states is thus the largest integer satisfying

$$N \le 1 + \frac{2L\sqrt{A_0}}{\pi}\,. \tag{8.73}$$

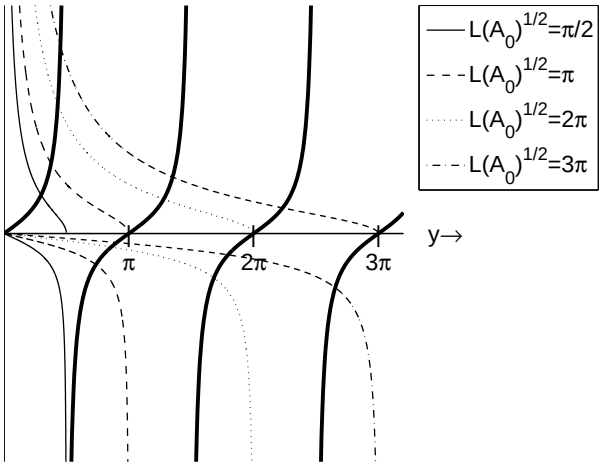

Fig. 8.9: The expressions in (8.72): the right-hand sides ($\tan y$) are shown as thick solid lines; the left-hand sides give a positive profile (first expression) and a negative one (second expression). Four cases are shown, clarifying how the number of intersections depends on the value of $L\sqrt{A_0}$.

The means that, for large N, the number of emerging solitons is proportional to the length scale of the initial profile, and to the square root of its amplitude.

This result holds more generally. For arbitrary initial profiles $q \leq 0$, decaying for $x \to \pm\infty$, the number of solitons is given by

$$\boxed{N \approx \frac{1}{\pi} \int_{-\infty}^{\infty} dx \, \sqrt{|q(x)|}\,,} \tag{8.74}$$

again for large N, see [90, p. 598]. The specifics of its form are thus immaterial to the number of emerging solitons; it is the overall size that matters.

8.5 Internal tides and solitons

Groups of internal solitons, like in Figure 8.1, have been observed at many locations in the oceans, and also in shallow seas. The passage of such groups is not a unique event; on the contrary, they usually appear every tidal period (an example is shown in Figure 8.15). This suggests that their origin is related to the internal tide. Two mechanisms have been identified by which an internal tide may give rise to internal solitons. In the first, a low-mode internal tide steepens by nonlinear effects, and at some point splits up into a train of solitons (Section 8.5.1). In the other, the starting point is an internal-tide *beam*, which necessarily involves several modes; if such a beam impinges on the (seasonal)

thermocline, it locally creates a wave-like disturbance, which may evolve into a train of internal solitons (Section 8.5.2).

8.5.1 Disintegration of an interfacial tide

To get an idea of how many solitons may arise from an internal tide, we apply the result (8.73) from Section 8.4.3, which was obtained for the initial profile (8.71), Figure 8.8. First we have to bring the KdV equation for the two-layer system, (8.11), or its short-hand form (8.14),

$$\frac{\partial \eta}{\partial t} + c_0 \frac{\partial \eta}{\partial x} + A\eta \frac{\partial \eta}{\partial x} + B \frac{\partial^3 \eta}{\partial x^3} \quad = \quad 0\,,$$

into the canonical form (8.64):

$$q_\tau - 6qq_y + q_{yyy} = 0\,,$$

(where we replaced x with y, for convenience). This is done via the transformations

$$\eta = -6\frac{B^{1/3}}{A} q - \frac{c_0}{A}\,, \qquad x = B^{1/3} y\,.$$

In (8.73) we found that the number of solitons emerging from a square well of amplitude A_0 and length $2L$ equals the largest integer N satisfying

$$N \quad \leq \quad 1 + \frac{2L\sqrt{A_0}}{\pi}\,.$$

Using the inverse transformation, we find

$$A_0 = \frac{c_0 - Aa_0}{6B^{1/3}}\,, \qquad L = \frac{l}{B^{1/3}}\,,$$

where $a_0 > 0$ and l are the depth and halflength of the initial profile in the two-layer system. Substituting the coefficients of the nonlinear and nonhydrostatic terms from (8.11), A and B, respectively, we find for the number of solitons:

$$N \quad \leq \quad 1 + \frac{2l}{\pi(h_1 h_2)^{1/2}} \left(1 - a_0 \frac{3(h_1 - h_2)}{2h_1 h_2}\right)^{1/2}. \tag{8.75}$$

This expression is remarkable for its absence of c_0, and hence of g', the important parameter determining the strength of the stratification at the interface. This absence is, however, only apparent, since the initial profile with amplitude $-a_0$ and halflength l does itself depend on the stratification, via the generation mechanism of internal tides. There is a further caveat, for in the transformation which led to (8.75) it must be assumed that $A \neq 0$ (the coefficient of the nonlinear term), so we may not afterward choose $h_1 = h_2$ in (8.75). Restricting its usage to $h_1 < h_2$, we find for realistic parameters such as $h_1 = 50\,\mathrm{m}$, $h_2 = 150\,\mathrm{m}$, $a_0 = 10\,\mathrm{m}$ and $2l = 5\,\mathrm{km}$, the following number of solitons: $N \approx 21$. This suggests that solitons should be quite ubiquitous in the ocean. They are,

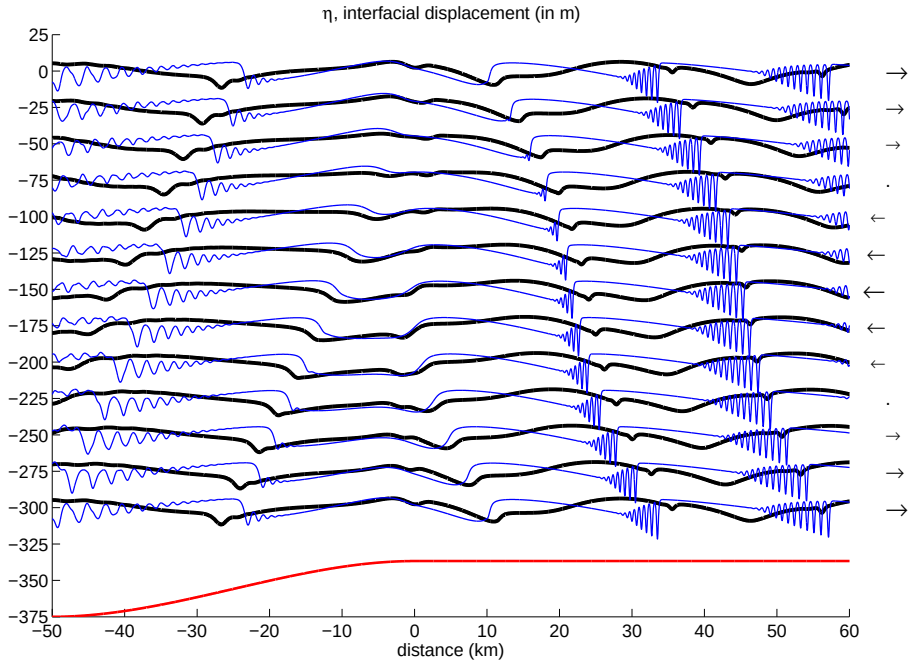

Fig. 8.10: Numerical model results simulating the generation of internal tides in the Celtic Sea (northern part of the Bay of Biscay), here during neap tides (i.e. weak forcing). Successive moments during a tidal cycle are shown in a downward sequence of profiles, at an offset of 25 m. At the bottom of the figure the continental slope is shown, whose real height is 3830 m. The actual simulation is shown as thick lines; clearly, no solitons emerge from the internal tide. However, if Coriolis effects are artificially switched off (as if the Celtic Sea were at the equator), then solitons do arise (blue thin lines). This demonstrates that Coriolis dispersion may prevent the internal tide from disintegrating into solitons. We note that during periods of stronger forcing (spring tides), internal solitons do appear even in the presence of rotation (not shown). The arrows on the right indicate the direction of the barotropic flow at each moment. Notice the effect of the barotropic flow on the propagation of internal tides and soliton packets over the continental shelf; when the flow is leftward, they are nearly blocked. After [25].

indeed, but trains of solitons usually consist of much fewer solitons. There is an effect, ignored in this chapter so far, which tends to diminish the number of solitons considerably: the Coriolis force.

This force also produces a dispersive effect, as is clear from (5.44). However, it differs in a fundamental way from non-hydrostatic dispersion: whereas the latter becomes stronger at shorter length scales, Coriolis dispersion becomes stronger at *longer* length scales. So, even if one would not expect a strong effect on the relatively short internal-soliton scales, Coriolis dispersion is still important because the solitons emerge from the much longer internal-tide scale. Specifically, Coriolis dispersion tends to counteract nonlinearity, in the sense

that it diminishes the latter's steepening effect. As a result, less solitons appear than would have been the case without Coriolis effects. It is not accidental that most of the largest solitons have been observed in tropical regions (such as the Andaman and Sulu Seas), where f is nearly zero.

Two examples from a numerical model are shown in Figure 8.10, one with and one without Coriolis effects, all other things being the same. This figure illustrates how internal solitons emerge from the internal tide, and how Coriolis effects affect this process. This numerical model solves a set of equations that combines the KdV effects of nonlinearity and nonhydrostatic dispersion with the generation mechanism of internal internal tides in a rotating system, as described in Chapter 7.

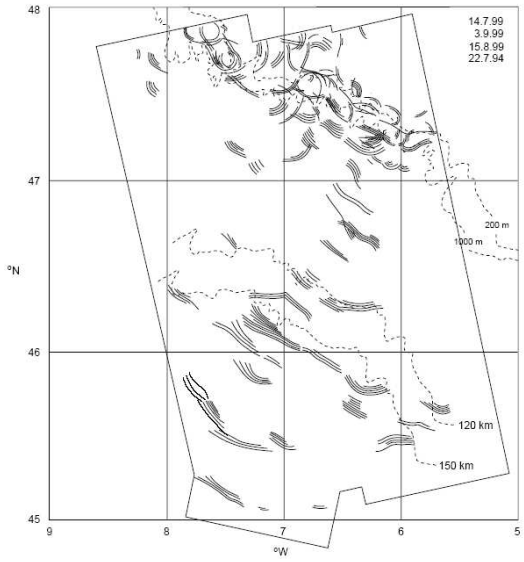

Fig. 8.11: Composite of internal solitary wave packets observed in SAR (Synthetic Aperture Radar) images. The packets are largely concentrated in *two* distinct regions: close to the shelf break (the dashed 200 m line), and near 150 km off the shelf break, in the deep central part of the basin. The latter are due to an impinging internal-tide beam coming from the deep ('local generation'). Significantly, all the observations were made during summer (July to early September), when the seasonal thermocline is well-developed, a sine qua non for internal solitons. From [62].

8.5.2 'Local generation' by internal-tide beams

In the previous section we have treated the internal tide as an interfacial wave. This is only one way in which internal solitons can arise from internal tides. A very different mechanism was discovered in the Bay of Biscay. Figure 8.11 shows a collection of renderings of SAR images. As in Figure 1.1, the stripes are surface manifestations of internal solitons. Near the continental shelf break (dashed

line), i.e. close to where internal tides are generated, trains of internal solitons appear. Further off the slope, oceanward, they tend to disappear. However, in the central bay, about 150 kilometers off the shelf break, *new* groups of internal solitons appear. The site of their generation turns out to coincide with the position where the internal tidal beam, after having reflected from the bottom, impinges on the seasonal thermocline. This is illustrated in another example, from the southern Bay of Biscay (Figure 8.12).

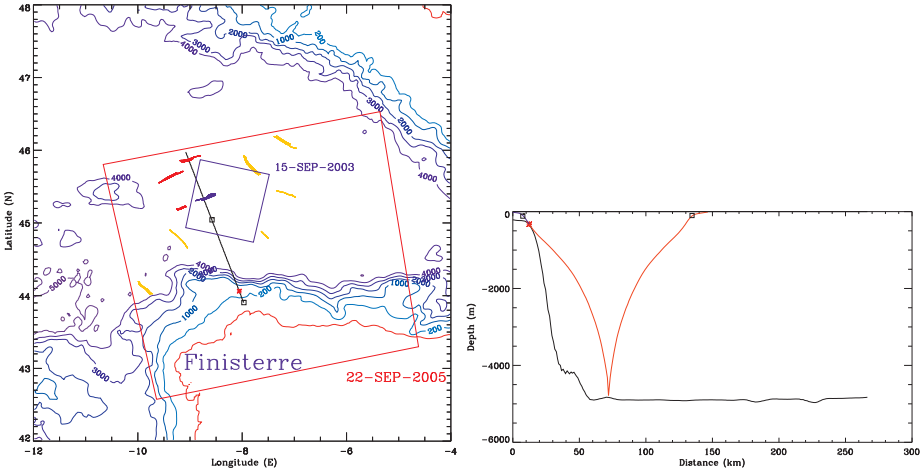

Fig. 8.12: Illustration of the process of local generation. One particular group of solitary waves (along the staight line in the left panel) arises near the location where the internal tidal beam hits the seasonal thermocline (right panel). From [10].

The mechanism of local generation can be regarded as a two-stage process. We showed above that even in linear theory, the internal tidal beam is strongly perturbed as it encounters the seasonal thermocline (Figure 7.10). The beam brings the seasonal thermocline into oscillation (Figure 7.11). So the first stage is essentially linear. It is followed by a second stage, in which nonlinear and nonhydrostatic effects come into play, allowing the depressions to evolve into internal solitons [26, 2].

8.5.3 Decay and dissipation

Little is yet known about the decay and dissipation of internal solitons. Some propagate all the way from the deep ocean into shallow harbours, causing noticeable surface oscillations there [32].

In any case, the generation of internal solitons implies a transfer of energy from the long internal-tide scale to the much shorter soliton scale, and thus forms one step in a chain towards ever shorter scales, down to the one at which viscous dissipation takes place. Since solitons have large amplitudes, and hence a strong shear, instabilities may develop as illustrated particularly clearly in Figure 8.13. By processes like these, turbulent mixing may occur in and around

the seasonal thermocline – a process that is also of biological significance as it brings nutrients upward from the nutrient-rich deeper layers into the upper layer where light penetrates and photosynthesis takes place; solitons thus act as a 'nutrient-pump'.

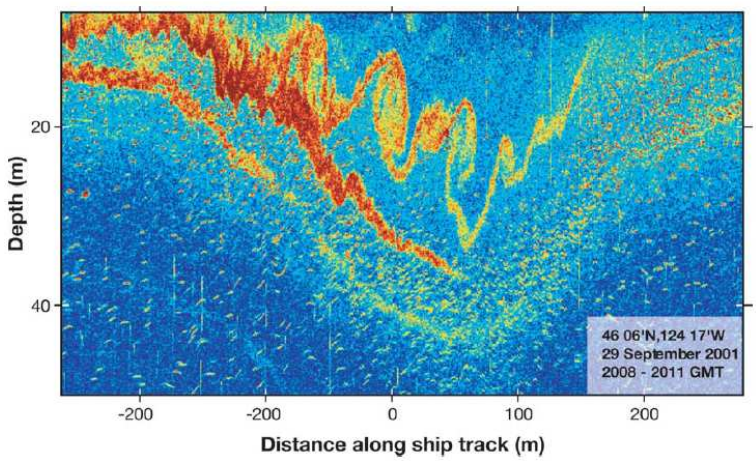

Fig. 8.13: An acoustic backscatter record showing an internal soliton accompanied by Kelvin-Helmholtz billows arising from shear instabilities; the wave is propagating from left to right. This observation was made off the Oregon coast. From [60].

8.6 Limitations of KdV

Apart from the neglect of Coriolis dispersion, the KdV equation has two other limitations. One is that waves are assumed to be long. One can think of configurations in which this assumption is not valid, for example a two-layer system with a very deep lower layer. For such a system, too, a soliton equation can be derived, the so-called Benjamin-Ono equation. To this end, we return to (5.44), with $f = 0$ (hence $q_u = k$),

$$\omega^2 = \frac{g'k}{\coth kd + \coth k(H-d)}\,. \tag{8.76}$$

We assume that waves are long with respect to the upper layer ($|kd| \ll 1$), but short with respect to the total water depth ($|kH| \to \infty$). This implies, of course, $H \gg d$. We can now use the fact that $\coth kH \to \mathrm{sgn}(k)$ for $|kH| \to \infty$; the denominator of (8.76), multiplied by kd, can thus be approximated as

$$kd\coth kd + kd\coth k(H-d) = 1 + \tfrac{1}{3}(kd)^2 + \cdots + kd\,\mathrm{sgn}(k) + \cdots \approx 1 + |k|d\,.$$

Selecting rightward propagating waves, we find

$$\omega = c_0 k(1 - \tfrac{1}{2}|k|d + \cdots)\,, \tag{8.77}$$

with $c_0^2 = g'd$. For $d \ll H$, the nonlinear term in (8.11) becomes $-(3c_0)/(2d)\eta\eta_x$. Thus, we expect the governing equation to be given by

$$\frac{\partial \eta}{\partial t} + c_0 \frac{\partial \eta}{\partial x} - \frac{3c_0}{2d}\,\eta \frac{\partial \eta}{\partial x} + \frac{c_0 d}{2\pi} \int_{-\infty}^{\infty} dx' \, \frac{\eta(t,x')}{x'-x} = 0\,, \tag{8.78}$$

where the integral is a Cauchy Principal Value integral (which in this particular form is also known as the Hilbert transform). The linearized form of (8.78) leads to the dispersion relation (8.77) via the identity

$$\int_{-\infty}^{\infty} dx' \, \frac{e^{ikx'}}{x'-x} = i\pi e^{ikx}\,\mathrm{sgn}(k)\,.$$

The Benjamin-Ono equation (8.78) admits an algebraic soliton solution,

$$\eta(t,x) = \frac{-Ar^2}{r^2 + (x - Ct)^2}\,,$$

with $C = c_0(1 + \frac{3}{8}A/d)$, $r = \frac{4}{3}d^2/A$ and $A > 0$. Here, too, the solitons appear as a depression. The Benjamin-Ono is largely of theoretical interest, because the observed solitons are usually better described by the KdV equation.

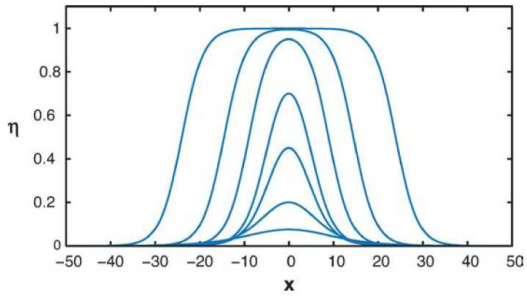

Fig. 8.14: Examples of solitary wave solutions (here as elevations) of the extended KdV equation; they are given by (8.80). As the wave amplitude approaches its maximum value (here normalized at 1), the wave crest becomes increasingly broad. From [40].

However, a sometimes severe limitation of the KdV equation is the underlying assumption of *weak* nonlinearity. Observed solitons may have amplitudes so large as to render this assumption inappropriate. An extended form of the KdV equation can be derived by proceeding to a higher order, which produces a cubic nonlinear term,

$$\frac{\partial \eta}{\partial t} + c_0 \frac{\partial \eta}{\partial x} + \frac{3}{2}\,\frac{h_1 - h_2}{h_1 h_2}\, c_0 \eta \frac{\partial \eta}{\partial x} + \frac{1}{6} c_0 h_1 h_2 \frac{\partial^3 \eta}{\partial x^3} = c_0 \frac{h_1^2 + 6h_1h_2 + h_2^2}{8(h_1h_2)^2} \frac{\partial}{\partial x}(\eta^3)\,. \tag{8.79}$$

Notice that the coefficient of the cubic term is always positive.[5] Equation (8.79)

[5] In the literature, the coefficient appears in various forms; in [12], for example, it is stated without recourse to the assumption that the difference in density between the layers is small. For $(\rho_2 - \rho_1)/\rho_2 \ll 1$, their coefficient reduces to the form given in (8.79).

is called the *extended* KdV equation. In the appendix to this chapter, we demonstrate that (8.79) admits solitons of the form

$$\eta = \frac{1}{c + a\cosh[b(x - Ct)]}\,. \tag{8.80}$$

This soliton solution departs from the classical KdV soliton in several ways; we here summarize its main properties (for proofs and details, see the Appendix).

As in the classical KdV equation, the amplitude a has the same sign as the quadratic nonlinear term; so, for $h_1 < h_2$, solitons will be depressions. However, a crucial difference is that the amplitude cannot now take arbitrary large values; it has an upper bound, which depends solely on h_1 and h_2. As the solitary wave approaches this upper bound, it becomes increasingly *broad* (as opposed to small-amplitude, classical KdV solitons, which become slightly narrower with increasing amplitude); the top becomes a broad plateau, terminated at each end by dissipationless bores, as illustrated in Figure 8.14.

Although the extended KdV equation contains a higher-order nonlinear term, it is based on a perturbation approach, and thus still subject to the assumption of weak nonlinearity. Yet, in practice, it turns out that the range of validity is widened considerably by the inclusion of the cubic term. Specifically, it is found that the soliton solution of the extended KdV equation, and the corresponding dependences of phase speed and length scale on amplitude, are fairly close to those found in fully nonlinear theories, as long as h_1 and h_2 are not very different [40]. The extended KdV equation then provides a satisfactory description of large-amplitude solitons. An example of such solitons is shown in Figure 8.15; they have amplitudes of about 25 m, whereas the thickness of the upper layer is only 7 m. (Notice that for thin upper layers the coefficient of the quadratic nonlinear term scales with $1/h_1$; the strength of nonlinearity is then determined by a/h_1.) They are due to a disintegration of the internal tide, as is clear from the fact that groups of solitons are separated by a tidal period.

Stanton & Ostrovsky compared observed solitons from Figure 8.15 with solutions from the KdV and extended KdV equations; they found that the former produces solitons that are much narrower than observed; the broader solitons from the extended KdV equation, on the other hand, give a good match [80].

They also found that observed amplitudes slightly exceed the theoretical maximum amplitude (which for their parameters is about 20 m). This discrepancy may be due to the simplified nature of the two-layer stratification adopted here, rather than being a limitation of the extended KdV equation *per se*. As a matter of fact, for other types of stratification $N(z)$, too, an extended KdV equation can be derived; it takes the same form, but with the coefficients defined differently [50]. One caveat should be noted. As will be shown in Section 9.2, for constant stratification the nonlinear terms vanish for a uni-modal solution. This means that one has either to drop the assumption of constant N, or (implicitly) include higher modes to arrive at a KdV equation.

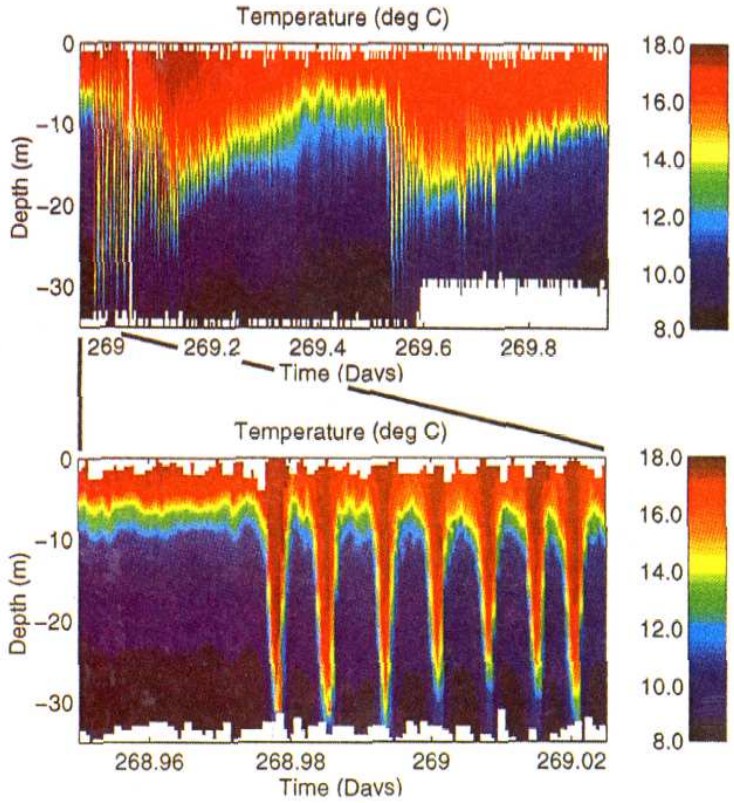

Fig. 8.15: An observation of two groups of solitons, separated by the semi-diurnal tidal period. The lower panel shows an enlargement, covering the first 100 minutes; the solitons have a period of about ten minutes. These measurements were made off the coast of Northern Oregon, over the continental shelf, at a water depth 147 m. The upper 35 m were covered by a Loose-tethered Microstructure Profiler (LMP), which measured temperature and salinity (the latter not shown here), and was raised and then allowed to free-fall through the water column; thus vertical profiles were obtained every 40 seconds. From [80].

Further reading

Good introductory papers on solitons in the ocean are those by Osborne & Burch [64] and Ostrovsky & Stepanyants [65]; much on theory and observations can also be found in the review report by Apel et al. [3]. On solitons in the atmosphere, see Rottmann & Grimshaw [73]. An overview on generalized KdV equations (including, for example, a dissipative term) can be found in Grimshaw [36]. More general reviews on long nonlinear internal waves are those by Grimshaw et al. [37] (on the effects of rotation), and Helfrich & Melville [40]. Many interesting pictures can be found on the soliton atlas website

http://www.internalwaveatlas.com/Atlas2_index.html.

Appendix: Form-preserving solutions of the extended KdV equation

In this appendix we derive form-preserving solutions of the extended KdV equation (8.79),[6] which we write in short-hand notation as

$$\frac{\partial \eta}{\partial t} + c_0 \frac{\partial \eta}{\partial x} + A\eta \frac{\partial \eta}{\partial x} + B\frac{\partial^3 \eta}{\partial x^3} = D\frac{\partial}{\partial x}(\eta^3)\,. \tag{8.81}$$

Seeking, then, solutions of the form $\eta = F(x - Ct)$, we obtain, after one integration,

$$-(C - c_0)F + \tfrac{1}{2}AF^2 + BF'' = DF^3 + \mu\,, \tag{8.82}$$

where primes denote derivatives to the argument $\xi = x - Ct$; we included a constant of integration μ, in order not to restrict ourselves to solutions that vanish at infinity. The point is that (8.82) admits a solution describing a stationary shock wave,

$$\boxed{F = a\tanh(b\xi) + c\,.} \tag{8.83}$$

This can be seen by substitution in (8.82), which gives

$$\begin{aligned} &-(C - c_0)(a\tanh b\xi + c) + \tfrac{1}{2}A(a\tanh b\xi + c)^2 - 2ab^2 B\,\mathrm{sech}^2 b\xi \tanh b\xi \\ &\qquad = D(a\tanh b\xi + c)^3 + \mu\,. \end{aligned}$$

Using $\mathrm{sech}^2 y = 1 - \tanh^2 y$, and gathering equal powers of tanh, we find

$$\begin{aligned} &-c(C - c_0) + \tfrac{1}{2}c^2 A - c^3 D - \mu + [-a(C - c_0) + acA - 2ab^2 B - 3ac^2 D]\tanh b\xi \\ &\qquad + [\tfrac{1}{2}a^2 A - 3a^2 cD]\tanh^2 b\xi + [2ab^2 B - a^3 D]\tanh^3 b\xi = 0\,. \end{aligned}$$

Hence,

$$c = \frac{A}{6D}\,; \quad b = a\Big(\frac{D}{2B}\Big)^{1/2}; \quad C = c_0 + \frac{A^2}{12D} - a^2 D\,; \quad \mu = \frac{A}{6D}\Big(a^2 D - \frac{A^2}{36D}\Big)\,.$$

The constant c depends only on the thickness of the layers, h_1 and h_2. We can choose the amplitude a arbitrarily. The wavenumber b then follows from the second expression; it is proportional to the wave amplitude. The phase speed C follows from the third expression; it contains a term proportional to the square of the wave amplitude. The essential role of the cubic nonlinear term, via D, is evident. The constant of integration μ, finally, follows from the fourth expression. Notice that the solution has a zero nett mass ($c, \mu = 0$) when the quadratic nonlinear term vanishes ($A = 0$), which happens when the two layers

[6] These solutions are mentioned in the literature, e.g. in [45], but with no derivation or proof being given; here we fill this gap.

are of equal thickness.

Seeking now localized solutions – i.e., F and its derivatives vanish at infinity – we must take $\mu = 0$ in (8.82). Multiplying by F' and integrating once (again leaving out the constant of integration), we obtain

$$-\tfrac{1}{2}(C-c_0)F^2 + \tfrac{1}{6}AF^3 + \tfrac{1}{2}B(F')^2 = \tfrac{1}{4}DF^4\,,$$

or, equivalently,

$$B^{1/2}F' = \pm F\left[(C-c_0) - \tfrac{1}{3}AF + \tfrac{1}{2}DF^2\right]^{1/2}.$$

It is convenient to introduce $F = 1/G$, so that

$$B^{1/2}G' = \mp\Big[(C-c_0)G^2 - \tfrac{1}{3}AG + \tfrac{1}{2}D\Big]^{1/2}. \tag{8.84}$$

In order to get rid of the square root on the right-hand side, we introduce

$$G = c + a\cosh(b\xi)\,. \tag{8.85}$$

(The idea is to bring the part in square brackets in a form proportional to $[\cosh^2 b\xi - 1]^{1/2}$, which equals $\sin b\xi$, and cancels against G'.) Substituting (8.85) in (8.84), we find for the part in square brackets,

$$\begin{aligned}[\cdots] &= (C-c_0)(c+a\cosh b\xi)^2 - \tfrac{1}{3}A(c+a\cosh b\xi) + \tfrac{1}{2}D \\ &= a^2(C-c_0)\cosh^2 b\xi + (2ac(C-c_0) - \tfrac{1}{3}aA)\cosh b\xi \\ &\quad + c^2(C-c_0) - \tfrac{1}{3}cA + \tfrac{1}{2}D\,. \end{aligned} \tag{8.86}$$

For the last expression to take the form

$$a^2(C-c_0)[\cosh^2 b\xi - 1]\,, \tag{8.87}$$

we must pose the following two requirements,

$$C - c_0 = \frac{A}{6c}\,; \qquad c^2(C-c_0) - \tfrac{1}{3}cA + \tfrac{1}{2}D = -a^2(C-c_0)\,. \tag{8.88}$$

Combining the two gives

$$c^2A - 3cD - a^2A = 0\,.$$

Hence

$$c = \frac{3D \pm (9D^2 + 4a^2A^2)^{1/2}}{2A}\,.$$

In order that $C > c_0$, the signs of A and c have to be the same, which implies that we must select the *plus*-sign in $\pm$:

$$c = \frac{3D + (9D^2 + 4a^2A^2)^{1/2}}{2A}\,. \tag{8.89}$$

We first consider the limit $D \to 0$, in which case we should find the earlier results of the classical KdV equation. In this limit, we have $c \to a$, so

$$F = \frac{1}{G} = \frac{1}{c + a\cosh(b\xi)} \to \frac{1}{a(1 + \cosh(b\xi))} = \frac{1}{2a\cosh^2(b\xi/2)}$$

(In the last step we used an identity for hyperbolic functions.) This is indeed the classical KdV soliton, in which $(2a)^{-1}$ serves as the amplitude; the corresponding expression for the phase speed C is then readily recovered from the first equation in (8.88).

Returning now to the general result, with D included, we have c given by (8.89). Substition of (8.85) in (8.84) gives

$$b = \mp\Big(\frac{C - c_0}{B}\Big)^{1/2} . \tag{8.90}$$

where we used $G' = ab\sinh(b\xi)$. (The sign in $\mp$ is immaterial because cosh is an even function.) This expression is, in fact, identical to that found for the classical KdV equation. To summarize, we have obtained a soliton solution of the form

$$\boxed{F = \frac{1}{c + a\cosh(b\xi)}\,,} \tag{8.91}$$

with c given by (8.89), the phase speed C by the first expression in (8.88), and the wavenumber b by (8.90).

An interesting limit is $a \to 0$, the large-amplitude limit. We then have, for sufficiently small ξ,

$$F \approx \frac{1}{c} \to \frac{1}{3D/A} = \frac{A}{3D}\,,$$

showing that the soliton becomes constant around its crest, giving the plateau-like shape depicted in Figure 8.14; this expression implies that solitons cannot have amplitudes larger than $A/(3D)$. The presence of such an upper bound forms a radical departure from the classical KdV soliton.

It is clear from (8.89) that c has the same sign as A (D being always positive), and so has a, otherwise singularities would occur. This means that the sign of A determines whether the soliton is a depression or an elevation, as in the classical KdV equation.

Chapter 9

Miscellaneous topics

In this chapter we discuss various topics. We saw already above, in Section 6.2, that an internal-wave beam becomes more intense (or the opposite), when it reflects from a sloping boundary. In Section 9.1 it is shown how, in a closed basin, a repetition of this process may lead to an increasingly focussed pattern, the so-called internal-wave attractor. Including nonlinear effects, one finds that reflection from boundaries is accompanied by the generation of higher harmonics (Section 9.3). More generally, interactions among waves of different frequencies may occur (Section 9.4), and these processes are thought to be reponsible for the distribution of energy throughout the internal-wave spectrum (Section 9.5).

9.1 Internal-wave attractors

In Chapter 6 it was explained how internal-wave energy propagates along characteristic coordinates, i.e. lines defined by $\xi_{\pm} = const$. Reflection from a boundary involves a change from one characteristic to the other: if the incident wave propagates along $\xi_{+} = const$, then the reflected one will propagate along $\xi_{-} = const$, and vice versa. In a closed basin, there will be successive reflections from the boundaries the wave encounters, giving rise to a 'web' of characteristics. These webs have some remarkable properties that have been confirmed in laboratory experiments. We outline them in this section.

Our starting point is the equation for linear internal-wave dynamics (6.2), now written as

$$\frac{\partial^2 w}{\partial z^2} - \alpha^2 \frac{\partial^2 w}{\partial x^2} \quad = \quad 0\,, \tag{9.1}$$

with

$$\alpha^2 \quad = \quad \frac{N^2 - \omega^2}{\omega^2 - f^2}\,.$$

In this section we assume N to be constant. This is, of course, a crude assumption with regard to the stratification in the ocean (depicted in Figure 1.6), and

this puts some limitations on the applicability of the theory (see Section 9.1.3).

We consider a basin of length $2L$ and height H, and introduce new spatial coordinates, indicated by primes,

$$x = Lx' \,; \quad z = \frac{L}{\alpha}\, z' \,.$$

With this, (9.1) becomes

$$\frac{\partial^2 w}{\partial z'^2} - \frac{\partial^2 w}{\partial x'^2} \quad = \quad 0 \,.$$

The characteristic coordinates associated with this equation are

$$\xi_\pm = \pm x' - z' \,.$$

So, in this scaled system, the characteristics make an angle of 45° with the vertical; the internal-wave energy thus propagates at this angle. The bottom, originally at $z = -H$, is now given by

$$z' = -\tau = -\frac{\alpha H}{L} \,.$$

We will regard H and L, as well as N and f, as fixed parameters; the scaled water depth τ then varies only with wave frequency ω, via α. The closed basin considered here is depicted in Figure 9.1; the wall on the right-hand side is tilted. The parameter defining the tilt is $s \;\; (= 1 - d)$, being the horizontal distance covered by the slope. For $s = 1$ we have the slope shown in Figure 9.1, while for $s = 0$ we get a rectangular basin of width 2 and height τ. We shall assume that the slope is supercritical, i.e. the slope τ/s is steeper than the characteristics:

$$\tau > s \,.$$

We now construct a web, starting from an arbitrary point at the surface, $(x_0, 0)$, see Figure 9.1. From this point, we follow the characteristic in the rightward direction (I); it meets the slope at a certain point (x_1, z_1). From there, a characteristic (II) departs downward (since the slope is supercritical), meeting the bottom at $(x_2, -\tau)$. Next we follow the upward characteristic (III), which encounters the vertical wall at $(-1, z_3)$. From there, finally, goes a characteristic IV towards the surface, arriving at $(x_4, 0)$. (In Figure 9.1 this point coincides with the starting point, but this is, of course, not in general the case.) Specifically, we can describe the characteristics by

$$\begin{array}{rlll}
\text{I:} & z' = -x' + b_I & \text{with} & b_I = x_0 = z_1 + x_1 \\
\text{II:} & z' = \;\; x' + b_{II} & \text{with} & b_{II} = z_1 - x_1 = -\tau - x_2 \\
\text{III:} & z' = -x' + b_{III} & \text{with} & b_{III} = -\tau + x_2 = z_3 - 1 \\
\text{IV:} & z' = \;\; x' + b_{IV} & \text{with} & b_{IV} = z_3 + 1 = -x_4 \,.
\end{array}$$

Combining the expressions for the b's, we find

$$x_4 = -2 + 2\tau - 2x_1 + x_0 \,.$$

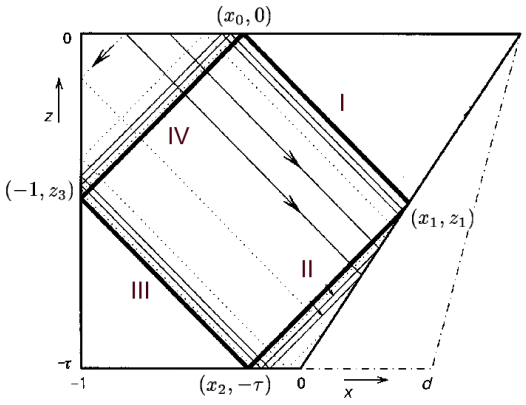

Fig. 9.1: A closed basin with a slope, filled with a stratified fluid. The dimensions have been scaled such that characteristics have a slope of 45° with the horizontal. The scaled depth is $\tau = \alpha H/L$, which depends on the wave frequency via α. The thick lines depict a closed web of characteristics. This is a special case; in general, starting from a horizontal position x_0 at the surface, one will reach a different point x_4 after traversing the basin once in a clockwise direction. From [54].

Moreover, since (x_1, z_1) lies both on characteristic I and on the slope, which is given by $z' = \tau(x' - 1)/s$, it follows that

$$x_1 = \frac{sx_0 + \tau}{\tau + s}\,.$$

Combining the last two expressions,

$$x_4 \;=\; 2\,\frac{(\tau - 1)(\tau + s) - \tau}{\tau + s} + x_0\,\frac{\tau - s}{\tau + s}\,. \tag{9.2}$$

This expression provides the connection between the starting point at the surface, x_0, and the arrival point x_4, reached after traversing the basin once in a clockwise direction. Of special interest are those points x_* for which the starting and arrival points coincide ($x_* = x_4 = x_0$); the web is then closed. This happens when

$$x_* \;=\; \frac{(\tau - 1)(\tau + s) - \tau}{s}\,. \tag{9.3}$$

9.1.1 Vertical wall

If the sloping wall is vertical, i.e. if $s = 0$, then (9.2) reduces to

$$x_4 = 2(\tau - 2) + x_0\,.$$

There is only *one* value of τ for which the web is closed, namely $\tau = 2$. The value of x_0 is then immaterial, indicating that the web is closed for every starting

point at the surface. The fact that there is only one value of τ, implies that there is only one frequency ω for which this situation occurs.

The solution corresponding to this case can be interpreted as a standing wave, of mode one. This can be seen as follows. The criterion for a standing wave is, for a basin of length $2L$: $m\lambda/2 = 2L$, for $m = 1, 2, 3 \cdots$. Here λ denotes the wavelength, which is given by $\lambda_n = 2\pi/k_n = 2\alpha H/n$, for $n = 1, 2, 3 \cdots$, where we used (5.14). Using the definition of τ, we can write this criterion as $\tau = 2n/m$. Here we have the lowest-mode case, with $n = m = 1$. The bottom line is that we have a discrete set of eigenvalues, which leads to discrete values of τ for which the web is closed.

9.1.2 Slope

The situation becomes very different if there is a slope; we examine the case $s = 1$, depicted in Figure 9.1. Closed webs occur when (9.3) is fulfilled,

$$x_* = (\tau - 1)(\tau + 1) - \tau \,.$$

We should of course require that $-1 < x_* < 1$; hence

$$1 < \tau < 2 \,.$$

There is now a *continuum* of frequencies for which a closed web occurs. These webs have a remarkable property: they form an *attractor*. Starting from some nearby point, $x_0 = x_* + \epsilon$, and traversing the basin once clockwise, the characteristic IV meets the surface at

$$x_4 = x_* + \epsilon \left(\frac{\tau - 1}{\tau + 1} \right) ,$$

as follows from (9.2). The factor in brackets lies between 0 and 1, implying that x_4 is nearer to x_* than is x_0. Repeating this procedure, one find that, after K cycles, the distance with x_* has become

$$\epsilon \left(\frac{\tau - s}{\tau + s} \right)^K .$$

In the limit $K \to \infty$, the point x_* is thus reached. Physically, this means that internal-wave energy, originally distributed over the basin, tends to converge to the closed web shown in Figure 9.1. The energy thus gets increasingly concentrated. As an aside, we note that the above reasoning can be reversed if we take the anti-clockwise direction; the roles of x_0 and x_4 are then interchanged and we find a defocusing of energy.

This notion of internal-wave atttractors in closed basins has been confirmed experimentally; an example is shown in Figure 9.2.

9.1.3 Discussion

The notion of an internal-wave attractor has thus been firmly established both theoretically and experimentally (for an insightful overview, see [53]). In the

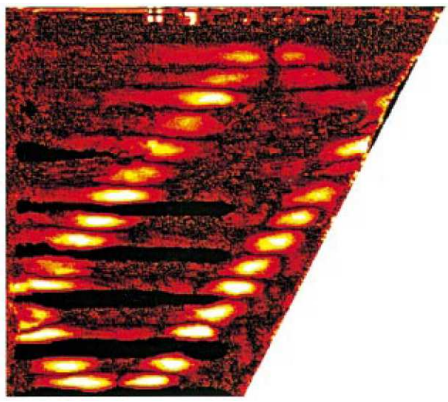

Fig. 9.2: The result of a laboratory experiment, with $\tau = 1.74$. The basin, filled with a constantly stratified fluid, was brought into a gentle vertical oscillation, at frequency ω_0; this creates instabilities, in the form of internal waves, of half the forcing frequency (i.e. $\omega = \omega_0/2$). Initially, they are spread over the entire basin, but by the process of focussing, described in this section, the energy finally gets concentrated near a closed web, the attractor, as predicted by theory. Using optical techniques, the isopycnal displacements are visualized; they are largest in the light areas. From [54].

ocean, however, they have not yet been observed. In fact, there are a number of factors at work that may make their appearance unlikely. First of all, N varies in the ocean, especially in the thermocline (see Figures 1.6 and 3.3b). These inhomogeneities cause internal-wave beams to suffer from internal reflections, as sketched in Figure 6.6. As a result, the beams become less intense; see, for example, Figure 7.10. These internal reflections counteract the focusing effect that would otherwise lead to attractors. Another factor lies in the required longevity of the waves; they may be dissipated, or be severely weakened by wave-wave interactions, before having traversed the basin a couple of times. Finally, basins need to have an appropriate shape for attractors to occur; in practice, channel-shape basins, for example, are not entirely closed, and internal-wave beams propagating in a cross-section of the channel may escape onto the shelves. Three-dimensional effects further add to the complexity of the problem.

Still, the discovery of internal-wave attractors calls for a revised view on internal-wave propagation in general. In Chapter 5 we examined internal waves using vertical modes; this is convenient because it leads to a description in terms of a discrete set of modes. It is tempting to extend this notion of discreteness to situations where separation of variables no longer applies: over sloping bottoms, notably. However, the suggestion from the theory of internal-wave attractors, as outlined above, is that the occurrence of a discrete set of modes is the exception rather than the rule; only for a very special geometry, namely rectangular basins, do discrete modes appear (Section 9.1.1). The general case is that where slopes occur, and there we find a continuum (here exemplified by a range of τ), along with a novel kind of wave forms.

9.2 Nonlinear effects – or the absence thereof: general remarks

We return to the set of nonlinear equations (4.17), written out in full, but under the Traditional Approximation (i.e. $\tilde{f} = 0$),

$$\frac{Du}{Dt} - fv = -\frac{1}{\rho_*}\frac{\partial p'}{\partial x} \tag{9.4a}$$

$$\frac{Dv}{Dt} + fu = -\frac{1}{\rho_*}\frac{\partial p'}{\partial y} \tag{9.4b}$$

$$\frac{Dw}{Dt} = -\frac{1}{\rho_*}\frac{\partial p'}{\partial z} + b \tag{9.4c}$$

$$u_x + v_y + w_z = 0 \tag{9.4d}$$

$$\frac{Db}{Dt} + N^2 w = 0\,. \tag{9.4e}$$

Recall that the material derivative D/Dt stands for $\partial/\partial t + \vec{u}\cdot\nabla$ and contains the advective terms, which render the equations nonlinear.

If we assume uniformity in the y-direction, i.e. $\partial/\partial y = 0$, we can introduce a streamfunction, defined via $u = \psi_z$ and $w = -\psi_x$. Eq. 9.4d) is then automatically satisfied. Furthermore, we can combine (9.4a) and (9.4c) to obtain

$$\nabla^2\psi_t + J(\nabla^2\psi, \psi) - fv_z + b_x \;=\; 0\,, \tag{9.5}$$

where $J(a,b) = a_x b_z - a_z b_x$ (the Jacobian), and $\nabla^2 = \partial_x^2 + \partial_z^2$. In terms of the streamfunction, (9.4b) and (9.4e) become

$$v_t + J(v,\psi) + f\psi_z \;=\; 0 \tag{9.6}$$

$$b_t + J(b,\psi) - N^2\psi_x \;=\; 0\,. \tag{9.7}$$

In Chapters 5 and 6 we have derived solutions of the linear equations. Surprisingly, some of them also satisfy the *non*linear equations! In the remainder of this section, we examine for which classes of solution this is the case.[1]

For constant N, we have the modal solutions from Section 5.2.2:

$$\psi \;=\; -\sum_n a_n k_n^{-1} \sin m_n z\, \sin(k_n x - \omega t) \tag{9.8}$$

$$v \;=\; \frac{f}{\omega}\sum_n a_n k_n^{-1} m_n\, \cos m_n z\, \cos(k_n x - \omega t) \tag{9.9}$$

$$b \;=\; \frac{N^2}{\omega}\sum_n a_n \sin m_n z\, \sin(k_n x - \omega t)\,, \tag{9.10}$$

[1] We should note at the outset that the approximation made in Section 4.3.1 removed a different kind of nonlinearity, namely $\rho\partial\vec{u}/\partial t + \cdots$ in the momentum equations, to which the following arguments do *not* apply.

where $m_n = n\pi/H$ denotes the vertical wavenumber. We will now select just *one* mode. Dropping the index n, we elaborate on the Jacobian $J(\nabla^2\psi, \psi)$. First, notice that $\nabla^2\psi = -(k^2 + m^2)\psi$; hence

$$J(\nabla^2\psi, \psi) = (\nabla^2\psi_x)\psi_z - (\nabla^2\psi_z)\psi_x = -(k^2 + m^2)[\psi_x\psi_z - \psi_z\psi_x] = 0\,.$$

Also, since $b \sim \sin mz \sin(kx - \omega t)$ and $\psi \sim \sin mz \sin(kx - \omega t)$ (leaving out the constant factors), it is clear that $J(b, \psi) = 0$ in (9.10).

However, with $v \sim \cos mz \cos(kx - \omega t)$, we find that

$$J(v, \psi) = v_x\psi_z - v_z\psi_x \sim km(\cos^2(kx - \omega t) - \cos^2 mz)\,,$$

which does not vanish. From this we can conclude that, in the absence of Coriolis effects (so that v plays no role), we are left with the equations (9.8) and (9.10), which are satisfied by a uni-modal solution, provided that N is constant. If N were variable, this would no longer be true: not even a uni-modal solution would in that case satisfy the nonlinear equations.

Next we consider the general solution in terms of the characteristics, (6.14), which also rests on the assumption of constant N. We will show in the next section that this is not a solution of the nonlinear equations. However, if we select just *one* the functions, F say, and take the other zero ($G = 0$), then the nonlinear equations (9.8)-(9.10) are automatically satisfied. This assumption implies that we are now considering a medium of *infinite* extension, since any boundary would engender a reflection and hence a non-zero G (see Section 6.2).

We use the expressions (6.17)–(6.19), with $G = 0$,

$$\begin{aligned} u &= \mu^{-1}F(\xi) \\ v &= -i\,\frac{f}{\omega}\mu^{-1}\,F(\xi) \\ b &= -i\,\frac{N^2}{\omega}F(\xi)\,, \end{aligned}$$

along with $w = F(\xi)$. For convenience, we have dropped the 'plus' in μ_+ and ξ_+. We recall that $\xi = \mu x - z$. Calculating now the Jacobians, we find,

$$\begin{aligned} J(\nabla^2\psi, \psi) &= -u\nabla^2 w + w\nabla^2 u = -\mu^{-1}F(\mu^2 + 1)F'' + F\mu^{-1}(\mu^2 + 1)F'' = 0 \\ J(v, \psi) &= uv_x + wv_z = -i\,\frac{f}{\omega}\mu^{-1}FF' + i\,\frac{f}{\omega}\mu^{-1}FF' = 0 \\ J(b, \psi) &= ub_x + wb_z = -i\,\frac{N^2}{\omega}FF' + i\,\frac{N^2}{\omega}FF' = 0\,. \end{aligned}$$

Hence a beam $F(\xi)$, taken in isolation (as in Figure 6.3, panel on the left), also fulfills the *non*linear equations! The same holds, of course, for G if $F = 0$. However, wherever F and G occur *together* (as in Figure 6.3, panel on the right), nonlinear effects will, in general, set in. This forms the topic of the next section.

9.3 Generation of higher harmonics

In Chapter 6 we discussed internal-wave beams, whose general form is (6.14)

$$w = F(\xi_+) + G(\xi_-)\,,$$

where F and G are arbitrary functions, and ξ_+ and ξ_- the characteristic coordinates. This was a solution of the linear equations. We now examine nonlinear effects. Specifically, two key properties will be demonstrated: 1) in the previous section it was shown that all nonlinear terms cancel if only *one* of the two, F or G, is present; now we show that nonlinear effects do arise at junctions of F and G; 2) these interactions between F and G generate higher harmonics, i.e. waves at frequencies that are a multiple of the basic frequency ω: 2ω, 3ω etc. Such junctions occur, for example, when an internal-wave beam reflects from a (sloping) bottom; the incident $F(\xi_+)$ is then reflected as $G(\xi_-)$, and in the region where the reflection occurs, the two interact. Notice that higher harmonics must travel *more steeply* than the basic wave, as follows from the dispersion relation (6.8), which shows that the steepness of energy propagation increases with frequency. These features have been demonstrated in laboratory experiments, see Figures 9.3 and 9.4.

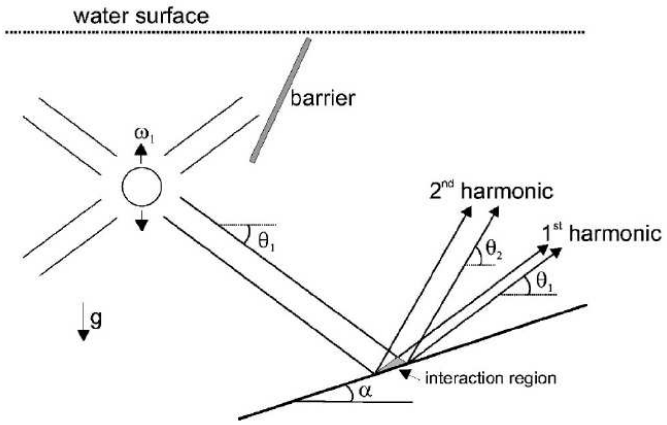

Fig. 9.3: Experimental setup: a wavemaker in a constantly stratified fluid, creating beams, one which reflects from the slope. In the region where the primary incident and reflected beams interact, higher harmonics are created. From [66].

Junctions of two crossing beams can also occur, of course, in the interior of the fluid. This may be due to the presence of different internal-wave sources, or due to internal reflections in regions of strongly varying N.

9.3.1 Formulation of the problem

An elegant derivation has been given by [83], which we generalize here by including Coriolis effects, under the Traditional Approximation. We thus start

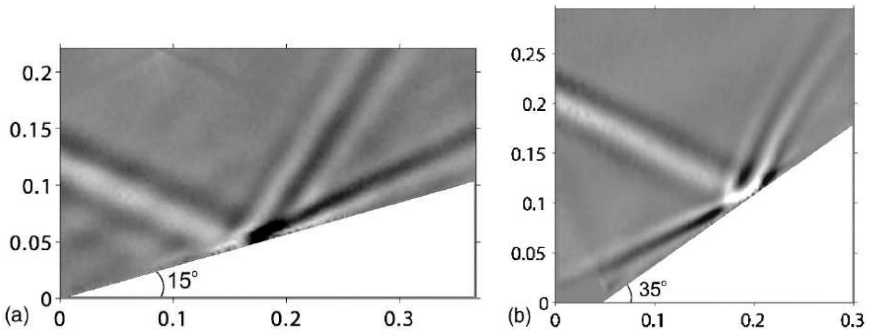

Fig. 9.4: Results from the laboratory experiments of Figure 9.3: from the region where interaction between the incident and reflected beam occurs, a steeper higher harmonic emanates. In **a**, a case of sub-critical reflection, for parameters $N = 1.28$ and $\omega = 0.536\,\mathrm{rad\,s^{-1}}$, with a slope of 15°; in **b**, super-critical reflection, with parameters $N = 1.16$ and $\omega = 0.538\,\mathrm{rad\,s^{-1}}$, and a slope at 35°. Horizontal and vertical scales are in meters. From [66].

with (9.5)–(9.7):

$$\begin{aligned} \nabla^2\psi_t + J(\nabla^2\psi,\psi) - fv_z + b_x &= 0 && (9.11)\\ v_t + J(v,\psi) + f\psi_z &= 0 && (9.12)\\ b_t + J(b,\psi) - N^2\psi_x &= 0\,, && (9.13) \end{aligned}$$

where ψ is the streamfunction, v the tranverse velocity, and b buoyancy. The buoyancy frequency N is assumed constant. Recall that the Jacobian is defined by $J(a,b) = a_x b_z - a_z b_x$, and $\nabla^2 = \partial_x^2 + \partial_z^2$.

We consider weakly nonlinear waves, monochromatic (with frequency ω) at lowest order, and write the fields in a formal expansion in which ϵ, a measure of the intensity of the wave, serves as the small parameter:

$$\begin{aligned} \psi &= \epsilon\{\Psi\exp(-i\omega t) + \text{c.c.}\} + \epsilon^2\{\Psi_0 + [\Psi_2\exp(-2i\omega t) + \text{c.c.}]\} + \cdots && (9.14)\\ v &= \epsilon\{V\exp(-i\omega t) + \text{c.c.}\} + \epsilon^2\{V_0 + [V_2\exp(-2i\omega t) + \text{c.c.}]\} + \cdots && (9.15)\\ b &= \epsilon\{\Gamma\exp(-i\omega t) + \text{c.c.}\} + \epsilon^2\{\Gamma_0 + [\Gamma_2\exp(-2i\omega t) + \text{c.c.}]\} + \cdots && (9.16) \end{aligned}$$

where we anticipate that mean fields (with index 0) and second harmonics (with index 2) will appear at order ϵ^2; the former is time-independent, the latter has frequency 2ω. (Notice that no complex conjugates are added to the mean fields; they will be found to be real by themselves.)

9.3.2 General solution

Lowest order

By substituting (9.14)–(9.16) into (9.11)–(9.13), we obtain, at lowest order,

$$\begin{aligned} -i\omega\nabla^2\Psi - fV_z + \Gamma_x &= 0 && (9.17)\\ -i\omega V + f\Psi_z &= 0 && (9.18)\\ -i\omega\Gamma - N^2\Psi_x &= 0\,, && (9.19) \end{aligned}$$

implying

$$(N^2-\omega^2)\Psi_{xx} - (\omega^2-f^2)\Psi_{zz} = 0\,, \qquad (9.20)$$

which is the familiar linear equation, (6.2). Its general solution is given by

$$\Psi = F(\xi_+) + G(\xi_-)\,, \qquad (9.21)$$

for arbitrary functions F and G, each describing propagation of wave-energy along one of the two characteristic coordinates, which are given by

$$\xi_+ = \mu_+ x - z\,; \quad \xi_- = \mu_- x - z\,,$$

with

$$\mu_\pm = \pm\Big(\frac{\omega^2-f^2}{N^2-\omega^2}\Big)^{1/2}\,.$$

The other fields can be expressed in terms of Ψ:

$$\Gamma = \frac{iN^2}{\omega}\Psi_x\,; \qquad V = -\frac{if}{\omega}\Psi_z\,. \qquad (9.22)$$

With this, the lowest-order problem has been fully solved.

Order ϵ^2; mean field

Gathering the time-independent terms resulting from the substitution of (9.14)–(9.16) in (9.11)–(9.13), we find

$$\begin{aligned} [J(\nabla^2\Psi,\Psi^*) + \text{c.c.}] - fV_{0,z} + \Gamma_{0,x} &= 0 && (9.23)\\ [J(V,\Psi^*) + \text{c.c.}] + f\Psi_{0,z} &= 0 && (9.24)\\ [J(\Gamma,\Psi^*) + \text{c.c.}] - N^2\Psi_{0,x} &= 0\,. && (9.25) \end{aligned}$$

We can use (9.22) to rewrite the Jacobian in (9.25), and then use the general identity $iJ(P_x,P^*) + \text{c.c.} = iJ(P,P^*)_x$ to obtain

$$\Psi_0 = \frac{i}{\omega}J(\Psi,\Psi^*)\,. \qquad (9.26)$$

Specifically, substitution of the lowest-order general solution (9.21) yields

$$\Psi_0 = \frac{2}{\omega}(\mu_+ - \mu_-)\text{Im}[F'(\xi_+)G'(\xi_-)^*]\,. \qquad (9.27)$$

(Primes denote derivatives with respect to the characteristic coordinates.) This expression confirms that no nonlinear contributions arise from an interaction of one plane internal wave (F, say) with itself; only junctions of plane waves, involving both F and G, provide nonlinear terms.

Equation (9.24) is now automatically satisfied, too. The remaining fields V_0 and Γ_0 cannot be uniquely determined because of geostropic degeneracy, as is clear from (9.23). This is a common problem in rectification studies; we do not pursue this point further, because the residual fields play no role in finding the second harmonics, which is the primary aspect of this section.

Order ϵ^2; second harmonic

Gathering the second harmonics $\exp(-2i\omega t)$ resulting from the substitution of (9.14)–(9.16) in (9.11)–(9.13), we find

$$\begin{aligned} -i\omega_2\nabla^2\Psi_2 + J(\nabla^2\Psi,\Psi) - fV_{2,z} + \Gamma_{2,x} &= 0 && (9.28)\\ -i\omega_2 V_2 + J(V,\Psi) + f\Psi_{2,z} &= 0 && (9.29)\\ -i\omega_2\Gamma_2 + J(\Gamma,\Psi) - N^2\Psi_{2,x} &= 0\,, && (9.30) \end{aligned}$$

where we introduced $\omega_2 = 2\omega$, the second harmonic. The set can be reduced to

$$(N^2-\omega_2^2)\Psi_{2,xx} - (\omega_2^2 - f^2)\Psi_{2,zz} = 2i\omega J(\nabla^2\Psi,\Psi) - fJ(V,\Psi)_z + J(\Gamma,\Psi)_x\,.$$

The left-hand side is identical to that in the lowest order, (9.20), but with ω replaced by the double frequency ω_2. The last two terms on the right-hand side can be simplified using the lowest-order eqs. (9.17)–(9.19), giving

$$(N^2-\omega_2^2)\Psi_{2,xx} - (\omega_2^2 - f^2)\Psi_{2,zz} = 3i\omega J(\nabla^2\Psi,\Psi)\,. \qquad (9.31)$$

The left-hand side of this equation describes the propagation of free waves at frequency 2ω; the right-hand side, the nonlinear forcing by the lowest-order terms. The general solution can be written

$$\Psi_2 = F_2(\xi_{2,+}) + G_2(\xi_{2,-}) + \Psi_{2,p}\,, \qquad (9.32)$$

where F_2 and G_2 are arbitrary functions, representing solutions in the absence of forcing; $\xi_{2,+}$ and $\xi_{2,-}$ are the characteristic coordinates defined similarly as ξ_+ and ξ_-, but now for frequency ω_2:

$$\xi_{2,+} = \mu_{2,+}x - z\,; \quad \xi_{2,-} = \mu_{2,-}x - z\,; \quad \mu_{2,\pm} = \pm\Big(\frac{\omega_2^2 - f^2}{N^2-\omega_2^2}\Big)^{1/2}\,.$$

The forcing, on the right-hand side of (9.31), is taken into account via $\Psi_{2,p}$, a particular solution to (9.31); an expression for $\Psi_{2,p}$ is derived below.

For later reference, we note that the right-hand side of (9.31) can be written in a more explicit form by using the lowest-order general solution (9.21), with $\mu = \mu_+ = -\mu_-$:

$$J(\nabla^2\Psi,\Psi) = -2\mu(1+\mu^2)\Big[F'''(\xi_+)G'(\xi_-) - G'''(\xi_-)F'(\xi_+)\Big]\,. \qquad (9.33)$$

Again we see that only products of F and G contribute.

9.3.3 Solutions for reflection from a uniform slope

We now introduce a uniform slope $z = \gamma x$, at which we require $\psi = 0$ (implying $w = \gamma u$). In the following sections we derive the solution for internal-wave reflection from the slope. The procedure is straightforward: we use the general solution obtained at first and second order, viz. (9.21), (9.27) and (9.32), and impose the boundary condition at each order.

Order ϵ; primary wave

We use the general solution (9.21), and require $\Psi = 0$ at the slope (thereby satisfying the boundary condition). This provides a coupling between F and G:

$$F([\mu_+ - \gamma]x) + G([\mu_- - \gamma]x) = 0 \quad \text{for all } x\,,$$

as is natural since one of them (G, say) now results from a reflection of the other (F). Hence, for all ξ_-,

$$G(\xi_-) = -F(\lambda\xi_-)\,, \tag{9.34}$$

with

$$\lambda = \frac{\mu_+ - \gamma}{\mu_- - \gamma}\,. \tag{9.35}$$

Without loss of generality, we may write F as

$$F(\xi_+) = \int_0^\infty dk\, a(k)\, e^{ik\xi_+}\,. \tag{9.36}$$

It is crucial that k be either exclusively positive or exclusively negative, since otherwise F cannot represent a purely incident wave. (We can ensure later that $F(\xi_+)$ actually does describe an incident wave by giving the frequency ω the appropriate sign.) The reflected wave is then represented by G in (9.34), so the total solution (9.21) becomes

$$\Psi = F(\xi_+) + G(\xi_-) = \int_0^\infty dk\, a(k)\Big[e^{ik\xi_+} - e^{ik\lambda\xi_-}\Big]\,. \tag{9.37}$$

Hereafter it will be understood that integrals are taken from zero to infinity without explicitly stating so.

Order ϵ^2; mean field

From (9.27), with (9.36) and (6.21), we obtain:

$$\Psi_0 = -\frac{4\lambda\mu}{\omega}\iint dk dk' a(k)a(k')kk' \sin(k\xi_+ - k'\lambda\xi_-)\,. \tag{9.38}$$

The boundary condition is automatically satisfied. (This can be seen by splitting the sine into the sum of its two halves, and interchanging k and k' in the second term; one then finds that the two terms cancel at the slope.)

Order ϵ^2; second harmonic

The right-hand side (9.33) can now be written

$$J(\nabla^2\Psi,\Psi) = 2\lambda\mu(1+\mu^2)\iint dkdk' a(k)a(k')kk'\Big[(k^2-(k'\lambda)^2\Big]e^{i(k\xi_+ + k'\lambda\xi_-)}\,.$$

Hence a particular solution to (9.31) is

$$\Psi_{2,p} = 6i\omega\lambda\mu\iint dkdk' S(k,k')e^{i(k\xi_+ + k'\lambda\xi_-)}\,, \tag{9.39}$$

with

$$S(k,k') = -\frac{(1+\mu^2)a(k)a(k')kk'\Big[k^2-(k'\lambda)^2\Big]}{\mu^2(N^2-\omega_2^2)(k-k'\lambda)^2-(\omega_2^2-f^2)(k+k'\lambda)^2}\,. \tag{9.40}$$

However, (9.39) does not by itself satisfy the boundary condition at the slope. We must add a solution of the homogeneous part of (9.31), which is given by F_2+G_2 in (9.32). Moreover, it must be associated with the reflected wave, so we take $F_2=0$, and select the appropriate form of $G_2(2,\xi_-)$; this will be the higher harmonic that travels away from the region of interaction. We thus arrive at

$$\Psi_2 = 6i\omega\lambda\mu\iint dkdk' S(k,k')\Big[e^{i(k\xi_+ + k'\lambda\xi_-)} - e^{iR(k,k')\xi_{2,-}}\Big]\,, \tag{9.41}$$

where

$$R(k,k') = \frac{\mu-\gamma}{\mu_{2,-}-\gamma}(k+k')\,.$$

9.3.4 Examples

We show an example in Figure 9.5, corresponding qualitatively to the experimental situation in Figure 9.4a. We evaluate the integrals from Section 9.3.3 numerically, choosing $a(k)=1$ in a certain range (k_1,k_2), and $a(k)=0$ elsewhere. In Figure 9.4, the wavemaker is a cylinder with a diameter of 12.5 mm; using this value, we find $k = 5.0\times10^2$ m^{-1}. We take this to be k_2, and choose $k_1 = k_2/2$. The strength of the harmonics depends on the amplitude of the primary wave (ϵ); however, for a qualitative comparison it suffices to show the normalized solutions of Ψ and Ψ_2. Also, we plot the streamfunction rather than the isopycnal displacements.

This qualitative comparison suggests that reflected beams are more confined in the theoretical results than in the laboratory experiments, probably due to viscous effects in the latter.

9.4 Wave-wave interactions

In the previous section we examined interactions between two waves of the same frequency. The concept of wave interaction is however more general. Here we heuristically outline the idea of resonant triads.

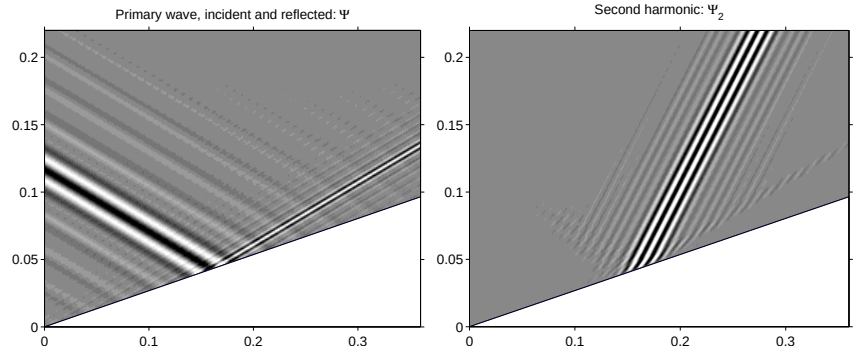

Fig. 9.5: On the left, the primary incident and reflected waves (9.37); the incident wave enters from the left. On the right, the nonlinearly generated second harmonic, (9.41). Parameters as in Figure 9.4a.

Suppose we have two waves with wavevectors $\vec{k}_1$ and $\vec{k}_2$, and frequencies ω_1 and ω_2. Each of them, $\exp i(\vec{k}_{1,2} \cdot \vec{x} - \omega_{1,2}t)$, is assumed to satisfy the linear wave equation,

$$\nabla^2 w_t + f^2 w_{zz} + N^2 \nabla_h^2 w = 0\,,$$

which means that each must satisfy the dispersion relation

$$H(\omega, \vec{k}) = (N^2 - \omega^2)(k^2 + l^2) - (\omega^2 - f^2)m^2 = 0\,.$$

Now, nonlinear terms like the Jacobians in (9.5)–(9.7), give rise 'forcing' terms to the linear remainder, symbolically written as $L(w)$,

$$\begin{aligned} L(w) &= \exp i(\vec{k}_1 \cdot \vec{x} - \omega_1 t)[\exp i(\vec{k}_2 \cdot \vec{x} - \omega_2 t) + \text{c.c.}] \\ &= \exp i([\vec{k}_1 + \vec{k}_2] \cdot \vec{x} - [\omega_1 + \omega_2]t) + \exp i([\vec{k}_1 - \vec{k}_2] \cdot \vec{x} - [\omega_1 - \omega_2]t)\,. \end{aligned}$$

If either the sum or the difference is itself a solution of the linear equation, i.e. either $H(\omega_1 + \omega_2, \vec{k}_1 + \vec{k}_2) = 0$ or $H(\omega_1 - \omega_2, \vec{k}_1 - \vec{k}_2) = 0$, then the right-hand side will act as a resonant forcing, giving rise to a rapid growth of a wave with wavevector $\vec{k}_3 = \vec{k}_1 \pm \vec{k}_2$ and frequency $\omega_3 = \omega_1 \pm \omega_2$ (the plus or minus chosen according to which of them satisfies the dispersion relation).

In such cases it is necessary to modify the solution from the start to allow for a 'slow' growth or decay; in other words, to include an amplitude factor depending on a long time-scale $T = \epsilon t$. At lowest order, this term acts simply as a constant factor; but at the next order, it gives rise to a time-derivative $\epsilon \partial/\partial T$, which should match the resonant forcing terms, appearing at the same order. This yields the amplitude equations describing the interaction among the three waves.

One important case is 'parametric subharmonic instability' in which energy is transferred from low to high wavenumbers of half the basic frequency. This mechanism may be at work near 'critical latitudes'. At 28.8°N/S, for example, half the semi-diurnal lunar frequency, $\frac{1}{2}\omega_{M_2}$, exactly matches the local Coriolis

parameter $|f|$. Equatorward of this latitude these diurnal tides can exist as a free internal wave. Numerical and observational evidence suggests that they are resonantly forced by M_2 in the vicinity of this latitude.

9.5 Internal-wave spectra

The previous sections described mechanisms that transfer internal-wave energy to other frequencies. As discussed in Section 1.3, the principal sources of internal waves – wind, and tidal flow over slopes – produce *low*-frequency waves. The transfer to other, mostly higher frequencies causes the internal-wave spectrum to be filled and smoothed; as a result, internal waves are found in the whole range of possible frequencies, $|f|$ to N. Internal-wave spectra deduced from observations usually share some general characteristics, which are described by the so-called Garrett-Munk spectrum (of which various versions exist). This spectrum has not been derived on theoretical grounds, but is rather based on – as one of its creators has put it – "rank empiricism" [61].

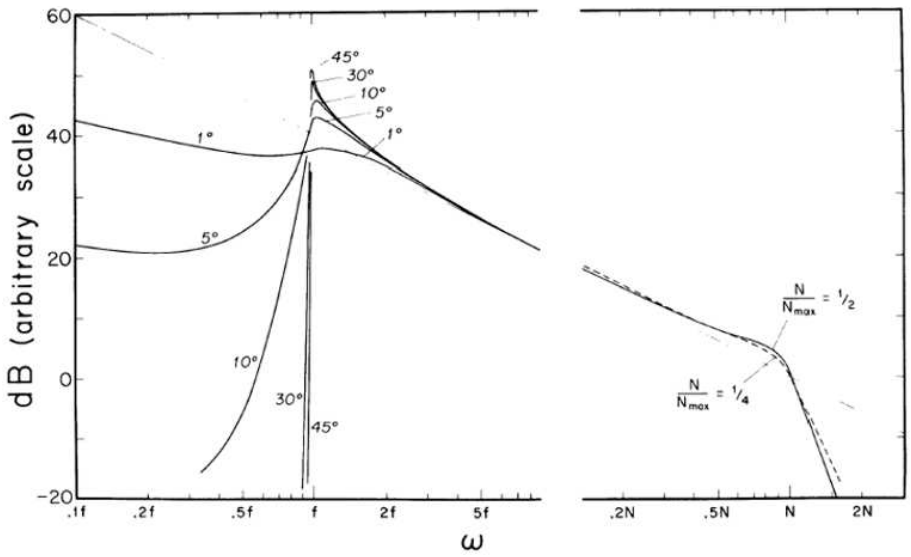

Fig. 9.6: The Garrett-Munk spectrum, at different latitudes, showing a peak at the inertial frequency (f), and a steady fall-off for higher frequencies. From [61].

The form given by [61] for the energy-density distribution is

$$E_{GM}(\omega, n, z) = b^2 N_0 N(z) E_0 B(\omega) H(n)\,, \tag{9.42}$$

which contains the following functions

$$B(\omega) = \frac{2f}{\pi\omega}\frac{1}{(\omega^2 - f^2)^{1/2}} \tag{9.43}$$

$$H(n) = \frac{(n^2 + n_*^2)^{-1}}{\sum_{m=1}^{\infty}(m^2 + n_*^2)^{-1}} \tag{9.44}$$

$$N(z) = N_0 \exp(z/b)\,. \tag{9.45}$$

The constant parameters are

$$b = 1.3 \text{ km}\,; \quad N_0 = 5.2 \times 10^{-3} \text{ rad s}^{-1}\,; \quad n_* = 3\,; \quad E_0 = 6.3 \times 10^{-5}\,.$$

The profile $N(z)$ gives a model profile of the stratification in the ocean below the seasonal thermocline; n is the modenumber. An example is shown in Figure 9.6. The important features are that most energy is contained at the lower frequencies, with a decay for higher frequencies at a power -2.

However, in observed internal-wave spectra, the inertial peak is often more pronounced than in the Garrett-Munk spectrum. As matter of fact, the inertial band – the generation and propagation of near-inertial waves – is not yet well understood. Other departures from the Garrett-Munk spectrum occur close to generation regions of internal tides; there one finds a strong peak at the dominant tidal frequency (usually M_2), as well as peaks at frequencies that are multiples of the basic frequency (e.g., M_4, M_6, etc). That they should be present is not surprising in the light of the analysis of the generation of higher harmonics in Section 9.3. An internal-wave spectrum derived from ADCP (current) measurements in the Bay of Biscay is shown in Figure 9.7; the tidal peak and corresponding higher harmonics are conspicuous.

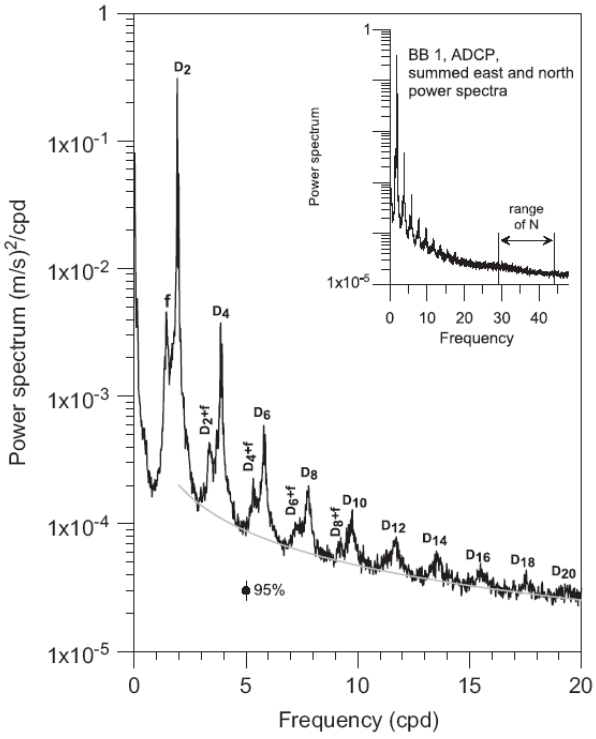

Fig. 9.7: Mean power spectrum of the horizontal velocity, obtained by summing the power spectra of the east and north velocity components. The inserted figure shows the entire internal-wave band (the arrows delineate the range of N). The main figure shows the low-frequency part, containing the inertial and tidal peaks; D_2 stands for "semi-diurnal". The grey line is added to show that the overall decay follows a $\omega^{-0.9}$ curve. From [85].

Bibliography

[1] M. J. Ablowitz and P. A. Clarkson. *Solitons, nonlinear evolution equations and inverse scattering*. Cambridge University Press, 1991.

[2] T. R. Akylas, R. H. J. Grimshaw, S. R. Clarke, and A. Tabaei. Reflecting tidal wave beams and local generation of solitary waves in the ocean thermocline. *J. Fluid Mech.*, 593:297–313, 2007.

[3] J. R. Apel, L. A. Ostrovsky, Y. A. Stepanyants, and J. F. Lynch. *Internal solitons in the ocean*. Technical Report WHOI-2006-04, Woods Hole., 2006.

[4] C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers*. McGraw-Hill, New York, 1978.

[5] V. Bjerknes. *Die Meteorologie als exakte Wissenschaft*. Vieweg, Braunschweig, 1913.

[6] L. M. Brekhovskikh. *Waves in layered media*. Academic Press, New York, 1960.

[7] D. E. Cartwright. *Tides: a scientific history*. Cambridge University Press, 1999.

[8] P. D. Craig. Solutions for internal tidal generation over coastal topography. *J. Mar. Res.*, 45:83–105, 1987.

[9] D. G. Crighton. Applications of KdV. *Acta Appl. Math.*, 39:39–67, 1995.

[10] J. C. B. Da Silva, A. L. New, and A. Azevedo. On the role of SAR for observing "local generation" of internal solitary waves off the Iberian Peninsula. *Can. J. Remote Sensing*, 33(5):388–403, 2007.

[11] E. M. De Jager. On the origin of the Korteweg-de Vries equation. *arXiv:math/0602661*, v1:1–23, 2006. [see http://arxiv.org/abs/math/0602661].

[12] V. D. Djordjevic and L. G. Redekopp. The fission and disintegration of internal solitary waves moving over two-dimensional topography. *J. Phys. Oceanogr.*, 8:1016–1024, 1978.

[13] J. A. Dutton. *Dynamics of atmospheric motion*. Dover, 1995. [NB: formerly titled *The ceaseless wind*].

[14] W. Eckhaus and A. van Harten. *The inverse scattering transformation and the theory of solitons: an introduction*. North-Holland Publishing Company, 1981.

[15] G. D. Egbert and R. D. Ray. Semi-diurnal and diurnal tidal dissipation from TOPEX/Poseidon altimetry. *Geophys. Res. Lett.*, 30, 17, 1907:doi:10.1029/2003GL017676, 2003.

[16] A. Einstein and L. Infeld. *The evolution of physics*. Simon and Schuster, New York, 1938.

[17] R. Feistel and E. Hagen. On the Gibbs thermodynamic potential of seawater. *Prog. Oceanog.*, 36:249–327, 1995.

[18] N. P. Fofonoff. Physical properties of sea-water. In M. N. Hill, editor, *The Sea, Vol. I*, pages 3–30. Wiley, New York, 1962.

[19] D. C. Fritts and M. J. Alexander. Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.*, 41(1):1003, doi:10.1029/2001RG000106, 2003.

[20] C. Garrett. Internal tides and ocean mixing. *Science*, 301:1858–1859, 2003a.

[21] C. Garrett. Mixing with latitude. *Nature*, 422:477–478, 2003b.

[22] C. Garrett and T. Gerkema. On the body-force term in internal-tide generation. *J. Phys. Oceanogr.*, 37(8):2172–2175, 2007.

[23] C. Garrett and E. Kunze. Internal tide generation in the deep ocean. *Annu. Rev. Fluid Mech.*, 39:57–87, 2007.

[24] C. Garrett and L. St. Laurent. Aspects of deep ocean mixing. *J. Oceanogr.*, 58:11–24, 2002.

[25] T. Gerkema. A unified model for the generation and fission of internal tides in a rotating ocean. *J. Mar. Res.*, 54(3):421–450, 1996.

[26] T. Gerkema. Internal and interfacial tides: beam scattering and local generation of solitary waves. *J. Mar. Res.*, 59(2):227–255, 2001.

[27] T. Gerkema, F. P. A. Lam, and L. R. M. Maas. Internal tides in the Bay of Biscay: conversion rates and seasonal effects. *Deep-Sea Res. II*, 51(25/26):2995–3008, 2004.

[28] T. Gerkema and V. I. Shrira. Near-inertial waves in the ocean: beyond the 'Traditional Approximation'. *J. Fluid Mech.*, 529:195–219, 2005.

[29] T. Gerkema and V. I. Shrira. Non-traditional reflection of internal waves from a sloping bottom, and the likelihood of critical reflection. *Geophys. Res. Lett.*, 33(L06611):doi:10.1029/2005GL025627, 2006.

[30] T. Gerkema, J. T. F. Zimmerman, L. R. M. Maas, and H. van Haren. Geophysical and astrophysical fluid dynamics beyond the Traditional Approximation. *Rev. Geophys.*, doi:10.1029/2006RG000220, 2008. [in press].

[31] J. W. Gibbs. *The Collected Works of J. Willard Gibbs. Volume I: Thermodynamics*. Yale University Press, New Haven, 1948.

[32] G. S. Giese, D. C. Chapman, M. G. Collins, R. Encarnacion, and G. Jacinto. The coupling between harbor seiches at Palawan Island and Sulu Sea internal solitons. *J. Phys. Oceanogr.*, 28:2418–2426, 1998.

[33] A. E. Gill. On the behavior of internal waves in the wakes of storms. *J. Phys. Oceanogr.*, 14:1129–1151, 1984.

[34] E. E. Gossard and W. H. Hooke. *Waves in the atmosphere*. Elsevier, Amsterdam, 1975.

[35] H. P. Greenspan. *The theory of rotating fluids*. Cambridge University Press, 1968.

[36] R. Grimshaw. Internal solitary waves. In R. Grimshaw, editor, *Environmental stratified flows*, pages 1–27. Kluwer, 2001.

[37] R. Grimshaw, L. A. Ostrovsky, V. I. Shrira, and Yu. A. Stepanyants. Long nonlinear surface and internal gravity waves in a rotating ocean. *Surv. Geophys.*, 19:289–338, 1998.

[38] R. H. J. Grimshaw. A note on the β-plane approximation. *Tellus*, 27(4):351–356, 1975.

[39] P. Groen. *The Waters of the Sea*. Van Nostrand, London, 1967.

[40] K. R. Helfrich and W. K. Melville. Long nonlinear internal waves. *Annu. Rev. Fluid Mech.*, 38:395–425, 2006.

[41] B. Helland-Hansen and F. Nansen. *The Norwegian Sea – its Physical Oceanography based upon the Norwegian researches 1900-1904. (Report on Norwegian Fishery and Marine Investigations, Vol. II, No. 2)*. Det Mallingske Bogtrykkeri, Kristiania, 1909.

[42] D. Holliday and M. E. McIntyre. On potential energy density in an incompressible, stratified fluid. *J. Fluid Mech.*, 107:221–225, 1981.

[43] P. M. Holligan, R. D. Pingree, and G. T. Mardell. Ocean solitons, nutrient pulses and phytoplankton growth. *Nature*, 314:348–350, 1985.

[44] E. L. Ince. *Ordinary differential equations*. Dover, 1956.

[45] T. Kakutani and N. Yamasaki. Solitary waves on a two-layer fluid. *J. Phys. Soc. Jap.*, 45(2):674–679, 1978.

[46] D. K. Kondepudi and I. Prigogine. *Modern thermodynamics: from heat engines to dissipative structures.* Wiley, 1998.

[47] W. Krauss. *Methoden und Ergebnisse der theoretischen Ozeanographie II: Interne Wellen.* Gebrüder Bornträger, Berlin, 1966.

[48] P. H. LeBlond and L. A. Mysak. *Waves in the ocean.* Elsevier, Amsterdam, 1978.

[49] B. LeCann. Barotropic tidal dynamics of the Bay of Biscay shelf: observations, numerical modelling and physical interpretation. *Cont. Shelf Res.*, 10(8):723–758, 1990.

[50] C. Lee and R. C. Beardsley. The generation of long nonlinear internal waves in a weakly stratified shear flow. *J. Geophys. Res.*, 1974.

[51] M. J. Lighthill. *Waves in fluids.* Cambridge University Press, 1978.

[52] C. C. Lin and L. A. Segel. *Mathematics applied to determinisitic problems in the natural sciences.* Macmillan, New York, 1974.

[53] L. R. M. Maas. Wave attractors: linear yet nonlinear. *Int. J. Bifurc. Chaos*, 15(9):2757–2782, 2005.

[54] L. R. M. Maas, D. Benielli, J. Sommeria, and F. P. A. Lam. Observation of an internal wave attractor in a confined, stably stratified fluid. *Nature*, 388:557–561, 1997.

[55] O. I. Mamayev. *Temperature-salinity analysis of world ocean waters.* Elsevier, Amsterdam, 1975.

[56] E. Merzbacher. *Quantum mechanics.* Wiley, New York, 3rd edition, 1998.

[57] J. W. Miles. The Korteweg-de Vries equation: a historical essay. *J. Fluid Mech.*, 106:131–147, 1981.

[58] Yu. Z. Miropol'sky. *Dynamics of internal gravity waves in the ocean.* Kluwer, Dordrecht, 2001. [The Russian original dates from 1981].

[59] R. M. Miura. The Korteweg-de Vries equation: a survey of results. *SIAM Rev.*, 18:412–459, 1976.

[60] J. N. Moum, D. M. Farmer, W. D. Smyth, L. Armi, and S. Vagle. Structure and generation of turbulence at interfaces strained by internal solitary waves propagating shoreward over the continental shelf. *J. Phys. Oceanogr.*, 33:2093–2112, 2003.

[61] W. Munk. Internal waves and small-scale processes. In B. A. Warren and C. Wunsch, editors, *Evolution of physical oceanography*, pages 264–291. MIT press, 1981.

[62] A. L. New and J. C. B. Da Silva. Remote-sensing evidence for the local generation of internal soliton packets in the central Bay of Biscay. *Deep-Sea Res.*, 49(5):915–934, 2002.

[63] J. Nycander. Generation of internal waves in the deep ocean by tides. *J. Geophys. Res*, 110(C10028):doi:10.1029/2004JC002487, 2005.

[64] A. R. Osborne and T. L. Burch. Internal solitons in the Andaman Sea. *Nature*, 208:451–460, 1980.

[65] L. A. Ostrovsky and Yu. A. Stepanyants. Do internal solitons exist in the ocean? *Rev. Geophys.*, 27(3):293–310, 1989.

[66] T. Peacock and A. Tabaei. Visualization of nonlinear effects in reflecting internal wave beams. *Phys. Fluids*, 17(061702):1–4, 2005.

[67] O. M. Phillips. *The dynamics of the upper ocean.* Cambridge University Press, 1966.

[68] G. L. Pickard and W. J. Emery. *Descriptive physical oceanography.* Pergamon Press, Oxford, 5th edition, 1990.

[69] R. D. Pingree and A. L. New. Abyssal penetration and bottom reflection of internal tide energy in the Bay of Biscay. *J. Phys. Oceanogr.*, 21(1):28–39, 1991.

[70] Lord Rayleigh. *The theory of sound.* Macmillan, London, 1929.

[71] M. J. Reeder, D. R. Christie, R. K. Smith, and R. Grimshaw. Interacting "Morning Glories" over Northern Australia. *Bull. Amer. Meteor. Soc.*, 76(7):1165–1171, 1995.

[72] J. Roberts. *Internal gravity waves in the ocean.* Marcel Dekker, Inc., New York, 1975.

[73] J. W. Rottman and R. Grimshaw. Atmospheric internal solitary waves. In R. Grimshaw, editor, *Environmental stratified flows*, pages 61–88. Kluwer, 2001.

[74] P. M. Saunders. Practical conversion of pressure to depth. *J. Phys. Oceanogr.*, 11:573–574, 1981.

[75] L. A. Segel. *Mathematics applied to continuum mechanics.* Macmillan, New York, 1977.

[76] J. B. Serrin. Mathematical principles of classical fluid mechanics. In S. Flügge, editor, *Handbuch der Physik, VIII/1*, pages 125–263. Springer Verlag, Berlin, 1959.

[77] H. L. Simmons, S. R. Jayne, L. C. St. Laurent, and A. J. Weaver. Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Modell.*, 6:245–263, 2004.

[78] A. Sommerfeld. *Vorlesungen über theoretische Physik, Band V: Thermodynamik und Statistik*. Dieterich'sche Verlagsbuchhandlung, Wiesbaden, 1952. [English translation: *Lectures on theoretical physics, Volume V: thermodynamics and statistical mechanics*, Academic Press, 1956].

[79] L. St. Laurent, S. Stringer, C. Garrett, and D. Perrault-Joncas. The generation of internal tides at abrupt topography. *Deep-Sea Res. I*, 50:987–1003, 2003.

[80] T. P. Stanton and L. A. Ostrovsky. Observations of highly nonlinear internal solitons over the continental shelf. *Geophys. Res. Lett.*, 25(14):2695–2698, 1998.

[81] R. B. Stull. Static stability – an update. *Bull. Amer. Meteor. Soc.*, 72(10):1521–1529, 1991.

[82] R. D. Susanto, L. Mitnik, and Q. Zheng. Ocean internal waves observed in the Lombok Strait. *Oceanogr.*, 18(4):80–87, 2005.

[83] A. Tabaei, T. R. Akylas, and K. G. Lamb. Nonlinear effects in reflecting and colliding internal wave beams. *J. Fluid Mech.*, 526:217–243, 2005.

[84] J. S. Turner. *Buoyancy effects in fluids*. Cambridge University Press, 1973.

[85] H. M. van Aken, H. van Haren, and L. R. M. Maas. The high-resolution vertical structure of internal tides and near-inertial waves measured with an ADCP over the continental slope in the Bay of Biscay. *Deep-Sea Res. I*, 54:533–556, 2007.

[86] F. van der Blij. Some details of the history of the Korteweg-de Vries equation. *Nieuw Archief voor Wiskunde*, 3(XXVI):54–64, 1978.

[87] G. Veronis. On properties of seawater defined by temperature, salinity, and pressure. *J. Mar. Res.*, 30(2):227–255, 1972.

[88] V. Vlasenko, N. Stashchuk, and K. Hutter. *Baroclinic tides: theoretical modeling and observational evidence*. Cambridge University Press, 2005.

[89] P. Welander. On the ocean heat engine, stiffness of chaotic systems and climate prediction. *Tellus*, 43AB:116–120, 1991.

[90] G. B. Whitham. *Linear and nonlinear waves*. Wiley, 1974.

[91] N. J. Zabusky and M. D. Kruskal. Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Lett*, 15(6):240–243, 1965.