# 12.333 Atmospheric and Oceanic Circulations

R. Alan Plumb

Spring 2003

# Contents

Lecture times: TR 11-1230

In this course, we will look at many important aspects of the circulation of the atmosphere and ocean, from length scales of meters to thousands of km and time scales ranging from seconds to years. We will assume familiarity with concepts covered in course 12.003 (Physics of the Fluid Earth). In the early stages of the present course, we will make somewhat greater use of math than did 12.003, but the math we will use is no more than that encountered in elementary electromagnetic field theory, for example. The focus of the course is on the *physics* of the phenomena we will discuss.

The assessment will be in three parts (with weight toward final grade):

1. Homework—five homework assignments, handed out approximately every two weeks and on material covered in the previous 2 weeks (40%).

2. Two mid-term quizzes (10% each)

3. A final, closed book, exam (40%)

There is no set text—notes will be handed out at each class and references for background or further reading will be pointed out wherever I think they will be useful.

# Chapter 1

# Shallow water gravity waves

## 1.1   Surface motions on shallow water

Consider two-dimensional ($x$-$z$) motions on a nonrotating, shallow body of water, of uniform density $\rho$, as shown in Fig. 1.1 below.
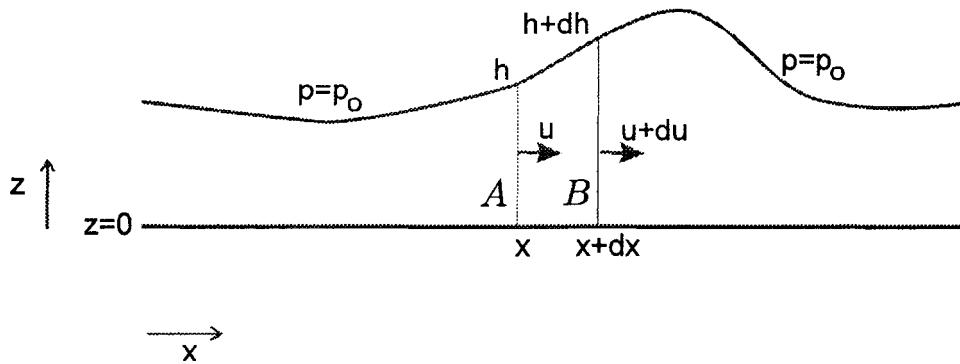


Figure 1.1: The shallow water system.

The flow is assumed to be inviscid and independent of the spatial dimension $y$ (into the paper). We shall *assume* that the water is so shallow that the flow velocity $u(x,t)$ is constant with depth. (We'll see later under what conditions this is reasonable; for now, let's just assume it to be true.) At the free surface, located at height $z = h(x,t)$, pressure is equal to atmospheric pressure $p_0$, assumed constant and uniform.

1

Consider the volume of water bounded by the vertical surfaces $A$ and $B$ in the figure. These surfaces are located at $x$ and $x + dx$ respectively. The mass of this volume, per unit length in $y$, is just $dm = \rho h\, dx$. Now, mass cannot be created or destroyed within the volume, so the only way it can change is because of the **fluxes** of mass across the interfaces $A$ and $B$. Consider Fig. 1.2.
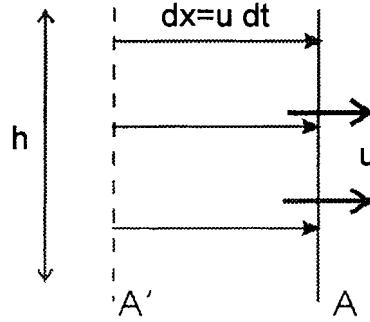


Figure 1.2: Illustrating the flux of mass across the interface $A$.

Since the velocity at $A$ is $u$, in a time interval $dt$ all the fluid between $A'$ and $A$ passes across $A$, where the distance between $A'$ and $A$ is $dx = u\, dt$. Thus, the area (i.e., the volume per unit length in $y$) passing across $A$ in this time is $hu\, dt$, and so the mass (per unit length in $y$) is $\rho hu\, dt$. Therefore the **mass flux**—the mass crossing $A$ per unit time, per unit length in $y$—is $\rho u(x) h(x)$. The mass flux across interface $B$ is $\rho u(x + dx) h(x + dx)$ (directed toward positive $x$, out of the volume). Therefore the rate of accumulation of mass (per unit length in $y$) within the volume defined by $AB$ is

$$\frac{\partial m}{\partial t} = \rho u(x) h(x) - \rho u(x + dx) h(x + dx)$$

$$= -\rho \frac{\partial (uh)}{\partial x}\, dx\ .$$

Since $m = \rho h\, dx$, the factors of $\rho\, dx$ cancel, leaving us with

$$\frac{\partial h}{\partial t} = -\frac{\partial (uh)}{\partial x}\ .$$

Differentiating the RHS by parts and rearranging, we arrive at the **equation of continuity**:

$$\frac{\partial h}{\partial t} + u \frac{\partial h}{\partial x} = -h \frac{\partial u}{\partial x}\ . \tag{1.1}$$

This equation expresses the local rate of change of surface height in terms of two contributions:

(i) by advection of height $-u\partial h/\partial x$

(ii) by **volume convergence** $-h\partial u/\partial x$ .

These two effects are depicted (both in a sense to increase $h$ locally) in Fig. 1.3.



advection                    convergence
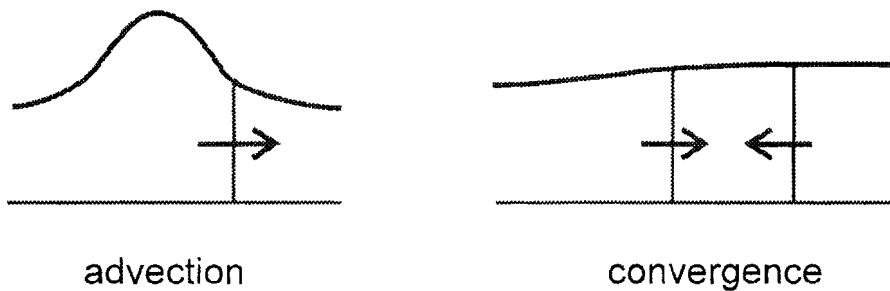
Figure 1.3: Contributions to $\partial h/\partial t$.

Now, in a similar way, consider the momentum balance of the water in the volume. We shall need to know the distribution of pressure $p$ within the water. To do this, we use the principle of **hydrostatic balance**, which states that the pressure increases with depth according to the overhead mass per unit area. Specifically (see Fig. 1.1), the pressure at any depth $h - z$ below the surface is related to surface pressure by

$$p(z,t) = p_0 + \int_z^h \rho g \, dz = p_0 + \rho g(h - z) \, , \qquad (1.2)$$

where $g$ is the acceleration due to gravity (and both $\rho$ and $g$ are constants). The second term on the RHS of (1.2) simply represents the mass of water per unit area above level $z$. Newton's law of motion applied to the volume gives

$$m\frac{du}{dt} = F \, ,$$

where $F$ is the net force (per unit length in $y$) applied to the volume. Since we are assuming friction to be negligible, the only such forces acting are pressure forces, which are as depicted in Fig. 1.4[1].
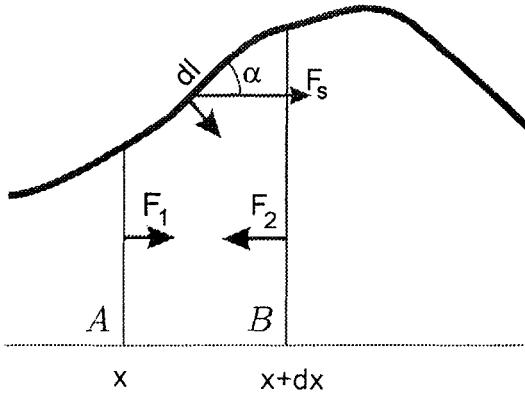
Figure 1.4: Forces acting on the fluid volume.

That acting on the volume across interface $A$ (tending to accelerate the volume in the positive $x$ direction) is equal to a force, per unit length in $y$, of $F_1 = \int_0^h p(x, z)\, dz$; that acting across interface $B$ (tending to accelerate the volume in the negative $x$ direction) is $F_2 = \int_0^h p(x + dx, z)\, dz$. However, there is a third component of the net force acting on the free surface, represented in the figure as $F_s$. Atmospheric pressure exerts a force $p_0 dl$ per unit length in $y$, where $dl$ is the volume's width along the surface. Because the surface is tilted, this has a nonzero component $p_0 dl \sin \alpha$ acting in the positive $x$-direction, where $\alpha$ is the angle of the interface. Since $dl = dx / \cos \alpha$, this contribution to the $x$-force is

$$F_s = p_0 \frac{\partial h}{\partial x} dx$$

(since $\tan \alpha = \partial h / \partial x$). Therefore the net force on the volume, per unit length in $y$, is

$$F = p_0 \frac{\partial h}{\partial x} dx + \int_0^h p(x, z)\, dz - \int_0^h p(x + dx, z)\, dz \ .$$

[1]We are in fact neglecting here one contribution to the force felt at the surface, that due to surface tension. Surface tension effects are negligible for motions of large horizontal scale (typically a few cm.), so this analysis is restricted to these large scales. Small-scale motions for which surface tension effects are important are known as *capillary waves*.

But, from (1.2), we have

$$
\begin{aligned}
\int_0^h p\, dz &= \int_0^h p_0\, dz + \rho g \int_0^h (h-z)\, dz\ , \\
&= p_0 h + \frac{1}{2}\rho g h^2\ .
\end{aligned}
$$

So

$$
\begin{aligned}
\int_0^h p(x,z)\, dz - \int_0^h p(x+dx,z)\, dz &= p_0 h(x) - p_0 h(x+dx) \\
&\quad + \frac{1}{2}\rho g h^2(x) - \frac{1}{2}\rho g h^2(x+dx) \\
&= -p_0 \frac{\partial h}{\partial x} dx - \rho g h \frac{\partial h}{\partial x} dx\ .
\end{aligned}
$$

Therefore the acceleration of the volume is given by

$$
m\frac{du}{dt} = F = -\rho g h \frac{\partial h}{\partial x} dx\ .
$$

Note that this is independent of surface pressure $p_0$ (the terms involving it have cancelled): the net force on the volume is entirely due to the pressure gradients within the water which, because of hydrostatic balance, are entirely due to gradients in surface height. Now, using our expression $m = \rho h\, dx$, the preceding equation gives us (cancelling the factors $\rho h\, dx$)

$$
\frac{du}{dt} = -g\frac{\partial h}{\partial x}
$$

Here the derivative $d/dt$ is the *material derivative*—this tells us how the velocity of the marked volume changes *as it moves around*. We need to convert this into a form that tells us how $u$ changes in fixed coordinates. Since $u = u(x,t) \equiv dx/dt$, we simply apply the chain rule to write

$$
\frac{du}{dt} = \frac{\partial u}{\partial t} + \frac{dx}{dt}\frac{\partial u}{\partial x} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}
$$

and thus to write our equation of motion in final form

$$
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -g\frac{\partial h}{\partial x}\ . \tag{1.3}
$$

Like (1.1), this links the local rate of change of velocity to two terms:

(i)  the pressure gradient term and

(ii)  the *advection of momentum.*

The two equations (1.1) and (1.3) give us two *predictive* equations in the two unknowns $u(x,t)$ and $h(x,t)$, and so in principle tell us all we need to know to determine how this system will evolve, given initial and boundary conditions. The equations are *nonlinear* (through the advective terms) and have complex properties in general, but they become quite simple under circumstances where they can be linearized.

## 1.2  Small-amplitude shallow-water surface waves

Suppose now our shallow water system is motionless ($u = 0$), with uniform depth $D$; this state trivially satisfies eqs. (1.1) and (1.3). Now suppose we perturb this state, such that $u(x,t) = u'(x,t)$ and $h(x,t) = D + h'(x,t)$, where the perturbation is small in the sense that

(i)  $|h'| \ll D$, and

(ii)  $|u'| \ll L/T$,

where $L$ and $T$ are respectively length and time scales for the motion. Now, eq. (1.1) becomes

$$\frac{\partial h'}{\partial t} + u'\frac{\partial h'}{\partial x} = -(D + h')\frac{\partial u'}{\partial x} \ ,$$

since the derivatives of $D$ are zero. We now replace $(D + h')$ by $D$ (using assumption (i)) and neglect the second term compared to the first (since $\partial h'/\partial t \sim |h'|\,/T$ and $u'\partial h'/\partial x \sim |u'|\,|h'|\,/L$, so the ratio of the latter to the former is $|u'|\,/LT$, which is small by assumption (ii)), leaving the **linearized** equation

$$\frac{\partial h'}{\partial t} = -D\frac{\partial u'}{\partial x} \ . \tag{1.4}$$

Similarly, (1.3) becomes

$$\frac{\partial u'}{\partial t} + u'\frac{\partial u'}{\partial x} = -g\frac{\partial h'}{\partial x} \ ;$$

again we can use assumption (ii) to neglect the second term, leaving

$$\frac{\partial u'}{\partial t} = -g\frac{\partial h'}{\partial x} \ . \tag{1.5}$$

We can now combine the two equations (1.4) and (1.5) to get a single equation for $h$, by combining $\partial/\partial t$ of (1.4):

$$\frac{\partial^2 h'}{\partial t^2} = -D\frac{\partial^2 u'}{\partial x \partial t}$$

with $\partial/\partial x$ of (1.5):

$$\frac{\partial^2 u'}{\partial x \partial t} = -g\frac{\partial^2 h'}{\partial x^2}$$

to give

$$\frac{\partial^2 h'}{\partial t^2} - gD\frac{\partial^2 h'}{\partial x^2} = 0 \ . \tag{1.6}$$

This is a **wave equation**, which describes how small-amplitude surface height perturbations evolve.

## 1.3    Background theory—nondispersive waves

### 1.3.1    Oscillations

Oscillations (*e.g.*, small amplitude oscillations of a simple pendulum) are often described by an equation of the form
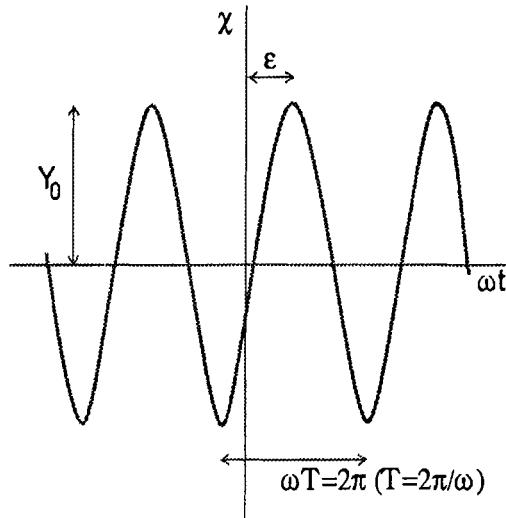
Figure 1.5: Characteristics of a simple oscillation.

$$\frac{d^2\chi}{dt^2} + \Omega^2\chi = 0 \tag{1.7}$$

where $t$ is time and $\chi$ is some system variable (angular displacement in the case of the simple pendulum). (1.7) has solutions of the form

$$\chi(t) = Re\left(Y\,e^{i\omega t}\right) = Y_0\cos(\omega t - \epsilon) \tag{1.8}$$

where frequency $\omega = \pm\Omega$, amplitude $Y_0$ and phase $\epsilon$ are real constants, and $Y = Y_0 e^{-i\epsilon}$ is a complex amplitude. Thus, as shown in Fig. 1.5, an oscillation is characterized by three constants: amplitude, frequency, and phase.

## 1.3.2    Nondispersive waves

Unlike such simple oscillations, waves are functions of both time and space. The simplest wave equation, of which (1.6) is an example, is of the form

$$\frac{\partial^2\chi}{\partial t^2} - c_0^2\frac{\partial^2\chi}{\partial x^2} = 0 \tag{1.9}$$

where $c_0$ is some constant. (In our case, $\chi$ represents surface height perturbations on shallow water of depth $D$ and $c_0 = \sqrt{gD}$. However, it could equally

well represent the electric or magnetic field *in vacuo*, with $c_0$ the speed of light; or pressure perturbations in a compressible fluid, with $c_0$ the sound speed.)

We can find solutions to (1.9) by *separating the variables*, writing

$$\chi(x, t) = Re\left[A(t)B(x)\right] .$$

More specifically, if we look for "wave-like" solutions for which $B(x) = e^{ikx}$, where $k$ is a real wavenumber (so $2\pi/k$ is wavelength)[2], then $d^2 B/dx^2 = -k^2 B$, so

$$\frac{\partial^2 \chi}{\partial x^2} = Re\left[A(t)\frac{d^2 B}{dx^2}(x)\right] = -k^2 Re\left[A(t)B(x)\right] ,$$

and (1.9) becomes

$$\frac{d^2 A}{dt^2} + k^2 c_0^2 A = 0 .$$

This has solutions like

$$A = \chi_+ e^{-i\omega t} \; ; \; A = \chi_- e^{+i\omega t}$$

where $\chi_+$ and $\chi_-$ are constant (complex) amplitudes and the frequency $\omega$ satisfies

$$\omega^2 = k^2 c_0^2 . \tag{1.10}$$

The full solution is

$$\chi(x, t) = Re\left[\chi_+ e^{i(kx - \omega t)} + \chi_- e^{i(kx + \omega t)}\right] . \tag{1.11}$$

Each of the two terms in (1.11) describes a **progressive wave** (Fig. 1.6):

---

[2]In general, any function of $x$ can be expressed as a Fourier integral of such waves.
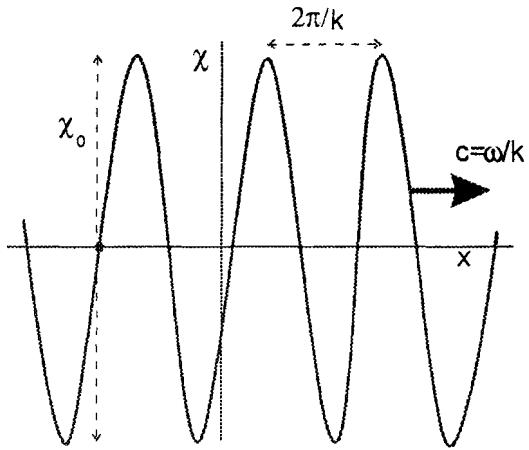
Figure 1.6: Characteristics of a progressive wave.

- at any instant, it is just a sinusoidal wave disturbance, of **wavelength** $2\pi/k$

- at any fixed location $x = x_0$, it is just an oscillation of the form $De^{\pm i\omega t}$, where $D = \chi_\pm e^{ikx_0}$ is its complex amplitude, of **period** $T = 2\pi/\omega$

- it **propagates** with **phase speed** $c = \omega/k = \pm c_0$. Note from (1.11) that $\chi$ is constant along characteristics with $kx \pm \omega t$ =constant, *i.e.*, $x = \mp\frac{\omega}{k}t$+constant.

Eq. (1.10) is the **dispersion relation** for the wave: for a given wavenumber $k$, it tells us the wave's frequency. This form is particularly simple, as shown in Fig. 1.7.
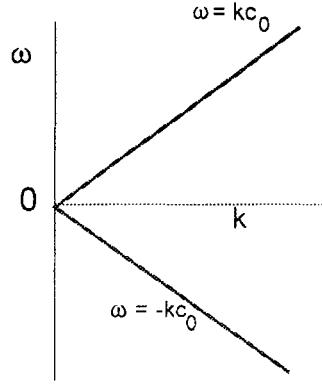
Figure 1.7: The dispersion relation for 1-D shallow water waves.

[Note that this is drawn for positive $k$ only—we may define $k$ positive, without loss of generality, as long as we do not try to constrain the sign of $\omega$.] The phase speed $c = \omega/k = \pm c_0$; the waves can propagate in either direction.

These waves are **nondispersive**, *i.e.*, their phase speed is independent of wavenumber. Thus, all waves, of any wavenumber, propagate at the same speed (in either direction), which means that non-sinusoidal disturbances propagate *without change of shape*. In fact, any function

$$\chi(x, t) = F(x \pm c_0 t) \tag{1.12}$$

is a solution to $(1.9)^3$. Eq. (1.12) just describes any shape of disturbance, including a localized one, that propagates at speed $c$ without changing its shape (Fig. 1.8).

---

$^3$To see this, note that if $X = x \pm c_0 t$, then $F = F(X)$ and the chain rule gives us

$$\frac{\partial \chi}{\partial x} = \frac{\partial F}{\partial x} = \frac{dF}{dX}\frac{\partial X}{\partial x} = \frac{dF}{dX} \; ;$$

$$\frac{\partial^2 \chi}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{dF}{dX}\right) = \frac{\partial X}{\partial x}\frac{d^2 F}{dX^2} = \frac{d^2 F}{dX^2} \; ;$$

$$\frac{\partial \chi}{\partial t} = \frac{\partial F}{\partial t} = \frac{dF}{dX}\frac{\partial X}{\partial t} = \pm c_0 \frac{dF}{dX} \; ;$$

$$\frac{\partial^2 \chi}{\partial t^2} = \frac{\partial}{\partial t}\left(\pm c_0 \frac{dF}{dX}\right) = \pm c_0 \frac{\partial X}{\partial t}\frac{d^2 F}{dX^2} = c_0^2 \frac{d^2 F}{dX^2} \; .$$

So, (1.9) is satisfied by (1.12).
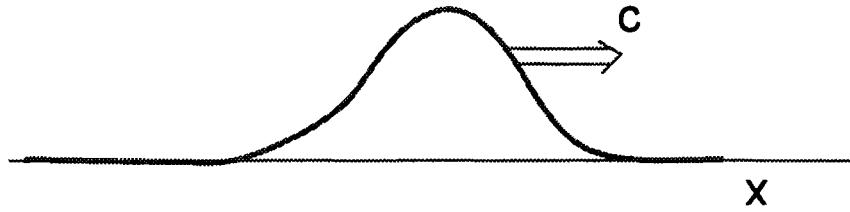
Figure 1.8: Nondispersive waves: arbitrary disturbances propagate without change of shape.

### 1.3.3   Two-dimensional waves

In two dimensions $(x, y)$, (1.9) is replaced by

$$\frac{\partial^2 \chi}{\partial t^2} - c_0^2 \nabla^2 \chi = 0 \tag{1.13}$$

where $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. We can now look for **plane wave** solutions of the form

$$\chi(x, y, t) = Re\left[A_0\, e^{i(kx + ly - \omega t)}\right] . \tag{1.14}$$

Then

$$\nabla^2 \chi = -\kappa^2 Re\left[A_0\, e^{i(kx + ly - \omega t)}\right]$$

where $\kappa = \sqrt{k^2 + l^2}$ is the total wavenumber. Therefore, substituting into (1.13) gives the dispersion relation for this case

$$\omega^2 = \kappa^2 c_0^2 . \tag{1.15}$$

[Note that the one-dimensional case we discussed above is just a special case of the two-dimensional problem, with $l = 0$.]

Eq. (1.14) describes a plane wave because $\chi$ is constant along lines of constant phase $kx + ly - \omega t = constant$, so at any instant in time, $kx + ly = constant$; see Fig. 1.9.
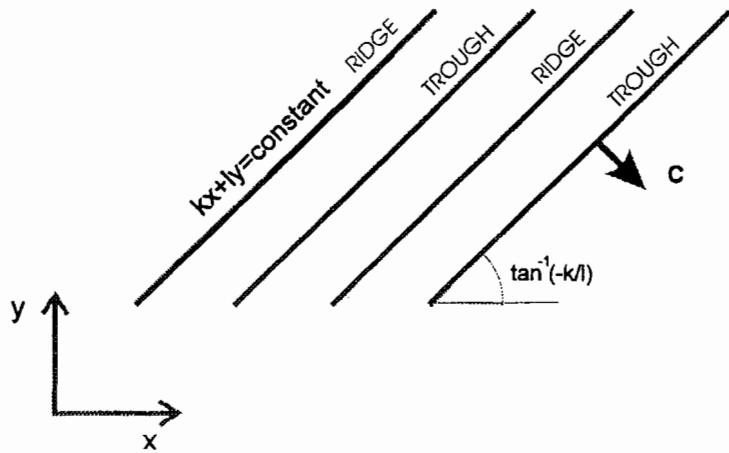
Figure 1.9:

The wave pattern moves at right angles to the phase lines, with speed $c$.

Plane waves are a special, and particularly simple, form of 2-D waves. Exactly what shape the wavefronts have will in general depend on the geometry of the system and of the process that generated the wave. If the source is very localized (*e.g.*, a stone dropped into water), the wavefronts will be circular, as shown in Fig. 1.10. Note that, far from the source (in the dashed rectangle), the wavefronts will look almost plane.
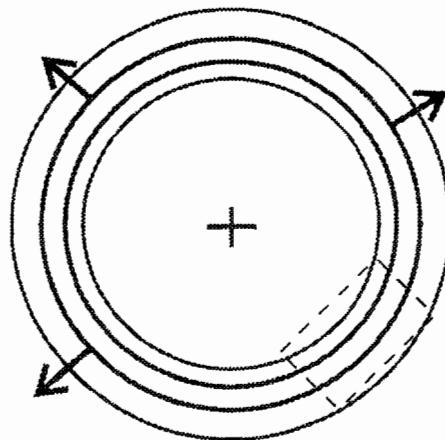


Figure 1.10: Circular wave fronts radiating from a localized source.

## 1.4    Motions within a wave

Returning to the 1-D problem, we can write the general solution, for a given wavenumber $k$, to (1.6) as

$$h'(x,t) = Re\left[H_+e^{ik(x-c_0t)} + H_-e^{ik(x+c_0t)}\right] , \qquad (1.16)$$

the first term representing a sinusoidal wave propagating to the right, the second one propagating to the left. Now, from (1.5),

$$u'(x,t) = Re\left[\frac{c_0}{D}H_+e^{ik(x-c_0t)} - \frac{c_0}{D}H_-e^{ik(x+c_0t)}\right] . \qquad (1.17)$$

So, for that mode propagating to the right (left), velocity and height perturbations are in phase (in antiphase), as shown in Fig. 1.11.
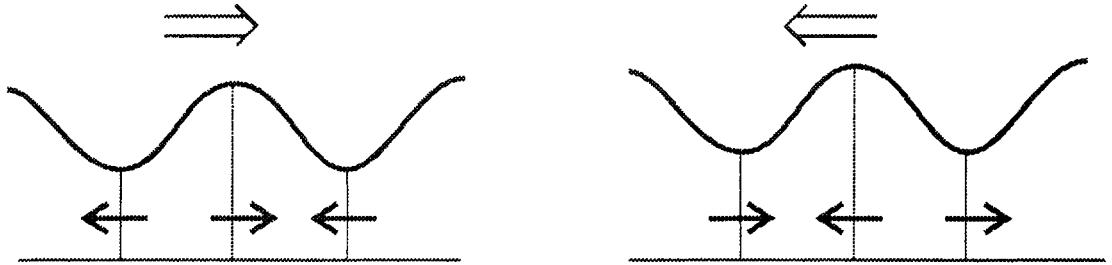


Figure 1.11: The relation between height and velocity perturbations for surface waves propagating to the right and left.

This means that convergence is occurring to the right of the height maximum for the wave propagating to the right, and to the left of the height maximum for one propagating to the left, which of course is consistent with the sense of propagation. (Note that the advective term in (1.1) vanished when we linearized.)

## 1.5    Surface wave reflection and modes

### 1.5.1    1-D reflection

In general, $H_\pm$ in eq. (1.16) are arbitrary constants, to be determined by initial and boundary conditions. Of course, if the waves are propagating

boundary itself, therefore, while both waves are present there, the two waves [see eq. (1.16)] interfere constructively in the $h$ field while they interfere destructively in $u$. So the height field perturbation actually amplifies while the wave packet is close to the boundary (Fig. 1.12(b)). (If you think this is getting something for nothing, note that the wave packet becomes laterally compressed during this time.) Subsequently, the entire incoming wave has been reflected and the wave packet propagates away to the left (c).

## 1.5.2  Modes in a bounded 1-D domain

Consider now a bounded domain, with coasts at $x = 0$ and $x = L$, at each of which $u' = 0$. The solutions (1.16) and (1.17) that satisfy these boundary conditions are

$$
\begin{aligned}
h'(x, t) &= H \cos kx \cos(kc_0 t - \epsilon) \ ; \\[2mm]
u'(x, t) &= \frac{c_0}{D} H \sin kx \sin(kc_0 t - \epsilon) \ ;
\end{aligned}
\tag{1.18}
$$

where $H$ is a real constant amplitude and $\epsilon$ is an arbitrary constant phase (which could be eliminated by a choice of origin for $t$). This is a solution *provided* $u'(L, t) = 0$, which requires the modal condition that the wavenumber satisfy $k = k_n$, where

$$
k_n = n \frac{\pi}{L} \ ,
\tag{1.19}
$$

where $n$ is a nonzero integer; the wave has $n$ half-wavelengths across the domain. Thus, the allowable wavenumber spectrum is quantized, as is the allowable frequency spectrum:

$$
\omega_n = k_n c_0 = n \frac{\pi c_0}{L} \ .
$$

---

$$
\begin{aligned}
u'(0, t) &= \tfrac{c_0}{D} Re \left[ H_+ e^{-ikc_0 t} - H_+^* e^{+ikc_0 t)} \right] \\
&= \tfrac{c_0}{D} Re \left[ \left( H_+ e^{-ikc_0 t} \right) - \left( H_+ e^{-ikc_0 t} \right)^* \right] = 0 \ ,
\end{aligned}
$$

since $Re\,(a - a^*) = 0$ for any $a$.

in an unbounded domain, the location of the sources will tell us which is nonzero (*e.g.*, if the only source for the wave is to the left, $H_- = 0$).

If, however, the domain is bounded, the wave may be reflected from the boundaries. Consider the semi-infinite domain bounded at its eastern side by a vertical coast at $x = 0$ (Fig. 1.12).
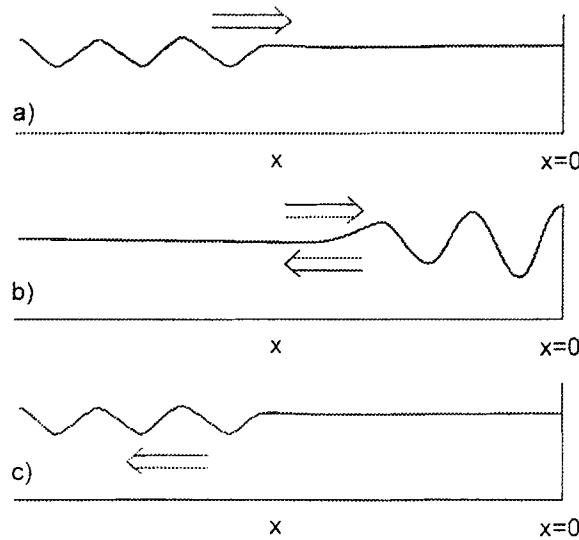


Figure 1.12: A shallow water wave reflecting from an eastern boundary.

A wave packet[4] generated at large negative $x$ propagates in toward the boundary (a). According to Fig. 1.11, it has a nonzero $u$ component in the peaks and troughs. After the wave has reached the boundary (b), it has to meet the boundary condition of zero motion normal to the coast (*i.e.*, $u = 0$ for all $t$), which a single wave component cannot do. The only way for the boundary condition to be met is for a second wave to be radiated from the boundary; in order for the $u$ component of this wave to cancel that of the incoming wave at the boundary *at all times*, it must have the same magnitude of frequency and therefore, from (1.10), the same wavenumber. In short, it must be the mirror-image wave, propagating to the left, with equal and opposite amplitude to that of the incoming wave. In terms of (1.17), $H_- = H_+^*$, where the asterisk denotes complex conjugate[5]. At the

---

[4] By "wave packet", I mean a wave with a finite number of, but many, wavelengths.

[5] This guarantees $u = 0$ at the boundary $x = 0$, since then
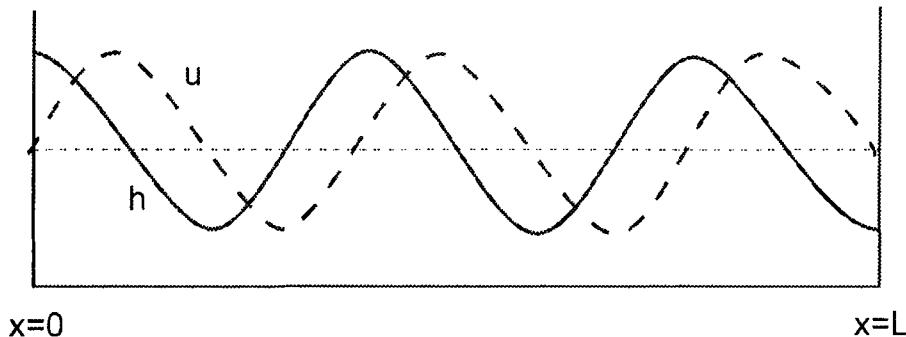
Figure 1.13: The n=5 1-D mode in a bounded domain.

So a finite 1-D domain of width $L$ supports a countably infinite number of discrete modes, the lowest frequency of which is $\pi c_0/L$, or a maximum period of $2L/c_0$. Fig. 1.13 shows the $u$ and $h$ structure of the $n = 5$ mode; the patterns oscillate without propagation.

Of course, these standing wave modes can, in terms of (1.16) and (1.17), simply be regarded as sums of two equal and opposite propagating waves, continuously being reflected from the boundaries.

## 1.5.3   Reflection of plane waves

Reflection of plane waves is only slightly less straightforward than that of 1-D waves. At a straight boundary they suffer specular reflection (equal angles of incidence and reflection), as shown in Fig. 1.14.
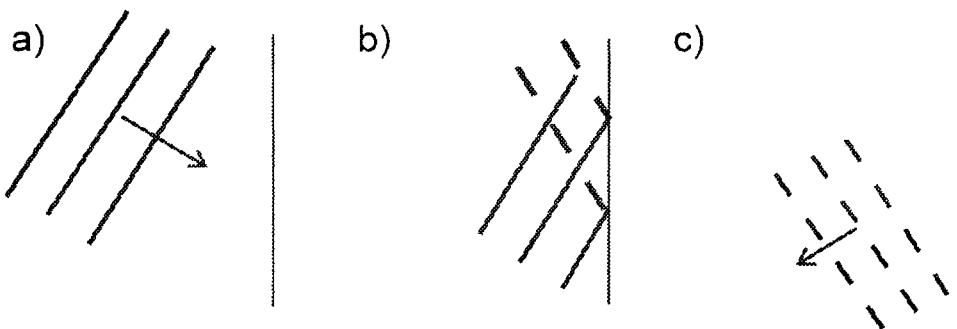


Figure 1.14: Reflection of plane waves.

[Note the interference between incident and reflected waves in (b). It is easy to generate almost-plane waves in a container (or a bathtub) and to see this effect.]

Modes also exist in 2-D containers with simple geometry (e.g., rectangular or circular). In a rectangular basin of dimensions $(L_x, L_y)$, modes are found with wavenumber components

$$k_m = m\frac{\pi}{L_x} \ ; \ l_n = n\frac{\pi}{L_y}$$

(where either $m$ or $n$, but not both, can be zero) with corresponding allowable frequencies

$$\omega_{mn} = c_0\sqrt{k_m^2 + l_n^2} \ .$$

Fig. 1.15 shows a (3,2) mode, which has period

$$\frac{2\pi}{\omega_{32}} = \frac{2L_xL_y}{c_0\sqrt{9L_y^2 + 4L_x^2}} \ .$$
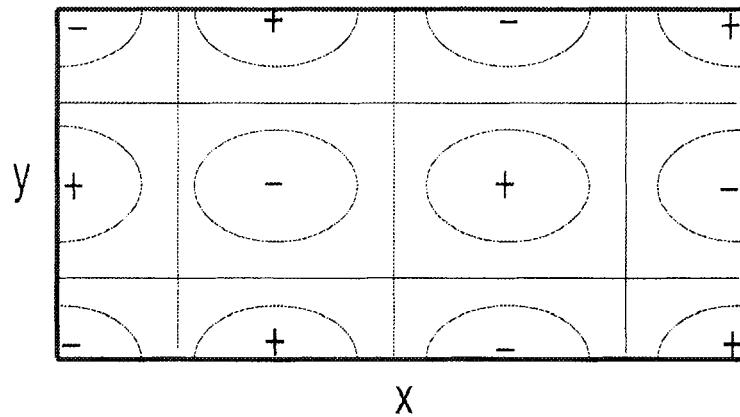


Figure 1.15: The structure of surface height displacement in a (3,2) mode in a rectangular basin.

# 1.6 Further reading

Elementary but brief descriptions of water waves (not confined to those on shallow water) can be found in:

*Waves, tides and shallow-water processes*, by the Open University Course Team, The Open University, Pergamon Press, 1989 *(*Chapter 1*)*.

*Elementary Fluid Dynamics*, by D.J. Acheson, Clarendon Press, Oxford, 1990.

Other treatments can be found in many fluid dynamics texts, but are usually *much* more advanced and more mathematical than these two. One particularly thorough treatment is in

*Waves in Fluids*, by James Lighthill, Cambridge University Press, 1978 (Chapter 3).

# Chapter 2

# Deep water gravity waves

## 2.1   Surface motions on water of finite depth

We now move on to consider motions in water that is not shallow, *i.e.*, we will allow the flow to vary with $z$ within the water:
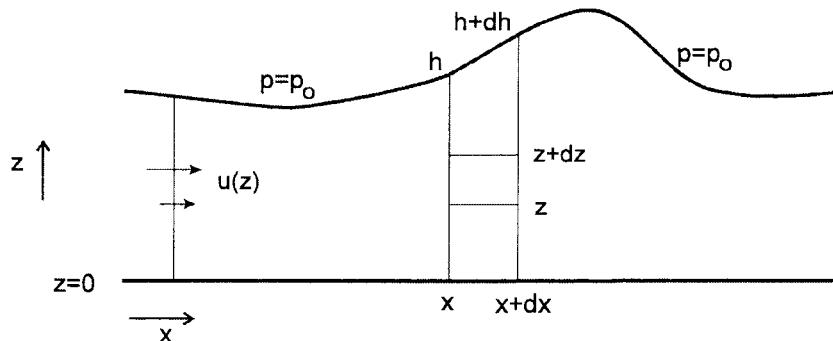


Figure 2.1: Configuration of deep water system.

As before, the system is nonrotating and inviscid, and the water density $\rho$ is assumed constant. Basically, we follow the same procedures as for the shallow water case. However, rather than consider balances for columns of water we must do so separately for elemental volumes $dx\, dz$, such as the box

in the figure. The horizontal and vertical momentum equations are[1]

$$\rho\frac{du}{dt} = \rho\frac{\partial u}{\partial t} + \rho u\frac{\partial u}{\partial x} + \rho w\frac{\partial u}{\partial z} = -\frac{\partial p}{\partial x} \; ;$$

$$\rho\frac{dw}{dt} = \rho\frac{\partial w}{\partial t} + \rho u\frac{\partial w}{\partial x} + \rho w\frac{\partial w}{\partial z} = -\frac{\partial p}{\partial z} - g\rho \; .$$

(2.1)

Thus, the processes tending to accelerate the flow in the $x$-direction are: the $x$-gradient of pressure and advection of $x$-momentum; in the $z$-direction: the $z$-gradient of pressure, gravity, and advection of $z$-momentum.

We now consider conservation of mass within the marked volume; its mass, per unit length in $y$, is $m = \rho\ dx\ dz$, which for this **incompressible** medium is constant with time.
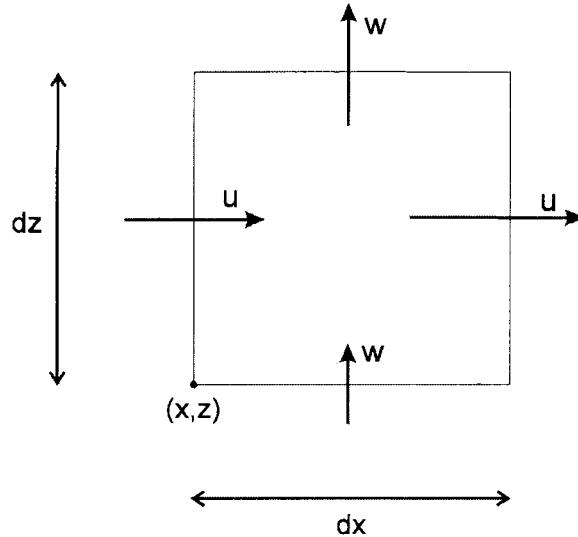


Figure 2.2: Mass continuity.

However, there are fluxes of mass across the volume, as shown in Fig. 2.2. Across the left face, into the box, is a mass flux, per unit length in $y$, of $\rho\ u(x, z)\ dz$; there is an outward flux of $\rho\ u(x + dx, z)\ dz$ across the right face. Similarly, there is a flux $\rho\ w(x, z)\ dx$ in through the lower face and

---

[1]Since the varaibles are now functions of $(x, z, t)$, the substantive derivative is now $\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{dx}{dt}\frac{\partial}{\partial x} + \frac{dz}{dt}\frac{\partial}{\partial z}$ .

a flux $\rho\, w(x, z + dz)\ dx$ out through the upper face. Altogether, the net **convergence** of the mass fluxes into the box is

$$
\begin{aligned}
C \;&=\; \rho\, dz\, [u(x, z) - u(x + dx, z)] + \rho\, dz\, [w(x, z) - w(x, z + dz)] \\
&=\; -\rho\, dx\, dz\, \left( \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} \right) = -\rho\, dx\, dz\, \nabla \cdot \mathbf{u}
\end{aligned}
$$

where $\mathbf{u}$ is the vector velocity $(u, w)$. Continuity of mass demands, since there is no change of mass within the box, that $C$ be zero, whence our **continuity equation**

$$
\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0 \; . \tag{2.2}
$$

This tells us that incompressible flow is **nondivergent**.

So we have 3 equations, the 2 momentum eqs. (2.1) and the continuity eq. (2.2), in three unknowns $u$, $w$, and $p$. We also have boundary conditions. At the lower boundary $z = 0$, there can be no normal motion, whence $w = 0$ there[2]. The motion will in general be nonzero at the free surface, since this can move. However, fluid on the surface (which is a **material surface**) must remain there—it cannot pass through, which tells us that a fluid parcel on the surface must move with the local fluid speed along the surface, or

$$
\frac{dh}{dt} = w\big|_{z=h} \; . \tag{2.3}
$$

Finally, we also know that the pressure in the air immediately above the surface is atmospheric pressure $p_0$ which, as before, we assume constant. Since pressure must be continuous across the surface (to be otherwise would imply an infinite pressure gradient and therefore infinite acceleration, which would be unphysical[3]). So our final boundary condition is

$$
p\big|_{z=h} = p_0 \; . \tag{2.4}
$$

---

[2] Note that we can impose no similar condition on $u$ at the lower boundary, since we have assumed inviscid flow and so cannot include viscous boundary effects.

[3] We are in fact neglecting surface tension here. If the surface is curved, as it will be in the presence of the wave, surface tension will exert an effective pressure on the fluid beneath. Such effects are significant only for waves of wavelength a few cm or less (and such waves are known as **capillary waves**).

## 2.2   Small-amplitude deep-water surface waves

As before, we investigate the properties of small-amplitude disturbances to a stationary basic state, described by

$$
\begin{aligned}
u &= w = 0 \,, \\
h &= D \,, \\
p &= p_0 + g\rho(D - z) \,.
\end{aligned}
\tag{2.5}
$$

Introducing small perturbations $u'$, $w'$, $h'$, and $p'$, the momentum eqs. (2.1) can then be written as

$$
\begin{aligned}
\frac{\partial u'}{\partial t} + \frac{1}{\rho}\frac{\partial p'}{\partial x} &= -u'\frac{\partial u'}{\partial x} - w'\frac{\partial u'}{\partial z} \,, \\
\frac{\partial w'}{\partial t} + \frac{1}{\rho}\frac{\partial p'}{\partial z} &= -u'\frac{\partial w'}{\partial x} - w'\frac{\partial w'}{\partial z} \,.
\end{aligned}
$$

Now, the terms in the RHS are *nonlinear*, in fact quadratic in the perturbation quantities; we neglect these on the grounds that the perturbations are small, leaving

$$
\frac{\partial u'}{\partial t} + \frac{1}{\rho}\frac{\partial p'}{\partial x} = 0 \,,
\tag{2.6}
$$

$$
\frac{\partial w'}{\partial t} + \frac{1}{\rho}\frac{\partial p'}{\partial z} = 0 \,.
\tag{2.7}
$$

The continuity eq. (2.2) becomes simply

$$
\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0 \,.
\tag{2.8}
$$

We now proceed to derive a single equation in a single unknown from these three. In fact, it is very simple to do so: taking the $x$-derivative of (2.6), the $z$-derivative of (2.7), and adding gives

$$
\frac{\partial}{\partial t}\left(\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z}\right) + \frac{1}{\rho}\left(\frac{\partial^2 p'}{\partial x^2} + \frac{\partial^2 p'}{\partial z^2}\right) = 0 \,.
$$

But, from (2.8), the term in the first parenthesis is zero, so

$$
\frac{\partial^2 p'}{\partial x^2} + \frac{\partial^2 p'}{\partial z^2} = 0 \,.
\tag{2.9}
$$

As before, we focus attention on wave-like disturbances, for which

$$p'(x, z, t) = Re\left[\widetilde{P}(z, t)e^{ikx}\right] \; ; \tag{2.10}$$

substitution into (2.9) then gives

$$\frac{\partial^2 \widetilde{P}}{\partial z^2} - k^2 \widetilde{P} = 0 \; , $$

which has solutions of the form

$$\widetilde{P}(z, t) = P_1(t)e^{kz} + P_2(t)e^{-kz} \; . \tag{2.11}$$

(Note that this tells us that the vertical scale of the motion is $k^{-1}$, which is formally the same as the horizontal length scale.) Now, the lower boundary condition tells us that $w' = 0$ at $z = 0$; therefore, from (2.7), $\partial p'/\partial z = 0$ there. So, using (2.10), we have

$$\frac{\partial \widetilde{P}}{\partial z}(0, t) = 0$$

whence, in (2.11), $P_1 = P_2 = P(t)/2$, say, and so (2.10) becomes

$$p'(x, z, t) = Re\left[P(t) \cosh kz \; e^{ikx}\right] \; . \tag{2.12}$$

It is now straightforward, from (2.6) and (2.7), to obtain the form of the velocity perturbations

$$u'(x, z, t) = Re\left[-i\frac{k}{\rho}Q(t) \cosh kz \; e^{ikx}\right] \; , \tag{2.13}$$

$$w'(x, z, t) = Re\left[-\frac{k}{\rho}Q(t) \sinh kz \; e^{ikx}\right] \; , \tag{2.14}$$

where $dQ/dt = P$.

Notice that, at this stage we have defined the spatial structure of the motions, but not the time dependence. To define the latter, we have to include consideration of the surface dynamics.
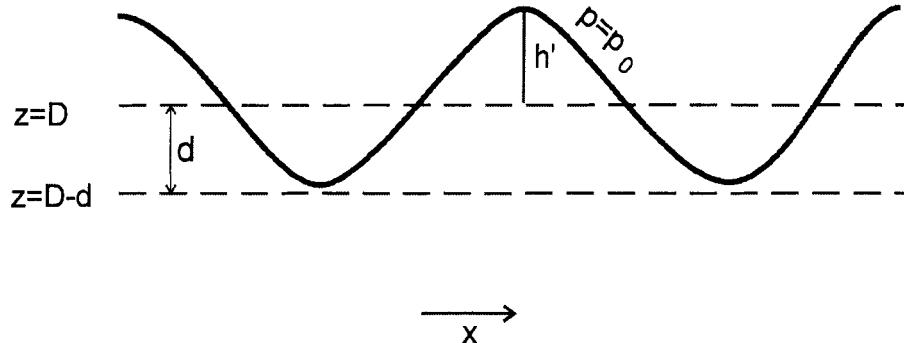
Figure 2.3: Near-surface details.

Consider the level $z = D - d$, where $d$ is just greater than $|h'|$, so the level lies just beneath the wave troughs, as shown in Fig. 2.3. (Since our equations are valid only within the water, we cannot choose $z = D$.) Now, since [eq. (2.4)] the surface pressure is $p_0$, the total pressure (background + perturbation) is, using hydrostatic balance,

$$p(x, D - d, t) = p_0 + g\rho(d + h') .$$

But, since the basic state pressure is, from (2.5), $p_0 + g\rho d$, the perturbation is $p'(x, D - d, t) = g\rho h'$; taking the limit $d \to 0$ (recall that $h'$ is arbitrarily small), we have

$$p'(x, D, t) = g\rho h' ,  \tag{2.15}$$

so, from (2.12),

$$h'(x, t) = \frac{1}{g\rho} Re \left[ P(t) \cosh kD \ e^{ikx} \right] .  \tag{2.16}$$

(Note that $p'(x, D, t)$ is nonzero—it is the pressure perturbation *at the surface* $z = D + h'$ that is zero.) The material surface condition (2.3), neglecting nonlinear terms, is

$$\frac{\partial h'}{\partial t} = w'(x, D, t) .$$

Using (2.15),

$$\frac{\partial p'}{\partial t}(x, D, t) = g\rho w'(x, D, t) .$$

Substituting from (2.12) and (2.14), we get

$$\frac{dP}{dt}(t) \ \cosh kD = -gk \ Q(t) \ \sinh kD$$

or, since $P = dQ/dt$,

$$\frac{d^2Q}{dt^2} + gk \ \tanh kD \ Q = 0 \ . \tag{2.17}$$

This has solutions

$$Q = Q_+ e^{-i\omega t} + Q_- e^{i\omega t} \tag{2.18}$$

provided $\omega$ satisfies the dispersion relation

$$\omega^2 = gk \ \tanh kD \ . \tag{2.19}$$

Eq. (2.19) can be rewritten as

$$\omega = \pm\frac{c_0}{D}\sqrt{kD \ \tanh kD} \ , \tag{2.20}$$

where $c_0 = \sqrt{gD}$ is the shallow water wave speed. (2.20) is plotted in Fig. 2.4.
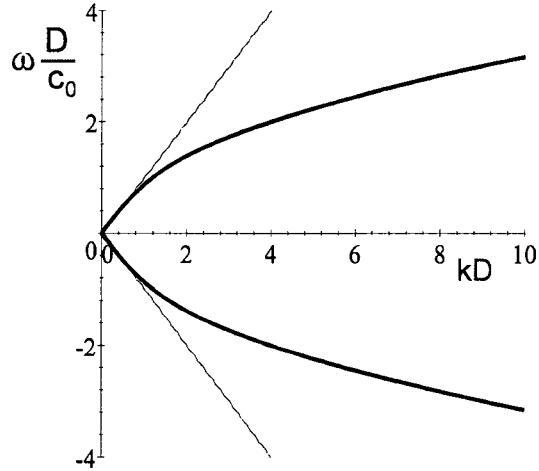


Figure 2.4: Dispersion relation for deep water gravity waves. Dashed lines show the shallow water relation.

Note the departure from shallow water behavior for $kD > 1$. In general, the most important difference is that, for deep water waves, the phase speed $c = \omega/k$ is *not* independent of wavenumber—such waves are known as **dispersive waves**.

## 2.2.1   The long wave (shallow water) limit

Note from (2.20) that, in the limit $kD \to 0$, $\omega^2 \to gDk^2$, so we recover the shallow water dispersion relation (and, in fact, shallow water dynamics) in this limit when the wavelength is much greater than the depth. However, $kD$ in fact has to be *very* small for this approximation to be valid: the wavelength $2\pi/k$ must be greater than 14 times the depth before the shallow water result for $c$ becomes within 3% of the correct value.

From (2.13), since $\cosh kz \to 1$ as $kD \to 0$ (note that $z \leq D$) the horizontal velocity becomes independent of $z$ in this limit, as we assumed in our shallow water analysis.

## 2.2.2   The short wave (deep water) limit

For $kD \to \infty$, $\tanh kD \to 1$, so (2.19) becomes

$$\omega = \pm\sqrt{gk} \tag{2.21}$$

which is independent of $D$. In fact, the whole problem becomes insensitive to the background depth in this limit, as the waves do not feel the bottom. The vertical structure functions $\cosh kz$ and $\sinh kz$ appearing in (2.13), (2.14) and (2.12) are

$$\cosh kz = \frac{1}{2}\left(e^{kz} + e^{-kz}\right),$$

$$\sinh kz = \frac{1}{2}\left(e^{kz} - e^{-kz}\right),$$

both of which can be approximated as $\frac{1}{2}e^{kz}$ (except close to the bottom) as $kD \to \infty$. This means that the solutions decay downwards from the surface as $\exp\left(-k\left(D - z\right)\right)$, and becoming vanishingly small at depth, as shown in Fig. 2.5. Thus, though the waves can propagate great distances horizontally, they remain **trapped** near the surface and do not penetrate deep into the water. So, the water depth becomes irrelevant. This is because all the
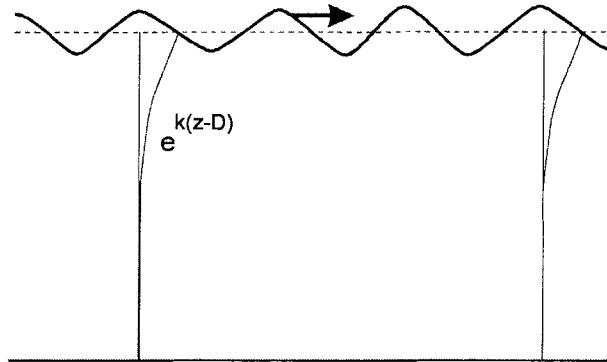
Figure 2.5: Surface waves on deep water.

dynamics of the waves—the effective elasticity that allows their existence—is tied up with the density discontinuity at the water surface. Thus, the surface acts as a **wave guide**, channeling the wave propagation.

## 2.3 Background theory—dispersive waves

### 2.3.1 Dispersion

The general dispersion relation for the frequency $\omega$ of 1-D waves of wavenumber $k$ can be written

$$\omega = \omega(k) \tag{2.22}$$

and the phase speed is

$$c = \frac{\omega(k)}{k} . \tag{2.23}$$

Clearly, $c$ is independent of $k$ for all $k$ only if $\omega(k)$ =constant$\times k$—this is the nondispersive case we discussed earlier and, as we saw, it implies that all disturbances, including localized ones, propagate without change of shape. This can be thought of in terms of Fourier components. Any non-sinusoidal disturbance can be described a sum of components of different wavenumber; if all these waves propagate at the same speed, so will the disturbance itself, and its shape will not change.

For many kinds of wave motion, however (including surface waves on deep water), $c$ varies with $k$, in which case the different wavenumber components will have different speeds, a phenomenon known as **wave dispersion**.

Therefore the way they interfere with one another will change with time—so the shape of the disturbance will change.

## 2.3.2 Group velocity

There is one particularly important aspect of dispersion, which concerns the way that modulations propagate on a wave train.

A monochromatic wave of the form

$$\chi(x,t) = Re\left[A\ e^{i[k_0 x - \omega(k_0)t]}\right]$$

with single wavenumber $k_0$ has the simple spectrum of a $\delta$-function $A\delta(k-k_0)$ [Fig. 2.6(a)], and behaves in the way we have discussed, propagating with a speed $c = \omega(k_0)/k_0$.
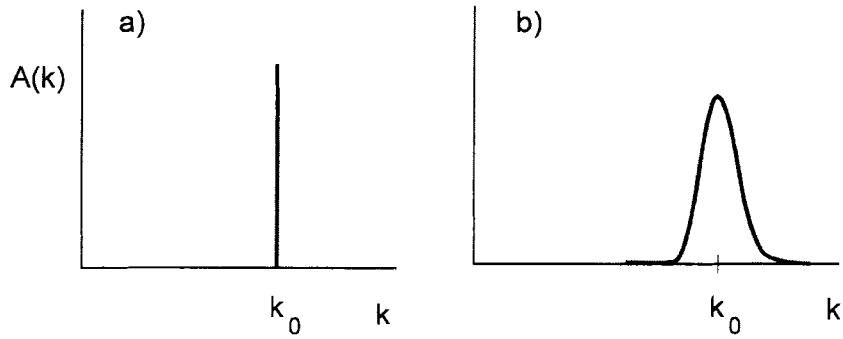


Figure 2.6: Wavenumber spectra for (a) a monochromatic wave, and (b) an almost-monochromatic wave.

Consider now an almost monochromatic wave, with a narrow spectrum [Fig. 2.6(b)]. This can be written

$$\chi(x,t) = Re \int_0^\infty A(k)\ e^{i[kx - \omega(k)t]}\ dk\ , \tag{2.24}$$

where $A(k)$ is nonzero only for wavenumbers in the vicinity of $k_0$. Writing $k = k_0 + \delta k$, we can rewrite this as

$$\chi(x,t) = Re \int_0^\infty A(k_0 + \delta k)\ e^{i[k_0 x - \omega(k_0)t]}\ e^{i[\delta k\ x - \delta\omega\ t]} dk\ ,$$

where $\delta\omega = \omega(k_0 + \delta k) - \omega(k_0) \simeq (\partial\omega/\partial k)\delta k$. At $t = 0$, this is simply

$$\chi(x, t) = F(x)\ e^{ik_0 x}$$

where

$$F(x) = Re \int_0^\infty A(k_0 + \delta k)\ e^{i\ \delta k\ x}\ dk$$

is the modulating envelope of the wave train, which has carrier wavenumber $k_0$, as illustrated in Fig. 2.7.
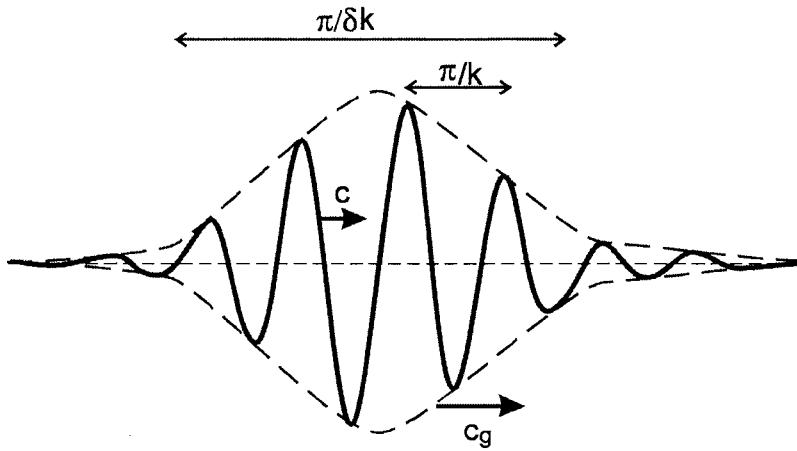


Figure 2.7: An almost-monochromatic wave packet, comprising many wavelengths. The phase of the carrier wave propagates at the phase speed, but the modulation envelope propagates at the group velocity.

Now, for $t > 0$, the wave packet behaves as

$$\chi(x, t) = \left\{ Re \int_0^\infty A(k_0 + \delta k)\ e^{i[\delta k(x - c_g t)]} dk \right\}\ e^{i[k_0(x - ct)]}$$

$$= F(x - c_g t)\ e^{i[k_0(x - ct)]}\ , \tag{2.25}$$

where

$$c_g = \left.\frac{\partial\omega}{\partial k}\right|_{k=k_0} \tag{2.26}$$

is the **group velocity**. Thus, while the carrier wave propagates at the phase speed, the modulation envelope propagates at the group velocity. This is an

important concept, as it is the latter velocity that governs the propagation of information, as we shall see.

Nondispersive waves have $\omega = c_0 k$, with constant phase speed $c_0$, and so their group velocity is the same as their phase velocity. But the group velocity of dispersive waves differs from the phase speed, so in a wave packet like that shown in Fig. 2.7 the wave crests will move at a different speed than the envelope. If $c > c_g$ (which, as we shall see, is the case for deep water waves), new wave crests appear at the rear of the wave packet, move forward through the packet, and disappear at its leading edge. We shall see some examples of this below.

In general, it is easy to get a feel for both phase and group propagation graphically from the dispersion relation, as shown in Fig. 2.8.
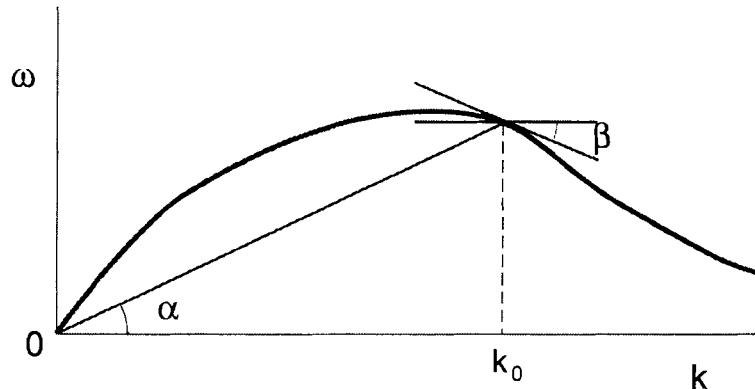


Figure 2.8: The disperion relation $\omega(k)$. The phase velocity at $k = k_0$ is $\tan \alpha$, the group velocity $\tan \beta$.

For any wavenumber $k$, the phase velocity is just given by the slope $\alpha$ of a line joining the point $(\omega, k)$ to the origin, while the group velocity is given by the slope $\tan \beta$ of the tangent to the curve at $(\omega, k)$.

## 2.3.3   Group velocity in multi-dimensional waves

For future reference, we note here that for 2- or 3-dimensional waves, the general dispersion relation is of the form

$$\omega = \omega(\mathbf{k}) \tag{2.27}$$

where $\mathbf{k} = (k, l, m)$ is the (2-D or 3-D) vector wavenumber. The phase velocity[4] is

$$\mathbf{c} = \left(\frac{\omega}{k}, \frac{\omega}{l}, \frac{\omega}{m}\right) , \tag{2.28}$$

and the group velocity

$$\mathbf{c}_g = \left(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial l}, \frac{\partial\omega}{\partial m}\right) . \tag{2.29}$$

As we shall see later (and a comparison of (2.28) and (2.29) implies), not only may $\mathbf{c}_g$ and $\mathbf{c}$ be of different magnitudes, they may also be *in different directions*.

## 2.4 Surface wave dispersion
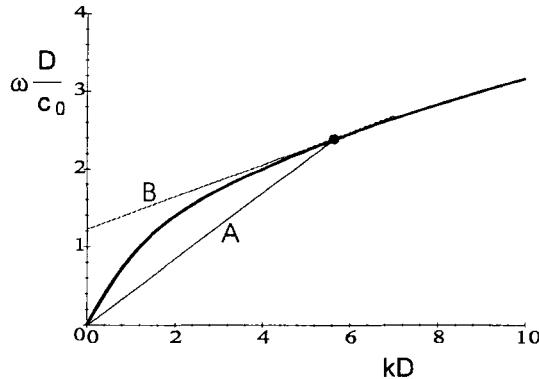


Figure 2.9: Phase speed $c$(slope of line $A$)and group velocity $c_g$ (slope of line $B$)for surface waves.

Returning now to surface waves, and the dispersion relation (2.19) shown in Fig. 2.4, we can see from Fig. 2.9 that $c_g \leq c$ for all wavenumbers (the slope of line $B$ never being greater than that of line $A$. This is shown more explicitly in Fig. 2.10.

---

[4]The phase velocity is in fact not a vector, even though it has magnitude and direction. It does not transform like a vector under rotation—this stems from the fact that phase propagation has no meaning along the phase lines.
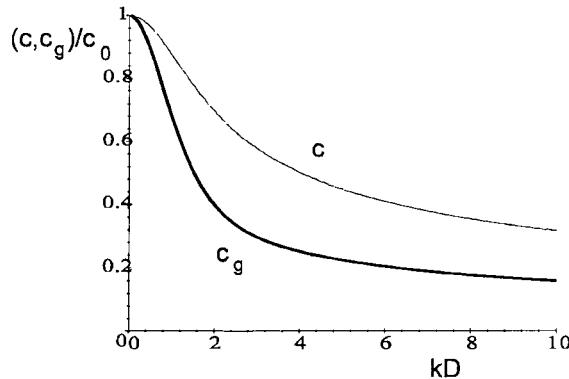
Figure 2.10: Scaled phase and group velocities for surface waves.

In fact, both group and phase speeds are greatest and equal in the long wave (shallow water) limit, when

$$c, \ c_g \to \sqrt{gD} \ , \quad kD \to 0 \ . \tag{2.30}$$

Note that in this limit $c$ becomes independent of $k$, i.e., the waves become nondispersive, as we saw in the shallow water case. However, as Figs. 2.9 and 2.10 make clear, $c$ and $c_g$ differ significantly for $kD \geq 1$. In the short wave (deep water) limit, in fact, from (2.21),

$$c \simeq 2c_g \to \sqrt{\frac{g}{k}} \ , \quad kD \to \infty \ . \tag{2.31}$$

The difference between nondispersive long waves and dispersive short waves is illustrated in the following. Consider an initial disturbance to the water surface of the form

$$h'(x,0) = \exp\left(-\left(\frac{x}{\Delta D}\right)^2\right) \ .$$

If $\Delta$ is large, this has a length scale long compared with $D$ (so it projects primarily onto waves with $kD < 1$) and, as shown in Fig. 2.11 for $\Delta = 4$, the waves emanating from the disturbance are essentially nondispersive.
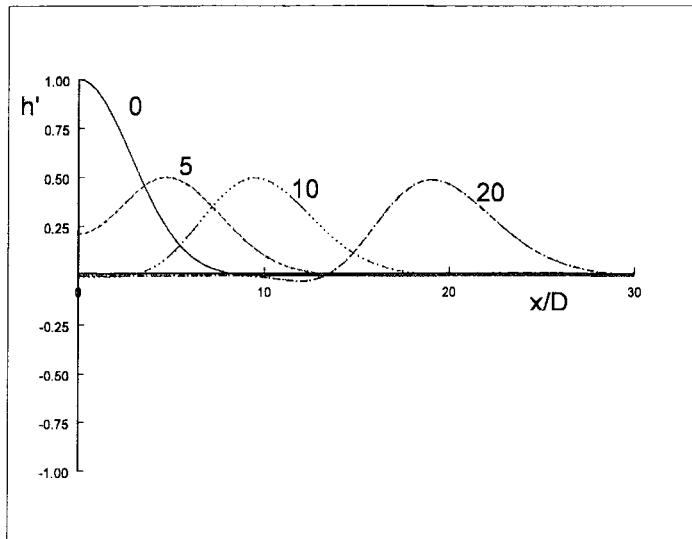
Figure 2.11: Dispersion of a large-scale initial disturbance $[\exp\left(-\left(x/4D\right)^2\right)$ on water of depth $D$. Behavior is symmetric about $x = 0$; only $x > 0$ is shown. Numbers on curves are time in units of $D/c_0$.

(There is just a hint of dispersion; note the negative tail at $t = 20D/c_0$. Note also that there is, for $t > 0$, an identical disturbance, not shown, in $x < 0$.) When $\Delta$ is smaller however, the initial disturbance has a smaller length scale and hence a greater projection onto the dispersive short waves. This is illustrated in Fig. 2.12, for which the initial disturbance has $\Delta = 0.5$.
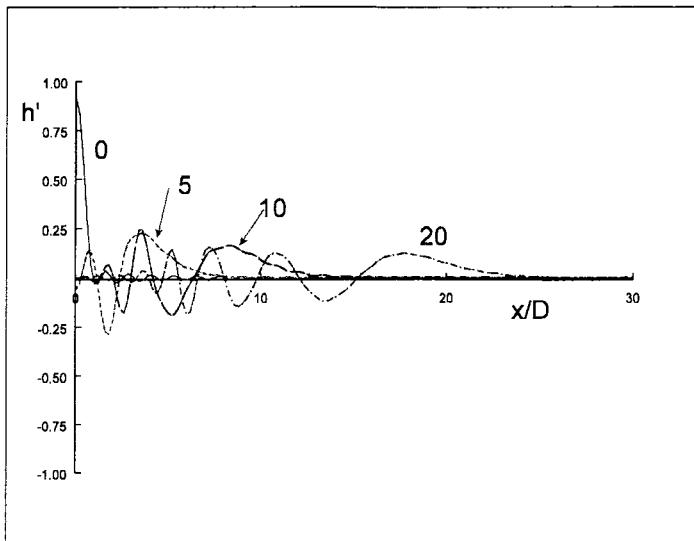
Figure 2.12: Same as the previous figure , but for a small-scale initial disturbance $[\exp(-(2x/D)^2)]$.

The dispersion of the resulting waves is evident. Note especially that (e.g., at $t = 20D/c_0$) the leading edge of the wave train has the largest wavenumber, consistent with our observation that the long waves have the greatest group velocity and, in fact, should travel a distance $x = 20D$ by this time). Wavelength becomes progressively shorter in the tail of the wave train.

Both examples can easily be reproduced in a container of water (one that is large enough to allow the dispersion to develop) or outdoors in a river or lake. If the water is sufficiently shallow, in response to a localized disturbance (e.g., from a small object dropped into the water) you will see a localized wave propagating away, with little or no apparent dispersion. If the water is deep, however, you will see a dispersing wave train, with the longest waves at the leading edge. If you look closely, you may be able to see the wave crests appearing in the rear and propagating forward through the wave train (since $c > c_g$)[5]. This effect can also be seen in a ship's wake (Fig. 2.13):

---

[5]Though in this case they never overtake the front of the wave train, since this consists of long waves for which $c = c_g$.
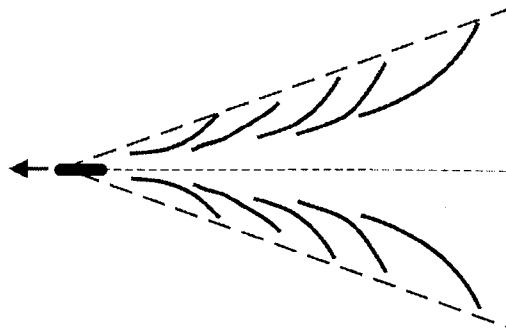
Figure 2.13: Waves in the wake of a ship.

individual wave crests can be seen propagating to the outside of the wake (changing wavenumber in the process), and disappearing there. (Incidentally, in deep water the half-angle of the "wedge" made by the wake is $\arcsin(\frac{1}{3}) = 19.5°$, independent of the speed of the ship. This can be shown to follow from the fact that $c_g/c = \frac{1}{2}$ for deep water; we will not pursue that here but the proof can be found in many texts, such as "Waves in Fluids", Lighthill, Cambridge U P, 1978; section 3.10.)

The results that $c_g \le c$ is made apparent in another common phenomenon, the wave train downstream of an obstacle in flowing water ($e.g.$, a river), as shown in Fig. 2.14.
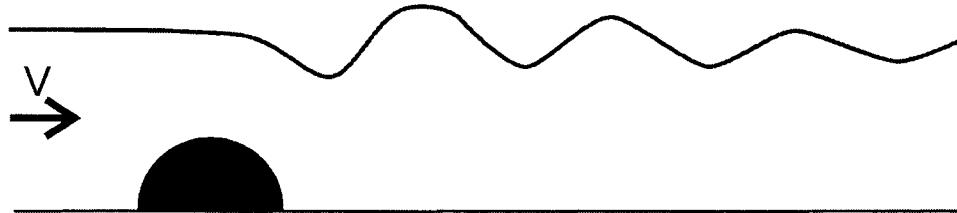


Figure 2.14: Disturbance produced by flow over an obstacle.

There are two key statements to be made:

(i) the wave train is **stationary**, and so has $c = 0$;

(ii) it is located *downstream*—there is no disturbance ahead of the obstacle.

Since our analysis has been for waves on stationary water, let's shift our frame of reference to move with the flow. So now the obstacle and its wave train are moving to the left with speed $V$. Since the wave crests are stationary with respect to the obstacle, they are also moving to the left with phase speed, in this frame of reference, $c' = -V$. Now, the energy is radiated from the moving obstacle radiates away at the group velocity, *which is never greater than* $|c'| = V$—so it must lag behind the moving obstacle. To put it another way, in this moving frame of reference, $|c'_g| \leq V$ (remember that $c_g$ is negative in this case). So, in the frame in which the obstacle is stationary, the group velocity is $c_g = c'_g + V \geq 0$. Thus, wave energy can only radiate downstream and there are no upstream effects[6].

There is another aspect to this problem that is illustrative of the general characteristics of waves in fluids. The obstacle is subject to a **force** associated with the flow across it. For a small, smooth obstacle, this force is not primarily frictional (though there is a component of that) but is mostly the result of **form drag**: the pressure on the upstream side of the obstacle is greater than that on the downstream side, and so the obstacle is subjected to a force to the right. (On Fig. 2.14, the free surface height is greater upstream of the peak of the obstacle than at a comparable position downstream; therefore there is a positive pressure gradient on the obstacle.) Simultaneously, of course, the water must be subjected to an equal force to the left (decelerating the flow). However, this is felt, *not* at the obstacle but downstream, within the wave train: this can happen because (like, *e.g.*, electromagnetic waves) the waves, which as we have seen are capable of transporting energy, can also effect **transport of momentum**. Thus, the drag on the obstacle is relayed to remote parts of the water. This has several ramifications for atmospheric and oceanic dynamics.

---

[6]The fact that this result is not trivial is underlined by the observation that it does not always apply. Capillary waves—those for which surface tension is crucial—*can* have $c_g > c$ and so can and do travel upstream. If the obstacle is small ($<$ a few cm in size) these waves are important though, as they have small wavelength, they dissipate quickly and so may be hard to see.

## 2.5  Particle motions within a wave

Consider now the motion of (neutrally buoyant) marked particles in the water, which is of course the same thing as the motion of the water itself. Let the instantaneous position of a particle be $(x, z) = (X, Z) + (\eta', \zeta')$ where $(X, Z)$ is the undisturbed position (where the particle would be in the absence of the wave), and $\eta'$ and $\zeta'$ are the small perturbations in position associated with the wave motion. From the definition of velocity as rate of change of position, we have

$$\frac{dx}{dt} = u \,,$$
$$\frac{dz}{dt} = w \,;$$

or (since $(X, Z)$ is fixed in time)

$$\frac{\partial \eta'}{\partial t} + u'\frac{\partial \eta'}{\partial x} + w'\frac{\partial \eta'}{dz} = u' \,,$$
$$\frac{\partial \zeta'}{\partial t} + u'\frac{\partial \zeta'}{\partial x} + w'\frac{\partial \zeta'}{dz} = w' \,.$$

Linearizing, we have

$$\frac{\partial \eta'}{\partial t} = u' \,;$$
$$\frac{\partial \zeta'}{\partial t} = w' \,.$$

Now, suppose we have a single propagating wave (single wavenumber, single frequency). Then, from (2.13), (2.14), and (2.18), we have

$$\left( \begin{array}{c} u' \\ w' \end{array} \right) = Re \left[ -\frac{k}{\rho}Q_0 \left( \begin{array}{c} i\,\cosh kz \\ \sinh kz \end{array} \right) e^{i(kx-\omega t)} \right] \,. \qquad (2.32)$$

Therefore

$$\left( \begin{array}{c} \eta' \\ \zeta' \end{array} \right) = Re \left[ \frac{k}{\omega\rho}Q_0 \left( \begin{array}{c} \cosh kz \\ i\,\sinh kz \end{array} \right) e^{i(kx-\omega t)} \right] \,. \qquad (2.33)$$

Note:

1. The displacements are oscillatory, so there is no net drift of the particles. Thus, even though the wave pattern propagates, fluid parcels do not: they merely oscillate about their mean position.
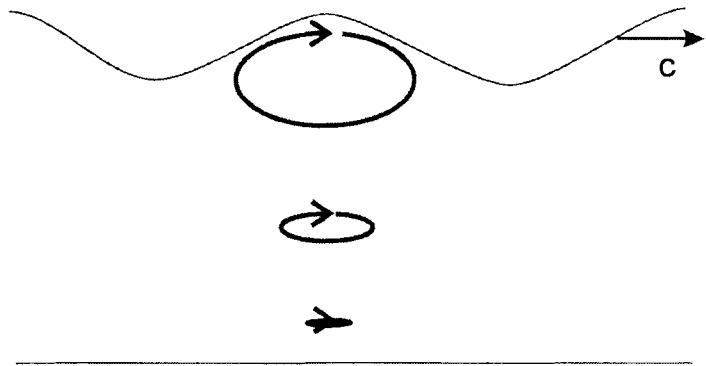
Figure 2.15: Parcel orbits in a water wave. Direction of motion is reversed for a wave traveling to the left.

2. Vertical and horizontal displacements are *in quadrature* ($\pi/2$ out of phase). The marked particles perform elliptical orbits, with, for real $Q_0 > 0$ (which we can insist on, by defining the time origin accordingly) and $k > 0$,

$$\begin{pmatrix} \eta' \\ \zeta' \end{pmatrix} = -\frac{kQ_0}{\omega\rho} \times \begin{bmatrix} \cosh kz \ \cos(kx - \omega t) \\ \sinh kz \ \sin(kx - \omega t) \end{bmatrix} ,$$

which implies that the parcels move clockwise around the orbit for $\omega > 0$ $(c > 0)$ , and anticlockwise for $\omega < 0$ $(c < 0)$. See Fig. 2.15. Note that the ratio of vertical to horizontal axes of the ellipse is $\tanh kz$, which increases from zero at the bottom (where the boundary ensures that $\zeta' \to 0$) to $\tanh kD < 1$ at the top. For waves in deep water $(kD \gg 1)$, the orbits become circular.

## 2.6 Wave generation by wind

It is common experience that waves on the ocean and lakes are usually weak on calm days, but strong on windy days, suggesting that much of the surface wave activity is somehow produced by the action of wind. It seems unlikely, though, that a wave of wavenumber $k$ and frequency $\omega$ is *directly* forced by winds as this would require that the wind itself (or the pressure fluctuations that accompany the wind) have a significant component at the same

frequency and wavenumber. There will always be some such component—especially in the turbulent atmospheric boundary layer, the winds have a rich spectrum—but, at the short wavelengths and high frequencies characteristic of surface water waves, these fluctuations are generally weak. Moreover, laboratory studies show that waves can be generated by blowing a *steady* air flow across a water surface. (Try blowing across a glass of water.)

The underlying process common to most wave generation is one of **instability**. That is, even though there may be no externally imposed "waviness" in the wind, there is a tendency to amplify any small perturbation on the water surface. Consider Fig. 2.16.
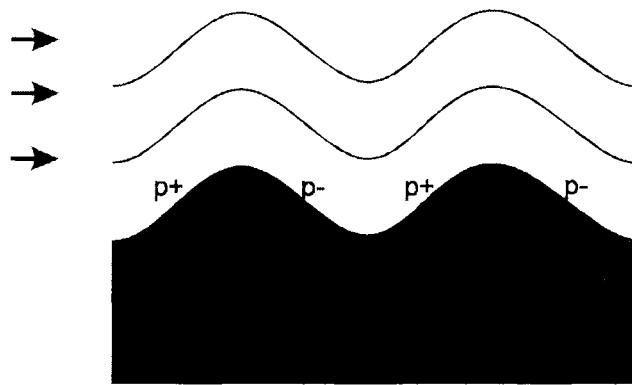


Figure 2.16: Wave generation by wind (schematic). A wave on the ocean surface disturbs the air flow in such a way a to produce a pressure perturbation at the surface that reinforces the wave.

In the absence of a surface wave, the air flow is uniform and the pressure on the water surface is uniform; there is thus no tendency to force waves in the water. However, in the presence of a small wave on the surface, the air flow is disturbed and, like the "rock-in-the-river" problem[7], a perturbed pressure gradient is produced at the water surface which has the same wavenumber (and frequency) as the surface perturbation. Provided the water-air system can get the phase relationships right (and it can) these perturbations reinforce will each other and grow, thus producing, eventually, a finite-amplitude surface wave from an initially infinitessimal perturbation.

---

[7]Except now the "rock" is the bump on the water surface and the "river" is the atmosphere.

## 2.7  Wave breaking

All our discussion thus far has been based on our linear (small-amplitude) approximation to the full problem. Using this approximation, we have been able to explain many of the commonly observed properties of surface water waves. There is, however, at least one familiar aspect of these waves that cannot be explained by linear theory: breaking, which is most commonly observed as waves run up a beach.
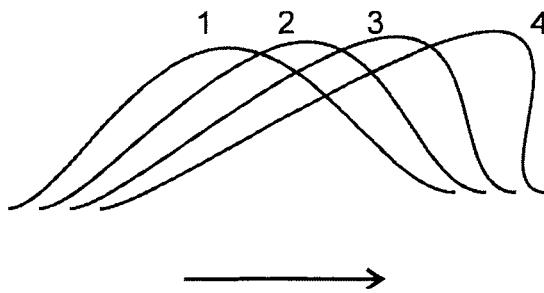


Figure 2.17: Steepening of a finite amplitude wave.

This happens for two reasons. First, as the waves run up the beach, the energy in the wave becomes focused into a shallower layer (once $D \leq k^{-1}$ or so), thus concentrating the energy and increasing the wave amplitude. Second, when $D \leq k^{-1}$ or so, the phase speed becomes dependent on depth, being greater where the water is deeper. In a finite amplitude wave, the water is deeper at the wave crest than in the wave trough. Hence the crests travel faster than the troughs; the crest therefore tend to catch up with and, eventually, overtake the troughs, as shown in Fig. 2.17. This produces the overturning of the wave that is familiar in breakers.

## 2.8  Further reading

See the suggestions given in Section 1.6.

# Chapter 3

# Internal Gravity Waves

## 3.1 Interfacial waves

We have thus far considered the dynamics of the air-water interface. The surface gravity wave motions that this interface permits owe their existence to the restoring force associated with the *density difference* across the interface. [Because we did not consider the effects of motions in the air—we neglected variations in atmospheric pressure—we implicitly assumed that $\rho_{air} \ll \rho_{water}$, which is a very good assumption.] In fact, similar waves are also possible at any *internal* interface in a fluid across which there is a density discontinuity, such as shown in Fig. 3.1. Suppose the densities above and below the interface are $\rho_1$, $\rho_2$, respectively (where $\rho_2 > \rho_1$). In general, further complexities are introduced by the different depths of the fluid layers; if we concentrate on layers of equal depth $D$, then the dispersion relation for the interfacial waves is the same as for the surface wave case (2.20), except that the shallow water wave speed, $\sqrt{gD}$ in the surface wave case, is replaced by

$$c_0 = \sqrt{g\frac{\rho_2 - \rho_1}{\rho_2 + \rho_1}D} \, . \tag{3.1}$$

[Note that this reduces to $\sqrt{gD}$ in the case $\rho_1 \ll \rho_2$.] The origins of the additional factor are not hard to see. The frequency of any oscillations on the interface depend on the restoring force acting on any deviations of the interface; this force is proportional to the density difference across the interface, $\rho_2 - \rho_1$. The frequency also depends on the *inertia* of the fluid, which is proportional to $\rho_2 + \rho_1$. If the density difference is small ($\rho_2 - \rho_1 \ll \rho_2 + \rho_1$),

1
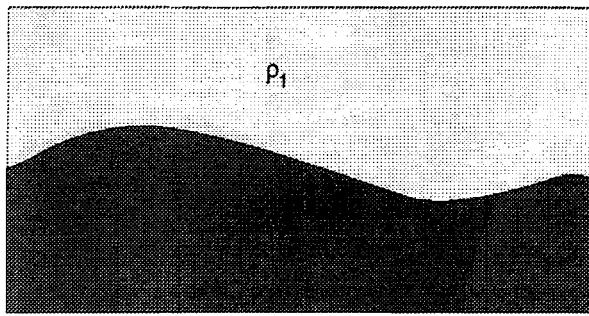
Figure 3.1: Interfacial waves on the interface between two fluids of different density.

the wave speed is much slower than that of surface waves[1].

Consider now the behavior of a fluid with many such layers, as shown in Fig. 3.2. Suppose something makes a disturbance on the surface. When we considered surface waves on deep water, we saw that there are motions within the water, extending a characteristic distance $k^{-1}$ below the surface. If there is a density interface withint this distance, that will be affected by these motions, and will become distorted by them. In turn, this will set up motions in the layer beneath that interface, which will perturb the layer below, etc., etc.. Thus, in addition to propagating horizontally along the interfaces, the disturbance will propagate *vertically* within the fluid. This is unlike the case of surface waves on a fluid of constant density; such internal waves can propagate vertically by virtue of the fluid's internal density structure. This is illustrative of the way fluids can often support three-dimensional wave propagation.

---

[1] One way of experiencing this is to gently rock a jar of oil-vinegar dressing to set up oscillations on the interface; when you find the resonance, the period will be much longer than if you repeat the experiment with a jar of oil or vinegar alone.

Figure 3.2: Interfacial (internal) waves in a fluid with many constant-density layers.

## 3.2 Internal waves in a fluid with continuous stratification

Most fluids—including the ocean and the atmosphere, do indeed have internal variations of density. Sometimes these variations occur sharply, but there is almost always a continuous variation of density, which supports internal waves in much the same way. In fact, if (see Fig. 3.3) we compare two fluids, one with many layers of slightly different density (which increases monotonically with depth), and the other with a continuous but otherwise similar density profile, it does not take much imagination to see that they would both behave very similarly; each density profile will support internal waves.

In fact, if $\rho$ varies linearly with $z$ in an incompressible fluid, the dispersion relation for plane waves of the form $w = Re\left[W_o e^{i(kx+ly+mz-\omega t)}\right]$ is

$$\omega = \pm N\sqrt{\frac{k^2 + l^2}{k^2 + l^2 + m^2}}, \tag{3.2}$$

where

$$N = \sqrt{\frac{g}{\rho}\frac{d\rho}{dz}} \tag{3.3}$$

is the "buoyancy frequency". Note that as $m \to 0$, $\omega \to \pm N$ (this actually corresponds to the case where air motions are exactly vertical), and that, in
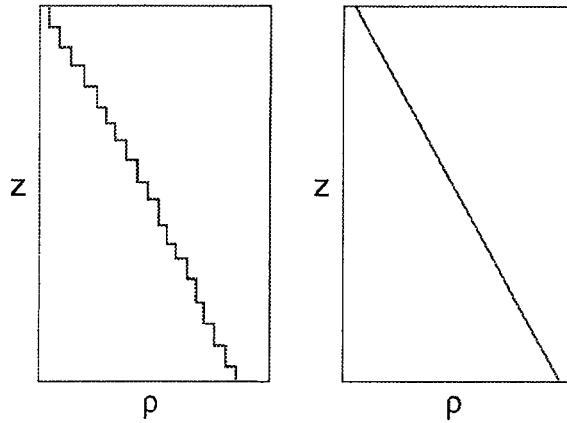
Figure 3.3: Density stratification of a fluid with (left) density steps and (right) continuous stratification.

general, $|\omega| \leq N$, so that the buoyancy frequency is the *maximum* frequency of these waves. The corresponding period, $2\pi/N$, may range from a few minutes in the atmosphere to several minutes to hours or days in the ocean.

## 3.3    Vertical density structure of the ocean

A typical vertical profile of ocean density is shown in Fig. 3.4.    The actual profile at any place and time will vary but the main characteristics are the same:

1. A "mixed layer" with the top few tens of meters, within which the density is almost uniform;

2. A "thermocline" at depths of around 100m, with a sharp density contrast (but note that its magnitude is only a few percent);

3. Below the thermocline, weaker but persistent gradients of density.

Such a profile is capable of supporting fast surface waves, slower interfacial waves on the thermocline, and much slower internal waves in the deep ocean. Internal waves are ubiquitous in the ocean.

Image removed due to copyright considerations.

## 3.4  Gravity waves in the Atmosphere

### 3.4.1  The vertical structure of a compressible atmosphere

Unlike the ocean, of course, the atmospheric density varies dramatically with height, primarily because of the compressibilty of air. We know that, in most situations (*i.e.*, unless vertical accelerations are significant, which only usually happens for small-scale motions), hydrostatic balance is satisfied:

$$\frac{\partial p}{\partial z} = -g\rho \; . \tag{3.4}$$

To determine how $\rho$ and $p$ vary with $z$, we need to invoke the **equation of state** (the relationship between pressure, density and temperature). For air, a good representation is the ideal gas law

$$pV = R^*T \tag{3.5}$$

where $V$ is the volume of one kilomole of air and $R^* = 8314.3$ J deg$^{-1}$kmol$^{-1}$ is the **universal gas constant**. Since $\rho = M/V$, where $M = 28.97$ kg is the mass of one kilomole of dry air (of mean molecular weight 28.97), eq. (3.5) may be written

$$p = \rho RT \; , \tag{3.6}$$

where $R = R^*/M = 287$ J deg$^{-1}$ kg$^{-1}$ is the **gas constant for air**.

Now, substituting from (3.6) into (3.4), we obtain

$$\frac{\partial p}{\partial z} = -g\frac{p}{RT} = -\frac{p}{H} \; ,$$  (3.7)

where

$$H = \frac{RT}{g}$$  (3.8)

is the pressure **scale height**. If $H$ is constant (isothermal atmosphere), for example, pressure decays exponentially with height, with $e$-folding scale $H$:

$$p = p_s e^{-\frac{z}{H}} \; ,$$  (3.9)

where $p_s$ is the surface pressure (at $z = 0$), and density likewise:

$$\rho = \frac{p}{RT} = \frac{p_s}{RT}e^{-\frac{z}{H}} \; .$$  (3.10)

If the atmosphere is not isothermal, but $T = T(z)$, $H = H(z)$ and

$$p = p_s \exp\left(-\int_0^z \frac{dz'}{H(z')}\right) \; ,$$  (3.11)

so $H$ is still the measure of the rate of decay of $p$, but in a local sense. For a typical value of $T = 270K$, $H \simeq 7.9$ km.

An example of a typical atmospheric temperature vs. height profile (at 35°N in April) is shown in Fig. 3.5. Within the troposphere ($z < 10$km at high latitudes, $z < 16$km in the tropics), temperature decreases with altitude at a rate of about 7K km$^{-1}$; in the stratosphere (up to $z \simeq 50$km), the temperature increases slowly with altitude.

## 3.5  Potential temperature and static stability

Consider the vertical displacement of air parcel, as shown in Fig. 3.6. The parcel $P$ is displaced from $z$ to $z + dz$, *i.e.*, from $p$ to $p + dp$, where, from (3.4),

$$dp = -g\rho(z)\, dz \; .$$
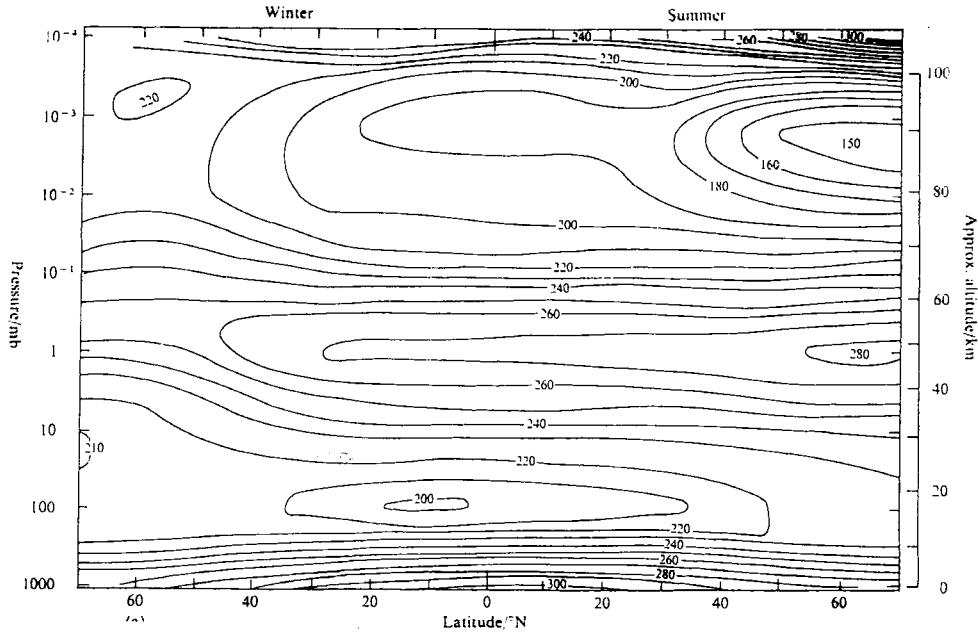
Figure 3.5: The observed, longitudinally averaged temperature distribution in northern summer. [After Houghton, "The Physics of Atmospheres", Cambridge Univ. Press, 1977.]

Since the pressure acting on the parcel changes during displacement, its density will also change, and the two are related to one another *and to temperature* through (3.6). In order to evaluate how density changes we need to know how the temperature changes, which in turn requires knowledge of the parcel's **heat budget** during displacement.

## 3.5.1 Thermodynamics of dry air

The first law of thermodynamics[2] states that the change in energy, $dq$, per unit mass of air undergoing temperature and density changes is

$$dq = c_v \, dT + p \, d\alpha \, , \tag{3.12}$$

where $c_v$ is the specific heat at constant volume and $d\alpha$ the change in specific volume (the volume of the unit mass). Since $\alpha = 1/\rho$, $d\alpha = -d\rho/\rho^2$.

---

[2] Good discussions of elementary atmospheric thermodynamics can be found in Chapter 2 of Wallace & Hobbs, *Atmospheric Science: an Introductory Survey*, (Academic Press, 1977) and Fleagle & Businger, *An Introduction to Atmospheric Physics*, (Academic Press, 1980).
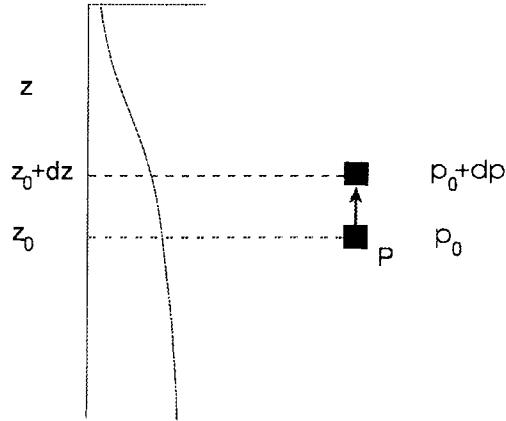
Figure 3.6: Vertical displacement of a compressible air parcel.

Therefore

$$p\,d\alpha = p\ d(\frac{1}{\rho}) = d(\frac{p}{\rho}) - \frac{1}{\rho}dp = R\,dT - \frac{1}{\rho}dp\ ,$$

where we have used the ideal gas law (3.6). Therefore (3.12) can be written

$$dq = c_p\,dT - \frac{1}{\rho}dp\ ,$$

where $c_p = c_v + R$ is the specific heat at constant pressure. To convert this into an equation for the change in heat content *per unit volume*, $dQ$, we just multiply by $\rho$ to give

$$dQ = \rho c_p dT - dp\ . \tag{3.13}$$

Hence, we can now write the first law in time derivative form (its customary form for meteorological application):

$$\frac{dT}{dt} - \frac{1}{\rho c_p}\frac{dp}{dt} = \frac{J}{\rho c_p}\ , \tag{3.14}$$

where $J = dQ/dt$ is the so-called **diabatic heating rate** per unit volume.
   Consider now the quatity

$$\theta = T\left(\frac{p_0}{p}\right)\kappa\ , \tag{3.15}$$

where $p_0$ is a constant (conventionally taken to be 100 kPa = 1000 mb) and $\kappa = R/c_p = 2/7$ for air. Then

$$
\begin{aligned}
d\theta &= dT \left(\frac{p_0}{p}\right) \kappa - \kappa T \frac{dp}{p} \left(\frac{p_0}{p}\right) \kappa \\
&= dT \left(\frac{p_0}{p}\right) \kappa - \frac{dp}{\rho c_p} \left(\frac{p_0}{p}\right) \kappa ,
\end{aligned}
$$

where we have used (3.6) to show $\kappa T/p = RT/pc_p = 1/\rho c_p$. Therefore (3.14) can be written

$$
\frac{d\theta}{dt} = \frac{J}{\rho c_p} \left(\frac{p}{p_0}\right) \kappa . \tag{3.16}
$$

Eq. (3.16) has the obvious advantage of being more concise than (3.14), but its great power—and the usefulness of the quantity $\theta$, which is known as **potential temperature**—becomes clearest under circumstances in which the diabatic heating $J$ can be neglected. The most important heating (or cooling) processes are:

(i) latent heating or cooling associated with condensation or evaporation of water. This is a very important process, which we will discuss in detail later.

(ii) radiation. On time scales of several days or longer, this is an important process, but is usually weak on shorter time scales.

(iii) conduction. This process is only important **very** close to the surface.

For dry motions, on sufficiently small time scales, and outside the boundary layer, it is usually valid to neglect the diabatic heating, in which case the motions are **adiabatic** and (3.16) becomes simply

$$
\frac{d\theta}{dt} = 0 . \tag{3.17}
$$

Potential temperature is thus **conserved** under adiabatic conditions[3]. Unlike temperature, the potential temperature does not change as an air parcel moves adiabatically to higher or lower pressure. Note that at $p = p_0$, $\theta =$

---

[3] $\theta$ is actually a measure of the **specific entropy** of air (which in fact is $c_p \ln \theta$, to within an arbitrary constant), which does not change under adiabatic processes.

$T$: so, if a parcel at some location in the atmosphere has temperature $T$ and potential temperature $\theta$ then, if $p \neq p_0$, $\theta$ and $T$ will be different. If we move the parcel adiabatically to the standard pressure, it will still have potential temperature $\theta$, but its temperature will now be $T = \theta$. Therefore, the physical meaning of $\theta$ is:

1. The potential temperature of an air parcel is the temperature it would have if moved adiabatically to standard pressure (1000 mbar).

## 3.5.2  Static stability

Now let's return to the vertically displaced air parcel of Fig. 3.6. If we assume the displacement is rapid (hours or less) and that there is no moisture condensation within the parcel, then we can assume the displacement to be adiabatic, so that $d\theta = 0$ as the parcel is displaced. Now, the parcel leaves height $z$ with initial density $\rho_i = \rho_e(z) = p(z)/RT_e(z)$, where $T_e$ and $\rho_e$ are the environmental temperature and density. At the final height, $z + dz$, the parcel has density

$$\rho_f = \frac{p(z + dz)}{RT_f}$$

and the environmental density is

$$\rho_e(z + dz) = \frac{p(z + dz)}{RT_e(z + dz)} \ .$$

Now the parcel will be **buoyant**—and therefore the displacement will continue to grow—if $\rho_f < \rho_e(z+dz)$, i.e., if $T_f > T_e(z+dz)$. If $T_f < T_e(z+dz)$, however, the parcel will be **negatively buoyant** and will return toward its original position: the environment will then be **statically stable** with respect to displacement. Now, since $d\theta = 0$ for the parcel, its temperature will change according to

$$dT = \frac{dp}{\rho c_p} = -\frac{g}{c_p} dz \tag{3.18}$$

where we have used hydorstatic balance (3.4). Therefore its final temperature is

$$T_f = T_i - \frac{g}{c_p} dz = T_e(z) - \frac{g}{c_p} dz \ ,$$

(assuming it left with environmental temperature). But the environmental temperature at this location is

$$T_e(z + dz) = T_e(z) + \frac{dT_e}{dz}dz \ ;$$

therefore the environment will be

$$
\begin{array}{llll}
\text{unstable} & if & \frac{dT_e}{dz} < -\frac{g}{c_p} & \left(\frac{d\theta_e}{dz} < 0\right) \\
\text{stable} & if & \frac{dT_e}{dz} > -\frac{g}{c_p} & \left(\frac{d\theta_e}{dz} > 0\right)
\end{array} \ . \tag{3.19}
$$

The critical value of temperature gradient

$$\frac{dT_e}{dz} = -\Gamma_{ad} = -\frac{g}{c_p} \tag{3.20}$$

is known as the **adiabatic lapse rate**. $c_p$ for air has a value of 1004 J $K^{-1}kg^{-1}$, so $\Gamma_{ad} = 9.8$ K $km^{-1}$. Usually (though not always), the actual lapse rate of temperature is less than this (typically 6-7 K $km^{-1}$) so the atmosphere is usually stable to dry displacements of this kind.

## 3.6 Internal waves in the atmosphere

### 3.6.1 The buoyancy frequency in a compressible atmosphere

A statically stable atmosphere, like a stably stratifed ocean, will support internal gravity waves. In fact, atmospheric internal waves are almost identical to those in the ocean—satisfying the same dispersion relation (3.2), for example[4]—but there is one major modification to be made. The buoyancy frequency for incompressible waves is proportional to vertical density gradient; in the atmosphere, as we have seen, it is not this that determines buoyancy, but the gradient of potential temperature. So, for the atmosphere, the buoyancy frequency [cf., eq. (3.3)] is defined by

$$N^2 = \frac{g}{\theta}\frac{d\theta}{dz} = \frac{g}{T}\left(\frac{dT}{dz} + \Gamma_{ad}\right) \ . \tag{3.21}$$

---

[4]There are some minor terms to be added to (3.2) in general, but in practice (3.2) is a good approximation.

Typically, in the troposphere, $N \simeq 0.01$ s$^{-1}$, corresponding to a period for vertical displacements of about 10 minutes (remember that this is a lower limit of the period of internal waves in general).

Like the ocean, the atmosphere is rich in internal waves (they can often be seen in clouds) though under most circumstances, their amplitudes are not very large in the lower atmosphere. One situation in which they are commonly large is when air flows over mountains—we shall look at such waves below.

Because of one other important effect of compressibility, these waves assume much greater importance in the upper atmosphere (especially in the mesosphere, above 50km altitude). As we have seen, such waves can prop-
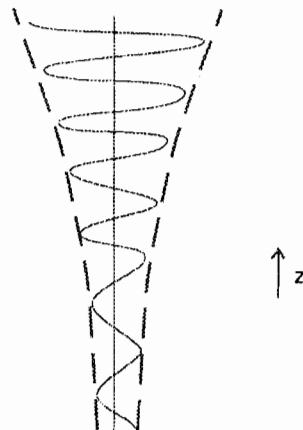


Figure 3.7: Schematic of vertically propagating internal waves.

agate vertically as well as horizontally; as they do, they encounter reduced environmental density. In order to conserve their energy (or something like it), they must increase their amplitude (Fig. 3.7) as they propagate to higher altitudes, rather like when ocean waves run up toward a beach into shallower water. The amplitude grows as something like $\rho^{-1/2}$. As a result, wave amplitudes are much larger in the upper atmosphere than in the lower atmosphere, even though it is in the latter that most of them originate.

## 3.6.2 Mountain waves

Air flowing over mountains produces a stationary wave train, just as a rock in a river produces a train of surface waves. In the former case, for mountains less than about 100km in width (we shall discuss large mountain ranges later) the wave train is comprised of internal waves which in this case are known as **lee waves**. A typical situation is shown in Fig. 3.8.
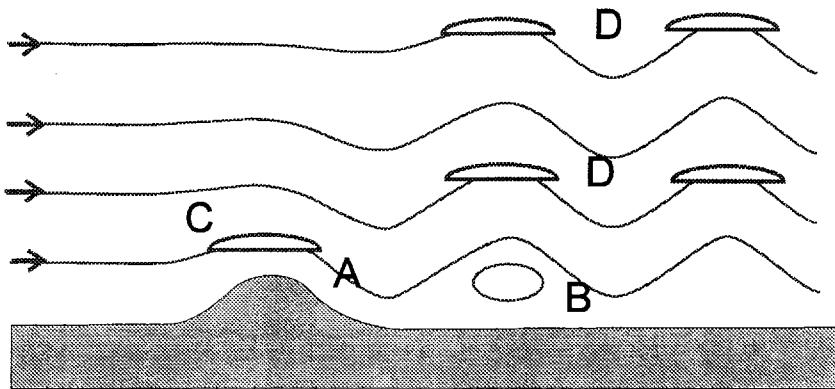


Figure 3.8: Schematic of air flow over a mountain range.

There are several noteworthy features of this flow:

(i) Like the rock-in-the-river problem, in situations where a wave train is produced, it exists downstream of the mountain, and for the same reasons. The wave train is stationary relative to the mountain. Consider the two-dimensional case with $y$-wavenumber $l = 0$. If the oncoming wind (which we assume to be uniform) is $U$, then *relative to the flow*, the mountain, and the wave train, have speed $-U$, whence, from (3.2),

$$\frac{\omega}{k} = -U = -\frac{N}{\sqrt{k^2 + m^2}}$$

(the minus sign has been chosen because the propagation is to the left). The $x$-component of group velocity relative to the flow is

$$c_{gx} = \frac{\partial \omega}{\partial k} = -\frac{Nm^2}{(k^2 + m^2)^{\frac{3}{2}}} = -U\,\frac{m^2}{k^2 + m^2} \geq -U$$

and therefore the group velocity relative to the mountain, $c_{gx} + U \geq 0$: there are no upstream effects[5].

(ii) Immediately in the lee of the mountain, (A on Fig. 3.8) there may be strong, warm, downslope winds. The air is warm because it has come from above, and (since the stratification is stable) $d\theta/dz > 0$, so air from aloft is warmer than surface air if the former is brought down to the surface. The air may also be warmed by latent heating associated with condensation in the upslope flow (see point (v), below).

(iii) Further downstream, there may be strong surface winds where (B on Fig. 3.8) the streamlines concentrate near the surface; these winds may occasionally be extremely strong, but may exist only in a narrow band parallel to the mountain range.

(iv) In regions above point A and above and just upstream of B, there is downward flow. Occasionally, this flow may be manifested as strong downdrafts that can be hazardous to aircraft operating out of or into airports downwind of large mountains.

(v) As the air is elevated over the mountain, condensation may occur, and **orographic clouds** are common (C on Fig. 3.8).

(vi) Clouds also frequently form at one or more levels in the peaks of the lee wave (D on Fig. 3.8). These lee-wave clouds are often seen with banded structure downstream of long ranges, but may also occur with less organization downstream of isolated mountains.

(vii) The lee waves propagate vertically, and so the form drag on the mountain may be communicated by the waves' momentum transport to high levels in the atmosphere. This process is significant enough to be included as an explicit parameterization in numerical weather prediction models, and, at very high levels, also has a dramatic effect on the circulation of the mesosphere.

Finally, we should note that our discussion of internal gravity waves has (for simplicity) been confined to waves on uniform background states (constant $N$ and $U$). In fact, the most dramatic mountain waves are found where

---

[5]There may be upstream effects for small $U$, when no wave train is produced and the flow cannot creep over the mountain, and when nonlinear effects we have not considered may be important.

$N$ and/or $U$ are very nonuniform, in which case **wave trapping**, and consequent amplification, may occur.

## 3.7 Further reading

Internal gravity waves are covered to some extent in many texts of geophysical fluid dynamics; a detailed but thorough treatment is given in Chapter 6 of A.E. Gill, *Atmosphere-Ocean Dynamics*, Academic Press, 1982.

## 3.8  Appendix to Ch. 3: Theory of internal gravity waves

### 3.8.1  Stable density stratification in an incompressible fluid

Consider the situation depicted in Fig. 3.9. The water is assumed to be
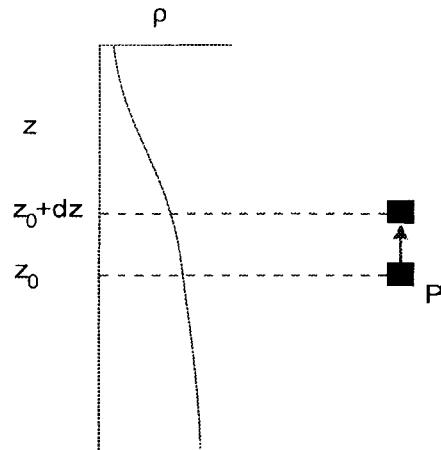


Figure 3.9: Displacement of a water parcel $P$ in stable stratification.

incompressible and to have density varying with depth only, $\rho = \rho(z)$. (N.B. Incompressibility means that density does not change in response to pressure variations; but it does depend on temperature and salinity, $\rho = \rho(T, S)$, so is not spatially constant.) A water parcel $P$, initially located at $z = z_0$, is displaced upward to $z = z_0 + dz$. The parcel initially had the same density as its environment, $\rho_P = \rho(z_0)$. Now, if we make the reasonable assumptions that there are no sources or sinks of salt within the parcel, and that it moves quickly enough to do so **adiabatically** (without loss or gain of heat), then it preserves its $T$ and $S$, and thus its density. So, after displacement, its density is still $\rho_P = \rho(z_0)$.

Now, its environment at its new location has density

$$\rho_e = \rho(z_0 + dz) \simeq \rho(z_0) + dz \frac{d\rho}{dz}(z_0) \, ,$$

for small displacement $dz$. Therefore the parcel will feel a **buoyancy** force causing it to rise further, or to return toward its initial location, depending on whether $\rho_e$ is greater than or less than $\rho_P$. We will defer discussion of the first possibility—the **unstable case**—until later; we now concentrate on the case of **stable stratification**, *viz.*,

$$\frac{d\rho}{dz} > 0 \, , \tag{3.22}$$

for which the buoyancy is negative and the associated restoring force tends to make the parcel motion oscillatory about its location of neutral buoyancy.

## 3.8.2 Small amplitude motions in an incompressible fluid with continuous stratification

So, we shall consider inviscid, adiabatic motions in an infinite, two-dimensional $(x - z)$ fluid, with density $\tilde{\rho}(x, z, t)$. As in Chapter 2, the equations of motion are

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + w \frac{\partial u}{\partial z} = -\frac{1}{\tilde{\rho}} \frac{\partial \tilde{p}}{\partial x} \, ,$$

$$\frac{dw}{dt} = \frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + w \frac{\partial w}{\partial z} = -\frac{1}{\tilde{\rho}} \frac{\partial \tilde{p}}{\partial z} - g \, ,$$

where $\tilde{p}$ is pressure. We note here that density variations in the ocean are small, and so we can write $\tilde{\rho} = \rho_{00} + \rho$, where $\rho$ is a small deviation from the constant reference density $\rho_{00}$. To be consistent, we also have to allow a reference pressure $p_{00}(z)$, in hydrostatic balance with $\rho_{00}$, such that

$$\frac{dp_{00}}{dz} = -g\rho_{00} \, ,$$

and so we write $\tilde{p} = p_{00} + p$. Then, relying on the smallness of $\rho$, we write

$$-\frac{1}{\tilde{\rho}} \frac{\partial \tilde{p}}{\partial x} \simeq -\frac{1}{\rho_{00}} \frac{\partial p}{\partial x} + \left\{ \frac{\rho}{\rho_{00}^2} \frac{\partial p}{\partial x} \right\} \, ,$$

$$-\frac{1}{\tilde{\rho}}\frac{\partial \tilde{p}}{\partial z} \simeq -\frac{1}{(\rho_{00}+\rho)}\frac{\partial (p_{00}+p)}{\partial z}$$

$$\simeq -\frac{1}{\rho_{00}}\frac{\partial (p_{00}+p)}{\partial z} + \frac{\rho}{\rho_{00}^2}\frac{\partial (p_{00}+p)}{\partial z}$$

$$\simeq g - \frac{1}{\rho_{00}}\frac{\partial p}{\partial z} + g\frac{\rho}{\rho_{00}} + \left\{\frac{\rho}{\rho_{00}^2}\frac{\partial p}{\partial z}\right\} .$$

Note that the terms in curly brackets are quadratic in departures from the reference state, so we neglect them, in which case the eqs. of motion become

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho_{00}}\frac{\partial p}{\partial x} ,$$

$$\frac{dw}{dt} = \frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + w\frac{\partial w}{\partial z} = -\frac{1}{\rho_{00}}\frac{\partial p}{\partial z} + g\frac{\rho}{\rho_{00}} .$$
(3.23)

We also have our incompressible continuity eq.

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0 .$$
(3.24)

To close the problem we need an equation for density. On the basis of our assumption that the motions are adiabatic, and that there are no sources or sinks of salt, it follows that parcels must conserve their density as they move around, *i.e.*,

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} + w\frac{\partial \rho}{\partial z} = 0 .$$
(3.25)

Now, we consider a steady, motionless basic state, in which $\rho = \rho_0(z)$ is a linear function of $z$ (for simplicity), such that $d\rho_0/dz = \Lambda$. The second of (3.23) tells us that the basic state pressure field $p_0(z)$ must be in hydrostatic balance with this density field:

$$\frac{dp_0}{dz} = -g\rho_0(z).$$
(3.26)

We now consider small amplitude perturbations to this state, such that

$$u = u'(x,z,t) ,$$
$$w = w'(x,z,t) ,$$
$$p = p_0(z) + p'(x,z,t) ,$$
$$\rho = \rho_0(z) + \rho'(x,z,t) .$$

Since the perturbations are small, we may neglect nonlinear terms like $u'\frac{\partial u'}{\partial x}$ and $w'\frac{\partial \rho'}{\partial z}$; therefore, our linearized perturbation equations become, from (3.23), (3.24), and (3.25),

$$\frac{\partial u'}{\partial t} + \frac{1}{\rho_{00}}\frac{\partial p'}{\partial x} = 0$$
$$\frac{\partial w'}{\partial t} + \frac{1}{\rho_{00}}\frac{\partial p'}{\partial z} = -g\frac{\rho'}{\rho_{00}} \tag{3.27}$$
$$\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0$$
$$\frac{\partial \rho'}{\partial t} - w'\Lambda = 0 \,.$$

With some juggling[6], these can be reduced to a single equation for $p'$:

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2 p'}{\partial x^2} + \frac{\partial^2 p'}{\partial z^2}\right) + N^2\frac{\partial^2 p'}{\partial x^2} = 0 \,. \tag{3.28}$$

The quantity $N$ in (3.28) has units of $t^{-1}$, and is defined by

$$N^2 = \frac{g\Lambda}{\rho_{00}} = \left(\frac{g}{\rho}\frac{d\rho}{dz}\right)_0 \,; \tag{3.29}$$

we will see its significance in a moment.

Clearly, (3.28) supports wavelike solutions of the form

$$p'(x,z,t) = Re\, P\, e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} = Re\, P\, e^{i(kx+mz-\omega t)} \,, \tag{3.30}$$

where $\mathbf{k}$ is wavenumber and $(k,m)$ its components in the $(x,z)$ directions, provided

$$\omega = \pm\frac{Nk}{\sqrt{k^2 + m^2}} \,. \tag{3.31}$$

Eq. (3.31) is our dispersion relation for internal gravity waves. It tells us that the wave frequency is independent of the *magnitude* of wavenumber, only on its direction; specifically, that

$$\omega = \pm N\sin\vartheta \tag{3.32}$$

---

[6] Take $\partial/\partial x$ of the 1st of (3.27) plus $\partial/\partial z$ of the 2nd, and use the 3rd to give $g\partial\rho'/\partial z = \partial(\partial^2 p'/\partial x^2 + \partial^2 p'/\partial z^2)/\partial t$; substitute this and the 1st equation into $\partial^2/\partial z\partial t$ of the 4th equation.

where $\vartheta = \arctan(k/m)$ is the angle the wavenumber vector makes with the vertical. For waves with wavenumber pointing horizontally (vertical wave crests), $\omega = \pm N$. So the quantity $N$ defined in (3.29)—which is thus known as the **buoyancy frequency** or the **Brunt-Väisälä frequency**—gives the frequency of such waves; in general (when $\vartheta$ is not $\pi/2$) it provides the scale for frequency, although it should be noted that in both ocean and atmosphere, very slow waves with $\omega \ll N$ (so $\vartheta \ll 1$) are common. $N$ is the upper limit of frequency for **propagating** waves, for which both components of wavenumber are real (if $\omega < N$, say because of external forcing at frequency $\omega$, at least one of $k$ and $m$ must be imaginary, and the disturbance will be **evanescent** in at least one direction).

Note from the 3rd of eqs. (3.27), together with (3.30) that

$$ku' + mw' = \mathbf{k} \cdot \mathbf{u}' = 0 \; ;$$

the motions are at right angles to the wavenumber, and thus along the phase lines: the wave motion is **transverse**. This is illustrated in Fig. 3.10. When $\vartheta = \pi/2$, the phase lines and motions are aligned vertically—so the
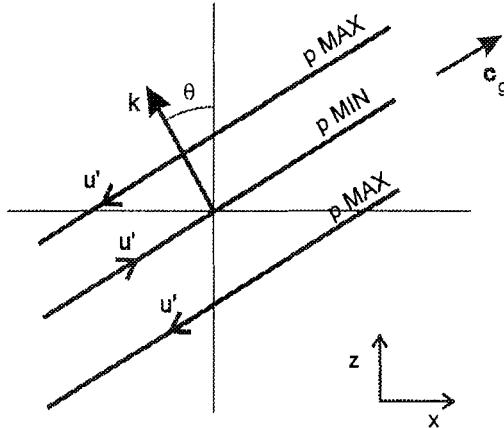


Figure 3.10: Phase lines and motions within a plane internal gravity wave.

oscillation of fluid parcels is just as we discussed at the beginning of this chapter; thus $N$ is the frequency of vertically-displaced parcels. For other values of $\vartheta$, the component of restoring force along the angle of displacement is what matters—hence (3.32).

# Chapter 4

# Tides

## 4.1   Tidal forcing

### 4.1.1   The "semi-diurnal" component

We need to consider how gravitational forces, due to the Sun or Moon, vary along the surface of the Earth. For simplicity in the following derivation, we shall focus on the Sun-Earth system (the Earth-Moon system produces the same result, but the analysis is alittle more complicated). We shall also neglect the inclination of the Earth's axis to its orbit, and consider only how the forcing varies along the equator; the geometry is shown in Fig. 4.1.
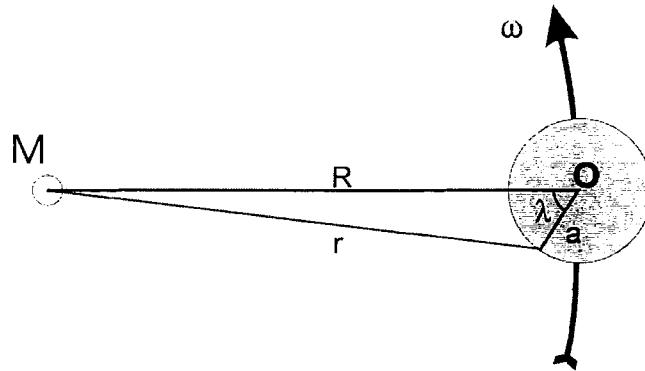


Figure 4.1: Gravitational tidal forcing.

Now, the gravitational potential at longitude $\lambda$ (measured relative to the

moving sub-solar or sub-lunar point on the Earth's surface, due to the tide-raising body (Sun or Moon) of mass $M$, is

$$\Phi_g = -\frac{GM}{r} = -\frac{GM}{\sqrt{R^2 - 2aR\cos\lambda + a^2}} \, ,$$

where $R$ is the distance of the tide-raising body from the center of the Earth, and $a$ the Earth's radius. Since $a/R \ll 1$, this can be approximated, correct to $O(a^2/R^2)$, as

$$\Phi_G \approx -\frac{GM}{R}\left(1 + \frac{a}{R}\cos\lambda - \frac{a^2}{2R^2}\left(1 - 3\cos^2\lambda\right)\right) \, .$$

Now, assuming that the center of the Earth's orbit coincides with the center of mass of $M$[1], the centrifugal potential is

$$\Phi_C = -\frac{1}{2}\omega^2 r^2 = -\frac{1}{2}\omega^2 R^2\left(1 - 2\frac{a}{R}\cos\lambda + \frac{a^2}{R^2}\right) \, ,$$

where $\omega$ is the angular velocity of the Earth in its orbit. Since the two components of force must balance at the Earth's center,

$$\omega^2 R = \frac{GM}{R^2} \, .$$

Therefore, the net variation of tidal potential around the equator is

$$\Phi_T = \Phi_G + \Phi_C = -\frac{3GM}{2R} - \frac{3GMa^2}{4R^3}\left(1 + \cos 2\lambda\right) \, . \tag{4.1}$$

The constant terms are, of course, irrelevant. The longitudinally-varying part describes a potential with wavenumber 2 around the globe: this is because the gravitational force decreases with distance, and the centrifugal force increases, so the former dominates at the subsolar (sublunar) point $P$, and the latter at $Q$, the antipodes of $P$ (see Fig. 4.2).

---

[1]This is clearly a very poor approximation for the Earth-Moon system; however, the end result is the same.
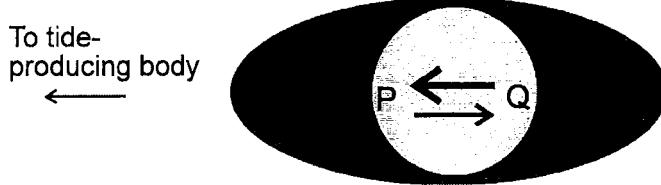
Figure 4.2: Illustrating the "semidiurnal" (wave 2) nature of the tidal potential.

Note that (because it is a *differential* measure of the gravitational field) the tidal forcing varies as $R^{-3}$.

The corresponding tidal force (per unit mass) is $-\nabla\Phi_T$; the horizontal component, along the Earth's surface, is the relevant one and this is just

$$F_T = -\frac{1}{a}\frac{\partial\Phi_T}{\partial\lambda} = -\frac{3GMa^2}{2R^3}\sin 2\lambda \ . \tag{4.2}$$

This is depicted in Fig. 4.3.
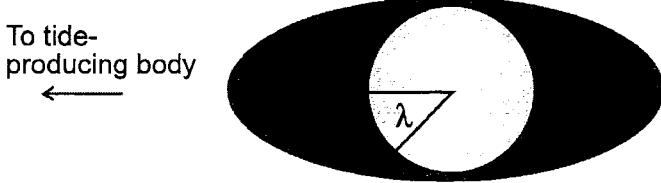


Figure 4.3: Tidal forces. The ellipse depicts the "equilibrium tide".

## 4.1.2 Lunar vs. solar forcing

Note that the magnitude of the tidal force depends on the properties of the tide-raising body as $M/R^3$. If the radius of the body is $b$, and its mean density $\rho$, then

$$\frac{M}{R^3} = \left(\frac{4\pi}{3}\right)\rho\left(\frac{b}{R}\right)^3 \ . \tag{4.3}$$

Now, because of the happy coincidence that the sun and moon subtend almost identical angles $\tan^{-1}(b/R)$ at the Earth, the ratio of their tidal forces is, by (4.3), approximately equal to their mean densities. As the lunar density

exceeds that of the sun (by a ratio of about 2:1), lunar tidal forces are greater than solar, and the dominant tide in most places on the Earth is lunar semidiurnal (period of about 12hr 25min). The solar forcing is by no means negligible, however, which is why the tide goes through its monthly modulation from the high "spring" tides, when lunar and solar forcings are in phase, to the weaker "neap" tides, when they are out of phase.

### 4.1.3   The "diurnal" component

Since the inclination of the Earth's axis to the Earth-moon and Earth-sun lines is not zero, the tidal forces are not purely semidiurnal. As shown in Fig. 4.4,
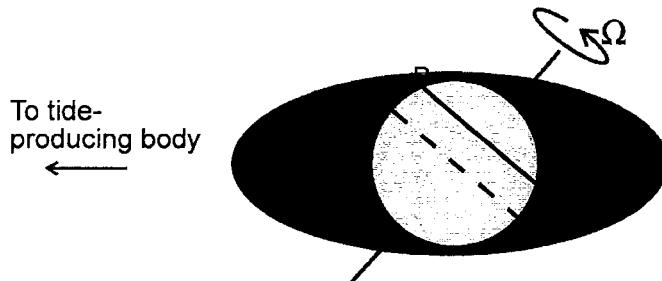


Figure 4.4: Illustrating the diurnal tidal component. Because the inclination of the Earth's axis is not zero, the high tide experienced at point $P$ is weaker than that experienced 12 (lunar) hours later at point $Q$.

the tilt of the rotation axis relative to the potential surfaces introduces a diurnal asymmetry: the tidal potential maximum at $Q$ is stronger than that at $P$. Thus, the tidal forcing has a diurnal, as well as semidiurnal, component. Note that the magnitude of this component will vary with period of a lunar month, as the orientation of the poles with respect to the Earth-moon line changes.

| quantity | value | units |
|----------|-------|-------|
| $G$ | $6.67 \times 10^{-11}$ | $Nm^2kg^{-1}$ |
| $M$ | $7.30 \times 10^{22}$ | kg |
| $R$ | $3.82 \times 10^8$ | m |
| $a$ | $6.38 \times 10^6$ | m |
| $g$ | $9.78$ | $ms^{-2}$ |

Table 4.1: Data for the tidal calculation

## 4.2 Tides in the ocean

### 4.2.1 The "equilibrium tide"

The total gravitational potential around the Earth includes, of course, that due to the Earth's own gravity, $\Phi = gz$. If we consider the lunar forcing only, then if the Earth were not rotating, the surface of the ocean would, in equilibrium, coincide with a geopotential surface, on which, using (4.1),

$$-\frac{3GMa^2}{4R^3}\cos 2\lambda + gz = \text{ constant} .$$

This surface is shown schematically (and much exaggerated!) in Fig. 4.3. Since $\cos 2\lambda$ varies from $-1$ to $1$, the extreme range (low to high tide) for the "equilibrium tide" is

$$Z_e = \frac{3GMa^2}{2gR^3} \tag{4.4}$$

Using the values from Table 4.1, we obtain the value[2] $Z_e = 0.545$m.

### 4.2.2 Tides in a global ocean

In reality, the tidal pattern remains fixed with respect to the Earth-moon axis, and so, relative to a point on the Earth's surface, it moves westward at a speed of $449$ms$^{-1}$ at the equator. Now, a typical ocean depth is about $D = 5$km; since the wavelength of the tidal oscillation is $2\pi a/2 \simeq 2 \times 10^4$km, and this is very much larger than $D$, we can use shallow water theory to

---

[2]The actual value for the equilibrium ocean tide is about 0.7 of this. We have neglected to allow for the fact that the solid earth itself is tidally distorted, and that the solid-earth ocean system then produces a wave 2 modulation of the local gravity field.

deduce that the phase speed of free waves is $\sqrt{gD} \simeq 220\text{ms}^{-1}$. Therefore, we can hardly assume that the tide is steady, since it moves faster than free waves in the ocean. So the tide is *dynamic*—we need to consider the dynamic, rather than the static, response to the tidal forcing.

A complete analysis of tides on a global ocean (without interruption by continents) is a classic (if unrealistic) problem. Apart from needing to take account of the spherical geometry, we would also need to include the effects of the Earth's rotation, which is a significant factor for motions with periods of about 12 hrs. To avoid these complications and to get some (limited) insight into tidal motions, we consider the non-rotating problem of one-dimensional (E-W) motions in a narrow channel around a latitude circle (see Fig. 4.5).
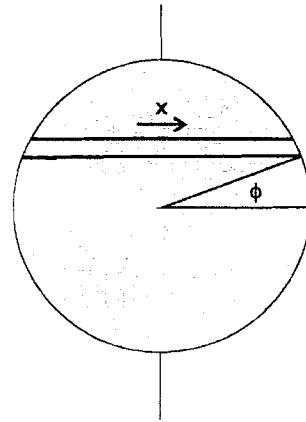


Figure 4.5: A narrow channel along a latitude circle, at latitude $\varphi$.

Since the channel is narrow, we can neglect curvature, and so use Cartesian coordinates, with $x$ the coordinate in the longitudinal direction. The channel length is $L = 2\pi a \cos\varphi$. The relevant equations for this channel for the long tidal motions are just the shallow water equations, modified to include the tidal potential $\Phi_T$:

$$\frac{du}{dt} = -g\frac{\partial h}{\partial x} - \frac{\partial \Phi_T}{\partial x} \; ;$$

$$\frac{dh}{dt} = -h\frac{\partial u}{\partial x} \; .$$

Since, from (4.4), we anticipate weak motions, we can reasonably linearize these equations about a state of uniform depth $D$ and no motion; we then

have

$$\frac{\partial u}{\partial t} = -g\frac{\partial h}{\partial x} - \frac{\partial \Phi_T}{\partial x} \; ;$$

$$\frac{\partial h}{\partial t} = -D\frac{\partial u}{\partial x} \; .$$

(4.5)

Now, we know that the forcing has zonal wavenumber 2 and period 0.5 lunar day, so we write

$$\Phi_T = \Phi_0 \cos(2\lambda) = Re\left[\Phi_0 e^{ik(x-ct)}\right]$$

where, from (4.1), $\Phi_0 = 3GMa^2/(4R^3)$, $k = \pi/L = 2/(a\cos\varphi)$ and $kc = 2\pi/(0.5\tau)$, where $\tau$ is the lunar day. If we therefore seek solutions of the form

$$\begin{pmatrix} u \\ h \end{pmatrix} = Re\left[\begin{pmatrix} U \\ H \end{pmatrix} e^{ik(x-ct)}\right] ,$$

(4.6)

then (4.5) give

$$cU = gH + \Phi_0 \; ;$$
$$cH = DU \; ;$$

and so

$$H = \frac{D}{(c^2 - c_0^2)}\Phi_0 \; ,$$

(4.7)

where $c_0 = \sqrt{gD}$ is the shallow water wave speed, as before.

In the limit $c \to 0$, this gives $H \to -D\Phi_0/c_0^2 = -\Phi_0/g$, i.e. the equilibrium tide, as we expect. Eq. (4.7) tells us:

(i) If $0 < c < c_0$, the tidal response is in phase with the equilibrium tide, and is larger;

(ii) If $c = c_0$, the response is **resonant**, since the system is being forced at its natural frequency;

(iii) If $c_0 < c < \sqrt{2}c_0$, the tidal response is larger than, **and out of phase with**, the equilibrium tide; and

(iv) If $c > \sqrt{2}c_0$, the response is smaller than, and out of phase with, the equilibrium tide.

Since, at the equator, the phase speed of the tidal forcing is 449ms$^{-1}$, while the wave speed is 220ms$^{-1}$ for an ocean of 5km depth, resonance would occur for a channel at latitude $\arccos(220/449) = 60.7°$. We cannot of course take these results literally latitude-by-latitude, as the whole spherical system is coupled together. While this exercise gives us some idea of how the local dynamics are tending to behave, it does not describe the actual tides very well, as we shall now see.

### 4.2.3  Tides in ocean basins

Tidal observations are of course made in many locations, but most of these are coastal. To get a picture of what we think the global structure of tides looks like, we have to resort to output from numerical models. Such a picture of tides is shown in 4.6.

Image removed due to copyright considerations.

The lines shown on Fig. 4.6 are of two types: **co-range** lines, which show the peak-to-peak amplitude (shown here with a contour interval of 0.25m, and labeled in meters), and **co-tidal** lines, which show the phase of the tide,

expressed as the time of high water in "lunar hours" (about 1hr 2min) after the moon passes the Greenwich meridian. Several features stand out.

1. The amplitude of the tide is in most places between 0.25 and 1.5m, *i.e.*, between one-half and three times the amplitude of the equilibrium tide.

2. The phase of the tide does not progress systematically eastward, as we assumed in the above example, except in parts of the Southern Ocean, which is the only part of the world where a disturbance can propagate right around a latitude circle, unobstructed by continents.

3. The greatest amplitudes are along the coasts, especially near gulfs. Correspondingly, there are regions of vanishingly small amplitude (so-called **amphidromic points**) in the middle of the ocean basins. The one exception to these statements is the maximum in the central equatorial Pacific Ocean.

4. The tide progresses systematically around each ocean basin (in fact, around the amphidromic points). For the most part, the progression is clockwise in the southern hemisphere and anticlockwise in the northern hemisphere.

There are two effects that make the tides look so different from our simple channel model. The most obvious is the presence of continents; the second is the Earth's rotation. One effect of the latter (we shall look at other effects later) is to allow waves to become trapped at the coasts.

## 4.2.4 Kelvin waves

Consider (*cf.* Fig. 4.7) shallow water behavior near a straight coast, in a rotating system.
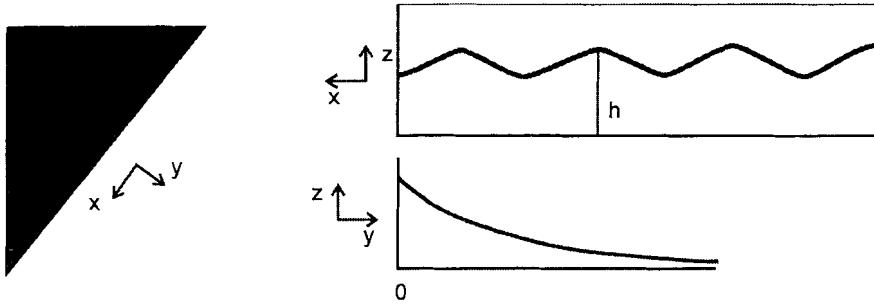
Figure 4.7: Schematic of the trapping of waves a coasts by the planetary rotation. Top figure is in a plane parallel to the coast, which runs along the x-axis; bottom figure normal to the coast.

The eqs. of motion then become (with $f$ the Coriolis parameter)

$$\frac{du}{dt} - fv = -g\frac{\partial h}{\partial x} ;$$

$$\frac{dv}{dt} + fu = -g\frac{dh}{dy} ;$$

$$\frac{dh}{dt} + D\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0 .$$

Assuming small amplitude perturbations to a basic state with no motion, and uniform depth $D$ gives

$$\frac{\partial u'}{\partial t} - fv' = -g\frac{\partial h'}{\partial x} ;$$

$$\frac{\partial v'}{\partial t} + fu' = -g\frac{\partial h'}{\partial y} ; \tag{4.8}$$

$$\frac{\partial h'}{\partial t} + D\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = 0 .$$

These eqs. have more than one kind of wavelike solution. One such solution— *the Kelvin wave*—is a little strange. The boundary condition at the coast $y = 0$ is $v = 0$: suppose there is a solution with $v = 0$ everywhere. Then

(4.8) become

$$\frac{\partial u'}{\partial t} = -g\frac{\partial h'}{\partial x} \; ;$$

$$fu' = -g\frac{\partial h'}{\partial y} \; ; \tag{4.9}$$

$$\frac{\partial h'}{\partial t} + D\frac{\partial u'}{\partial x} = 0 \; .$$

We have left ourselves with 3 eqs in 2 unknowns, which would normally suggest that we are on the wrong track. However, note that the 1st and 3rd of (4.9) are exactly the same two eqs we get in the one-dimensional, nonrotating case. So, just as in the nonrotating case, we get solutions of the form

$$\left( \begin{array}{c} u' \\ h' \end{array} \right) = \mathrm{Re}\left\{ \left( \begin{array}{c} U(y) \\ H(y) \end{array} \right) \exp\left[ik\left(x - ct\right)\right] \right\}$$

where $c = \sqrt{gD}$ and where $U = gH/c = cH/D$. However, we have the further constraint of the 2nd of (4.9), which gives

$$\frac{dH}{dy} = -\frac{f}{g}U = -\frac{f}{c}H \; ,$$

which, for constant $f$, gives[3]

$$H = const \times \exp\left( -\frac{f}{c}y \right) \; . \tag{4.10}$$

The effects of rotation for these Kelvin waves is therefore to *trap* the waves along the coastline, with an e-folding distance of $c/f$. Otherwise, the motions are entirely parallel to the coast everywhere, and the wave travel at the speed of nonrotating shallow water gravity waves. However, there is one further important implication of (4.10). Nonrotating gravity waves can propagate in either direction. But a physically meaningful solution must *decay* away from the coast (it cannot grow indefinitely as $y \rightarrow +\infty$) so we must have $f/c > 0$. In the northern hemisphere ($f > 0$), then, $c > 0$: the wave can only propagate in one direction, with the coast to the right of the direction of propagation (to the left, in the southern hemisphere where $f < 0$).

---

[3] We will later consider an important case for which $f$ is not constant.

Image removed due to copyright considerations.

In the presence of coasts, the tide takes on the characteristics of the kelvin wave. Thus, the tide will tend to propagate anticlockwise around amphidromic points in northern hemisphere ocean basins, and clockwise in the northern hemisphere. Locally, this effect may be counteracted by the tendency of the "open ocean" tide to follow the moon westward, and by interaction between adjacent amphidromic points.

## 4.2.5  Tides in inlets and bays

Similar behavior is seen on a smaller scale in smaller bodies of water. Figure 4.8 shows the tide in the North Sea. Note the large amplitude, as compared with typical open ocean values, and the similar anticlockwise propagation around the coasts. The propagation is much slower here, consistent with the

shallower water.

In such small bodies of water, the effects of gravitational forcing acting directly on the water body are small compared to the indirect effects of open ocean forcing. That is to say, tides in coastal seas and bays are driven primarily by the open ocean tide at the mouth of the bay, rather like driving an organ pipe at a specific frequency by externally playing a note at the end of the pipe. In some cases, this can lead to large amplitudes, by at least two processes. One is simply focusing: if the bay becomes progressively narrower along its length, the tide will be confined to a narrower channel as it propagates, thus concentrating its energy. There are suggestions of this in Fig. 4.8, in the English Channel at the bottom of the figure.

Image removed due to copyright considerations.

The second processes is constructive interference between the incoming tide and a component reflected from the coast. Fig 4.9 shows a more spectacular example, the tide in the Gulf of Maine.

The tidal range in the Gulf of Maine is about 3ft at its entrance, but it increases substantially towards the coast and most dramatically in the Bay of Fundy at the NE corner of the Gulf, where the tide exceeds 30ft[4]. We

---

[4]The mean tide at Burntcoat Head, at the head of the Bay of Fundy, is 38.4ft (11.8m),
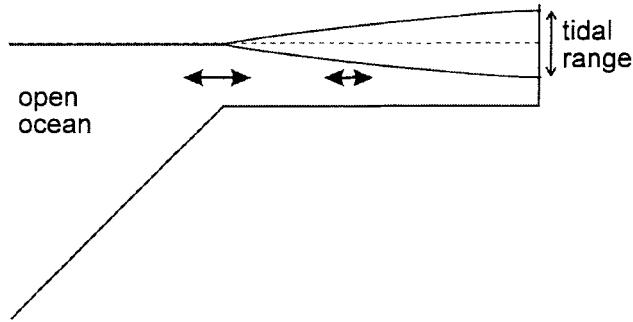
Figure 4.10: Schematic of a quarter-wavelength resonance in a bay.

have seen that a simple reflection can amplify wave amplitude at the coast by a factor of 2, but not 10 or more. What seems to be happening is that the Gulf of Maine/Bay of Fundy system is resonating at the tidal period. This is illustrated in Fig 4.10. Just as in the organ pipe problem, the bay is forced by the tidal currents at its mouth; if the geometry of the bay is such that it takes one-quarter period for a wave to propagate its length, it will support a quarter-wavelength mode at the forcing period, leading to large tides at the head of the bay.

the highest mean tide in the world.

## 4.3 Atmospheric tides

The atmosphere also has tides, if by "tides" one means motions that fluctuate with diurnal, semidiurnal (terdiurnal, ...) period. Fig. 4.11 shows surface

Image removed due to copyright considerations.

pressure at a tropical location and a northern midlatitude location. There is little evidence for a diurnal and semidiurnal component in middle latitudes, for two reasons: the signal (which can, in fact, be extracted by analysis of long time series) is weak there, and there is a large "synoptic" variability of pressure associated with day-to-day weather events, which masks the tidal signal. In the tropics, conversely, the day-to-day variability is small (for reasons we shall see later), and the tidal signal is stronger. Fig. 4.11 shows a tidal range of about 5mm Hg ($\simeq$ 7hPa) at Batavia. This rather small amplitude makes the atmospheric tide a curiosity, rather than an important

phenomenon, at the surface, though its amplitude is much larger in the upper atmosphere.

Like the ocean tide, the atmospheric tidal signal (at the surface) is predominantly semidiurnal but, unlike the ocean tide, it is *solar* semidiurnal: its phase remains fixed with respect to the (solar) clock. [This is just evident in Fig. 4.11, and more clear with a longer time series.] Since we deduced from eq. (4.3) that the lunar gravitational forcing dominates the solar, this seems curious. Moreover, calculations show that the atmospheric tide to be expected from gravitational forcing—lunar or solar—is much weaker than observed. Something else must be forcing the "tide."

That "something else" is heating. The atmosphere is of course subject to diurnally-varying solar heating, whose effects in driving large-scale motions far outweigh gravitational forcing. Even though the thermal forcing is "diurnal" it is not a single harmonic (since, for example, nighttime cooling via IR radiation varies much less through the night than does solar forcing during the day). The diurnal variation of net heating is shown schematically in Fig. 4.12. We could expand the heating $J(\lambda - \Omega_s t)$ (the whole heating pattern, to
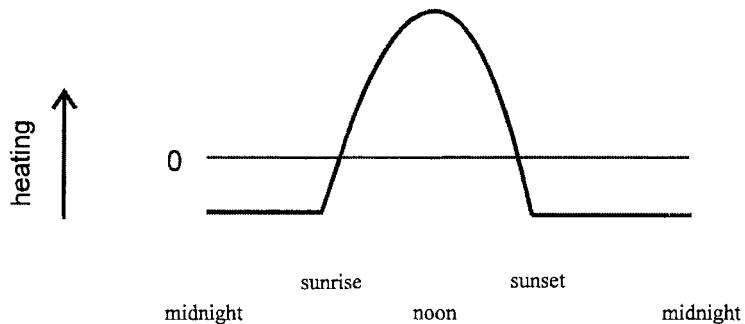


Figure 4.12:

a first approximation, it propagates around the world at the angular speed $\Omega_s$ of the subsolar point) as a series

$$J(\lambda - \Omega_s t) = \sum_{n=1}^{\infty} J_n \cos\left[n\left(\lambda - \Omega_s t\right)\right] ,$$

where $n = 1, 2, 3, ...$ corresponds to the diurnal, semidiurnal, terdiurnal, ... components, all of which will be nonzero. Nevertheless, this does not

solve our problem as, with any reasonable representation of $J$, the diurnal component is the largest. So why is the observed tide semidiurnal?

For many decades, it was thought that the answer had to lie in the resonance of the atmosphere at the semidiurnal period[5]. This was hard to prove, as in order to calculate the atmosphere's resonant frequencies, it is necessary to know its thermal structure, and little was known above altitudes of about 15km until the 1940s. Calculations then showed the resonance hypothesis to be untenable.

So what is happening? It turns out that the most important region of forcing of the thermal tide is in the stratosphere, at altitudes above around 30km, through absorption of insolation by ozone. At these altitudes, the diurnal component is larger than the semidiurnal. However, in order to reach the surface, the tide must *propagate* there—the tide is a wave motion, albeit a forced one. Now, we saw from (3.2) that internal gravity waves must satisfy

$$\omega^2 = \frac{N^2 k^2}{k^2 + m^2} \, ,$$

from which we deduced that $\omega^2$ cannot exceed $N^2$. In the tidal case, $\omega$ and $k$ are given by the forcing; the only free parameter is the vertical wavenumber $m$, which satisfies:

$$m^2 = k^2 \left( \frac{N^2 - \omega^2}{\omega^2} \right) \, . \tag{4.11}$$

So we can now be more specific and state, not that $\omega$ cannot exceed $N$, but that the vertical wavenumber is real—and hence vertical propagation can occur—only if $\omega^2 < N^2$. Otherwise, the wave is trapped in the vertical.

Now, this criterion is no problem for tides—their frequencies are very much less than $N$. However, we noted earlier that the effects of the Earth's rotation are important for tides; when rotation is included, (4.11) becomes

$$m^2 = k^2 \left( \frac{N^2 - \omega^2}{\omega^2 - f^2} \right) \, , \tag{4.12}$$

where $f = 2\Omega \sin \varphi$ ($\Omega$ =Earth rotation rate, $\varphi$ =latitude) is the Coriolis parameter. Thus, vertical propagation requires

$$f^2 < \omega^2 < N^2 \, . \tag{4.13}$$

---

[5]Atmospheric tides have, in fact, aroused the interest of some powerful minds. It was Laplace who suggested that they are thermally forced; Lord Kelvin put forward the resonance hypothesis.

Since $\omega = n\Omega$ (with an error of $1/365$, because of the Earth's orbit of the sun), and since $f$ reaches its maximum value of $2\Omega$ at the poles, $f^2 \leq N^2$ everywhere for the semidiurnal and higher components. For the diurnal component, however, $f < N$ only within $30^o$ of the equator; at higher latitudes, the tide cannot propagate downward. For this reason, much of diurnal forcing is extremely inefficient at producing a surface response, and the semidiurnal component dominates there.

## 4.4  Further reading

An elementary discussion of ocean tides is given in:
*Waves, tides and shallow-water processes*, by the Open University Course Team, The Open University, Pergamon Press, 1989.

A comprehensive discussion of the mathematical theory of tides on a global ocean is given in the classical text:
*Hydrodynamics*, H. Lamb, Cambridge U.P., 1916.

For a discussion of atmospheric tides (including an historical account of the resonance theory) see Chapter 9 of:
*Dynamics in Atmospheric Physics*, R.S. Lindzen, Cambridge U.P., 1990.

# Chapter 5

# Large-scale motions on a rotating Earth

## 5.1 The equations of motion on a rotating plane

In an inertial frame of reference, the equation of motion (momentum) is

$$\rho \frac{d\mathbf{u}}{dt} = -\nabla p - \rho \nabla \Phi + \mathbf{F} \tag{5.1}$$

where $\mathbf{u}$ is the vector velocity, $\Phi = gz$ the gravitational potential, and $\mathbf{F}$ represents any applied external body force or frictional forces acting per unit volume. For a tidal problem, $\mathbf{F}$ would represent the gravitational tidal force; in the ionized upper atmosphere, it could include forces involved in moving ions across the Earth's magnetic field lines. In almost all cases, however, such effects are negligible, and the only "force" acting is friction. Even this is negligible, except very close to the Earth's surface (for the atmosphere), or in the surface and bottom (benthic) boundary layers of the ocean.

Expressed relative to a frame rotating with the planetary rotation rate $\boldsymbol{\Omega}$, equation (5.1) is

$$\rho \left( \frac{d\mathbf{u}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{u} + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) \right) = -\nabla p - \rho \nabla \Phi + \mathbf{F} \,,$$

the 2nd and 3rd terms on the LHS representing the Coriolis and centrifugal terms, respectively, and $\mathbf{r}$ is the position vector measured from the planetary center. It is conventional to simplify things a little by absorbing the
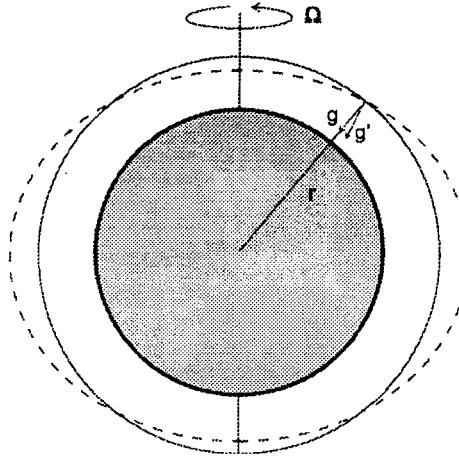
Figure 5.1: Geometry of a spherical surface (solid) and the geoid (dashed) through a point at location r.

centrifugal term into the gravitational potential. One can do this easily, since

$$\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) = \boldsymbol{\nabla}(\frac{1}{2}\Omega^2 r^2) \ ,$$

where $\Omega = |\boldsymbol{\Omega}|$ and $r = |\mathbf{r}|$; hence one can absorb this term into the definition of $\Phi$ (writing $\widetilde{\Phi} = \Phi - \frac{1}{2}\Omega^2 r^2$), leaving

$$\rho\left(\frac{d\mathbf{u}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{u}\right) = -\boldsymbol{\nabla}p - \rho\boldsymbol{\nabla}\widetilde{\Phi} + \mathbf{F} \ . \tag{5.2}$$

We now have to regard gravity ($\boldsymbol{\nabla}\widetilde{\Phi}$) not as $g$, pointing downward relative to the spherical surface through $\mathbf{r}$, but as $g'$, pointing downward relative to the **geoid** through $\mathbf{r}$ (see Fig. 5.1). So in order for gravity to remain vertical we must, in principle, use slightly non-spherical coordinates; in practice, the geoid is so close to being spherical that we can ignore this complexity without introducing significant error. Thus, we ignore the "twiddle" on $\Phi$ in (5.2).

## 5.2 Rapid rotation

The material time derivative on the LHS of (5.2) can be written

$$\frac{d\mathbf{u}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{u} = \frac{\partial\mathbf{u}}{\partial t} + (\mathbf{u} \cdot \boldsymbol{\nabla})\mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} \ .$$

If we assume that a typical magnitude for velocity is $U$, that the distance on which velocity varies is typically $L$, and the time on which it changes is typically $T$, the 3 terms on the RHS have typical magnitudes

$$\frac{U}{T} \; ; \; \frac{U^2}{L} \; ; \; 2\Omega U \; .$$

For motions that are nearly **steady**, in the sense that $2\Omega T \gg 1$, and are **slow** in the sense that $Ro \equiv U/(2\Omega L) \ll 1$, the third term (the Coriolis term) is dominant. (The dimensionless number $Ro$ is known as the **Rossby number** of the flow.) Now, the rotation rate $\Omega = 2\pi/(1\text{day}) = 7.2722\ 10^{-5}\text{s}^{-1}$. The first condition requires $T \gg 0.08\text{day}$, an excellent approximation for large scale motions in the atmosphere ($T \geq 1\text{day}$) and even better for large-scale motions in the ocean. As for the second condition, for a synoptic system in the atmosphere, $U \simeq 30\text{ms}^{-1}$, while $L \simeq 1000\text{km}$, so $Ro \simeq 0.2$; for an oceanic eddy, $U \simeq 0.1\text{ms}^{-1}$, $L \simeq 100\text{km}$, so $Ro \simeq 0.01$ (and for larger scale motions, it is smaller than this). So the assumption $Ro \ll 1$ is quite good for large scale motions in the atmosphere and excellent for the ocean.

If we assume the Coriolis term to dominate the LHS, therefore, and further assume that the motions are **inviscid** (so that $\mathbf{F} = 0$, an excellent approximation outside boundary layers), eq. (5.2) becomes

$$2\boldsymbol{\Omega}{\times}\rho\mathbf{u} = -\boldsymbol{\nabla}p - \rho\boldsymbol{\nabla}\Phi \; .$$

Taking the curl (since $\nabla \times \nabla a = 0$, for any $a$), we get

$$\nabla \times (2\boldsymbol{\Omega}{\times}\rho\mathbf{u}) = -\nabla \times (\rho\boldsymbol{\nabla}\Phi) \; .$$

Using the vector identity

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = \mathbf{A}\,(\nabla \cdot \mathbf{B}) - \mathbf{B}\,(\nabla \cdot \mathbf{A}) - (\mathbf{A}{\cdot}\nabla)\,\mathbf{B} + (\mathbf{B} \cdot \nabla)\,\mathbf{A} \; ,$$

and noting that $\boldsymbol{\Omega}$ is constant, we get

$$\nabla \times (2\boldsymbol{\Omega}{\times}\rho\mathbf{u}) = 2\boldsymbol{\Omega}\,(\nabla \cdot [\rho\mathbf{u}]) - 2\,(\boldsymbol{\Omega} \cdot \nabla)\,\rho\mathbf{u} \; .$$

But continuity of mass gives

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho\mathbf{u} = 0 \; .$$

So, for steady flow, $\nabla \cdot [\rho\mathbf{u}] = 0$. Thus, using the further vector identity

$$\nabla \times (c\mathbf{A}) = \nabla c \times \mathbf{A} + c\nabla \times \mathbf{A} ,$$

we obtain

$$2 (\mathbf{\Omega} \cdot \nabla) \rho\mathbf{u} = \nabla\rho \times \nabla\Phi .$$

If the flow is hydrostatic, then $\partial p/\partial z = -g\rho$, or, since $d\Phi = g\,dz$, $\partial p/\partial\Phi = -\rho$, whence $\nabla\Phi = -\rho^{-1}\nabla p$, and then

$$2 (\mathbf{\Omega} \cdot \nabla) \rho\mathbf{u} = -\frac{1}{\rho}\nabla\rho \times \nabla p .$$

Now, if the flow is also **barotropic**, by which we mean that density is a function of pressure only (i.e., no density variations along the almost-horizontal pressure surfaces), $\rho = \rho(p)$, whence $\nabla\rho = \frac{d\rho}{dp}\nabla p$, and so $\nabla\rho \times \nabla p = \frac{d\rho}{dp}(\nabla p \times \nabla p) = 0$. Thus we arrive at **the Taylor-Proudmann theorem**:

$$(\mathbf{\Omega} \cdot \nabla) \rho\mathbf{u} = 0 :$$

*For slow, steady, inviscid, barotropic motions in a rotating system, the momentum density vector ($\rho\mathbf{u}$) is constant along the direction parallel to the axis of rotation.*

Now, neither the atmosphere nor ocean are truly barotropic (they would be much less interesting if they were) but, nevertheless, many aspects of their dynamics can be captured in models that are barotropic (or nearly so), which is where we start.

## 5.3 Two-dimensional rotating flow

### 5.3.1 The barotropic equations of motion

We now investigate the properties of two-dimensional flow. If the atmosphere or ocean is assumed to be barotropic, the flow is independent of the direction along the rotation axis; given that they are also **thin**, in the sense that their depths are very much less than the Earth's radius, it is a very good approximation to assume that this implies that the flow is independent of $z$, the coordinate vertical to the Earth's surface (see Fig. 5.2). If we adopt
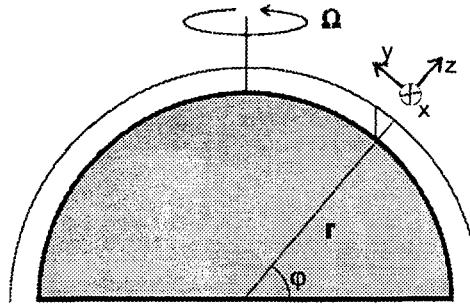
Figure 5.2: Coordinates for a shallow atmosphere or ocean.

local Cartesian coordinates[1] $(x, y, z)$ as shown on the Figure, the flow is in the $(x - y)$ plane, and $w = 0$. The $x$ and $y$ components of (5.2), are then

$$\frac{du}{dt} - fv = -\frac{1}{\rho}\frac{\partial p}{\partial x} + G_x$$

$$\frac{dv}{dt} + fu = -\frac{1}{\rho}\frac{\partial p}{\partial y} + G_y \qquad (5.3)$$

where $f = 2\Omega \sin\varphi$,with $\varphi$ being latitude, and $\mathbf{G} = (G_x, G_y) = \mathbf{F}/\rho$ is the applied (frictional) force expressed in units of acceleration. Note that the coefficient $f$ appearing in the Coriolis term is twice the vertical ($z$-) component of the rotation rate. This coefficient is known as the **Coriolis parameter**. Note that $f$ is a function of latitude, the importance of which we shall see later.

Eqs. (5.3) give us 2 equations in the 3 unknowns $u$, $v$, and $w$ ($\rho$ is assumed to be known as a function of $p$). We close the system with the equation of continuity $\nabla \cdot \mathbf{u} = 0$, or

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 . \qquad (5.4)$$

## 5.3.2 Vorticity and the barotropic vorticity equation

One of the difficulties of working with momentum (or velocity) of a parcel in fluid mechanics stems from the pressure forces to which the parcel is

---

[1]We are assuming here that the region of interest is a small part of the whole globe, otherwise it is necessary to use spherical coordinates (of course).

subjected, which are continuously changing the parcel's momentum in complicated ways (since pressure is not fixed, but itself evolves with the flow). However, while pressure gradients can change a parcel's *momentum*, they cannot change its *spin*, because, as we have seen, for barotropic flow for which[2] $\rho = \rho(p)$,

$$\nabla \times \left(\frac{1}{\rho}\nabla p\right) = 0 \ .$$

So, if we take the curl of the momentum equations, the pressure gradient term disappears. If we do this for eqs. (5.3), by taking $\partial/\partial x$ of the second minus $\partial/\partial y$ of the first, we get

$$\frac{\partial}{\partial x}\left(\frac{dv}{dt}\right) - \frac{\partial}{\partial y}\left(\frac{du}{dt}\right) + f\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + v\frac{df}{dy} = Z \ , \qquad (5.5)$$

where $Z = \partial G_y/\partial x - \partial G_x/\partial y$ is the (vertical component of the) curl of the frictional force per unit mass, and note that $f = f(y)$, since it is a function of latitude only. Now, from (5.4), the third term vanishes; moreover, a little mathematical juggling [expand the total derivatives, and use (5.4)] shows that

$$\frac{\partial}{\partial x}\left(\frac{dv}{dt}\right) - \frac{\partial}{\partial y}\left(\frac{du}{dt}\right) = \frac{d}{dt}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \ .$$

The term inside the bracket on the RHS is the vertical component of the **vorticity**, defined by

$$\boldsymbol{\xi} = \nabla \times \mathbf{u} \ ; \qquad (5.6)$$

its vertical component is

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \ . \qquad (5.7)$$

Since the flow in this barotropic problem lies within horizontal planes, only the vertical component is nontrivial[3].

The vorticity is a *local* measure of the spin of the fluid motion. For example if the fluid (relative to the rotating frame, remember) is in solid body rotation about the origin with angular frequency $\omega$, then (see Fig. 5.3)

$$u = -\omega y; v = \omega x \ ;$$

---

[2]Note that this includes an fluid of constant density.

[3]In large-scale meterology and oceanography, the general term "vorticity" is often used to mean the vertical component, unless specified otherwise.
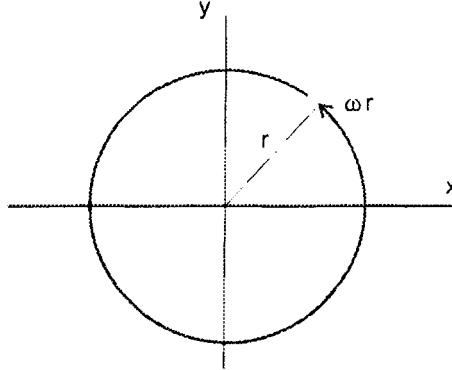
Figure 5.3: Rotation about the origin; the velocity at position $r = (x, y)$ is $U = \omega r$.

so the vorticity is $\zeta = 2\omega$—twice the rotation rate (anticlockwise being positive).

To return to (5.5), then, we have

$$\frac{d\zeta}{dt} = -v\frac{df}{dy} + Z \ . \tag{5.8}$$

This equation states that the time derivative *following the motion* of the vorticity is (in this barotropic case) given by two terms. The second represents the creation or destruction of vorticity by viscous torques (curl of the frictional force per unit mass), while the first represents advection of $f$, the Coriolis parameter. But we have already seen that $f = 2\Omega \sin\varphi$ is twice the vertical component of the planetary rotation rate; so looking down on the planet at latitude $\varphi$, an observer in an inertial frame would say the rotation rate of the fluid is not $\omega$, but $\Omega \sin\varphi + \omega$, and hence that the **absolute vorticity** of the flow—that observed from a nonrotating frame—is not $\zeta$ but

$$\zeta_a = f + \zeta \ . \tag{5.9}$$

Now, since $f$ is a function of $y$ only, its material derivative is

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + u\frac{\partial f}{\partial x} + v\frac{\partial f}{\partial y} = v\frac{\partial f}{\partial y}$$

and so the **barotropic vorticity equation** can be written

$$\frac{d\zeta_a}{dt} = Z \ . \tag{5.10}$$

Therefore:

> *In inviscid barotropic flow, the absolute vorticity is conserved following the motion.*

For most purposes away from boundary layers, the inviscid limit is a relevant one, so this theorem is profoundly useful for barotropic flows. (As we shall see, it needs modification for non-barotropic flows.) Put very simply, it says that if a fluid parcel is at position $x_o$ and has absolute vorticity $\zeta_{ao}$ at time $t_0$, and moves without viscous influence to position $x_1$ at time $t_1$, we know its absolute vorticity is still $\zeta_{a0}$—and we know this without needing to know anything about the path the parcel took in the intervening period. (So absolute vorticity is a **tracer**, and behaves just like, say, a dye marker.) Contrast this with velocity: to know how the velocity changed between $t_0$ and $t_1$, we would need to know its path, and the history of the pressure gradient along this path.

Now, one might object that absolute vorticity is not as interesting as velocity—that we may know what it is, but that that knowledge is not useful in telling us about what we want to know. However, if we know the distribution of $\zeta_a$ at any time, we know the distribution of $\zeta$ (since we know $f$) and, from that, we can determine the flow. To see this, first note from the continuity equation (5.4) that we can satisfy this by defining velocity in terms of a **stream function** $\psi$, such that $\mathbf{u} = -\hat{\mathbf{z}} \times \nabla\psi$, or

$$u = -\frac{\partial \psi}{\partial y} \; ; \; v = \frac{\partial \psi}{\partial x} \; , \tag{5.11}$$

which guarantees that $\nabla \cdot \mathbf{u} = 0$. Since $\mathbf{u}$ is normal to $\nabla\psi$, it is directed along contours of $\psi$, as shown in Fig. 5.4. Moreover, $\psi$ is a measure of the **flux** of fluid since the net amount of fluid passing per unit between the two streamlines $A$ and $B$, on which the streamfunction is (say) $\psi$ and $\psi + \delta\psi$ is $|\mathbf{u}|\,\delta l$, where $\delta l$ is the distance $AB$ between the streamlines. But, from (5.11), $|\mathbf{u}| = |\delta\psi|\,/\delta l$, so the flux (which in this 2-dimensional case has units of area per unit time) between the streamlines is just $|\delta\psi|$. Note that this flux is constant along the streamlines, so the velocity is large where the streamlines are close together, and weak where they are far apart—as is obvious from (5.11).
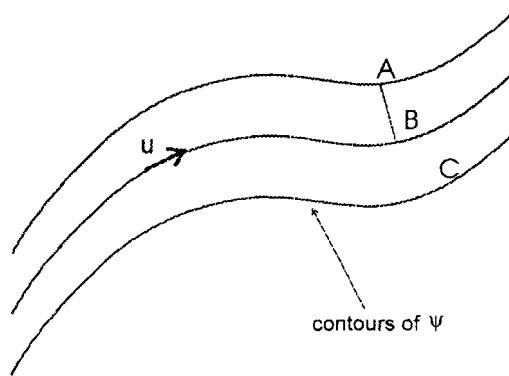
Figure 5.4: Flow is along streamlines (lines of constant $\psi$).

Now, in terms of streamfunction, it is obvious from (5.7) and (5.11) that the vorticity can be written
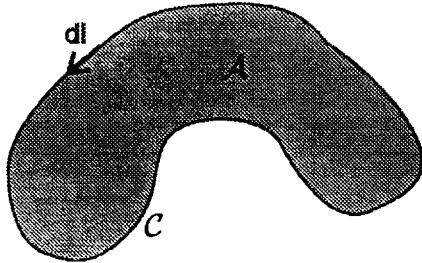
$$\zeta = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} \equiv \nabla_2^2 \psi \qquad (5.12)$$

where $\nabla_2^2$ is the two-dimensional (horizontal) Laplacian operator. So, if we know the vorticity distribution at any time [and note that (5.12) is a *diagnostic*, not a *predictive*, equation] we can calculate the stream function—and hence the velocities—from that knowledge. Note that since (5.12) is a second-order, elliptic, equation, we need appropriate boundary conditions to determine the solution. So all the information of dynamical importance is implicit in the vorticity distribution; hence the importance of (5.10). In principle, then, (5.10) can be used to predict how the absolute vorticity distribution changes, then (5.9) tells us the vorticity distribution; then, assuming we know the boundary conditions, (5.12) can be solved for the stream function, and hence the velocity.

Note the analogy between (5.12) and the equation for electric potential $V$ in the presence of a two-dimensional charge distribution $q(x, y)$; stream function $\psi$ is analogous to potential $V$, vorticity $\zeta$ to charge $q$.

A concept related to vorticity is **circulation**. The circulation $C$ around a closed contour $\mathcal{C}$ (see Fig. 5.5) is simply defined as

$$C = \oint_{\mathcal{C}} \mathbf{u} \cdot \mathbf{dl} \,, \qquad (5.13)$$

Figure 5.5: The contour $C$ in the definition of circulation.

where the integral is around the contour and **dl** the linear increment along $C$. But, from Stokes' theorem,

$$C = \oint_C \mathbf{u} \cdot \mathbf{dl} = \int_{\mathcal{A}} (\nabla \times \mathbf{u}) \cdot \hat{\mathbf{z}} \, dA = \int_{\mathcal{A}} \zeta \, dA \,, \tag{5.14}$$

where $dA$ is the area element and $\mathcal{A}$ the area enclosed by $C$. Thus, *the circulation around a closed contour is equal to the integrated vorticity enclosed by that contour.*

**Example—the flow around a point vortex.**

Suppose there is a **point vortex**, for which $\zeta(x,y) = Z_0 \delta(x - x_0) \delta(y - y_0)$ [so $\zeta = 0$ everywhere except at $(x_0, y_0)$]. Since we can anticipate the problem to have circular symmetry, we move into polar coordinates, with $(x_0, y_0)$ as the origin (see Fig, 5.6). In polar coordinates, the Laplacian is

$$\nabla_2^2 \equiv \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \vartheta^2}$$

so that, if we look for symmetric solutions for which $\psi = \psi(r)$ then, everywhere except $r = 0$, $\nabla_2^2 \psi = 0$ or

$$\frac{1}{r} \frac{d}{dr} \left( r \frac{d\psi}{dr} \right) = 0 \,.$$

The general solution is

$$\psi = A + B \ln r$$
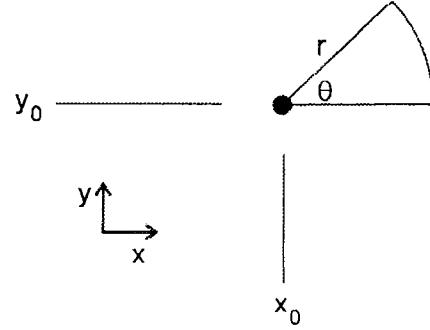
Figure 5.6: Geometry of point vortex example.

where $A$ and $B$ are constants. The constant $A$ is irrelevant (as only the gradients of $\psi$ have physical meaning); the velocity is

$$\mathbf{u} = -\widehat{\mathbf{z}} \times \nabla \psi = \left( \frac{\partial \psi}{\partial r}, -\frac{1}{r}\frac{\partial \psi}{\partial \vartheta} \right) = \left( \frac{B}{r}, 0 \right) .$$

To determine $B$, we note that this circulation around any contour enclosing the point vortex is

$$C = \int \zeta \, dA = \iint Z_0 \delta(x - x_0)\delta(y - y_0) \, dx \, dy = Z_0 .$$

But if we choose a circular contour at radius $r$, then

$$C = \oint_C \mathbf{u} \cdot \mathbf{dl} = \int_0^{2\pi} u(r) \, r \, d\vartheta = 2\pi r \, u,$$

where $u$ is the azimuthal velocity (see Fig 5.7), and so $B = Z_0/(2\pi)$. So the solution is

$$\psi(r) = \frac{Z_0}{2\pi} \ln r ;$$
$$\mathbf{u} = \left( \frac{Z_0}{2\pi r}, 0 \right) .$$

One important property of fluid flow—and rotating flow in particular—that this example makes clear is that the circulation is **nonlocal**: even a
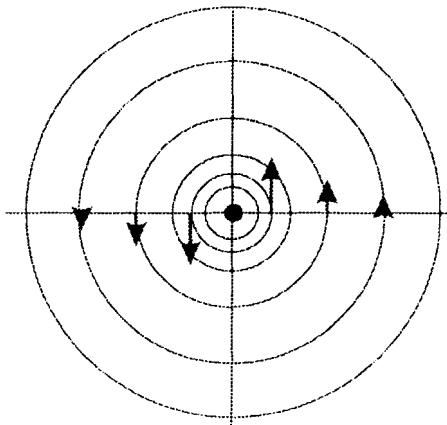
Figure 5.7:  Circulation around a cyclonic point vortex (northern hemi-sphere).

localized vorticity will induce a *remote* circulation, just as electrical charges induce a remote field. Amongst other things, this means that one cannot in general think about fluid dynamics in terms of local, fluid parcel arguments, since the flow at the location of the parcel depends on the behavior of all other parcels.

## 5.4   Further reading

This material is covered in several geophysical fluid dynamics texts.  The most suitable is Chapters 1 and 4 of:

"An Introduction to Dynamic Meteorology", J.R. Holton, Academic Press, 1979 (2nd edition).

# Chapter 6

# Rossby waves and planetary scale motions

## 6.1 Observed planetary scale waves in the atmosphere

Fig. 6.1 shows (solid contour; interval 4hPa) a typical northern hemisphere surface pressure map. It shows a rich structure, mostly of "synoptic scale" systems, especially small low-pressure storm systems. These have a range of sizes and intensities; there is a particularly large and vigorous storm over Iceland.

If we look, at the same time, at upper air charts, we see the influence of these storm systems weakening in the analysis. Fig 6.2 shows the height of the 500hPa pressure surface (solid contours; interval 60m) at the same time. The intense surface features are much less obvious here. Rather, the midlatitude jet is apparent[1] in the belt of tight height gradient around the hemisphere. However, there are strong wavy perturbations of the jet, usually of larger scale than the features that dominated the surface analysis (except over N America, where "synoptic" scale features are apparent at 500hPa also.) In terms of zonal wavenumber (the number of wavelengths around a latitude circle), the large-scale upper level disturbances have typical wavenumbers 1-4. These scales are referred to as *planetary*, and the wave motions on these scales as *planetary waves*. These waves migrate both east-

---

[1]Through geostrophic balance, the tight height gradient implies rapid flow along the height contours.

1

D052    ANALYSIS SFC/1000-500 THICKNESS 00Z WED 26 MAR 1997

Figure 6.1: Surface pressure analysis (solid contours), 26 March 1997.

Figure 6.2: 500hPa analysis (solid contours), 26 Mar 1997.

ward and (sometimes) westward; they also include a substantial *stationary* component. This latter fact is evident from Fig. 6.3, which shows the N

JANUARY NH 700 MB HEIGHT    JANUARY NH 500 MB HEIGHT    JANUARY NH 300 MB HEIGHT
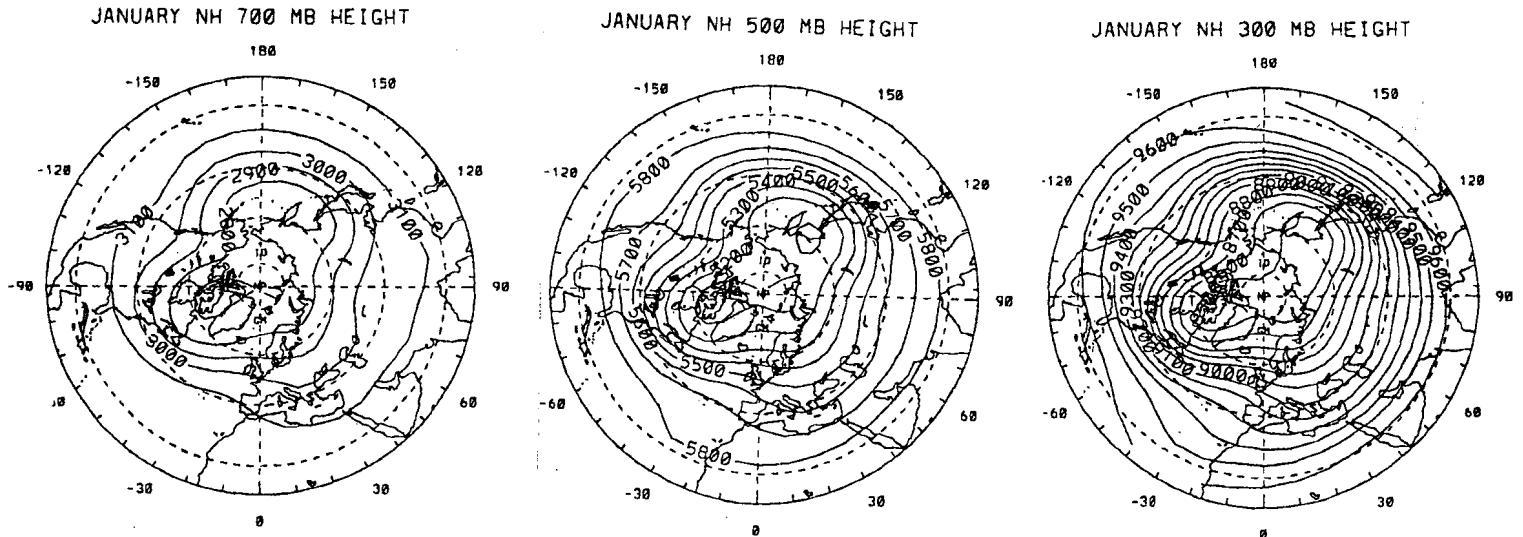
Figure 6.3: Long-term January mean heights, 300 - 700hPa.

hemisphere geopotential height in the lower (700hPa), middle (500hPa), and upper (300hPa) troposphere, averaged over 12 Januarys. The averaging has suppressed any signal from the mobile, synoptic scale storm systems, as well as from mobile planetary waves. What remains is the stationary component. As can be seen from the figure, this component is substantial. Note:

1. The time-averaged flow (along the height contours) departs significantly

from zonality;

2. in some regions, the mean flow departs greatly from being eastward, e.g. near the east coast of N America, where storm systems will tend to be steered by the mean wind to move up the coast, and to the west of N America and Europe, where a southwesterly fetch in the prevailing winds is an ameliorating influence on wither climate;

3. the wave phase is stationary, despite the mean almost-westerly flow: why are the waves not "blown away" by the wind?

4. these waves are vertically coherent, illustrating the Taylor-Proudmann effect, and giving us some hope that a barotropic analysis will be adequate to reveal the underlying dynamics.

## 6.2 Theory of Rossby waves
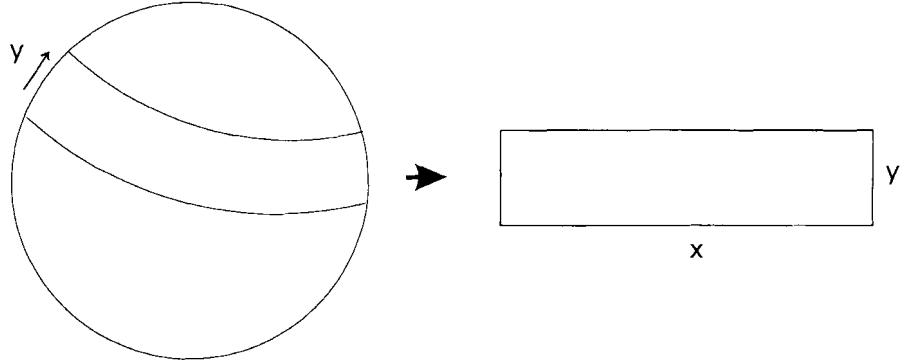
### 6.2.1 The $\beta$-plane

We saw in the derivation of the barotropic vorticity equation the potential importance of the fact that the Coriolis parameter varies with latitude, a consequence of spherical geometry. However, dealing with spherical geometry is (a little) more complicated than with planar geometry, so it is common to represent a strip of the sphere—limited in latitude but going all the way around the world in longitude—as a plane, as in Fig. 6.4. We consider a strip centered on longitude $\phi_0$, and define a $y$ coordinate $y = a\,(\phi - \phi_0)$, and an $x$ coordinate $x = a\lambda$, where $\lambda$ is longitude. Since $f = f\,(\phi) = 2\Omega\sin\phi$, in the $(x, y)$ system it becomes $f = f(y)$. Assuming that the width of the strip is small enough, we can approximate $f(\phi)$ as a Taylor series about the central latitude:

$$f(\phi) \simeq f(\phi_0) + (\phi - \phi_0)\left(\frac{df}{d\phi}\right)(\phi_0) + \dots$$

where

$$
\begin{aligned}
f(\phi_0) &= 2\Omega\sin\phi_0 \;; \\
df.d\phi_0 &= 2\Omega\cos\phi_0 \;.
\end{aligned}
$$

Figure 6.4: The $\beta$-plane.

Substituting for $y$, we get

$$f(y) = f_0 + \beta y \, , \tag{6.1}$$

where $f_0 = 2\Omega \sin \phi_0$ and

$$\beta = \frac{2\Omega}{a} \cos \phi_0 \, .$$

Note that, for a latitude of $\pi/4$, $\beta = 1.617 \times 10^{-11} \text{m}^{-1}\text{s}^{-1}$. Note that, though the sign of $f$ changes from N to S hemisphere, $\beta$ is always positive (since $f$ always increases northward).

## 6.2.2 Small amplitude barotropic waves on a motionless basic state

Neglecting viscous effects, the barotropic vorticity equation (??) becomes

$$\frac{d\zeta_a}{dt} = \frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta_a = 0 \, ;$$

absolute vorticity is conserved following the flow. Suppose now the motions of interest are small amplitude disturbances to a motionless basic state on a

$\beta$-plane, for which $\zeta_a = f$ (relative motion is zero, hence relative vorticity is zero) where $f$ is given by (6.1). Since there is no perturbation to $f$, we have

$$\begin{aligned} (u, v) &= (u', v') \,, \\ \zeta_a &= f + \zeta' \,, \end{aligned}$$

where the primes denote the perturbations. Neglecting terms quadratic in the primed quantities, we have (since $f$ is a function of $y$ only)

$$\frac{\partial \zeta'}{\partial t} + \mathbf{u}' \cdot \nabla f = \frac{\partial \zeta'}{\partial t} + \beta v' = 0 \,.$$

Since $\zeta = \nabla^2 \psi$, $\zeta' = \nabla^2 \psi'$, and, with $v' = \partial \psi'/\partial x$ (from the definition of streamfunction), we can easily get a single equation for $\psi'$:

$$\frac{\partial}{\partial t} \left( \frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2} \right) + \beta \frac{\partial \psi'}{\partial x} = 0 \,.$$

If we look for solutions of the form
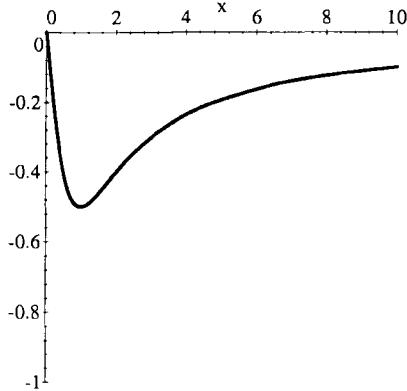
$$\psi' = Re \left[ \Psi \exp i \left( kx + ly - \omega t \right) \right] \,,$$

we get the dispersion relation for Rossby waves:

$$\omega = -\frac{\beta k}{k^2 + l^2} \,. \tag{6.2}$$

This function is plotted in Fig. 6.5. [Note that $\omega l/\beta = -x/(x^2 + 1)$, where $x = k/l$.] Note that:

1. $\omega/k = -\beta/\left(k^2 + l^2\right) < 0$: the phase speed is negative. So the phase of Rossby waves (on a motionless state) always propagates *westward*;

2. since $\omega$ is a nonlinear function of $\mathbf{k}$, Rossby waves are *dispersive*.

3. From Fig. 6.5 it is clear that $\partial \omega/\partial k > 0$ for $k/l > 1$, and $\partial \omega/\partial k < 0$ for $k/l < 1$: the group velocity of Rossby waves is eastward for zonally short waves, westward for zonally long waves.

4. The magnitude of the group velocity (judge by the slope of Fig. 6.5) is, typically, greater for the westward-propagating long waves than for the eastward-propagating short waves.

Figure 6.5: The function $-x/(x^2+1)$.

## 6.2.3 Typical values

At 45 degN,

$$\beta = \frac{2\Omega}{a}\cos 45\deg = \frac{2\pi\sqrt{2}}{86400 \times 6.37 \times 10^6} = 1.6145 \times 10^{-11}\text{m}^{-1}\text{s}^{-1}\ .$$

A typical midlatitude disturbance might have a half-wavelength of 5000km in both directions, so

$$k = l \simeq \frac{\pi}{5\times 10^6} = 6.28 \times 10^{-7}\text{m}^{-1}$$

and then

$$\omega = -\frac{1.6145 \times 10^{-11}}{2 \times 6.28 \times 10^{-7}} = -1.29 \times 10^{-5}\text{s}^{-1}\ .$$

The period is

$$\frac{2\pi}{|\omega|} = \frac{2\pi}{1.29 \times 10^{-5}} = 4.87 \times 10^5\text{s}$$

$$\simeq \frac{4.87 \times 10^5}{86400} = 5.6\text{d}\ .$$

The westward phase speed is

$$c = -\frac{\omega}{k} = \frac{1.29 \times 10^{-5}}{6.28 \times 10^{-7}} = 20.5\text{ms}^{-1}\ .$$

So the typical periods and phase speeds (relative to a stationary atmosphere) for these *planetary scale* Rossby waves are of order (days) and comparable with wind velocities, and so are meteorologically significant.

### 6.2.4 Mechanism of Rossby wave propagation

From (6.2), it is clear that the propagation of Rossby waves (indeed, the existence of the waves themselves) is dependent on the existence of the planetary vorticity gradient, $\beta$. In fact, had we allowed the basic state to have relative vorticity, it would have been the gradient of the mean absolute vorticity, rather than just $\beta$, that appeared in (6.2). How does a basic state vorticity gradient lead to waves? Consider Fig. 6.6. We assume that there
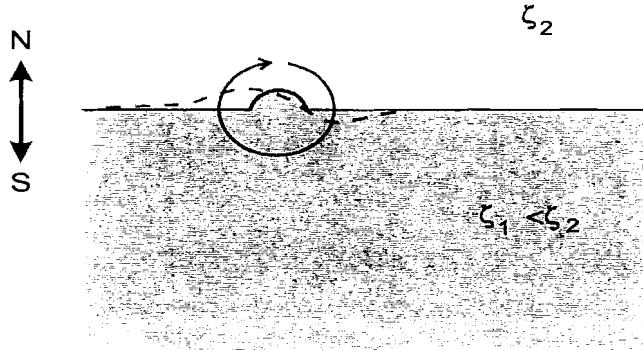


Figure 6.6:

are two regions of uniform vorticity, separated initially by a straight E-W boundary. North of the boundary, the absolute vorticity is $\zeta_2$; to the south, it is $\zeta_1$. Since the Coriolis parameter increases northward, we specify that $\zeta_1 < \zeta_2$. Now let's perturb the interface as shown on the figure, locally and northward. There is now a perturbation in the vorticity field, which is zero everywhere except in the bulge in the interface, where the vorticity perturbation is $\zeta_1 - \zeta_2 < 0$: the anomaly is negative (and therefore clockwise). Just as perturbing an array of electric charges would induce an anomalous electric field, this vorticity perturbation will induce an anomalous circulation. In fact, the streamfunction of the perturbed circulation is $\psi'$ where $\nabla^2\psi' = \zeta'$.

So the problem of determining the circulation is essentially the same as that for the circulation around a point vortex in Section **??**: hence the induced circulation will be clockwise, decaying as $1/r$ from the vorticity anomaly, much as depicted schematically in the figure.

Now, because absolute vorticity is conserved following the flow, it is simply advected by the circulation. The effect of the induced circulation on the vorticity distribution will be to advect the interface as shown: northward to the west, southward to the east. As the initial perturbation was northward, the perturbation itself tends to move toward the west—this is the westward phase propagation we noted from (6.2). The spreading, and changing of shape of the perturbation—manifested, amongst other things, by the developing southward perturbation to the east—is a manifestation of the dispersion we also noted.

## 6.3 Rossby waves in westerly flow

### 6.3.1 Dispersion relation: stationary waves and dispersion

The planetary scale waves observed in the atmosphere do not always show phase propagation westward, even though they are indeed Rossby waves. Some propagate to the east, some to the west, and as we saw earlier, there is substantial part of the planetary wave field that is stationary. The reason of course is that, unlike the simple preceding theory, the midlatitude atmosphere has mean westerly flow. In uniform flow, the preceding results for phase and group velocity should be interpreted as applying *relative to the background flow*, so the short waves (slow phase velocity relative to the flow) actually propagate to the east; only for sufficiently long waves is the westward Rossby wave propagation strong enough to overcome advection.

In a uniform background eastward flow $U$, the dispersion relation becomes[2]

$$\omega = Uk - \frac{\beta k}{k^2 + l^2} \ . \tag{6.3}$$

Just as for the "rock in the river" problem, it is possible to have *stationary*

---

[2]Relative to the moving flow, the phase velocity, from (6.2), is $\tilde{c} = -\beta/\left(k^2 + l^2\right)$; so relative to the ground, $c = \tilde{c} + U$, whence $\omega = ck = \left(\tilde{c} + U\right)k = Uk - \beta k/\left(k^2 + l^2\right)$.

*waves* for which the frequency is zero, provided $U > 0$. This happens when

$$k^2 + l^2 = \kappa_s^2 \,, \tag{6.4}$$

where $\kappa_s = \sqrt{\beta/U}$ is known as the *stationary wavenumber*. For typical midlatitude values $U = 30\text{ms}^{-1}$, $\beta = 1.5 \times 10^{-11}\text{m}^{-1}\text{s}^{-1}$, $\kappa_s^{-1} \simeq 1400\text{km}$, so such waves have typical wavelength $2\pi/\kappa_s \simeq 9000\text{km}$, which at $45°$ latitude corresponds approximately to zonal wavenumber 3. From (6.3), the zonal component of group velocity is

$$c_{gx} = U + \beta\frac{(k^2 - l^2)}{(k^2 + l^2)^2} \,;$$

given (6.4) and some manipulation, it follows that, for stationary waves with $k^2 + l^2 = \kappa_s^2 = \beta/U$,

$$c_{gx}(\omega = 0) = 2k^2\frac{U}{\beta} \;:$$

the zonal group velocity is eastward.

## 6.3.2 Forced stationary waves

We are now equipped to understand a simple representation of atmospheric stationary waves. The fact that these waves have a rather special value of phase velocity—zero—tells us that there is something special about forcing them: the forcing itself must be stationary. In fact, there are many ways such waves could be forced: by flow over very large-scale mountain ranges (the Himalaya, the Rockies, Antarctica, primarily), by geographically fixed regions of heating (which affect vorticity by ways we will discuss later), and by other, more subtle, means. The details of the waves produced by localized, stationary forcing depend on the nature of the forcing; however, in light of the above, there are some general things we can say, specifically:

1. The stationary wave will be located to the east of the forcing (since the group velocity has an eastward component), and

2. the length scale of the response will be determined by the inverse of stationary wavenumber.

These features are apparent in explicit solutions such as illustrated in Fig. 6.7.
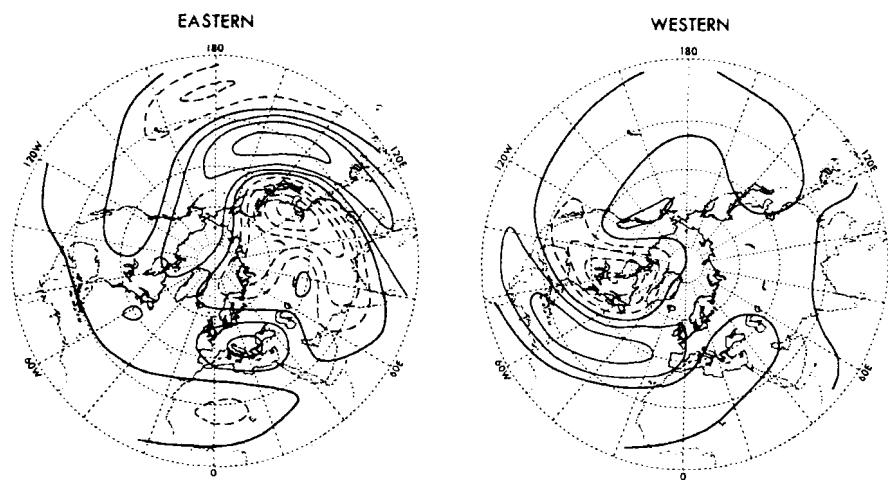
EASTERN WESTERN

Figure 6.7: Flow over a localized mountain. Numerical solutions for the perturbation streamfunction $\psi'$ for flow over (left) mountains in the eastern hemisphere (Tibet, mostly, with a small contribution from the Alps) and (right) the western hemisphere (mostly the Rockies). ote the Rossby waves propagating "downstream" (eastward) of the mountains.

## 6.3.3   Vertical structure

The theory developed thus far has been based on the assumption that the flow is barotropic. In reality, there are density variations in the atmosphere, which allow the existence of *baroclinic* (*i.e.*, non-barotropic) motions. The vertical structure of the Rossby wave train produced by a localized mountain is shown in Fig.  6.8.      In this figure, we can see two components of
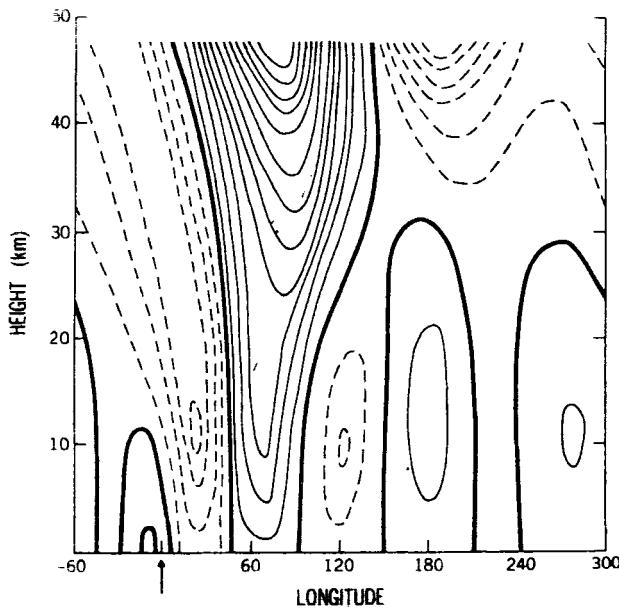


Figure 6.8: Perturbation streamfunction as a function of longitude and height for a 3D calculation of the response to flow over an isolated mountain (location marked by an arrow).

the response: a *surface wave*, which is trapped near the surface, just like an ocean surface wave is trapped at the ocean surface; and a *vertically-propagating component*. The former behaves very much like the barotropic waves we have been discussing. The latter is a wave that propagates in all three directions, including upward [cf. internal gravity waves, Chapter 3]. (In fact, the winds must be westerly aloft for vertical propagation, which restricts this behavior to the winter half-year.) Waves that propagate to great

heights reach large amplitude: because the atmospheric density decreases with height, an upward propagating wave becomes "focussed" into less and less mass the higher it goes, and thus must increase its velocity perturbations to compensate. These planetary Rossby waves dominate the meteorology of the winter stratosphere.

## 6.4 Rossby waves in the ocean

The ocean supports Rossby waves, just as the atmosphere does, obeying the same dispersion relation, and for the same reasons. (In fact, for barotropic motions, the theory does not discriminate between atmosphere and ocean.) The coastal boundaries of the ocean prevent a sustained east-west circulation (except in the Southern Ocean) and sustained east-west propagation of the waves themselves, so in practice there are many differences. The presence of coasts means that ocean basins can support trapped modes, for one thing. However, much of the large-scale variability of the ocean can be described as Rossby waves, albeit in a less organized way than for the atmosphere. However, there is one central aspect of ocean dynamics that may not appear to involve Rossby waves, but in fact does: the existence of western boundary currents.

### 6.4.1 Western intensification

It is evident from (6.2) and the ensuing discussion that Rossby wave behavior is zonally asymmetric. In particular, we saw that the group propagation of long Rossby waves is fast and westward, while that of short waves is slow and eastward. As illustrated in Fig. 6.9, this has dramatic consequences for ocean dynamics. Any large-scale disturbance in mid-ocean will generate Rossby waves; the larger scale of these will propagate rapidly westward. Before long, they will reach the western boundary of the ocean where they will be reflected. Unlike gravity (or light) waves, the reflected waves will not simply be a mirror image of the incident waves: the reflected waves must have an eastward component of group velocity and so must be of short zonal wavelength. Moreover, they will propagate relatively slowly, more so than the incoming waves. Thus, there will be a kind of "traffic jam" at the western boundary—information can get in more readily that it gets out. The information that accumulates there will involve motions of small zonal scale.
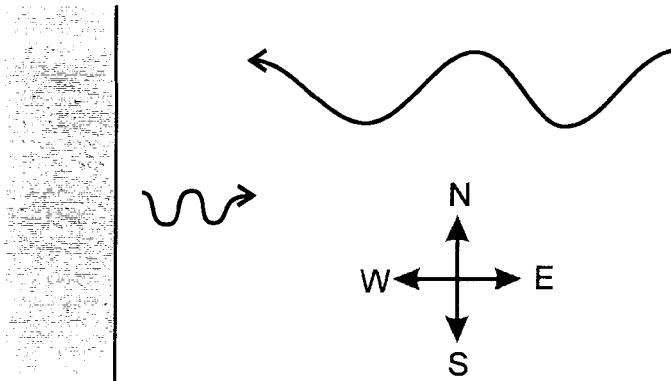
Figure 6.9:

This is the underlying dynamical reason for the existence of strong boundary currents on the western, rather than eastern, sides of the ocean. The underlying reason for the east-west asymmetry is $\beta$, the northward gradient of planetary vorticity.

# 6.5 Vorticity and potential vorticity in a fluid of varying depth

Now consider pseudo-barotropic motion in a fluid of varying depth. By "pseudo"-barotropic we mean that the horizontal flow is independent of the vertical coordinate (thus satisfying the Taylor-Proudmann theorem) but, because of depth variations, cannot be exactly nondivergent. So the system we will consider is an inviscid shallow water system, with a base that is not necessarily flat, as shown in Fig. 6.10. The system is assumed to be rotating with *uniform* Coriolis parameter $f$.

Our rotating shallow water equations are

$$
\begin{aligned}
\frac{du}{dt} - fv &= -g\frac{\partial h}{\partial x} \\
\frac{dv}{dt} + fu &= -g\frac{\partial h}{\partial y} \\
\frac{dH}{dt} &= -H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right).
\end{aligned}
\tag{6.5}
$$
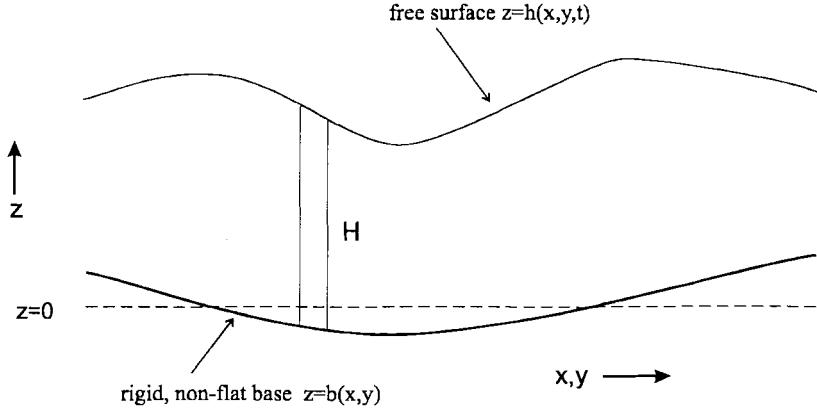
Figure 6.10: Shallow water model with varying depth.

Here, $H(x, y, t) = h - b$ is the *total* depth, where $h(x, y, t)$ is the height of the free surface and $b(x, y)$ the height of the bottom boundary. Note that the continuity equation involves $H$ rather than $h$, because the total mass convergence into the column is $\rho H \nabla \cdot u$, and the rate of change of column mass (following the flow) is $dH/dt$, rather than $dh/dt$.

Now, let's form our vorticity equation in the usual way, by taking $\partial/\partial x$ of the 2nd eq. $-\partial/\partial y$ of the 1st. As before,

$$
\begin{aligned}
\frac{\partial}{\partial x}\left(\frac{dv}{dt}\right) - \frac{\partial}{\partial y}\left(\frac{du}{dt}\right) &= \frac{\partial}{\partial x}\left(\frac{\partial v}{\partial t}\right) - \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial t}\right) \\
&\quad + \frac{\partial}{\partial x}\left(u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right) - \frac{\partial}{\partial y}\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right) \\
&= \frac{d}{dt}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) + \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial v}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\frac{\partial u}{\partial y} \\
&= \frac{d}{dt}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) + \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \\
&= \frac{d\zeta}{dt} + \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\zeta .
\end{aligned}
$$

Then we get

$$
\frac{d\zeta_a}{dt} + \zeta_a\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0 ,
$$

where $\zeta_a = f + \zeta$ is absolute vorticity, as before. So absolute vorticity is not conserved in this system: it can change whenever the divergence is nonzero (we'll see why). But using the 3rd equation of (6.5), the divergence is just

$$\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = -\frac{1}{H}\frac{dH}{dt} \ .$$

Substituting,

$$\frac{d\zeta_a}{dt} - \frac{\zeta_a}{H}\frac{dH}{dt} = H\frac{d}{dt}\left(\frac{\zeta_a}{H}\right) = 0 \ ,$$

and so

$$\frac{d}{dt}\left(\frac{\zeta_a}{H}\right) = 0 \ . \tag{6.6}$$

What (6.6) tells us is that, although absolute vorticity is not conserved, there is a quantity that is conserved following the flow: this quantity is

$$P = \frac{\zeta_a}{H}$$

and is known as the *potential vorticity*. What it means can be seen in the following. Suppose, as shown in Fig. 6.11, that a cylindrical column, initially with absolute vorticity $\zeta_a$ and length $H$, is stretched along its length. Mass continuity demands that the column must contract laterally as it is stretched; angular momentum conservation then dictates that the fluid must spin faster. Eq. (6.6) tells us that $\zeta_a$ increases in proportion to $H$: this process is known as *vortex stretching*.

## 6.6 Rossby waves in a fluid of varying depth

Consider now perturbations to a otherwise motionless fluid (so $\zeta = 0$ in the absence of perturbations) contained between sloping surfaces, as in Fig. 6.12. The column depth, $H(y)$ is a linear function of $y$, and we assume the perturbation velocities to be small, so that we can linearize. The potential vorticity equation (6.6) is

$$\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)\left(\frac{\zeta_a}{H}\right) = 0 \ ,$$
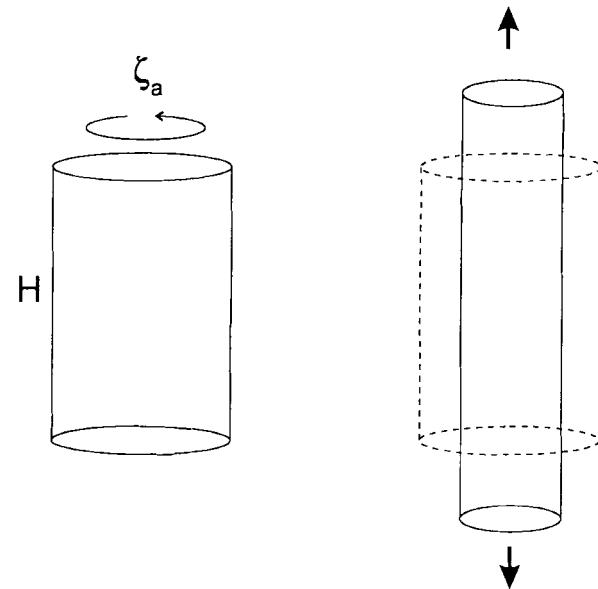
Figure 6.11: Illustrating vortex stretching.
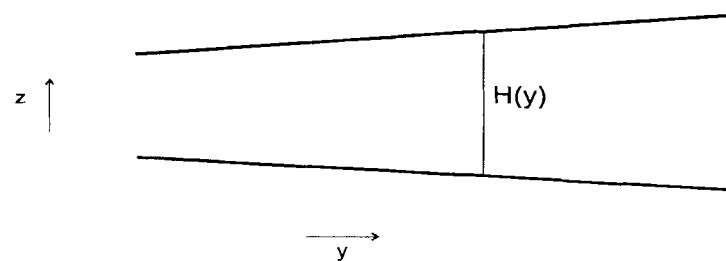


Figure 6.12:

whence

$$\frac{1}{H}\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)\zeta_a - \frac{\zeta_a}{H^2}\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)H = 0 .$$

Since $H = H(y)$ and $\zeta_a = f + \zeta'(x, y, t)$, this linearizes to give

$$\frac{\partial \zeta'}{\partial t} + v'\tilde{\beta} = 0 ,$$

where

$$\tilde{\beta} = -\frac{f}{H}\frac{dH}{dy} . \tag{6.7}$$

Thus, the vorticity equation becomes precisely equivalent to that in the exactly barotropic case on a $\beta$-plane, with in this case $\tilde{\beta}$—a measure of the gradient of fluid depth—replacing the gradient of $f$. Thus, *e.g.*, a sloping ocean bottom can give rise to Rossby waves, called "topographic Rossby waves", just as can the curvature of the Earth.

In fact, in the case of the Earth's curvature, the two effects are just another way of saying the same thing. Each is illustrated in Fig. 6.13. On the left, we take a traditional view of the atmosphere (or ocean), which is assumed to be contained within a spherical shell of depth $D$. The "vertical" is defined to be the local upward normal to the surface, and the component of planetary vorticity in this direction is $2\Omega\sin\phi = f$, the Coriolis parameter. Since the thickness of the fluid in the vertical direction is $D$, the potential vorticity is

$$P = \frac{f}{D} = \frac{2\Omega\sin\phi}{D} ,$$

and its gradient is

$$\frac{1}{a}\frac{dP}{d\phi} = \frac{2\Omega}{aD}\cos\phi = \frac{1}{D}\frac{df}{dy} = \frac{\beta}{D} .$$

In this view, the depth of the fluid column, $D$, never changes, so conservation of potential vorticity $P$ implies conservation of absolute vorticity $\zeta_a$. If a fluid column is moved northward to where $f$ is greater, $\zeta_a = f + \zeta$ is conserved by $\zeta$ decreasing as $f$ increases—so a northward displacement induces anticyclonic (negative) relative vorticity.

In the second view, we define the direction of the Earth's rotation vector to be the "vertical". The component of planetary vorticity in this direction
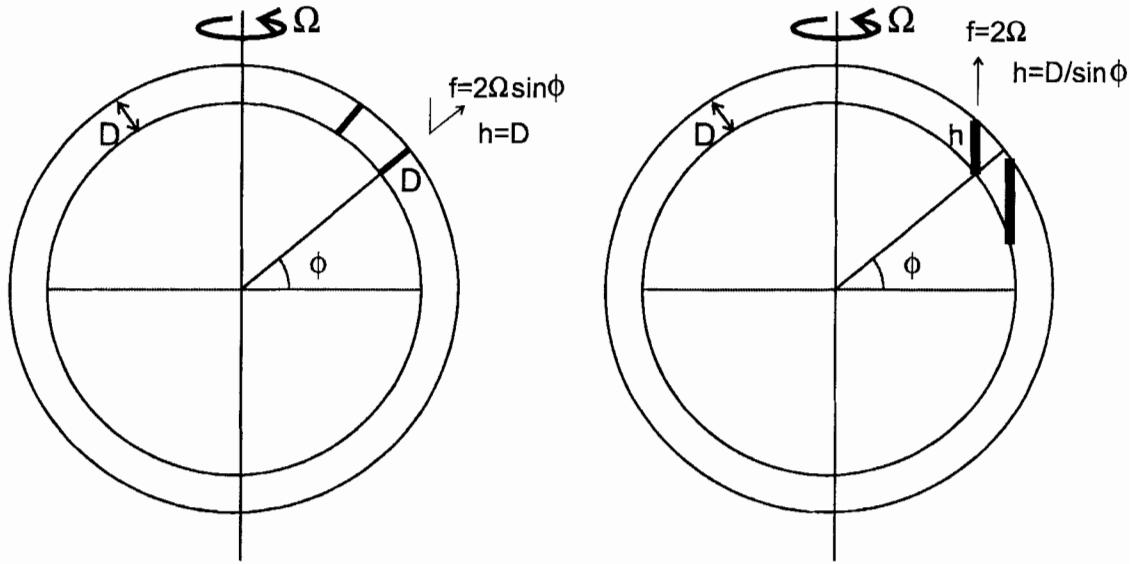
Figure 6.13: Illustrating the equivalence between the two forms of beta in spherical geometry.

is just $2\Omega$, which is of course constant. But the thickness of the atmospheric shell in this is not constant, but is $h = D/\sin\phi$. So the potential vorticity is

$$P = \frac{\zeta_a}{h} = \frac{2\Omega\sin\phi}{D},$$

just the same! And its gradient is

$$\frac{1}{a}\frac{dP}{d\phi} = \frac{2\Omega}{a}\frac{d}{d\phi}\left(\frac{1}{h}\right) = \frac{2\Omega}{a}\frac{d}{d\phi}\left(\frac{\sin\phi}{D}\right) = \frac{\beta}{D}.$$

So the PV gradient is (of course) exactly the same as in the first case, but we see it differently. In this viewpoint, the planetary vorticity is everywhere $2\Omega$, but as fluid columns move north or south, their length changes. A northward displacement produces a contraction of the column: in response (in order to conserve $P$) the absolute vorticity $2\Omega + \zeta$ must decrease, so $\zeta$ must become anticyclonic (negative).

# 6.7 GFD experiment: topographic Rossby waves in the lee of a ridge
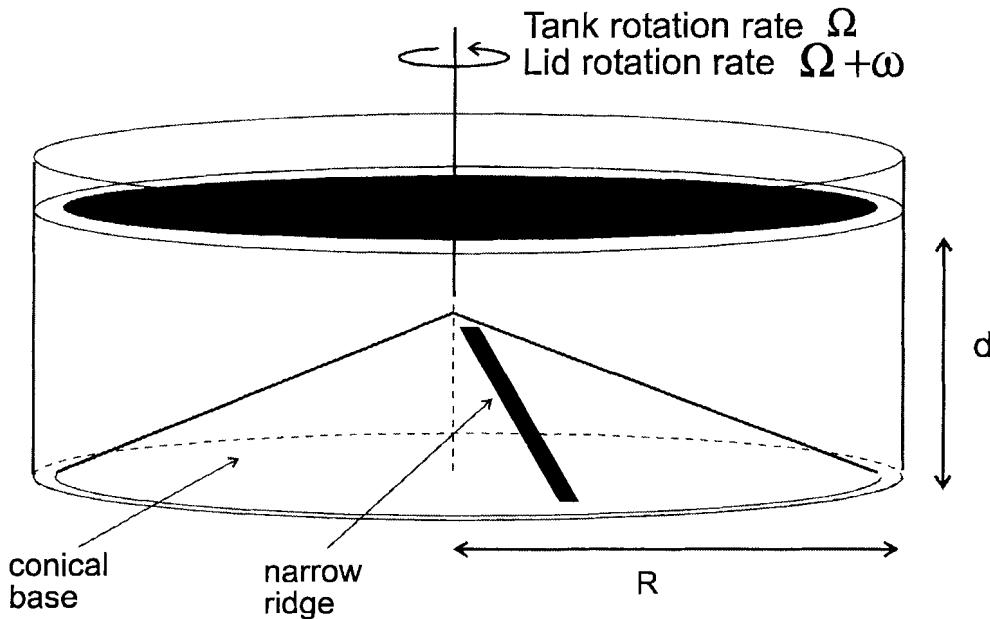


Figure 6.14: Schematic of the tank experiment.

Fig 6.14 show the set-up. A cylindrical tank, on a turntable rotating at rate $\Omega$, is fitted with a conical base; since the deepest water is at the outer rim, that corresponds to the equator. The effective $\beta$ in this setup is

$$\tilde{\beta} = \frac{2\Omega}{H}\frac{dH}{dr} = \frac{2\Omega}{R}\frac{\delta H}{H} \, .$$

A lid rotates cyclonically relative to the tank at rate $\omega$. This drives flow (of angular velocity $\sim \omega/2$) in the tank, over a small, straight ridge on the conical base. We expect this to produce a train of stationary Rossby waves of total wavenumber

$$\kappa \simeq \sqrt{\frac{\tilde{\beta}}{U}}$$

where $U = (R/2)(\omega/2)$ is the flow at radius $R/2$. So we expect the magnitude of the wavelength to be

$$\frac{2\pi}{\kappa} = 2\pi\sqrt{\frac{U}{\tilde{\beta}}} = \pi\sqrt{\frac{R\omega R}{2\Omega}\left(\frac{H}{\delta H}\right)} = \pi R \times \sqrt{\left(\frac{\omega}{2\Omega}\right)\bigg/\left(\frac{\delta H}{H}\right)}.$$

We will have $\omega \approx 0.1\Omega$, and $\delta H \approx H/2$, so we expect

$$\frac{2\pi}{\kappa} \approx \pi R \times 0.3 .$$

Since, at mid-channel, a wave of zonal wavenumber one has wavelength $\pi R$, this will give us something like zonal wavenumber 3. (We will update these numbers when we do the experiment.)

## 6.8 Further reading

Observational and theoretical aspects of Rossby waves are covered in several geophysical fluid dynamics texts, including

"An Introduction to Dynamic Meteorology", J.R. Holton, Academic Press, 1979 (2nd edition).

# Chapter 7

# Baroclinic instability and midlatitude storms

## 7.1  Three-dimensional geostrophic flow

### 7.1.1  In geometric coordinates $(x, y, z)$

In Cartesian, geometric coordinates, the equations of motion and of hydrostatic balance are

$$
\begin{aligned}
\frac{du}{dt} - fv &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + \mathcal{F}_x \,, \\
\frac{dv}{dt} + fu &= -\frac{1}{\rho}\frac{\partial p}{\partial y} + \mathcal{F}_y \,, \\
\frac{\partial p}{\partial z} &= -g\rho \,,
\end{aligned}
\tag{7.1}
$$

where $(\mathcal{F}_x, \mathcal{F}_y)$ are the $(x, y)$ components of friction. The continuity equation is

$$
\frac{d\rho}{dt} + \rho \nabla . \mathbf{u} = 0 \,.
\tag{7.2}
$$

For small Rossby number ($U/fL \ll 1$, where $U$ and $L$ are magnitudes for the flow and for spatial scales), the wind can be determined from *geostrophic balance*:

$$
\mathbf{u} = \frac{1}{f\rho}\widehat{\mathbf{z}} \times \nabla p \,,
\tag{7.3}
$$

or, in its components,

$$
\begin{aligned}
u &= -\frac{1}{f\rho}\frac{\partial p}{\partial y} \; ; \\
v &= \frac{1}{f\rho}\frac{\partial p}{\partial x} \; ; \\
w &= 0 \; .
\end{aligned}
\tag{7.4}
$$

Note that

$$
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \; :
$$

the geostrophic wind is nondivergent[1], which means the geostrophic flow is quasi-2D, and has some parallels with barotropic (precisely 2D) flow.

Since $\mathbf{u} \cdot \nabla p = 0$, the flow is *normal to the pressure gradient*, along the isobars. Thus, the isobars are *streamlines of the geostrophic flow*. In fact, from (7.4) we can define a *geostrophic streamfunction*, $\psi = p/(f\rho)$, which has the same properties as barotropic streamfunction.

## 7.1.2 In pressure coordinates (x,y,p)

In pressure coordinates (which are more useful for compressible atmospheres than height coordinates) the equations become

$$
\begin{aligned}
\frac{du}{dt} - fv &= -g\frac{\partial z}{\partial x} + \mathcal{F}_x \; , \\
\frac{dv}{dt} + fu &= -g\frac{\partial z}{\partial y} + \mathcal{F}_y \; , \\
\frac{\partial z}{\partial p} &= -\frac{1}{g\rho} \; ,
\end{aligned}
\tag{7.5}
$$

where the $x-$ and $y-$ derivatives should be understood as applying at constant pressure. One of the great simplifications of pressure coordinates is that the continuity equation is

$$
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial \omega}{\partial p} = \nabla_p \cdot \mathbf{u} = 0 \; ,
\tag{7.6}
$$

where $\omega = dp/dt$ is the pressure coordinate equivalent of vertical velocity.

---

[1] This is strictly true only if variations in $f$ are negligible, which means that the length scale of the motions must be much less than the Earth's radius.

Now geostrophic balance (7.3) becomes, in pressure coordinates,

$$\mathbf{u} = \frac{g}{f}\widehat{\mathbf{z}}_p \times \nabla_p z \,, \tag{7.7}$$

where $\widehat{\mathbf{z}}_p$ is the upward unit vector in pressure coordinates and $\nabla_p$ denotes the gradient operator in pressure coordinates. In component form,

$$(u, v) = \left(-\frac{g}{f}\frac{\partial z}{\partial y}, \frac{g}{f}\frac{\partial z}{\partial x}\right) \,.$$

Note that, like $p$ contours on surfaces of constant $z$, $z$ contours on constant $p$ are streamlines of the geostrophic flow.

### 7.1.3 Thermal wind balance

. Taking the $p$-derivative of the $x$-component of (7.7) gives

$$\frac{\partial u}{\partial p} = -\frac{g}{f}\frac{\partial^2 z}{\partial p \partial y} = -\frac{g}{f}\left(\frac{\partial}{\partial y}\left[\frac{\partial z}{\partial p}\right]\right)_p = \frac{1}{f}\frac{\partial}{\partial y}\left(\frac{1}{\rho}\right)_p \,.$$

Since $1/\rho = RT/p$, its derivative *at constant pressure* is

$$\frac{\partial}{\partial y}\left(\frac{1}{\rho}\right)_p = \frac{R}{p}\left(\frac{\partial T}{\partial y}\right)_p \,,$$

whence

$$\frac{\partial u}{\partial p} = \frac{R}{fp}\left(\frac{\partial T}{\partial y}\right)_p \,. \tag{7.8}$$

Similarly, for $v$ we find

$$\frac{\partial v}{\partial p} = -\frac{R}{fp}\left(\frac{\partial T}{\partial x}\right)_p \,. \tag{7.9}$$

Thus, horizontal gradients of temperature must be accompanied by vertical gradients of wind.

### 7.1.4 Thermodynamic equation

In eq. (3.14), we had

$$\frac{dT}{dt} - \frac{1}{\rho c_p}\frac{dp}{dt} = \frac{J}{\rho c_p} \,,$$

where $J$ is the **diabatic heating rate** per unit volume. Now, just as in geometric coordinates where the natural definition of vertical velocity is $w = dz/dt$, in pressure coordinates "vertical velocity" becomes[2] $\omega = dp/dt$. Then (3.14) can be written

$$\frac{dT}{dt} - \frac{1}{\rho c_p}\omega = \frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} - \omega S = \frac{J}{\rho c_p} , \qquad (7.10)$$

where $S = \partial T/\partial p - 1/\rho c_p < 0$ for a stable atmosphere, and $J$ is the diabatic heating rate per unit volume. Note that $J = 0$ for adiabatic motions.

Equivalently, defining potential temperature $\theta = T\,(p_0/p)^\kappa$, as in (3.15), (7.10) can be written

$$\frac{d\theta}{dt} = \left(\frac{p_0}{p}\right)^\kappa \frac{J}{\rho c_p} . \qquad (7.11)$$

For adiabatic motions $(J = 0)$, $\theta$ is conserved following the flow.

# 7.2 Structure of synoptic storm systems

The typical midlatitude synoptic storm system, such as those seen in Fig. 6.1, are mobile systems of both low and high pressure (though only the low pressure systems are usually associated with storms) that dominate the meteorology of the lower atmosphere, especially in winter. A typical northern hemisphere pattern may look like that shown in Fig. 7.1. Typical length

---

[2]Since $p$ varies most strongly in the vertical,

$$\begin{aligned}\frac{dp}{dt} &\equiv \frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + w\frac{\partial p}{\partial z}\\[4pt] &\simeq w\frac{\partial p}{\partial z} = -wg\rho ,\end{aligned}$$

assuming hydrostatic balance. So $\omega$ and $w$ are opposite in sign—*e.g.*, $\omega < 0$ is upward motion (toward lower pressure). Note also that

$$\begin{aligned}S &= \frac{\partial T}{\partial p} - \frac{1}{\rho c_p}\\[4pt] &= -g\rho\left(\frac{\partial T}{\partial z} + \frac{g}{c_p}\right) .\end{aligned}$$

scales are a few hundred km (with high pressure systems being typically larger than low pressure systems). From (7.4), it follows that the geostrophic flow, along the pressure contours, is *cyclonic* (in the same sense as the Earth's rotation) around the low and *anticyclonic* around the high pressure center.

Surface pressure is always within 10% of 1000hPa; in middle latitudes, typical storms may have pressure anomalies of 20hPa (cyclones) or 10hPa (anticyclones). If we regard each eddy as circular, then we may represent, crudely, the pressure structure (departure from 1000hPa) of a cyclonic eddy as $p' \simeq P_0 \exp(-r^2/2L^2)$, where $r$ is the distance from the center, and $L$ the radius at which $p'$ falls off by $1/\sqrt{e}$ from its central value. The azimuthal component of wind is, from (7.4),

$$
\begin{aligned}
u &= \frac{1}{f\rho}\frac{\partial p'}{\partial r} \\
&= -\frac{P_0}{f\rho}\frac{r}{L^2}\exp(-r^2/2L^2) \ .
\end{aligned}
$$

The maximum wind is at $r = L$, where

$$
|u|_{\text{max}} = \frac{P_0}{f\rho L}e^{-\frac{1}{2}} \ .
$$

Air has STP has density 1.293 kg m$^{-3}$; at 45° latitude, $f \simeq 1.0 \times 10^{-4}\text{s}^{-1}$. Hence, using $P_0 = 2000\text{Pa}$, and a size $L = 500\text{km}$, we find

$$
|u|_{\text{max}} \simeq 20\text{ms}^{-1} \ .
$$

This is a typical maximum wind in the lower *free* troposphere. Within the frictional boundary layer (where, of course, we live) surface friction slows the flow, and makes it *sub-geostrophic*, spiraling into low pressure cyclones and out of high pressure anticyclones, as in Fig. 7.2. The low-level inflow into cyclones produces, through "Ekman pumping" in the Ekman boundary layer, upwelling within the cyclones, which are therefore associated with clouds and rain; in anticyclones, descending air makes for clear skies. But note that most of the "weather" associated with cyclones is associated with thermal fronts embedded within the cyclone; we shall discuss these later.

# 7.3  Cyclogenesis and energetics

Where do these synoptic systems come from? They are mobile, and thus not attached to any surface features, so it seems unlikely that they are pro-
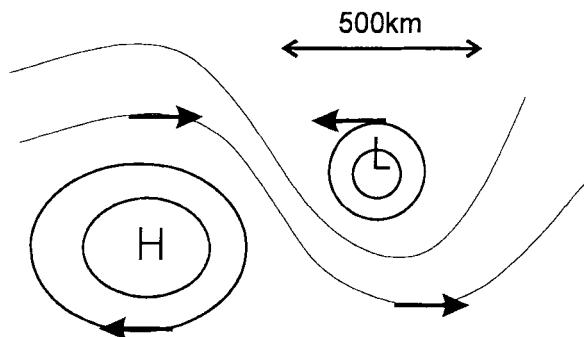
Figure 7.1: Schematic of the surface isobars around synoptic systems.
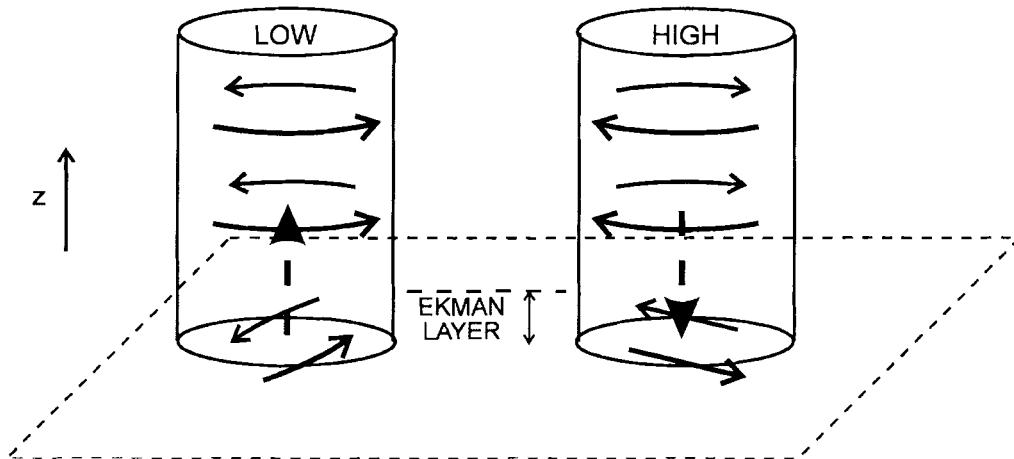


Figure 7.2: Schematic of Ekman inflow (in low pressure systems) and outflow (in high pressure systems).
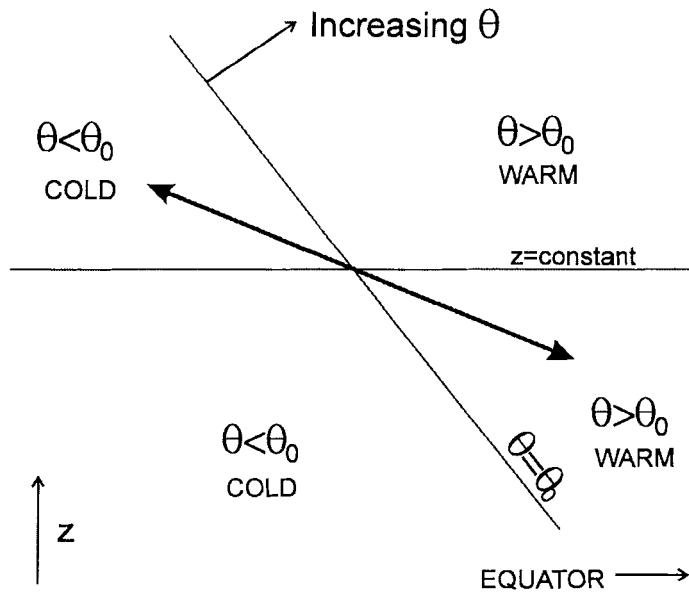
Figure 7.3: The "wedge of instability."

duced by external forcing. Rather, they are produced by a process known as *baroclinic instability*. The presence of horizontal temperature gradients in the atmosphere implies the existence of *available potential energy*, since the isentropes (surface of constant $\theta$) can be re-arranged to reduce the potential energy, as shown in Fig. 7.3. Exchanging air in the direction of the arrow will reduce the potential energy by moving warm (light) air upward and cold (heavy) air downward. The potential energy lost must appear as kinetic energy of the motion.

We can formalize this by considering the development of eddies on a basic state that is initially independent of $x$; for simplicity (in fact, to avoid some unnecessary complications), we assume that the basic flow is in fact exactly zonal, and that the flow is a function of $p$ only. Note, of course, from thermal wind balance (7.8), that the basic state temperature must then be a function of $y$ as well as $p$:

$$
\begin{aligned}
(u, v, w) &= (U_0(p), 0, 0) \; ; \\
T &= T_0(y, p) \; ; \\
\theta &= \theta_0(y, p) \; .
\end{aligned}
\tag{7.12}
$$

We now consider perturbations to this state. To keep things simple, we will assume the perturbations to be small so that we can linearize in the usual way. Of course, the real, fully developed weather systems we are interested in are not of small amplitude. If they grow through a linear instability, they grow from infinitesimal amplitude (when our linear assumptions are justifiable), reaching finite amplitude only as they mature. Detailed calculations (which are beyond the scope of what we are trying to do here) show that, for a typical midlatitude atmospheric state, waves with wavelengths of around 1000km will grow the fastest, with growth times (*e*-folding times) of typically 2-3 days.

The more limited question we are going to ask is where these systems get their energy from in the first place, and what characteristics they must have to allow them to extract energy from the basic state. Perturbing the state (7.12), therefore, and taking the inviscid eqns. of motion (7.5),

$$
\begin{aligned}
\frac{\partial u'}{\partial t} + U_0 \frac{\partial u'}{\partial x} + -fv' &= -g \frac{\partial z'}{\partial x} , \\
\frac{\partial v'}{\partial t} + U_0 \frac{\partial v'}{\partial x} + fu' &= -g \frac{\partial z'}{\partial y} , \\
\frac{\partial z'}{\partial p} &= -\frac{RT'}{gp} .
\end{aligned}
\tag{7.13}
$$

Now, the kinetic energy of the perturbation motions per unit volume is $\rho \left( u'^2 + v'^2 \right)$; therefore their globally integrated K.E. is

$$
\begin{aligned}
K &= \int \int \int_0^\infty \rho \left( u'^2 + v'^2 \right) \, dx \, dy \, dz \\
&= \frac{1}{g} \int \int \int_0^{p_0} \left( u'^2 + v'^2 \right) \, dx \, dy \, dp .
\end{aligned}
\tag{7.14}
$$

Now, we can form an equation for $dK/dt$ by taking $u' \times$ the first of (7.13)$+v' \times$ the second:

$$
\left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \frac{1}{2} \left( u'^2 + v'^2 \right) = -g \left( u' \frac{\partial z'}{\partial x} + v' \frac{\partial z'}{\partial y} \right) .
$$

Using the continuity equation, we can write

$$
\begin{aligned}
\left( u' \frac{\partial z'}{\partial x} + v' \frac{\partial z'}{\partial y} \right) &= \mathbf{u}' \cdot \nabla_p z' - \omega' \frac{\partial z'}{\partial p} \\
&= \nabla_p \cdot \left( \mathbf{u}' z' \right) - \omega' \frac{\partial z'}{\partial p} .
\end{aligned}
$$

Using hydrostatic balance, $\partial z'/\partial p = RT'/gp$, so

$$\frac{1}{g}\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)\frac{1}{2}\left(u'^2 + v'^2\right) = \nabla_p \cdot (\mathbf{u}'z') - \frac{R}{gp}\omega'T' \ .$$

Integrating over the whole system,

$$\frac{1}{g}\iiint \left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)\frac{1}{2}\left(u'^2 + v'^2\right) dx\ dy\ dp = \iiint \nabla_p \cdot (\mathbf{u}'z')\ dx\ dy\ dp$$
$$- \iiint \frac{R}{gp}\omega'T'\ dx\ dy\ dp\ .$$

But

1. $\iiint \frac{\partial}{\partial t}\left[\frac{1}{2}\left(u'^2 + v'^2\right)\right] dx\ dy\ dp = \frac{dK}{dt}$ ;

2. $\iiint \frac{\partial}{\partial x}\left[\frac{1}{2}\left(u'^2 + v'^2\right)\right] dx\ dy\ dp = \iint \left[\frac{1}{2}\left(u'^2 + v'^2\right)\right]_{x_1}^{x_2} dy\ dp$, where $x_1$, $x_2$, are the limits of integration. But, since the system is periodic ($360°$ is the same as $0°$), $[\_]_{x_1}^{x_2} = 0$. Hence $\iiint \frac{\partial}{\partial x}\left[\frac{1}{2}\left(u'^2 + v'^2\right)\right] dx\ dy\ dp = 0$.

3. $\iiint \nabla_p \cdot (\mathbf{u}'z')\ dx\ dy\ dp = \iint \mathbf{u}'z' \cdot \mathbf{n}\ dA$, where the integral is over the area bounding the system (the entire atmosphere) and $\mathbf{n}$ is the unit normal to the boundary. Since the only boundaries are the Earth's surface $p = p_s$ and the top of the atmosphere $p = 0$,

$$\iiint \nabla_p \cdot (\mathbf{u}'z')\ dx\ dy\ dp = \iint [\omega'z']_{p=p_s}\ dx\ dy - \iint [\omega'z']_{p=0}\ dx\ dy\ .$$

But, at the top of the atmosphere, $p = 0$ and $\omega' = 0$; at the surface, which is geometrically fixed $z' = 0$. So $\iiint \nabla_p \cdot (\mathbf{u}'z')\ dx\ dy\ dp = 0$.

Therefore

$$\frac{dK}{dt} = -\iiint \frac{R}{gp}\omega'T'\ dx\ dy\ dp\ . \tag{7.15}$$

Hence, the disturbance kinetic energy can grow only if, on average $\omega'T' < 0$ (*upward heat flux*). Since upward motion (toward lower pressure) means $\omega < 0$, this means that warm (light) air must rise and cold (dense) air sink in a developing storm.

That much may seem obvious, but there is more. Consider now the thermodynamic equation; assuming adiabatic motion[3], $d\theta/dt = 0$. The perturbed equation is therefore

$$\frac{\partial \theta'}{\partial t} + U_0 \frac{\partial \theta'}{\partial x} + v' \frac{\partial \theta_0}{\partial y} + \omega' \frac{\partial \theta_0}{\partial p} = 0 \ .$$

Multiplying by $\theta'$ and integrating over the atmosphere:

$$\iiint \left( \frac{\partial}{\partial t} + U_0 \frac{\partial}{\partial x} \right) \frac{1}{2} \theta'^2 \, dx \, dy \, dp = - \iiint \left[ v' \theta' \frac{\partial \theta_0}{\partial y} + \omega' \theta' \frac{\partial \theta_0}{\partial p} \right] \, dx \, dy \, dp \ .$$

Now, as above, it follows that

$$\iiint U_0(p) \frac{\partial}{\partial x} \left[ \frac{1}{2} \theta'^2 \right] \, dx \, dy \, dp = \iint \left\{ \int \frac{\partial}{\partial x} \left[ \frac{1}{2} \theta'^2 \right] \, dx \right\} U_0(p) \, dy \, dp$$

$$= \iint \left[ \frac{1}{2} \theta'^2 \right]_{x_1}^{x_2} U_0(p) \, dy \, dp$$

$$= 0 \ ,$$

so

$$\frac{d}{dt} \iiint \frac{1}{2} \theta'^2 \, dx \, dy \, dp = - \iiint \left[ v' \theta' \frac{\partial \theta_0}{\partial y} + \omega' \theta' \frac{\partial \theta_0}{\partial p} \right] \, dx \, dy \, dp \qquad (7.16)$$

Note that, if the disturbance is to grow from infinitesimal amplitude, the l.h.s must be positive (since $\theta'^2$ is positive definite).

Thus, from the two constraints (7.15) and (7.16), we have that, on average, $\omega' \theta' < 0$ and $\mathbf{u}' \theta' \cdot \nabla_p \theta_0 < 0$: so the vertical component must be upward BUT the vector component must be *downgradient*, toward lower basic state $\theta_0$. In order to achieve this—see Fig. 7.4—we need the vector $(v' \theta', \omega' \theta')$ to lie within the "wedge of instability" between the horizontal and mean isentropic surfaces, as previously depicted in Fig. 7.3. Fig. 7.4 has been drawn on the easily defensible assumptions that the atmosphere is statically stable against convection—so that $\theta_0$ increases upward—and that the basic state

---

[3]In fact, of course, rain—consequent on condensation and on release of latent heat—is a feature of intense storms, and thus the motions will not be adiabatic. However, this effect is not essential to the mechanism of storm development.
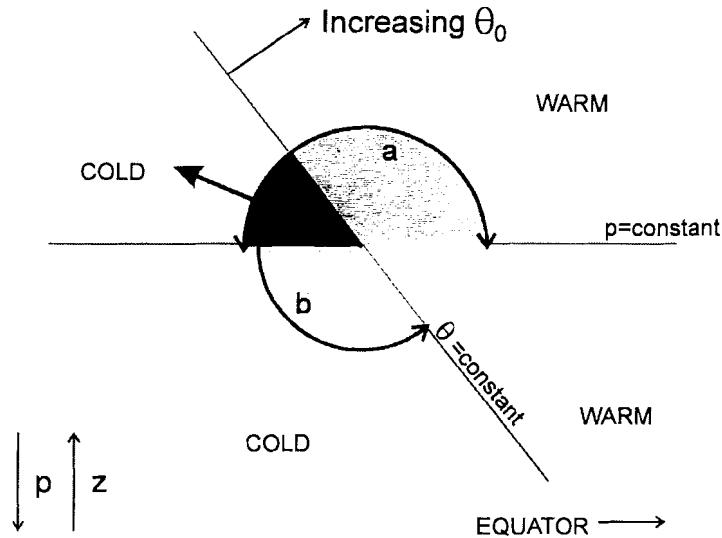
Figure 7.4: The "wedge of instability." The heat flux $(v'\theta', \omega'\theta')$ must be both upward (region $a$) and downgradient (region $b$) and thus must lie within the overlap of the two regions, as shown by the arrow.

temperature decreases toward the pole—so that $\theta_0$ does the same. Hence the gradient vector $\nabla\theta_0$ is as drawn.

As depicted in Fig. 7.4, the average of the *poleward heat flux*, $v'\theta'$, must be poleward. This follows from eqs. (7.15) and (7.16):

1. $\partial\theta_0/\partial p < 0$ ($\theta$ increases upward) in a stable atmosphere; and

2. $\theta' = T'(p_0/p)^\kappa$, whence $\omega'\theta' = \omega'T'(p_0/p)^\kappa$, and we saw above that the integral of the latter must be negative; so

3. the second term on the r.h.s. of (7.16) must be negative (allowing for the minus sign).

4. So the wave can grow only if the first term on the r.h.s. is sufficiently positive.

5. Since $\partial\theta_0/\partial y < 0$ (temperature decreases poleward), growing disturbances must have $v'\theta' > 0$.

Therefore, growing disturbances must *transport heat poleward* in order to extract energy from the basic state. Since the reason the basic state has available potential energy (APE) for the disturbance to extract is the presence of a horizontal temperature gradient, it is not surprising that the disturbances must transport heat down this gradient. By transporting heat poleward, the disturbances tend to warm the higher latitudes, thus reducing the APE. The lost APE appears, of course, as KE of the disturbances.

## 7.4 Vertical structure of growing disturbances

The simplest growing disturbances can be represented as being wavelike in the longitudinal direction, with height perturbation something like

$$z'(x,p) = \mathcal{Re}\ F(p)\ e^{i(kx-\omega t)}$$

(remember $\omega$ is now complex when the wave is growing). We have neglected any $y$-variation here; this is not particularly realistic, but it suffices to illustrate the point.

Suppose first that there is *no vertical phase tilt* of the disturbance, as depicted in Fig. 7.5. The 'L' and 'H' denote the locations of low and high
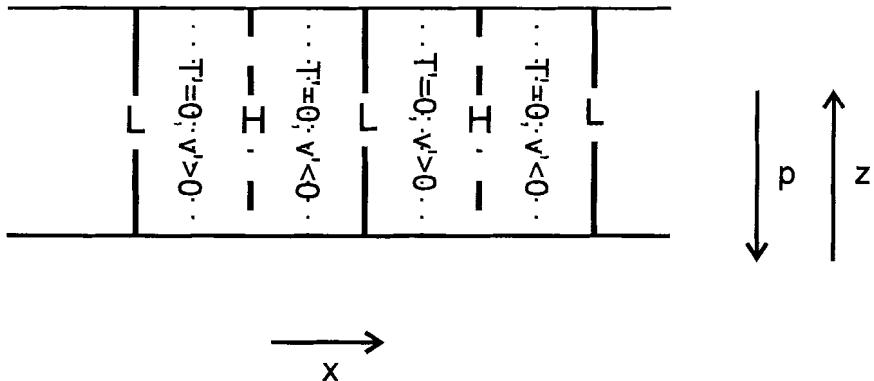


Figure 7.5:

height ($z'$) perturbations, respectively. The dotted lines show where $z' = 0$ at all heights—since we have specified no phase tilt. Then from the hydrostatic eq., (7.5), it follows that $T' = 0$ there (where $z' = 0$) also. But, geostrophic

balance, eq. (7.7), tells us that $v' = (g/f)\partial z'/\partial x$ is an *extremum* at this location. Therefore $v'$ and $T'$ are out of phase (in quadrature, in fact) so that, on average, $\overline{v'T'} = 0$. This clearly cannot be the structure of a growing disturbance.

Consider now a disturbance whose phase slopes westward with height; see Fig. 7.6. The maximum northward (/southward) flow is still to the
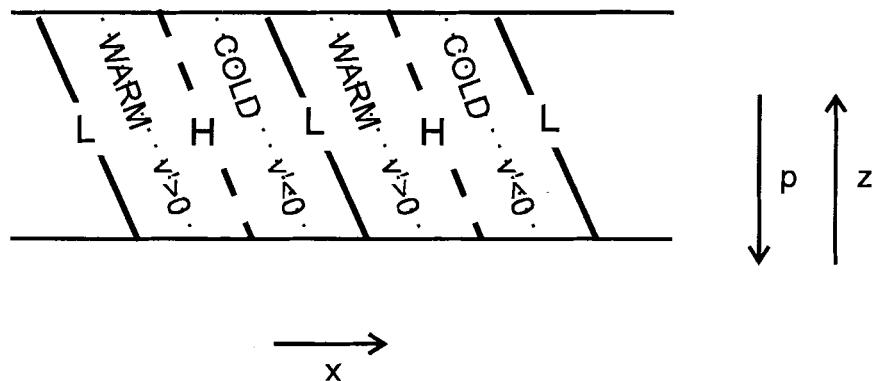


Figure 7.6:

east of the low (/high) height perturbation (northern hemisphere), but now there is a temperature anomaly there. Where $v' > 0$, $z'$ increases with $z$ (decreases with $p$), so the temperature perturbation is *warm* there; so warm air is moving north. Similarly, 180° further east, cold air is moving south. Hence $\overline{v'T'} > 0$ for this configuration, showing that this wave structure will lead to growth.

If the disturbance is tilted eastward with height, one finds $\overline{v'T'} < 0$, so there is no growth in this case.

An example of this is shown in Fig. 7.7. On 19 Dec 1964, a weak surface low lies to the east of an upper level (500hPa) trough (marked 'A' on the figure)—so the low tilts westward with height. This is favorable for growth, as illustrated by the explosive development that followed over the next 12

Image removed due to copyright considerations.

hrs.

## 7.5 Fronts

Even though baroclinic systems get their energy by reducing the overall temperature gradient, it is common experience that they often have *fronts*—bands of strong near-surface temperature gradient—embedded within them. These are usually marked on surface synoptic charts by thick lines barbed with symbols: triangles for cold fronts, semicircles for warm fronts, and both for occluded fronts (*q.v.*). The barbs are put on the side towards which the front is moving. We saw some examples in Fig. 7.7; more typical structures are evident 12 hrs later, shown in Fig. 7.8. Notice how the surface pressure

Image removed due to copyright considerations.

field is distorted by the presence of the fronts: there is usually strong curvature in the surface isobars (contours of constant pressure) in conjunction

with a pressure trough at the frontal position.

## 7.5.1 Frontogenesis

How do fronts form in the first place? We know that baroclinic systems must develop where there is a temperature gradient (and therefore available potential energy), but why do tight gradients form? the answer is that gradient-tightening is inevitable, in the presence of flow *deformation*. Since $w = 0$ at the surface, we can think about the effects of the flow on surface temperature purely in terms of horizontal advection (neglecting non-adiabatic effects, for now). In general, we can characterize the horizontal shear in flow by the four terms

$$
\begin{pmatrix}
\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\
\frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}
\end{pmatrix} .
$$

Alternatively, we can express these by the four independent linear combinations

$$
\begin{pmatrix}
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} & \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \\
\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} & \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}
\end{pmatrix} .
$$

We have already encountered two of these combinations: the divergence $\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$ (which is zero for geostrophic flow), and the vorticity $\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$. the other two terms are expressions of the deformation.

Consider what the flow does to a material box of surface air, of dimension $\delta x \times \delta y$. First, consider the evolution of the area of the box. The area evolves
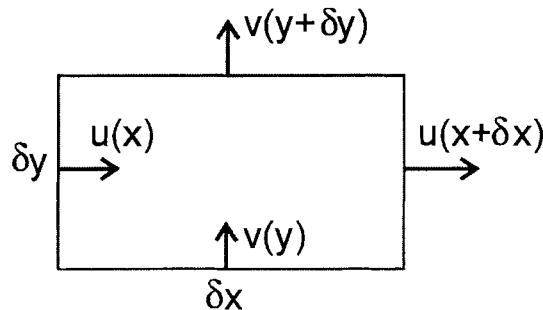


Figure 7.9:

as (see Fig. 7.9)

$$\frac{d}{dt}(\delta x\, \delta y) \;=\; \delta x\, (v(x,y+\delta y) - v(x,y)) + \delta y\, (u(x+\delta x,y) - u(x,y))$$

$$\;=\; \delta x\, \delta y\, \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) .$$

So the area is preserved in a nondivergent flow.

We know that the vorticity does—it rotates the fluid elements. This leaves the deformation. Consider the "pure deformation" flow described by the streamfunction $\psi = -Kxy$, where $K$ is a constant, shown in Fig. 7.10. Such a flow is nondivergent, and irrotational (zero vorticity). However, it
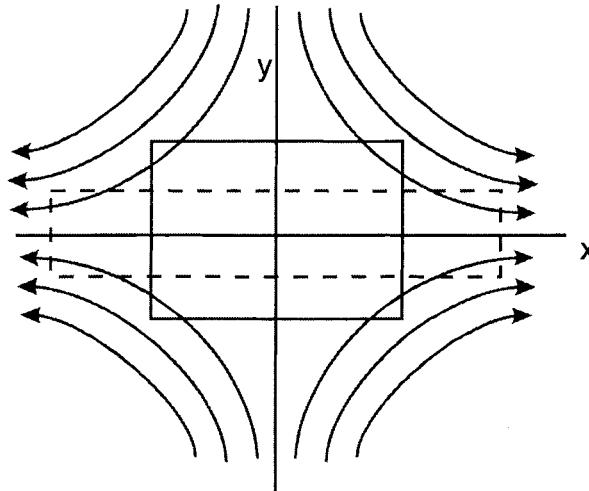


Figure 7.10:

has deformation, since

$$\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \;=\; \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = -2\frac{\partial^2 \psi}{\partial x \partial y} = 2K \;;$$

$$\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \;=\; \frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial y^2} = 0 \;.$$

Consider (Fig. 7.10) what happens to a material element such as the solid rectangle on the figure. With the given flow configuration, the box will be

stretched in the $x$-direction (the *axis of dilation*), as shown; since area is conserved, it must simultaneously contract in the $y$-direction (the *axis of contraction*). Thus, in this flow, any temperature gradient in the $y$-direction will be intensified by the flow, thus forming a temperature front.

Now consider a developing baroclinic wave. If the temperature gradient
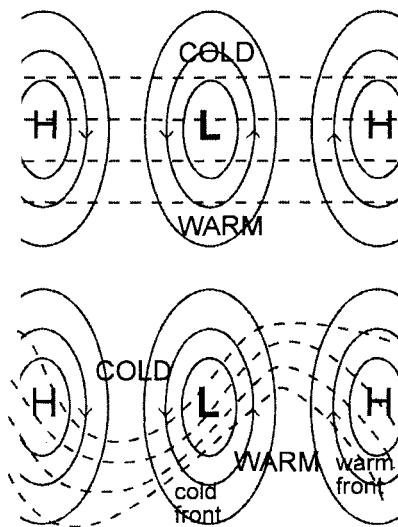


Figure 7.11:

is initially N-S (top of Fig. 7.11), it will be deformed (by the deformation between low and high) and twisted (by the cyclonic vorticity around the low), much as shown in the bottom of Fig. 7.11. Thus, it will form a warm front ahead of the low, and a trailing cold front behind.

## 7.5.2 Frontal evolution

The "textbook" picture of frontal evolution is as depicted in Fig. 7.12, The fronts form a "warm sector", usually to the south of the low pressure center (in the northern hemisphere). The warm sector moves around the storm a little, and contracts; sometimes the fronts merge near the storm center to form an "occlusion." A secondary low pressure center may from at the point of occlusion.

Image removed due to copyright considerations.

## 7.5.3 Frontal structure and weather

Fronts may be most intense near the ground, but they do extend vertically. As shown in Fig. 7.13, they slope with height, with the warm air overlying
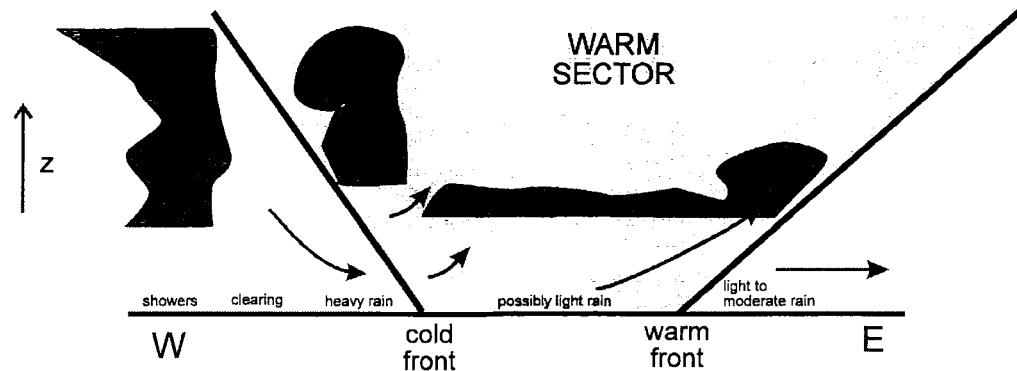


Figure 7.13: A schematic cross-section through the warm sector of a mid-latitude cyclone, running approximately west-east. Frontal motion is to the east; arrows denote air movement.

the denser cold air. At the leading warm front, the warm air rides up over the front. If the air is moist, this will produce clouds and, if the system is energetic enough, rain (or perhaps snow); note that the precipitation at the front, though it may be formed in the warm air, will fall through cold air—the precipitation is formed aloft, and so will be ahead of the surface front. Within the warm sector, there is weak upwelling, so this sector will often be completely cloud covered and there may be extensive rain or drizzle. At the cold front, the cold air undercuts the warm air, pushing the latter upward. At the front itself, there may be heavy rain or snow; some time later, there is a clearing in the subsiding air behind the front. Later still, convection may occur, which in some systems can be intense: cold air is moving over ground that is warm, following passage of the warm sector, and so the temperature structure is often convectively unstable. Immediately behind the front, convection may be suppressed by the subsidence. Once this abates, convection may set in. (Intense thunderstorms often follow the passage of summertime cold fronts.)

At the apex of the warm sector, the fronts may occlude. Often, this takes the form of the cold/warm front intersection leaving the ground, as

the warm sector gets squeezed aloft. Then air at the ground is everywhere cold, as in Fig. 7.14. The warm air slides up the occlusion, from the warm
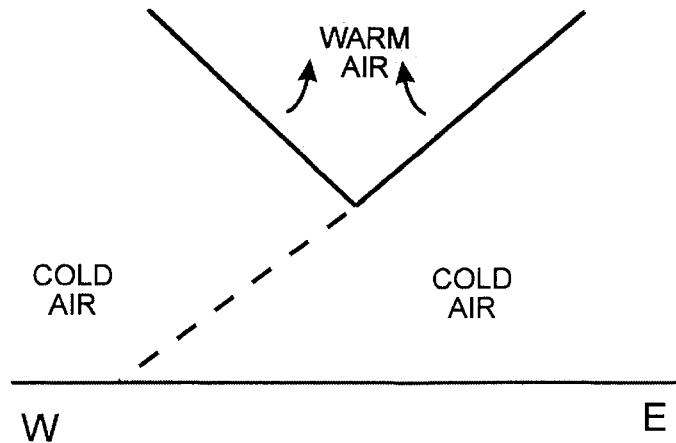


Figure 7.14: Cross-section (nominally W-E) through an occlusion.

sector. Precipitation is frequent in such situations; in winter it is frequently snow (in New England, this is the classic snowstorm), as the precipitation falls through a deep layer of cold air.

# 7.6 Climatology of synoptic systems: storm tracks

Synoptic-scale storm systems are not uniformly distributed over the globe; they concentrate in middle latitudes of the winter hemisphere and, even there, are more common in some longitudes than others. Fig. 7.15 shows the average distribution of eddy kinetic energy density $\frac{1}{2}\left(u'^2 + v'^2\right)$ and r.m.s. geopotential variance $\sqrt{z'^2}$ for northern winter. There are two main "storm tracks" where these quantities are large: across the N Pacific and N Atlantic Oceans.

The reason for this structure is the planetary scale structure of the background state. Storms tend to be generated in the regions of strongest *baroclinicity*—temperature gradient—off the E coasts of Asia and North America. Once formed, the storm motion is steered by the planetary scale flow which, as we saw in Fig. 6.3, is southwesterly (northeastward) across these oceans.

Image removed due to copyright considerations.

Thus, the synoptic scale systems are controlled by the planetary wave flow; in turn, the synoptic systems influence the planetary scale flow also. Eventually, the storms dissipate, though Pacific storms may propagate across the Atlantic and beyond before they die.

The situation in southern hemisphere winter is shown in Fig. 7.16; the main southern storm track extends across the southern Atlantic and Indian oceans.

## 7.7 Further Reading

Holton, J.R., "An Introduction to Dynamic Meteorology", Academic Press, 1979.

James, I.N., "Introduction to Circulating Atmospheres", Cambridge University Press, 1994.

Wallace, J.M., and P.V. Hobbs, "Atmospheric Science: An Introductory Survey", Academic Press, 1977.

Image removed due to copyright considerations.

# Chapter 8

# The equatorial atmosphere and ocean

## 8.1 Tropical meteorological maps

In the extratropics, we have got used to summarizing synoptic meteorological situations through the use of maps of pressure (at fixed height) or of height (at fixed pressure). Such maps are useful because the geostrophic wind relationship allows us to determine wind speed and direction from this information alone. However, this kind of map is much less useful in the tropics, for two reasons:

1. Pressure (or height) variations become weak in the tropics. From the geostrophic relationship

$$\mathbf{u} = \frac{1}{f\rho}\mathbf{k} \times \nabla p \,,$$

    the typical magnitude of wind speed is $U \sim \delta p / (f\rho L)$, where $\delta p$ is the magnitude of pressure variations over distance $L$. Near the equator, $f = 2\Omega \sin\varphi \to 0$; since (as a matter of observation, as well as physical common sense) $U$ does not become infinite, we must have $\delta p \to 0$. So *pressure variations are weak in the tropics*: so there is little to plot on a pressure map.

2. The validity of the geostrophic relationship requires that the Rossby number $R = U/fL$ be small. As $f \to 0$ in the tropics, this becomes

1

less valid. So, even if we did plot maps of the (weak) tropical pressure structure, we could not use these maps to deduce winds as we do in middle latitudes.

3. A corollary of the previous point is that tropical winds are not necessarily even approximately nondivergent.

The upshot of all this is that pressure (or height maps) are not as much use in the tropics. To see winds, it is more revealing to plot the winds directly, either (nowadays) as vector (arrow) wind plots, or as maps of:

1. **Streamlines**—lines that follow the direction of the flow. Note that these are not *contours*, of streamfunction or of anything else. Their spacing is of no significance and, unlike contour plots, these lines can converge into, or diverge out of, a point where the flow is divergent (or convergent).

2. **Isotachs**—contours of wind speed.

Such maps take some getting used to; we shall see some examples in what follows.

## 8.2    The Trade Wind circulation

The climatological mean upper and lower tropospheric winds in Jan and July are shown in Figs 8.1 and 8.2.    Note the presence, across the Pacific and Atlantic oceans just north of the equator, and across the Indian ocean south of the equator in January, of lines of *low-level convergence* in the flow. These regions are known collectively as the **intertropical convergence zone (ITCZ)**, which is associated with frequent and extensive deep convection. A schematic is shown in Fig. 8.3 of the low level flow over the Pacific east of the date line. (The Atlantic region is similar, also the Indian Ocean region in northern winter, except that the latitude is reversed).    A typical latitude-height cross-section is shown in Fig. 8.4.

The zone of deep cumulonimbus (Cb) and rainfall is located near the maximum sea surface temperature (SST) but is significantly narrower (we'll see the SST distribution in Fig. 8.8, below). The subsidence near the subtropical highs produces a low level inversion—the **trade inversion**, capping

Image removed due to copyright considerations.

## 8.1. THE TRADE WIND CIRCULATION

Image removed due to copyright considerations.
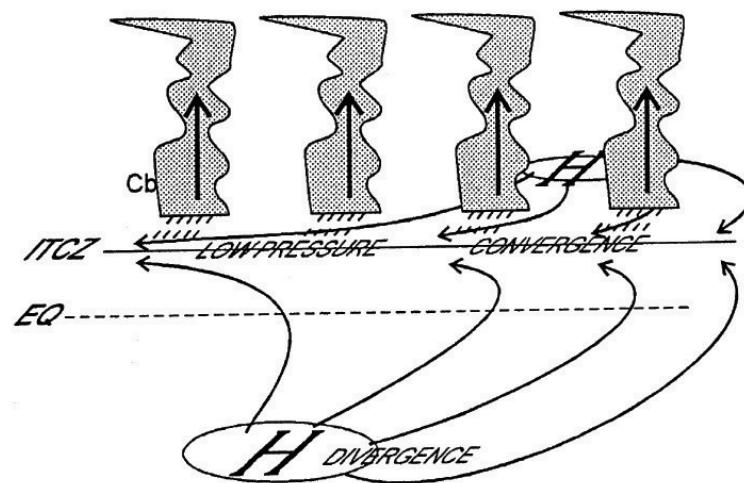
Figure 8.3: Schematic of the trade wind/ITCZ circulation
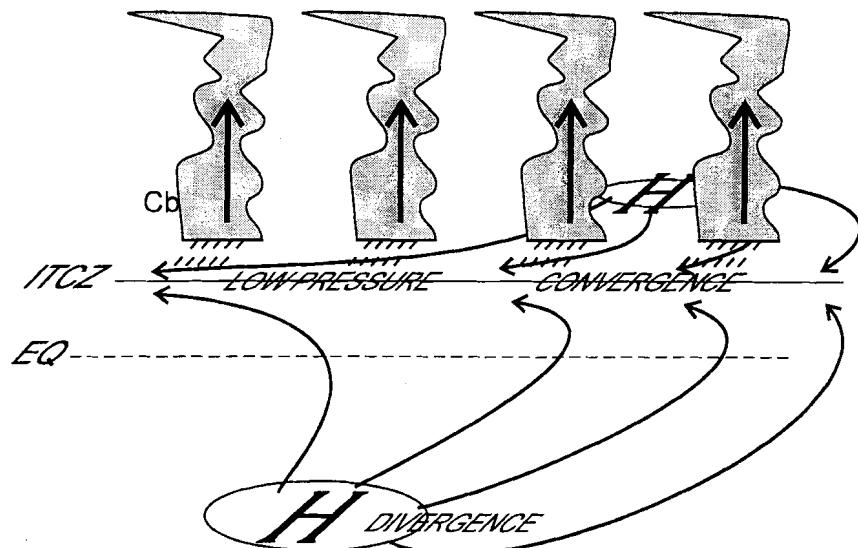
Figure 8.3: Schematic of the trade wind/ITCZ circulation
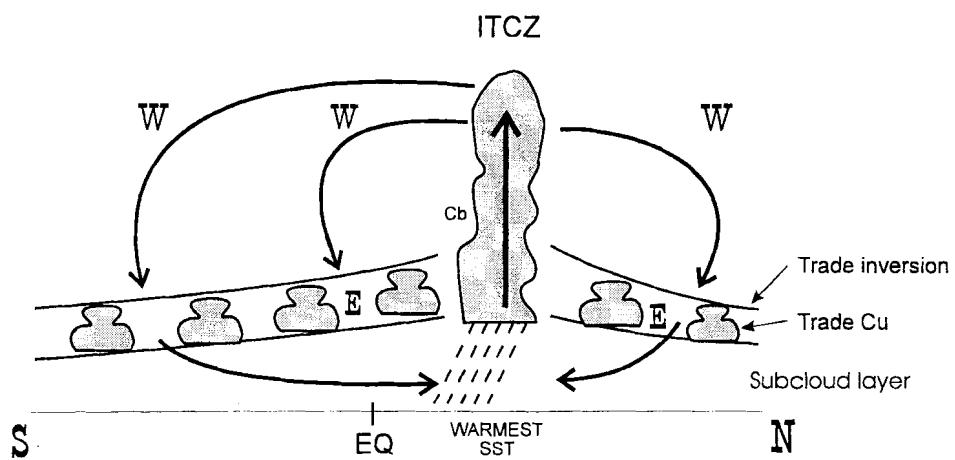


Figure 8.4: Schematic N-S cross-section of the ITCZ.

a layer of shallow "**trade cumulus**", over the ocean; **deserts** over land. Because these structures are long in the E-W direction, it seems sensible to try (at first) to understand the flow in terms of an axisymmetric (no E-W variation) theory (such as that for the tropical Hadley circulation, of which these structures form a part). However, the presence of upper-level westerlies over the equatorial regions in N winter and immediately south of the equator in N summer is inconsistent with this, as we will now see.

For inviscid flow under zonal symmetry, angular momentum, whose density is

$$m = \Omega a^2 \cos^2 \varphi + ua \cos \varphi \ ,$$

is conserved. Now, suppose all the ascending air within the ITCZ leaves the boundary layer at altitude $\varphi_0$, and that, because of boundary layer friction, $u \approx 0$ within the boundary layer. Then the angular momentum density of the air ascending within the ITCZ is

$$m_0 = \Omega a^2 \cos^2 \varphi_0 \ .$$

Once the air has left the boundary layer, friction becomes negligible and angular momentum is conserved, so that $m = m_0$ all along the streamlines of the flow until it subsides and enters the boudary layer again. In the upper troposphere, therefore, $m = m_0$, whence

$$\Omega a^2 \cos^2 \varphi + ua \cos \varphi = \Omega a^2 \cos^2 \varphi_0 \ ,$$

i.e.,

$$u = \Omega a \frac{[\cos^2 \varphi_0 - \cos^2 \varphi]}{\cos \varphi} \ .$$

Specifically, at the equator $\varphi = 0$,

$$u_{eq} = -\Omega a \sin^2 \varphi_0 \ .$$

Thus, if $\varphi_0 = 0$, $u_{eq} = 0$; otherwise, $u_{eq} < 0$, *i.e.*, easterly.

Thus, while an axisymmetric model does represent many features of the observed tropical circulation, it is clearly incomplete.

## 8.3 The "Walker Circulation" of the equatorial Pacific atmosphere

### 8.3.1 Observations; the atmosphere

To understand the failure of axisymmetric models, we need to consider the zonally asymmetric nature of the tropical circulation and the processes driving it. The observed situation is summarized in the following maps of rainfall (Fig. 8.5), and OLR [1] (Fig. 8.6). Note the following:

1. Latent heat release over the Pacific (illustrated in rainfall and/or OLR maps) is concentrated in the ITCZ, in the South Pacific convergence (SPC), over continental tropical America and, especially, in the far western equatorial Pacific around Indonesia and Melanesia.

2. There are upper tropospheric westerlies / lower tropospheric easterlies over the equatorial Pacific and the opposite pattern (in N winter) over Indian ocean.

3. The circulation in the equatorial ($x$-$z$) plane (Fig. 8.7) shows upwelling over the west Pacific and S America but downwelling over the east Pacific (and east Atlantic)—a suggestion of an overturning circulation in the longitude-height plane, driven by localized thermal driving over Indonesia? (*cf.*, circulation in the latitude-height plane—the Hadley circulation—driven by localized heating in the ITCZ.)

### 8.3.2 Observations; the ocean

Why is the west Pacific wet, the east Pacific dry? The distribution and seasonal variation of sea surface temperature (SST) is shown in Fig. 8.8). Note how the rainfall / OLR pattern mirrors (to a reasonable approximation) the location of the warm water. Note also the narrow strip of cold water in

---

[1] Outgoing Longwave Radiation (OLR) is measured from satellites; it is the total IR emission from the Earth. Assuming balckbody radiation, the OLR is a measure of the temperature of the emitting layer. For radiation from the surface or from low cloud, the emitting layer is warm, and OLR high. For radiation from high clouds—and only the tops of deep convective clouds are opaque enough—the emitting layer is cold. Thus, low OLR corresponds to regions of deep convection and thus of heavy rainfall. In fact, there is a good quantitative correspondence between rainfall and negative OLR anomalies.

Image removed due to copyright considerations.
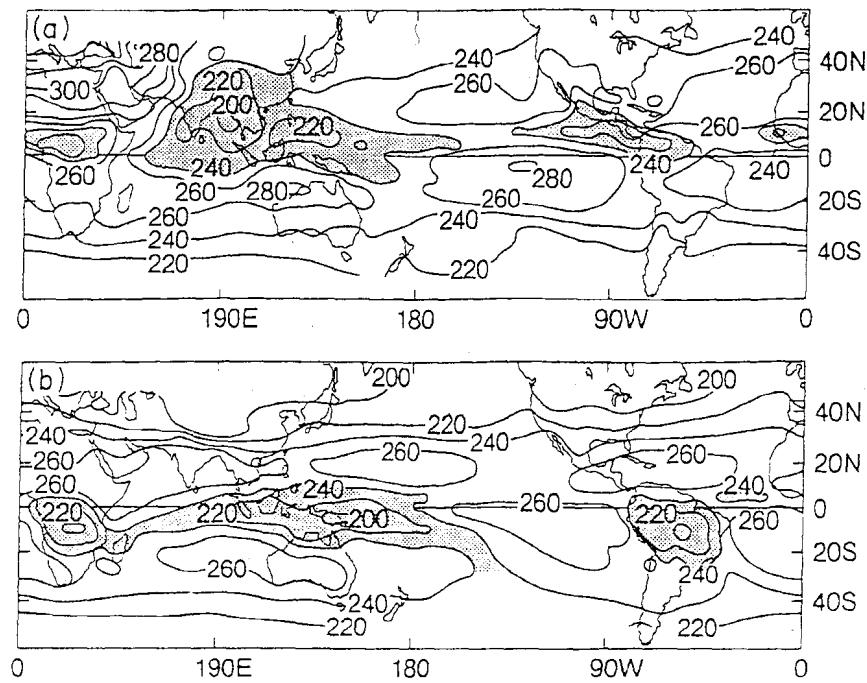
Figure 8.6: Outgoing longwave radiation (OLR) in January and July. (Contour interval: 50Wm$^{-2}$.)

Image removed due to copyright considerations.

Image removed due to copyright considerations.

the equatorial east Pacific, separating bands of warmer water to the north and south. Note the similar pattern of SST in the Atlantic.

Why does the SST look like this? The equatorial ocean is subjected to a primarily easterly wind stress. In response to this wind stress, the primary balance of the mixed layer to a westward wind stress is simply geostrophic:

$$-fv = \tau/\rho H$$

where $\tau$ is the wind stress and $H$ the depth of the mixed layer. Thus, there is (for $\tau < 0$) a northward Ekman flow north of the equator, and a southward flow to the south. This induces *upwelling* along the equator, which brings cold deep water to the surface. This is illustrated in Fig. 8.9.

Image removed due to copyright considerations.

But this is not the whole story; the westward wind stress also drives a westward surface flow (because the upwelling raises the thermocline near the equator, requiring a westward flow in geostrophic balance). The surfaces waters are thus advected westward, becoming warmer because of heat input from above and because the westward flow produces a convergence in the

western ocean which opposes the tendency for upwelling there. So, though the thermocline is elevated in the east and SST is cold there, it is depressed in the west where the SST is much greater. This is shown in the temperature cross-section of Fig. 8.10.
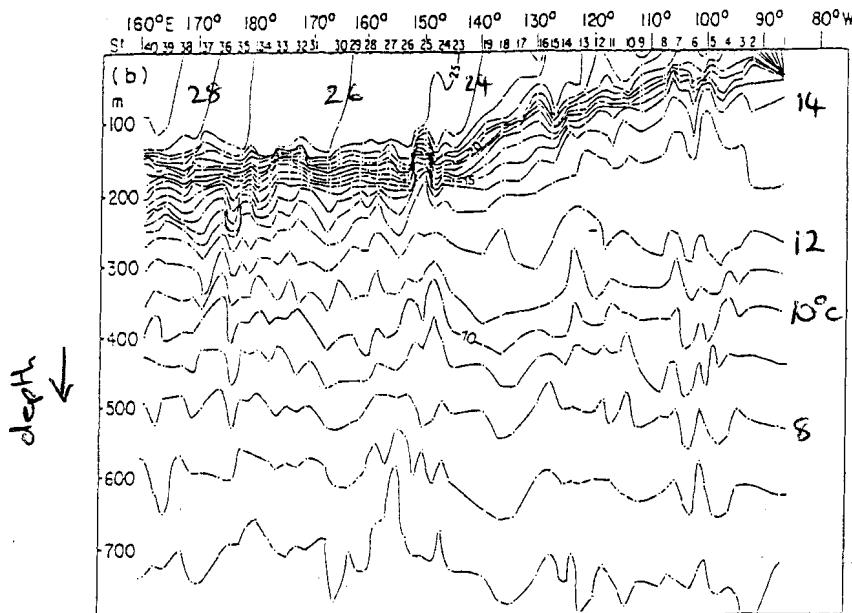


Figure 8.10: Longitude-depth cross-section of temperature across the equatorail Pacific ocean.

Note also that, because of the EW boundaries, there is a "piling up" of water in the west which produces an E-W pressure gradient; below the surface (away from the direct effects of the wind stress) this pressure gradient drives an *eastward* "jet", the **equatorial undercurrent**. This current (and other aspects of the circulation) is confined tightly to the equator—even though the wind stress distribution is broad—because of the governing equatorial dynamics.

## 8.3.3 Theory of the Walker circulation

In the simplest picture of the zonally symmetric Hadley circulation—such as we depicted in Fig. 8.4—we consider the circulation driven by a latitudi-

nally localized, but zonally symmetric, heating. As we have seen, the dominant heating—latent heating, implicit in the rainfall and OLR patterns—is far from zonally symmetric, having a strong maximum in the far western equatorial Pacific. This suggests an analogous, simple model for the Walker circulation—the circulation driven by a longitudinally localized heating on the equator.

Since the prevailing winds are weak near the equator, we neglect any background flow. If we assume the problem to be linear (which will be accurate for a sufficiently small amplitude disturbance) then the problem becomes one of wave propagation (see Fig. 8.11): the only way the circulation can extend
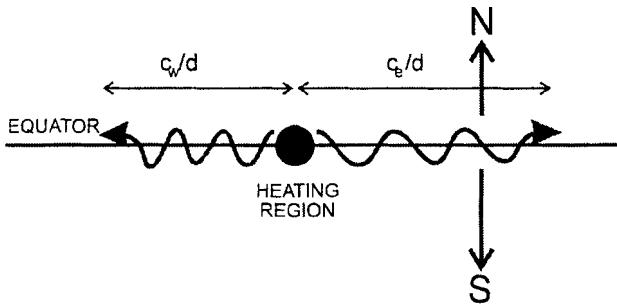


Figure 8.11:

beyond the localized heating region is through propagation of information. As the information propagates, it will be dissipated, at rate $d$, say. If the fastest wave propagation (group velocity) eastward is $c_e$, then we expect the information to propagate a distance $c_e/d$ before being dissipated; therefore the forced circulation should extend a distance $c_e/d$ east of the forcing. Similarly, the circulation will extend a distance $c_w/d$ to the west, where $c_w$ is the fastest westward wave speed. Rossby wave dynamics has taught us not to expect that $c_e = c_w$.

In fact, equatorial dynamics are rather special. Rossby waves still exits there, though they are a special type that are confined to an equatorial "wave guide"; like middle latitude Rossby waves, the long waves (which are the most important) have westward group velocity. In addition, however, there is a new (to us) and rather special class of waves that exist within this wave guide, known as **equatorial Kelvin waves**.

Consider a shallow water sysytem of mean depth $D$; with the addition of

rotation,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - fv = -g\frac{\partial h}{\partial x} .$$

Note that the second term vanishes under our assumption of linearity about a motionless basic state. Now, in our midlatitude analyses we have used the "beta-plane" approximation. We can do the same here, writing (for $\varphi \ll 1$)

$$f = 2\Omega\sin\varphi \simeq 2\Omega\varphi = \beta_0 y$$

where $y = a\varphi$ and

$$\beta_0 = \frac{2\Omega}{a} = 2.29 \times 10^{-11}\text{m}^{-1}\text{s}^{-1}$$

is the equatorial beta parameter. Making this substitution, and adding the $y$-momentum equation and the linearized continuity equation,

$$
\begin{aligned}
\frac{\partial u}{\partial t} - \beta_0 yv &= -g\frac{\partial h}{\partial x} ; \\
\frac{\partial v}{\partial t} + \beta_0 yu &= -g\frac{\partial h}{\partial y} ; \\
\frac{\partial h}{\partial t} + D\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) &= 0 .
\end{aligned}
\tag{8.1}
$$

Equations (8.1) actually describe many kinds of wave motion, including Rossby waves, but there is one special kind of solution that turns out to be very important. Let's look for solutions that have $v = 0$, i.e., flow that is exactly E-W and vertical. The second of (8.1) immediately gives

$$\beta_0 yu = -g\frac{\partial h}{\partial y} ;
\tag{8.2}$$

there is geostrophic balance in the NS direction, despite the fact that we are at the equator. The first of (8.1) gives

$$\frac{\partial u}{\partial t} = -g\frac{\partial h}{\partial x} ,
\tag{8.3}$$

which is precisely the equation for small amplitude nonrotating motions—the Coriolis term has dropped out of the $u$−equation because $v = 0$. Now, for solutions of the form

$$\begin{pmatrix} u(x,y,t) \\ h(x,y,t) \end{pmatrix} = Re \begin{pmatrix} U(y) \\ H(y) \end{pmatrix} e^{ik(x-ct)} ,$$

(8.3) gives

$$cU = gH \ ,$$

whence (8.2) gives

$$\frac{dH}{dy} = -\frac{\beta_0}{c}yH \ .$$

This has solutions

$$H(y) \sim \exp\left(-\frac{\beta_0}{2c}y^2\right)$$

which is a physically reasonable solution that satisfies boundedness as $y \to \pm\infty$ only if $c > 0$. So these motions are trapped near the equator, with a characteristic length scale $L = \sqrt{c/\beta_0}$, and propagate *eastward*. These are equatorial Kelvin waves. For the kinds of baroclinic motions we are interested in here[2], $c \sim 20\text{ms}^{-1}$ and $L \sim 1000$km (about 10° of latitude).

So, we have Rossby waves to transmit information westward and Kelvin waves to transmit it eastward. For a dissipation rate of $d = 1/(5\text{days})$, the
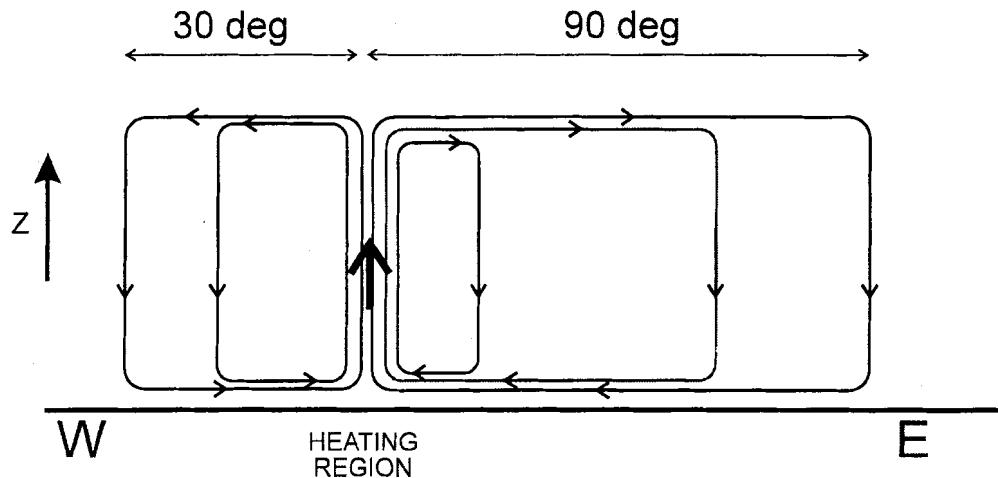


Figure 8.12:

Kelvin waves will reach a distance of $c/d \simeq 10000$km, about 90° of longitude. It turns out that the fastest Rossby wave group velocity is only one-third that of the Kelvin wave so, in the context of Fig. 8.11, the circulation extends only

---

[2]Kelvin waves are nondispersive, so this value is the same for phase and group velocities.

one-third as far to the west as it does to the east. So we expect a circulation something like that shown in Fig. 8.12.

The horizontal structure, from a complete linear solution, is shown in the Fig. 8.13  ($u, v$ and $p$ shown are for the lower level reponse; upper levels are

Image removed due to copyright considerations.

opposite in sign). Note:

- the different structures and different zonal length scales (factor of 3) of the Kelvin and Rossby components to east and west.

- The twin low level cyclones / upper level anticyclones straddling the equator just west of forcing.

## 8.4   Monsoons

### 8.4.1   Seasonal variations over the tropics

We saw the characterisitcs of the tropical circulation in Figs. 8.1 and 8.2. the correponding cloud distributions are shown in Fig. 8.14.   As we noted

Image removed due to copyright considerations.

earlier, over the oceans there is a near-equatorial trough of low pressure—the Intertropical Convergence Zone (ICTZ)—that shows very clearly in the cloudiness (and rainfall). Over the contintents, however, the low pressure and cloudiness (and rainfall) is displaced well into the summer hemisphere. In places where there is ocean equatorward of the subtropical land, these sumertime regions of cloudiness are associated with very heavy rain and characteristic wind patterns—this is the monsoon circulation.

Image removed due to copyright considerations.

2. In northern summer, there is strong northward flow across the equator in the Indian ocean (especially in the west), eastern Atlantic, and eastern Pacific. In southern summer, southward cross-equatorial flow across the Indian Ocean to the date line.

3. In northern summer, low-level westerlies in the northern tropics over the Indian Ocean, west Africa, and (weakly) eastern Pacific. In southern summer, low-level westerlies in the southern tropics across the eastern Indian Ocean and north of Australia.

4. In northern summer, an intense upper-level anticyclone across south Asia at about 30 degN, with an intense easterly jet on its southern flank, extending from south-east Asia to the western Atlantic. There are much weaker upper-level anticyclones over northern Mexico and the eastern north tropical Pacific, though without much of an associated easterly jet. In southern summer, upper level anticyclones over northern Australia/Melanesia, with an easterly flow equatorward. There are other anticyclones over south tropical Africa and South America, though without much, if anything, to indicate an easterly jet (nor low-level westerlies).

Characteristics of the JJA circulation in the Indian Ocean region are summarized in Fig. 8.16. The cross-equatorial flow at low levels is concentrated in the "Findlater jet" on the eastern flank of the East African mountains[3]. Note also the strong upper-level anticyclone over south Asia and easterly jet equatorward of this.

## 8.5 Monsoon depressions and breaks

Within any monsoon season, rainfall is very variable: on synoptic time scales (a few days), as "monsoon depressions" propagate across th emonsoon region; and on time scales of 1-3 weeks, the so-called "active/break" cycle.

In practice, most rain falls in association with monsoon depressions. In the Indian monsoon, these usually propagate westward across the subcontinent, producing daily variability of rainfall through the monsoon season. They form over the Bay of Bengal, and move northwestward. In the vertical,

---

[3]This jet is actually below the level of the highlands, and has some similarites with western boundary currents in the ocean.
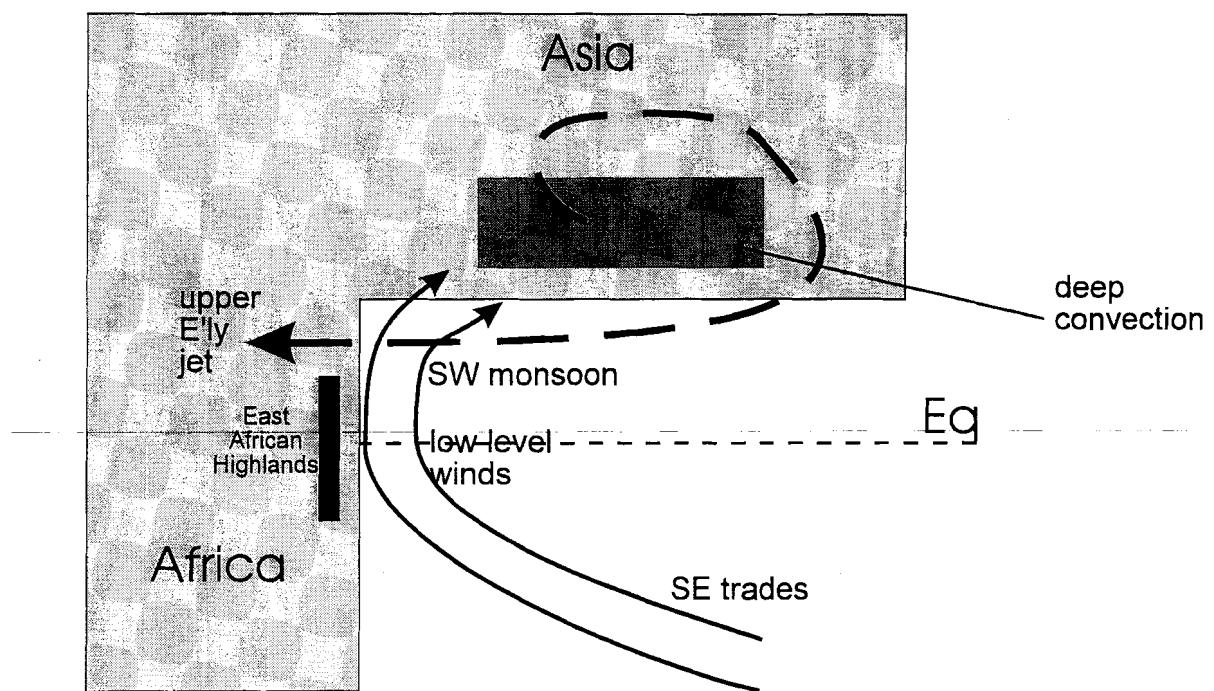
Figure 8.16:

they show a characteristic southwestward tilt (Fig. 8.17).    This structure is

Image removed due to copyright considerations.

reminiscent of the structure of midlatitude baroclinic cyclones; we saw that they must tilt westward in order to extract potential energy from the background flow.  In fact, while there is still a range of opinions, it now seems likely that monsoon depressions are also produced in part from baroclinic instability (like midlatitude cyclones) but with an important contribution from latent heat release associated with their intense rainfall (like tropical cyclones).

There is an equally pronounced variability on time scales of 10-20 days, as illustrated in Fig. 8.18.    The wet/dry periods are referred to as active monsoon / monsoon breaks. This variability appears to be a manifestation of the northward propagation of monsoon convection across the subcontinent, with new convective regions forming to the south.

Image removed due to copyright considerations.

# Chapter 9

# El Niño and the Southern Oscillation

## 9.1 Interannual fluctuations of the Walker circulation: the "Southern Oscillation"

In some years, the actual tropical circulation, especially in the Pacific region, is quite different from the climatological picture. This phenomenon, which was detected about 60 years ago by Walker and given the name "Southern Oscillation", shows up very clearly in anti-phased fluctuations of surface pressure between the west and east Pacific. The extraordinary anticorrelation in monthly mean surface pressure at Darwin (on the north central coast of Australia) and Tahiti is shown in Fig. 9.1. The relationship actually extends over a wide area; Fig. 9.2 shows the spatial structure of the temporal correlation[1] of annual-mean SLP with that of Darwin. The correlation reveals a

---

[1] If the SLP at location $\mathbf{x}$ and time $t$ is $p(\mathbf{x}, t)$, the *correlation coefficient*, $C(\mathbf{x}_0, \mathbf{x})$, between the time series of SLP at a reference location $\mathbf{x}_0$ and any other location $\mathbf{x}'$ is

$$C(\mathbf{x}_0, \mathbf{x}) = \frac{\overline{p'(\mathbf{x}_0, t)p'(\mathbf{x}, t)}}{\sqrt{\overline{p'^2(\mathbf{x}_0, t)}}\sqrt{\overline{p'^2(\mathbf{x}, t)}}},$$

where the overbar denotes the time average over the entire record and $p' = p - \bar{p}$ is the departure from that average. Note that, if $p'(\mathbf{x}, t) = \alpha p'(\mathbf{x}_0, t)$, where $\alpha$ is a constant, $C = sign(\alpha)$. If the two time series are perfectly correlated ($\alpha > 0$), $C = +1$; if perfectly anti-correlated ($\alpha < 0$), $C = -1$. If they are uncorrelated, $\overline{p'(\mathbf{x}_0, t)p'(\mathbf{x}, t)} = 0$ and so $C = 0$.

1

Images removed due to copyright considerations.

trans-Pacific dipole, with structure roughly similar to that of the Walker cell. In fact, what we are seeing here are interannual fluctuations of the Walker circulation: along with these pressure variations are variations in rainfall and in the strength of the easterly Trade winds across the tropical Pacific basin and beyond. As a measure of these fluctuations, it has become conventional to define a "Southern Oscillation Index" (SOI) as

$$SOI = 10 \times \frac{SLP_{Tahiti} - SLP_{Darwin}}{\sigma},$$

where $\sigma$ is the standard deviation of the pressure difference time series. The time series of the index is shown in Fig. 9.3; note the existence of dramatic and apparently isolated "events" (*e.g.*, 1940/41, 1982/83, 1997/98) but also periods of fluctuation (*e.g.*, 1968-77).

These fluctuations are strongest in the near-equatorial Pacific region, but in fact have a significant influence on the climate in other regions (*e.g.*, note the wave-like feature over N. America in Fug. 9.2). Fig. 9.4 shows annual rainfall at several tropical and subtropical locations. Note the tendency for certain anomalies—drought in eastern Australia, Indonesia/Melanesia, and as far as India and southeast Africa, and unusually strong rains in the central Pacific and equatorial Africa—to coincide with El Niño events.

## 9.2 SST variations: El Niño and La Niña

Manifestations are not, however, confined to the atmosphere. A phenomenon known as "El Niño" has been known for centuries to the inhabitants of the west coast of equatorial S America. Amongst other things, this comprises unusual warmth of the (usually cold) surface waters in the far eastern equatorial Pacific, poor fishing and unusual rains. Fig 9.5 shows a time series of SST in the far eastern equatorial Pacific. These show clear interannual fluctuations, on a typical time scale of a few years, with anomalously warm years occurring maybe twice per decade. Notice that the warm years (*e.g.*, 1983, 1998) tend to coincide with, or immediately follow, periods of strongly negative *SOI* (*cf.*, Fig. 9.3).

Some clues as to what is happening in the ocean are revealed by Fig. 9.6. Note the persistent W-to-E decrease of SST we noted before, the persistence of the warm waters in the west, and the annual development of very cold water in the east in the second half of the year, associated with

Image removed due to copyright considerations.

Image removed due to copyright considerations.

Image removed due to copyright considerations.

Monthly Mean SST  2°S to 2°N Average
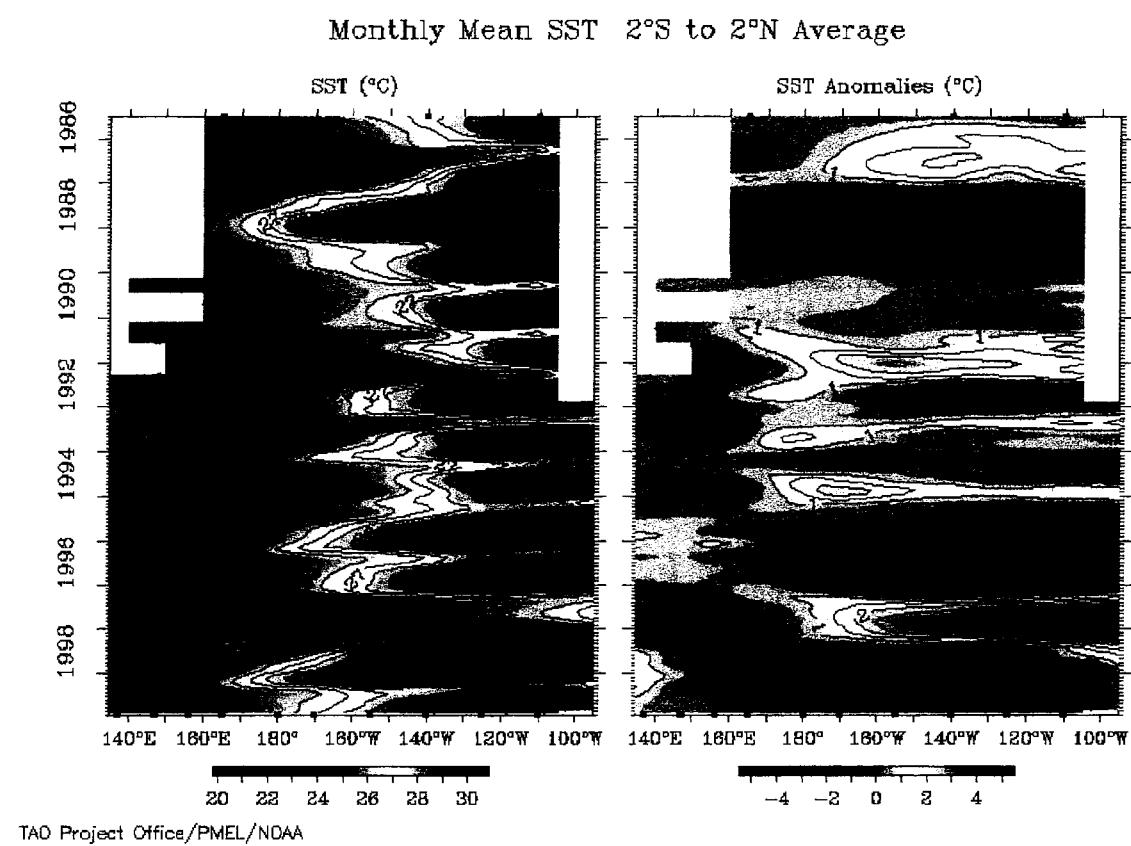
Figure 9.6: SST (left) and SST anomaly (departure from average for the time of year), as functions of time and longitude across the equatorial Pacific.

upwelling of cold water from depth. But, as Fig. 9.6 shows, the extent of
this development varies from year to year. Occasionally, the development is
unusually weak and in extreme cases (*e.g.*, 1997) hardly occurs at all. At such
times, the eastern ocean, though still no warmer than the western equatorial
Pacific waters, is very much warmer than normal for that time of year. It
is these warm events that are referred to as "El Niño"[2]. In most of these
cases the failure of the cold tongue in the eastern ocean is accompanied by
an eastward encroachment of warm water from the west, so that the SSTs
are anomalously high all the way from the eastern side almost to the date
line.

## 9.3    The coupled phenomenon

The "El Niño" phenomenon, like the SO, is irregular but has typical period-
icity of a few (2-5, usually) years. In fact, it is evident from Figs. 9.3 and
9.5 that periods of negative SOI correspond with warm periods in the east
Pacific. This is shown more clearly in Fig. 9.7. Note the extemely strong
anticorrelation.

## 9.4    Theory of ENSO

We have seen earlier that the ocean-atmosphere system is as depicted schemat-
ically in Fig. 9.8. The Walker circulation in the atmosphere is sustained by
the east-to-west gradient in SST. The ocean is driven by the wind stress as-
sociated with the easterly Trade winds. But, of course, the strength of the
Trades is determined in part by the strength of the Walker circulation: the
system is circular (Fig. 9.9), with the potential for positive feedback: change
one component, and the whole system responds in such away as to reinforce
the change.

### 9.4.1    What the observations suggest

The "big picture" of what happens during a warm ENSO event is illustrated
in Fig. 9.10.    In "normal" conditions, there is a strong E-W tilt of the
thermocline and a corresponding E-W gradient of SST, with cold upwelled

---

[2]The opposite, cold, phase (*e.g.*, 1998-99) is known as "La Niña."
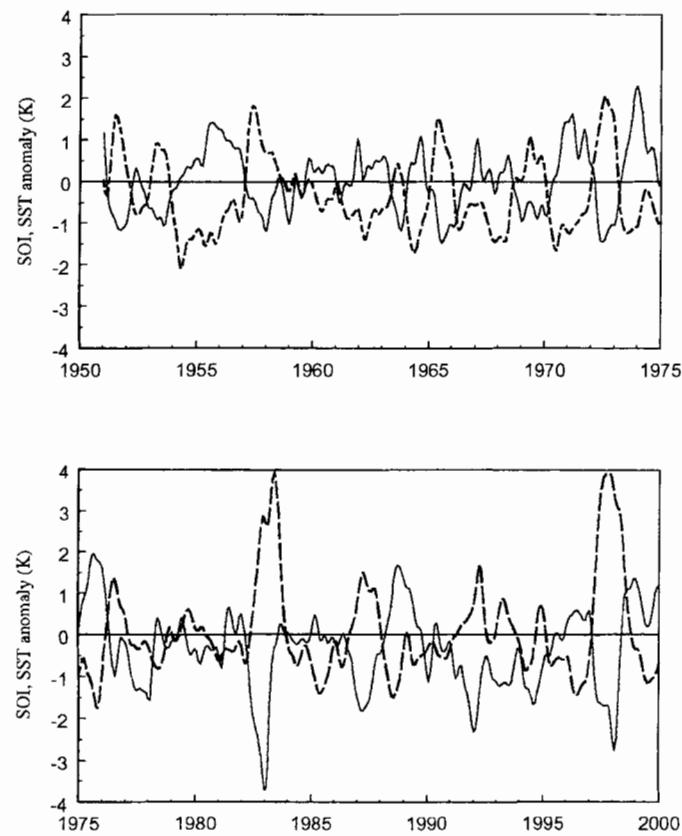
Figure 9.7: Monthly mean SOI index (red) and "Nino1+2" SST anomalies.
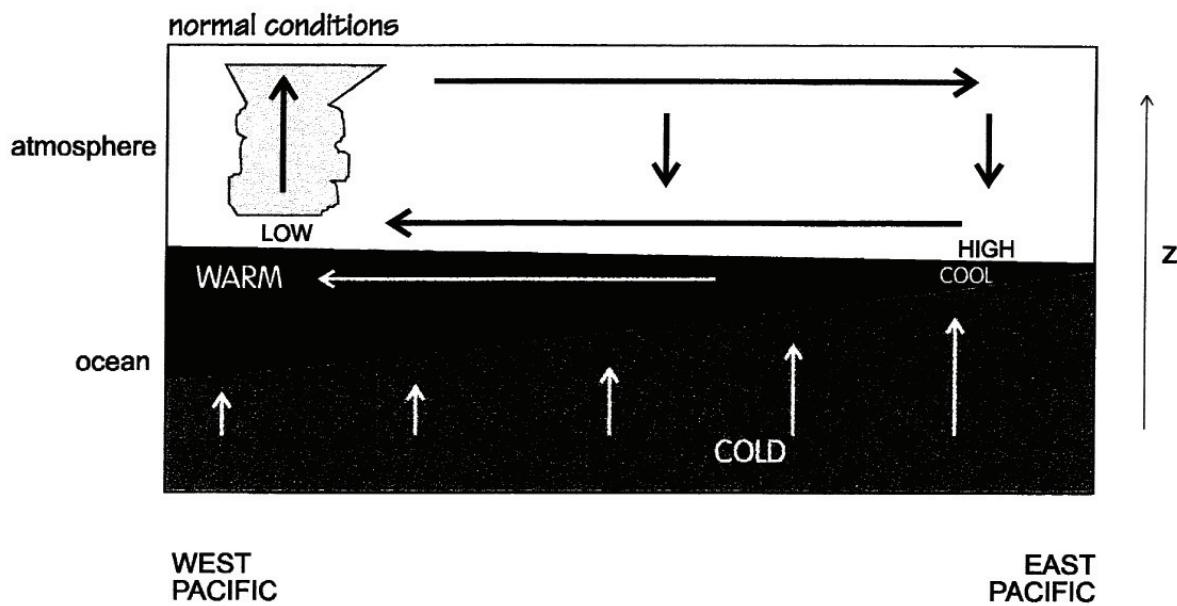
Figure 9.8: Schematic of the tropical Pacific Ocean-atmosphere system.

Image removed due to copyright considerations.

**Normal Conditions**

Convective Loop

Equator

120°E                    80°W

**El Niño Conditions**

Increased
Convection

Equator

120°E                    80°W

*NOAA/PMEL/TAO*
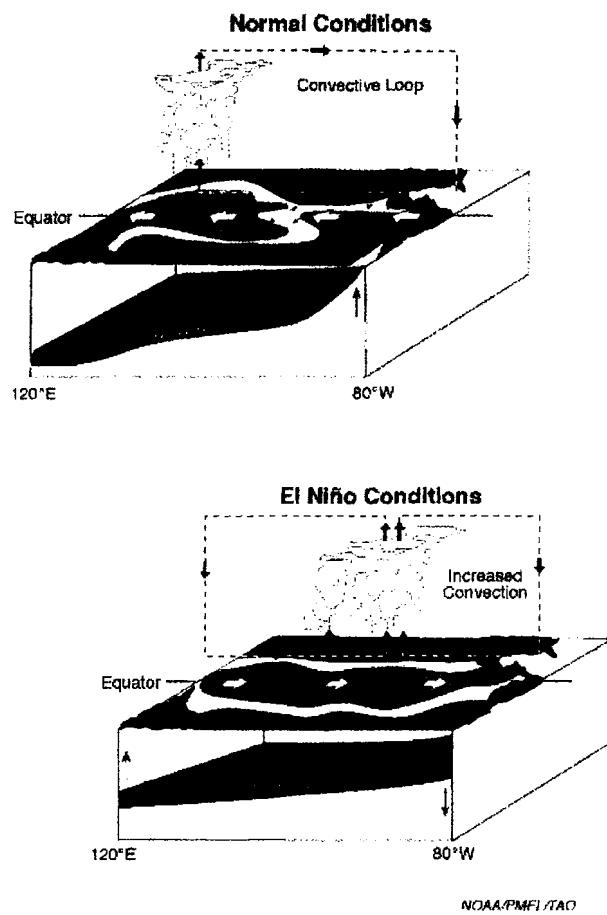
Figure 9.10: A schematic of the ocean-atmosphere behavior in the tropical Pacific basin under "normal" conditions and during a warm event.

TAO Monthly Mean SST (°C) and Winds (m s⁻¹)
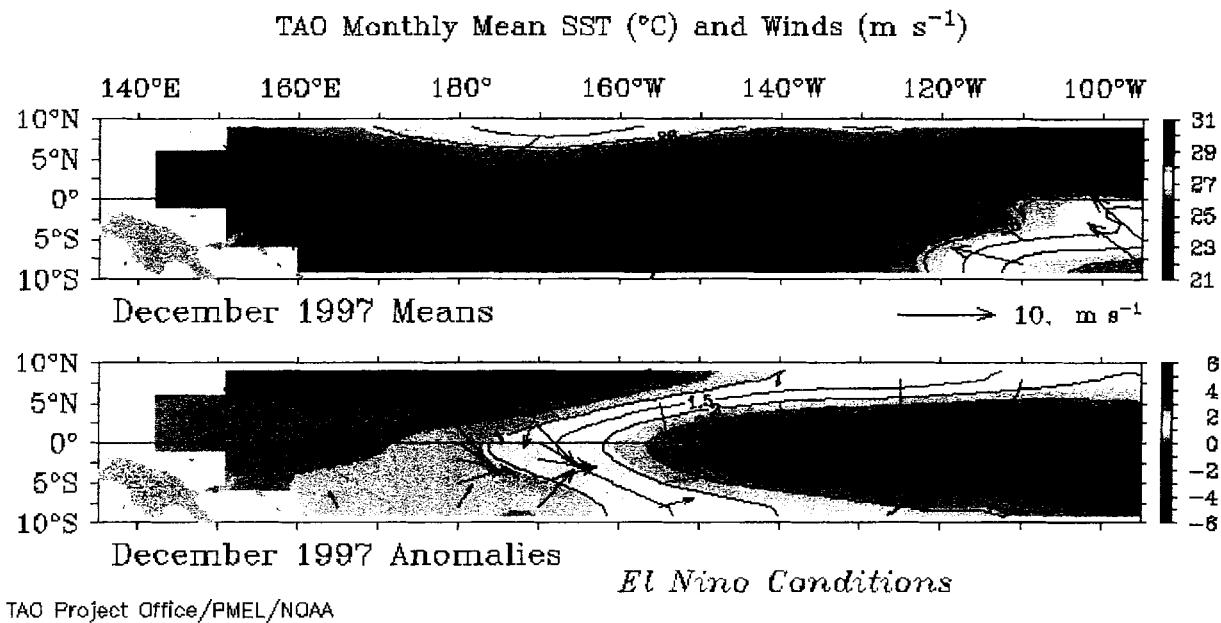


Figure 9.11: SST, winds (and anomalies) during the warm phase (El Niño) in Dec 1997.

water to the east and warm water to the west. Atmospheric convection over the warm water drives the Walker circulation, reinforcing the easterly trade winds over the equatorial ocean. During a warm El Niño event, the warm pool spreads eastward, associated with a relaxation of the tilt of the thermocline. Atmospheric convection also shifts east, moving the atmospheric circulation pattern with it. This leads to a weakening or, in a strong event, a collapse of the easterly trade winds, at least in the western part of the ocean. These features are illustrated by Figs. 9.11 and 9.12, during the two extreme phases of the phenomenon. Note especially how the Trades were weak during the warm event of 1997 and strong during the cold event of 1998.

## 9.4.2    The ocean forces the atmospheric behavior

**Remark 1** *The atmospheric fluctuations manifested as the Southern Oscillation seem to be an atmospheric response to the changed lower boundary conditions associated with El Nino SST fluctuations*

TAO Monthly Mean SST (°C) and Winds (m s$^{-1}$)
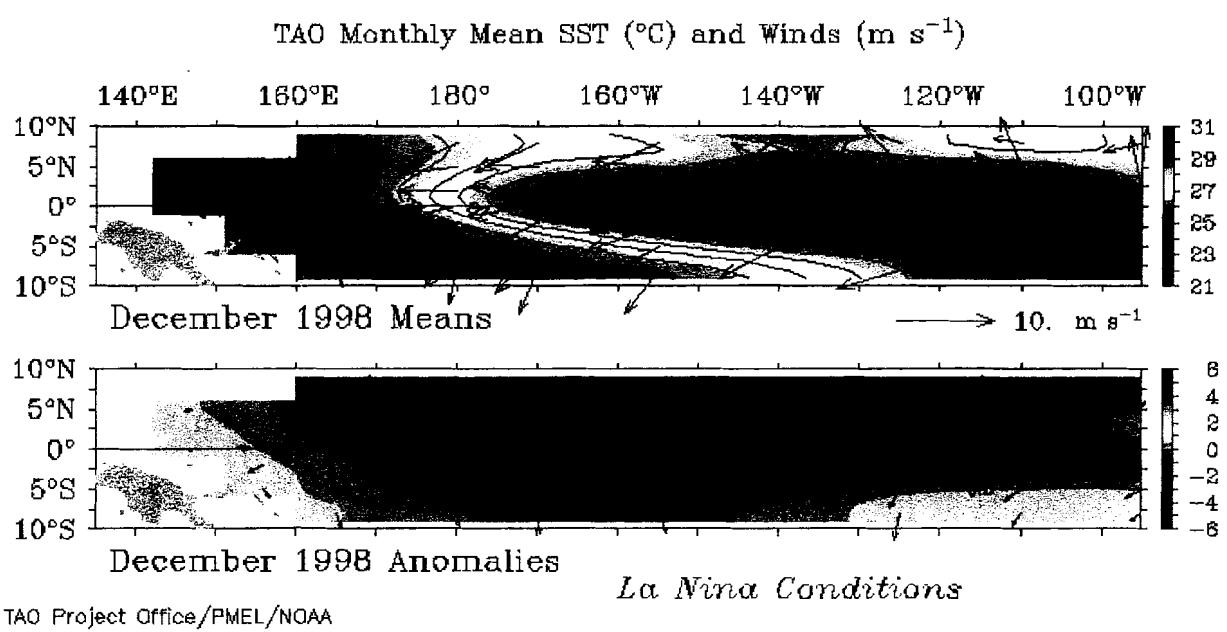


Figure 9.12: SST, winds (and anomalies) during the cold event (La Niña) of December 1998.

This has been demonstrated in a whole range of models, from the very simplest to full, three-dimensional general circulation models (GCMs). In simple terms we should expect (on the basis of our simple model) that the Walker circulation would be reduced (and the Pacific Trades tend to collapse) if the E-W contrast in SST is reduced as it is during El Nino. Specifically, one would expect to see equatorial wind anomalies in response to a shift of the heating region to be much as observed. There have been many studies using sophisticated atmospheric general circulation models (GCMs). These experiments quite successfully reproduced the Southern Oscillation, given the SST evolution as input.

### 9.4.3   The atmosphere forces the oceanic behavior

**Remark 2** *The oceanic fluctuations manifested as El Nino seem to be an oceanic response to the changed wind stress distribution associated with the Southern Oscillation.*

This was first argued by Bjerknes, who suggested that the collapse of the trades in the west Pacific in the early stages of an El Nino would drive (see Fig. 9.13) an oceanic Kelvin wave (of thermocline depression) eastward; this
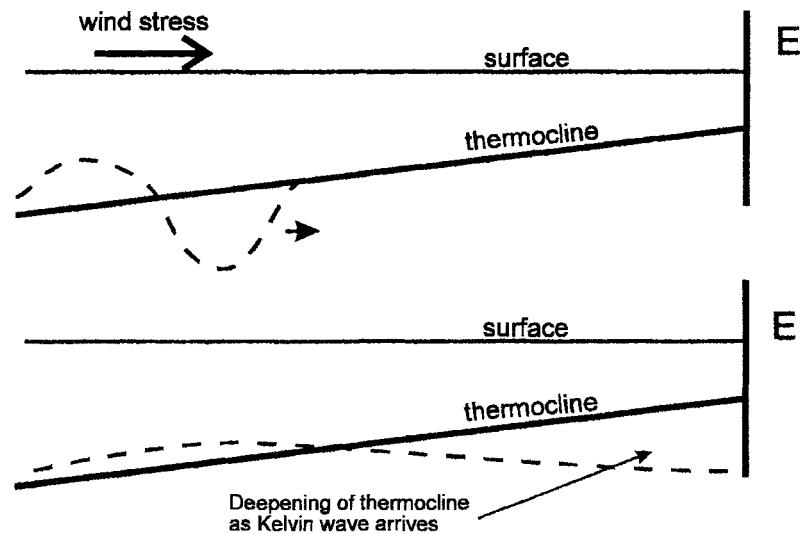


Figure 9.13:

would deepen the thermocline in the east Pacific some two months later (the speed of the relevant Kelvin waves in the equatorial ocean is about $2ms^{-1}$). This would raise the SST in the east (the upwelling continues but warmer water is being upwelled to the surface). The basic postulate—that the ocean responds to the atmosphere—has been confirmed in ocean models forced by "observed" wind stresses.

### 9.4.4 ENSO is a coupled atmosphere-ocean phenomenon

**Remark 3** *The El Nino - Southern Oscillation phenomenon arises spontaneously as an oscillation of the coupled ocean-atmosphere system*

Bjerknes first suggested that what we now call ENSO is a single phenomenon and a manifestation of ocean-atmosphere coupling. The results noted above appear to confirm that the phenomenon depends crucially on feedback between ocean and atmosphere. This is demonstrated in coupled models, in which ENSO-like fluctuations may arise spontaneously. Studies have been done with coupled models of varying complexity; such models, given the right parameters, spontaneously produce ENSO-like oscillations.

## 9.5 Further reading

A good, basic discussion of all the issues presented here (and several of the figures), as well as discussion of the impact of El Nino, can be found in:

Philander: *El Nino, La Nina and the Southern Oscillation.* Academic Press, 1990. (The later chapters are at an advanced level.)

An interesting series of articles on the large 1982-83 El Nino was published in *Science* on 16 Dec 1983. For discussion of the ocean and atmosphere, see the articles by Cane and by Rasmusson and Wallace.

Information about past and current behavior can be found on many web sites, such as **http://www.elnino.noaa.gov**