# ATMOSPHERIC AND OCEANIC FLUID DYNAMICS

*Supplementary Material for 2nd Edition*

*Geoffrey K. Vallis*

# Contents

In the main text, sections that are more advanced or that contain material that is peripheral to the main narrative are marked with a black diamond, ✦. Sections that contain material that is still not settled or that describe active areas of research are marked with a dagger, †.

*What's new, Pussycat?*

# Preface

March 8, 2014

*Major changes from July 2013 release:*

*(i)* A chapter on the stratosphere has been added.
*(ii)* The sections on wave–mean-flow interaction have been extended.
*(iii)* Numerous corrections have been made throughout.

THE DOCUMENT you are reading contains draft material for the second edition of *Atmospheric and Oceanic Fluid Dynamics* (AOFD). The publication of that book is still a year or two away.

This document contains new or revised material on the following:

*(i)* The material on waves has been extended and consolidated, and most of it has been moved from Part I into Part II. Part II now begin with a chapter on wave basics and Rossby waves.
*(ii)* A chapter on gravity waves has been added.
*(iii)* Material on wave–mean-flow interaction has been revised, including among other things Rossby wave absorption near critical layers.
*(iv)* A chapter on linear dynamics at low latitudes (equatorial waves and the Matsuno–Gill problem) has been added.
*(v)* A chapter on stratospheric dynamics has been added.
*(vi)* A chapter on the circulation equatorial ocean has been added. To this will be added a few sections on El Niño

To this will be added a chapter on moist dynamics and the tropical atmosphere, if it can be made coherent, hopefully in late 2014. In addition there will be a number of corrections and more minor changes; for example, there will be some new material on the oceanic overturning circulation and the sections on the Southern Ocean and on thickness-weighted averaging and the TEM in the primitive equations will be re-written.

I would appreciate any comments you, the reader, may have whether major or minor. Suggestions are also welcome on material to include or omit. There is no need, however, to comment on typos in the text — these will be cleaned up in the final version. However, please do point out typos in equations and, perhaps even more importantly, *thinkos,* which are sort of typos in the brain.

### An Introductory Version

As the second edition of the book will perforce be rather long (about 1000 pages), it may not be appropriate for graduate students who do not plan a career in dynamics. Thus, I expect to prepare a shorter 'student edition', which would have the advanced or more arcane material omitted and some of the explanations simplified. The resulting would likely be about 500 pages.

### Problem Sets

One omission in the first edition is numerically-oriented problems that graphically illustrate some phenomena using Matlab or Python or similar. If you have any such problems or would like to develop some that could be linked to this book, please let me know. Additional problems of a conventional nature would also be welcome.

Thank you!
Geoff Vallis
g.vallis at exeter.ac.uk

# Part I

# FUNDAMENTALS OF GEOPHYSICAL FLUID DYNAMICS

*To begin at the beginning:*
*It is spring, moonless night in the small town,*
*starless and bible-black, the cobblestreets silent*
*and the hunched, courters'-and-rabbits' wood limping invisible down*
*to the sloeblack, slow, black, crowblack, fishingboat-bobbing sea.*

Dylan Thomas, *Under Milk Wood*, 1954.

CHAPTER 1

# Equations of Motion

HAVING NOTHING BUT A BLANK SLATE, we begin by establishing the governing equations of motion for a fluid, with particular attention to the fluids of Earth's atmosphere and ocean. These equations determine how a fluid flows and evolves when forces are applied to it, or when it is heated or cooled, and so involve both dynamics and thermodynamics. And because the equations of motion are nonlinear the two become intertwined and at times inseparable.

## 1.1 TIME DERIVATIVES FOR FLUIDS

The equations of motion of fluid mechanics differ from those of rigid-body mechanics because fluids form a continuum, and because fluids flow and deform. Thus, even though the same relatively simple physical laws (Newton's laws and the laws of thermodynamics) govern both solid and fluid media, the expression of these laws differs between the two. To determine the equations of motion for fluids we must clearly establish what the time derivative of some property of a fluid actually means, and that is the subject of this section.

### 1.1.1 Field and Material Viewpoints

In solid-body mechanics one is normally concerned with the position and momentum of identifiable objects — the angular velocity of a spinning top or the motions of the planets around the Sun are two well-worn examples. The position and velocity of a particular object are then computed as a function of time by formulating equations of the form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = F(\{x_i\}, t), \tag{1.1}$$

where $\{x_i\}$ is the set of positions and velocities of all the interacting objects and the operator $F$ on the right-hand side is formulated using Newton's laws of motion. For example, two massive point objects interacting via their gravitational field obey

$$\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} = \boldsymbol{v}_i, \qquad \frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = \frac{Gm_j}{(\boldsymbol{r}_i - \boldsymbol{r}_j)^2}\hat{\boldsymbol{r}}_{i,j}, \qquad i = 1, 2;\ \ j = 3 - i. \tag{1.2}$$

We thereby predict the positions, $\boldsymbol{r}_i$, and velocities, $\boldsymbol{v}_i$, of the objects given their masses, $m_i$, and the gravitational constant $G$, and where $\hat{\boldsymbol{r}}_{i,j}$ is a unit vector directed from $\boldsymbol{r}_i$ to $\boldsymbol{r}_j$.



https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

In fluid dynamics such a procedure would lead to an analysis of fluid motions in terms of the positions and momenta of different fluid parcels, each identified by some label, which might simply be their position at an initial time. We call this a *material* point of view, because we are concerned with identifiable pieces of material; it is also sometimes called a *Lagrangian* view, after J.-L. Lagrange. The procedure is perfectly acceptable in principle, and if followed would provide a complete description of the fluid dynamical system. However, from a practical point of view it is much more than we need, and it would be extremely complicated to implement. Instead, for most problems we would like to know what the values of velocity, density and so on are at *fixed points* in space as time passes. (A weather forecast we might care about tells us how warm it will be where we live and, if we are given that, we do not particularly care where a fluid parcel comes from, or where it subsequently goes.) Since the fluid is a continuum, this knowledge is equivalent to knowing how the fields of the dynamical variables evolve in space and time, and this is often known as the *field* or *Eulerian* viewpoint, after L. Euler.[1] Thus, whereas in the material view we consider the time evolution of identifiable fluid elements, in the field view we consider the time evolution of the fluid field from a particular frame of reference. That is, we seek evolution equations of the general form

$$\frac{\partial}{\partial t}\varphi(x, y, z, t) = \mathcal{G}(\varphi, x, y, z, t), \tag{1.3}$$

where the field $\varphi(x, y, z, t)$ represents all the dynamical variables (velocity, density, temperature, etc.) and $\mathcal{G}$ is some operator to be determined from Newton's laws of motion and appropriate thermodynamic laws.

Although the field viewpoint will often turn out to be the most practically useful, the material description is invaluable both in deriving the equations and in the subsequent insight it frequently provides. This is because the important quantities from a fundamental point of view are often those which are associated with a given fluid element: it is these which directly enter Newton's laws of motion and the thermodynamic equations. It is thus important to have a relationship between the rate of change of quantities associated with a given fluid element and the local rate of change of a field. The material or advective derivative provides this relationship.

### 1.1.2 The Material Derivative of a Fluid Property

A *fluid element* is an infinitesimal, indivisible, piece of fluid — effectively a very small fluid parcel of fixed mass. The *material derivative* is the rate of change of a property (such as temperature or momentum) of a particular fluid element or finite mass of fluid; that is, it is the total time derivative of a property of a piece of fluid. It is also known as the 'substantive derivative' (the derivative associated with a parcel of fluid substance), the 'advective derivative' (because the fluid property is being advected), the 'convective derivative' (convection is a slightly old-fashioned name for advection, still used in some fields), or the 'Lagrangian derivative' (after Lagrange).

Let us suppose that a fluid is characterized by a given velocity field $\boldsymbol{v}(\boldsymbol{x}, t)$, which determines its velocity throughout. Let us also suppose that the fluid has another property $\varphi$, and let us seek an expression for the rate of change of $\varphi$ of a fluid element. Since $\varphi$ is changing in time and in space we use the chain rule,

$$\delta\varphi = \frac{\partial\varphi}{\partial t}\delta t + \frac{\partial\varphi}{\partial x}\delta x + \frac{\partial\varphi}{\partial y}\delta y + \frac{\partial\varphi}{\partial z}\delta z = \frac{\partial\varphi}{\partial t}\delta t + \delta\boldsymbol{x}\cdot\nabla\varphi. \tag{1.4}$$

This is true in general for any $\delta t$, $\delta x$, etc. The total time derivative is then

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{\partial\varphi}{\partial t} + \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}\cdot\nabla\varphi. \tag{1.5}$$

If this equation is to represent a material derivative we must identify the time derivative in the second term on the right-hand side with the rate of change of position of a fluid element, namely

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

its velocity. Hence, the material derivative of the property $\varphi$ is

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{\partial \varphi}{\partial t} + \boldsymbol{v} \cdot \nabla \varphi. \tag{1.6}$$

The right-hand side expresses the material derivative in terms of the local rate of change of $\varphi$ plus a contribution arising from the spatial variation of $\varphi$, experienced only as the fluid parcel moves. Because the material derivative is so common, and to distinguish it from other derivatives, we denote it by the operator $\mathrm{D}/\mathrm{D}t$. Thus, the material derivative of the field $\varphi$ is

$$\boxed{\frac{\mathrm{D}\varphi}{\mathrm{D}t} = \frac{\partial \varphi}{\partial t} + (\boldsymbol{v} \cdot \nabla)\varphi.} \tag{1.7}$$

The brackets in the last term of this equation are helpful in reminding us that $(\boldsymbol{v} \cdot \nabla)$ is an operator acting on $\varphi$. The operator $\partial/\partial t + (\boldsymbol{v} \cdot \nabla)$ is the *Eulerian representation of the Lagrangian derivative as applied to a field*. We use the notation $\mathrm{D}/\mathrm{D}t$ rather generally for Lagrangian derivatives, but the operator may take a different form when applied to other objects, such as a fluid volume.

*Material derivative of vector field*

The material derivative may act on a vector field $\boldsymbol{b}$, in which case

$$\frac{\mathrm{D}\boldsymbol{b}}{\mathrm{D}t} = \frac{\partial \boldsymbol{b}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{b}. \tag{1.8}$$

In Cartesian coordinates this is

$$\frac{\mathrm{D}\boldsymbol{b}}{\mathrm{D}t} = \frac{\partial \boldsymbol{b}}{\partial t} + u\frac{\partial \boldsymbol{b}}{\partial x} + v\frac{\partial \boldsymbol{b}}{\partial y} + w\frac{\partial \boldsymbol{b}}{\partial z}, \tag{1.9}$$

and for a particular component of $\boldsymbol{b}$, $b^x$ say,

$$\frac{\mathrm{D}b^x}{\mathrm{D}t} = \frac{\partial b^x}{\partial t} + u\frac{\partial b^x}{\partial x} + v\frac{\partial b^x}{\partial y} + w\frac{\partial b^x}{\partial z}, \tag{1.10}$$

and similarly for $b^y$ and $b^z$. In Cartesian tensor notation the expression becomes

$$\frac{\mathrm{D}b_i}{\mathrm{D}t} = \frac{\partial b_i}{\partial t} + v_j\frac{\partial b_i}{\partial x_j} = \frac{\partial b_i}{\partial t} + v_j \partial_j b_i, \tag{1.11}$$

where the subscripts denote the Cartesian components, repeated indices are summed, and $\partial_j b_i \equiv \partial b_i/\partial x_j$. In coordinate systems other than Cartesian the advective derivative of a vector is not simply the sum of the advective derivative of its components, because the coordinate vectors themselves change direction with position; this will be important when we deal with spherical coordinates. Finally, we remark that the advective derivative of the position of a fluid element, $\boldsymbol{r}$ say, is its velocity, and this may easily be checked by explicitly evaluating $\mathrm{D}\boldsymbol{r}/\mathrm{D}t$.

### 1.1.3 Material Derivative of a Volume

The volume that a given, unchanging, mass of fluid occupies is deformed and advected by the fluid motion, and there is no reason why it should remain constant. Rather, the volume will change as a result of the movement of each element of its bounding material surface, and in particular will change if there is a non-zero normal component of the velocity at the fluid surface. That is, if the volume of some fluid is $\int \mathrm{d}V$, then

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \mathrm{d}V = \int_S \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{S}, \tag{1.12}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

where the subscript $V$ indicates that the integral is a definite integral over some finite volume $V$, although the limits of the integral will be functions of time if the volume is changing. The integral on the right-hand side is over the closed surface, $S$, bounding the volume. Although intuitively apparent (to some), this expression may be derived more formally using Leibniz's formula for the rate of change of an integral whose limits are changing. Using the divergence theorem on the right-hand side, (1.12) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \mathrm{d}V = \int_V \nabla\cdot\boldsymbol{v}\,\mathrm{d}V. \tag{1.13}$$

The rate of change of the volume of an infinitesimal fluid element of volume $\Delta V$ is obtained by taking the limit of this expression as the volume tends to zero, giving

$$\lim_{\Delta V\to 0}\frac{1}{\Delta V}\frac{\mathrm{D}\Delta V}{\mathrm{D}t} = \nabla\cdot\boldsymbol{v}. \tag{1.14}$$

We will often write such expressions informally as

$$\frac{\mathrm{D}\Delta V}{\mathrm{D}t} = \Delta V\nabla\cdot\boldsymbol{v}, \tag{1.15}$$

with the limit implied.

Consider now the material derivative of some fluid property, $\xi$ say, multiplied by the volume of a fluid element, $\Delta V$. Such a derivative arises when $\xi$ is the amount per unit volume of $\xi$-substance — the mass density or the amount of a dye per unit volume, for example. Then we have

$$\frac{\mathrm{D}}{\mathrm{D}t}(\xi\Delta V) = \xi\frac{\mathrm{D}\Delta V}{\mathrm{D}t} + \Delta V\frac{\mathrm{D}\xi}{\mathrm{D}t}. \tag{1.16}$$

Using (1.15) this becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}(\xi\Delta V) = \Delta V\left(\xi\nabla\cdot\boldsymbol{v} + \frac{\mathrm{D}\xi}{\mathrm{D}t}\right), \tag{1.17}$$

and the analogous result for a finite fluid volume is just

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \xi\,\mathrm{d}V = \int_V\left(\xi\nabla\cdot\boldsymbol{v} + \frac{\mathrm{D}\xi}{\mathrm{D}t}\right)\mathrm{d}V. \tag{1.18}$$

This expression is to be contrasted with the Eulerian derivative for which the volume, and so the limits of integration, are fixed and we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \xi\,\mathrm{d}V = \int_V \frac{\partial\xi}{\partial t}\,\mathrm{d}V. \tag{1.19}$$

Now consider the material derivative of a fluid property $\varphi$ multiplied by the mass of a fluid element, $\rho\Delta V$, where $\rho$ is the fluid density. Such a derivative arises when $\varphi$ is the amount of $\varphi$-substance per unit mass (note, for example, that the momentum of a fluid element is $\rho\boldsymbol{v}\Delta V$). The material derivative of $\varphi\rho\Delta V$ is given by

$$\frac{\mathrm{D}}{\mathrm{D}t}(\varphi\rho\Delta V) = \rho\Delta V\frac{\mathrm{D}\varphi}{\mathrm{D}t} + \varphi\frac{\mathrm{D}}{\mathrm{D}t}(\rho\Delta V). \tag{1.20}$$

But $\rho\Delta V$ is just the mass of the fluid element, and that is constant — that is how a fluid element is defined. Thus the second term on the right-hand side vanishes and

$$\frac{\mathrm{D}}{\mathrm{D}t}(\varphi\rho\Delta V) = \rho\Delta V\frac{\mathrm{D}\varphi}{\mathrm{D}t} \qquad \text{and} \qquad \frac{\mathrm{D}}{\mathrm{D}t}\int_V \varphi\rho\,\mathrm{d}V = \int_V \rho\frac{\mathrm{D}\varphi}{\mathrm{D}t}\,\mathrm{d}V, \tag{1.21a,b}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

**Material and Eulerian Derivatives**

The material derivatives of a scalar ($\varphi$) and a vector ($\boldsymbol{b}$) field are given by:

$$\frac{\mathrm{D}\varphi}{\mathrm{D}t} = \frac{\partial\varphi}{\partial t} + \boldsymbol{v}\cdot\nabla\varphi, \qquad \frac{\mathrm{D}\boldsymbol{b}}{\mathrm{D}t} = \frac{\partial\boldsymbol{b}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{b}. \tag{D.1}$$

Various material derivatives of integrals are:

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \varphi\,\mathrm{d}V = \int_V \left(\frac{\mathrm{D}\varphi}{\mathrm{D}t} + \varphi\nabla\cdot\boldsymbol{v}\right)\mathrm{d}V = \int_V \left(\frac{\partial\varphi}{\partial t} + \nabla\cdot(\varphi\boldsymbol{v})\right)\mathrm{d}V, \tag{D.2}$$

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \mathrm{d}V = \int_V \nabla\cdot\boldsymbol{v}\,\mathrm{d}V, \tag{D.3}$$

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \rho\varphi\,\mathrm{d}V = \int_V \rho\frac{\mathrm{D}\varphi}{\mathrm{D}t}\,\mathrm{d}V. \tag{D.4}$$

These formulae also hold if $\varphi$ is a vector. The Eulerian derivative of an integral is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \varphi\,\mathrm{d}V = \int_V \frac{\partial\varphi}{\partial t}\,\mathrm{d}V, \tag{D.5}$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \mathrm{d}V = 0 \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\int_V \rho\varphi\,\mathrm{d}V = \int_V \frac{\partial\rho\varphi}{\partial t}\,\mathrm{d}V. \tag{D.6}$$

where (1.21b) applies to a finite volume. That expression may also be derived more formally using Leibniz's formula for the material derivative of an integral, and the result also holds when $\varphi$ is a vector. The result is quite different from the corresponding Eulerian derivative, in which the volume is kept fixed; in that case we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \varphi\rho\,\mathrm{d}V = \int_V \frac{\partial}{\partial t}(\varphi\rho)\,\mathrm{d}V. \tag{1.22}$$

Various material and Eulerian derivatives are summarized in the shaded box above.

## 1.2 THE MASS CONTINUITY EQUATION

In classical mechanics mass is absolutely conserved and in solid-body mechanics we normally do not need an explicit equation of mass conservation. However, in fluid mechanics fluid flows into and away from regions, and fluid density may change, and an equation that explicitly accounts for the flow of mass is one of the equations of motion of the fluid.

### 1.2.1 An Eulerian Derivation

We will first derive the mass conservation equation from an Eulerian point of view; that is to say, our reference frame is fixed in space and the fluid flows through it.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

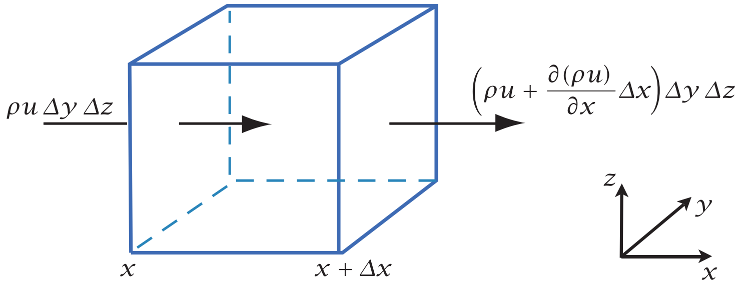


**Fig. 1.1** Mass conservation in an Eulerian cuboid control volume. The mass convergence, $-\partial(\rho u)/\partial x$ (plus contributions from the $y$ and $z$ directions), must be balanced by a density increase, $\partial\rho/\partial t$.

### *Cartesian derivation*

Consider an infinitesimal, rectangular cuboid, control volume, $\Delta V = \Delta x \Delta y \Delta z$ that is fixed in space, as in Fig. 1.1. Fluid moves into or out of the volume through its surface, including through its faces in the $y$–$z$ plane of area $\Delta A = \Delta y \Delta z$ at coordinates $x$ and $x + \Delta x$. The accumulation of fluid within the control volume due to motion in the $x$-direction is evidently

$$\Delta y \Delta z[(\rho u)(x, y, z) - (\rho u)(x + \Delta x, y, z)] = -\frac{\partial(\rho u)}{\partial x}\bigg|_{x,y,z} \Delta x\, \Delta y\, \Delta z. \tag{1.23}$$

To this must be added the effects of motion in the $y$- and $z$-directions, namely

$$-\left[\frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z}\right]\Delta x\, \Delta y\, \Delta z. \tag{1.24}$$

This net accumulation of fluid must be accompanied by a corresponding increase of fluid mass within the control volume. This is

$$\frac{\partial}{\partial t}(\text{density} \times \text{volume}) = \Delta x\, \Delta y\, \Delta z \frac{\partial \rho}{\partial t}, \tag{1.25}$$

because the volume is constant. Thus, because mass is conserved, (1.23), (1.24) and (1.25) give

$$\Delta x\, \Delta y\, \Delta z \left[\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z}\right] = 0. \tag{1.26}$$

The quantity in square brackets must be zero and we therefore have

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0.} \tag{1.27}$$

This is called the *mass continuity equation* for it recognizes the continuous nature of the mass field in a fluid. There is no diffusion term in (1.27), no term like $\kappa\nabla^2\rho$. This is because mass is transported by the macroscopic movement of molecules; even if this motion appears diffusion-like any net macroscopic molecular motion constitutes, by definition, a velocity field.

### *Vector derivation*

Consider an arbitrary control volume $V$ bounded by a surface $S$, fixed in space, with by convention the direction of $\boldsymbol{S}$ being toward the outside of $V$, as in Fig. 1.2. The rate of fluid loss due to flow through the closed surface $S$ is then given by

$$\text{fluid loss} = \int_S \rho \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{S} = \int_V \nabla \cdot (\rho \boldsymbol{v})\, \mathrm{d}V, \tag{1.28}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

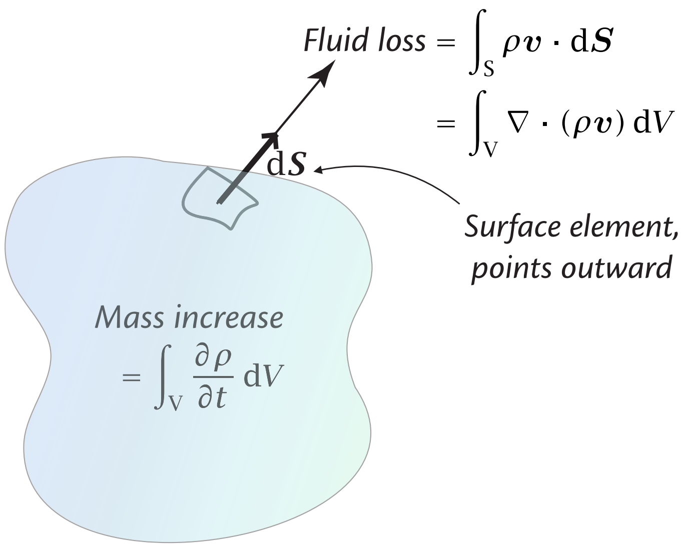


**Fig. 1.2** Mass conservation in an arbitrary Eulerian control volume $V$ bounded by a surface $S$. The mass increase, $\int_V (\partial \rho/\partial t)\, \mathrm{d}V$ is equal to the mass flowing into the volume, $-\int_S (\rho \boldsymbol{v}) \cdot \mathrm{d}\boldsymbol{S} = -\int_V \nabla \cdot (\rho \boldsymbol{v})\, \mathrm{d}V$.

using the divergence theorem.

This must be balanced by a change in the mass $M$ of the fluid within the control volume, which, since its volume is fixed, implies a density change. That is

$$\text{fluid loss} = -\frac{\mathrm{d}M}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_V \rho\, \mathrm{d}V = -\int_V \frac{\partial \rho}{\partial t}\, \mathrm{d}V. \tag{1.29}$$

Equating (1.28) and (1.29) yields

$$\int_V \left[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) \right] \mathrm{d}V = 0. \tag{1.30}$$

Because the volume is arbitrary, the integrand must vanish and we recover (1.27).

### 1.2.2 Mass Continuity via the Material Derivative

We now derive the mass continuity equation (1.27) from a material perspective. This is the most fundamental approach of all since the principle of mass conservation states simply that the mass of a given element of fluid is, by definition of the element, constant. Thus, consider a small mass of fluid of density $\rho$ and volume $\Delta V$. Then conservation of mass may be represented by

$$\frac{\mathrm{D}}{\mathrm{D}t}(\rho \Delta V) = 0. \tag{1.31}$$

Both the density and the volume of the parcel may change, so

$$\Delta V \frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \frac{\mathrm{D}\Delta V}{\mathrm{D}t} = \Delta V \left( \frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \nabla \cdot \boldsymbol{v} \right) = 0, \tag{1.32}$$

where the second expression follows using (1.15). Since the volume element is arbitrary, the term in brackets must vanish and

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \nabla \cdot \boldsymbol{v} = 0. \tag{1.33}$$

After expansion of the first term this becomes identical to (1.27). This result may be derived more formally by rewriting (1.31) as the integral expression

$$\frac{\mathrm{D}}{\mathrm{D}t} \int_V \rho\, \mathrm{d}V = 0. \tag{1.34}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

Expanding the derivative using (1.18) gives

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \rho \,\mathrm{d}V = \int_V \left( \frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \nabla \cdot \boldsymbol{v} \right) \mathrm{d}V = 0. \tag{1.35}$$

Because the volume over which the integral is taken is arbitrary the integrand itself must vanish and we recover (1.33). Summarizing, equivalent partial differential equations representing conservation of mass are:

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \nabla \cdot \boldsymbol{v} = 0, \qquad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0. \tag{1.36a,b}$$

### 1.2.3 A General Continuity Equation

The derivation of a continuity equation for a general scalar property of a fluid is similar to that for density, except that there may be an external source or sink, and potentially a means of transferring the property from one location to another differently than by fluid motion, for example by diffusion. If $\xi$ is the amount of some property of the fluid per unit volume (the volume concentration, sometimes simply called the concentration), and if the net effect per unit volume of all non-conservative processes is denoted by $Q_{[v,\xi]}$, then the continuity equation for concentration may be written:

$$\frac{\mathrm{D}}{\mathrm{D}t}(\xi \Delta V) = Q_{[v,\xi]} \Delta V. \tag{1.37}$$

Expanding the left-hand side and using (1.15) we obtain

$$\frac{\mathrm{D}\xi}{\mathrm{D}t} + \xi \nabla \cdot \boldsymbol{v} = Q_{[v,\xi]}, \qquad \text{or} \qquad \frac{\partial \xi}{\partial t} + \nabla \cdot (\xi \boldsymbol{v}) = Q_{[v,\xi]}. \tag{1.38}$$

If we are interested in a tracer that is normally measured per unit mass of fluid (which is typical when considering thermodynamic quantities) then the conservation equation would be written

$$\frac{\mathrm{D}}{\mathrm{D}t}(\varphi \rho \Delta V) = Q_{[m,\varphi]} \rho \Delta V, \tag{1.39}$$

where $\varphi$ is the tracer mixing ratio or mass concentration — that is, the amount of tracer per unit fluid mass — and $Q_{[m,\varphi]}$ represents non-conservative sources of $\varphi$ per unit mass. Then, since $\rho \Delta V$ is constant we obtain

$$\frac{\mathrm{D}\varphi}{\mathrm{D}t} = Q_{[m,\varphi]} \qquad \text{or} \qquad \frac{\partial (\rho\varphi)}{\partial t} + \nabla \cdot (\rho \varphi \boldsymbol{v}) = \rho Q_{[m,\varphi]}, \tag{1.40}$$

using the mass continuity equation, (1.36), to obtain the equation on the right. The source term $Q_{[m,\varphi]}$ is evidently equal to the rate of change of $\varphi$ of a fluid element. When this is so, we often write it simply as $\dot{\varphi}$, so that

$$\frac{\mathrm{D}\varphi}{\mathrm{D}t} = \dot{\varphi}. \tag{1.41}$$

A tracer obeying (1.41) with $\dot{\varphi} = 0$ is said to be *materially conserved*. If a tracer is materially conserved except for the effects of sources or sinks, or diffusion terms, then it is sometimes (if rather loosely) said to be an 'adiabatically conserved' variable, although adiabatic properly means with no heat exchange. If those sources and sinks are in the form of the divergence of a flux with $\varphi$ satisfying $\rho \mathrm{D}\varphi/\mathrm{D}t = \nabla \cdot \boldsymbol{F}_\varphi$ or equivalently, using the mass continuity equation, $\partial(\rho\varphi)/\partial t + \nabla \cdot (\rho \boldsymbol{v} \varphi) = \nabla \cdot \boldsymbol{F}_\varphi$, then $\varphi$ is said to be a *conservative* variable because, with no flux boundary conditions, $\int \rho\varphi \,\mathrm{d}V = \text{constant}$. Although momentum as a whole is conserved, momentum is not a materially conserved variable, as we are about to see.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

## 1.3 THE MOMENTUM EQUATION

The momentum equation is a partial differential equation that describes how the velocity or momentum of a fluid responds to internal and imposed forces. We will derive it using material methods, with a very heuristic treatment of the terms representing pressure and viscous forces.

### 1.3.1 Advection

Let $\boldsymbol{m}(x, y, z, t)$ be the momentum-density field (momentum per unit volume) of the fluid. Thus, $\boldsymbol{m} = \rho\boldsymbol{v}$ and the total momentum of a volume of fluid is given by the volume integral $\int_V \boldsymbol{m}\,\mathrm{d}V$. Now, for a fluid the rate of change of a momentum of an identifiable fluid mass is given by the material derivative, and by Newton's second law this is equal to the force acting on it. Thus,

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \rho\boldsymbol{v}\,\mathrm{d}V = \int_V \boldsymbol{F}\,\mathrm{d}V, \tag{1.42}$$

where $\boldsymbol{F}$ is the force per unit volume. Now, using (1.21b) (with $\varphi$ replaced by $\boldsymbol{v}$) to transform the left-hand side of (1.42), we obtain

$$\int_V \left(\rho\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} - \boldsymbol{F}\right)\mathrm{d}V = 0. \tag{1.43}$$

Because the volume is arbitrary the integrand itself must vanish and we obtain

$$\rho\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \boldsymbol{F} \quad \text{or} \quad \frac{\partial\boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = \frac{\boldsymbol{F}}{\rho}, \tag{1.44a,b}$$

having used (1.8) to expand the material derivative.

We have thus obtained an expression for how a fluid accelerates if subject to known forces. These forces are, however, not all external to the fluid itself: a stress arises from the direct contact between one fluid parcel and another, giving rise to pressure and viscous forces, sometimes referred to as *contact* forces. Because a complete treatment of these forces would be very lengthy, and is available elsewhere, we treat them informally and intuitively.

### 1.3.2 Pressure and Viscous Forces

*Pressure*

Within or at the boundary of a fluid the pressure is the normal force per unit area due to the collective action of molecular motion. Thus

$$\mathrm{d}\widehat{\boldsymbol{F}}_p = -p\,\mathrm{d}\boldsymbol{S}, \tag{1.45}$$

where $p$ is the pressure, $\widehat{\boldsymbol{F}}_p$ is the pressure force and $\mathrm{d}\boldsymbol{S}$ an infinitesimal surface element. If we grant ourselves this intuitive notion, it is a simple matter to assess the influence of pressure on a fluid, for the pressure force on a volume of fluid is the integral of the pressure over its boundary and so

$$\widehat{\boldsymbol{F}}_p = -\int_S p\,\mathrm{d}\boldsymbol{S}. \tag{1.46}$$

The minus sign arises because the pressure force is directed inwards, whereas $\boldsymbol{S}$ is a vector normal to the surface and directed outwards. Applying a form of the divergence theorem to the right-hand side gives

$$\widehat{\boldsymbol{F}}_p = -\int_V \nabla p\,\mathrm{d}V, \tag{1.47}$$

where the volume $V$ is bounded by the surface $S$. The pressure force per unit volume, $\boldsymbol{F}_p$, is therefore just $-\nabla p$, and the force per unit mass is $\nabla p/\rho$. The force is evidently non-zero only if the pressure varies in space and for this reason it is more properly known as the *pressure-gradient* force.

**Table 1.1** Experimental values of viscosity for air, water and mercury at room temperature and pressure.

| | $\mu$ ( kg m$^{-1}$ s$^{-1}$) | $\nu$ ( m$^2$ s$^{-1}$) |
|---|---|---|
| Air | $1.8 \times 10^{-5}$ | $1.5 \times 10^{-5}$ |
| Water | $1.1 \times 10^{-3}$ | $1.1 \times 10^{-6}$ |
| Mercury | $1.6 \times 10^{-3}$ | $1.2 \times 10^{-7}$ |

*Viscosity*

Viscosity, like pressure, is a force due to the internal motion of molecules. The effects of viscosity are apparent in many situations — the flow of treacle or volcanic lava are obvious examples. In other situations, for example large-scale flow in the atmosphere, viscous effects are negligible. However, for a constant density fluid viscosity is the *only* way that energy may be removed from the fluid, so that if energy is being added in some way viscosity must ultimately become important if the fluid is to reach an equilibrium in which energy input equals energy dissipation. When tea is stirred in a cup, it is viscous effects that cause the fluid to eventually stop spinning after we have removed our spoon.

A number of textbooks show that, for most Newtonian fluids, the viscous force per unit volume is approximately equal to $\mu\nabla^2\boldsymbol{v}$, where $\mu$ is the viscosity. Obtaining this expression involves making an incompressibility assumption and is not exact, but it is in fact a good approximation for most liquids and gases. With this term and the pressure-gradient force the momentum equation becomes,

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\frac{1}{\rho}\nabla p + \nu\nabla^2\boldsymbol{v} + \boldsymbol{F}_b, \tag{1.48}$$

where $\nu \equiv \mu/\rho$ is the *kinematic viscosity* and $\boldsymbol{F}_b$ represents external body forces (per unit mass) such as gravity, $\boldsymbol{g}$. Equation (1.48) is sometimes called the Navier–Stokes equation. For gases, dimensional arguments suggest that the magnitude of $\nu$ should be given by

$$\nu \sim \text{mean free path} \times \text{mean molecular velocity}, \tag{1.49}$$

which for a typical molecular velocity of 300 m s$^{-1}$ and a mean free path of $7 \times 10^{-8}$ m gives the not unreasonable estimate of $2.1 \times 10^{-5}$ m$^2$ s$^{-1}$, within a factor of two of the experimental value (Table 1.1). Interestingly, the kinematic viscosity is smaller for water and mercury than it is for air.

### 1.3.3 Hydrostatic Balance

The vertical component — the component parallel to the gravitational force, $\boldsymbol{g}$ — of the momentum equation is

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g, \tag{1.50}$$

where $w$ is the vertical component of the velocity and $\boldsymbol{g} = -g\mathbf{k}$. If the fluid is static the gravitational term is balanced by the pressure term and we have

$$\frac{\partial p}{\partial z} = -\rho g, \tag{1.51}$$

and this relation is known as *hydrostatic balance,* or hydrostasy. It is clear in this case that the pressure at a point is given by the weight of the fluid above it, provided $p = 0$ at the top of the fluid. It might also appear that (1.51) would be a good *approximation* to (1.50) provided that vertical accelerations, $\mathrm{D}w/\mathrm{D}t$, are small compared to gravity, which is nearly always the case in the atmosphere and ocean. While this statement is true if we need only a reasonable approximate value of

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

the pressure at a point or in a column, the satisfaction of this condition is *not* sufficient to ensure that (1.51) provides an accurate enough pressure to determine the horizontal pressure gradients responsible for producing motion. We return to this point in Section 2.7.

## 1.4 THE EQUATION OF STATE

In three dimensions the momentum and continuity equations provide four equations, but contain five unknowns — three components of velocity, density and pressure. Obviously other equations are needed, and an *equation of state* is an expression that diagnostically relates the various thermodynamic variables to each other. The *conventional* equation of state, or the *thermal* equation of state, is an expression that relates temperature, pressure, composition (the mass fraction of the various constituents) and density, and we may write it, rather generally, as

$$p = p(\rho, T, \varphi_n), \tag{1.52}$$

where $\varphi_n$ is mass fraction of the $n$th constituent. An equation of this form is not the most fundamental equation of state from a thermodynamic perspective, an issue we visit later, but it connects readily measurable quantities.

For an ideal gas (and the air in the Earth's atmosphere is very close to ideal) the thermal equation of state is

$$p = \rho R T, \tag{1.53}$$

where $R$ is the gas constant for the gas in question and $T$ is temperature. $R$ is a specific constant, and is related to the universal gas constant $R^*$ by $R = R^*/\overline{\mu}$, where $\overline{\mu}$ is the mean molar mass (molecular weight in kg/mol) of the constituents of the gas. Equivalently, $R = n_m k_B$, where $k_B$ is Boltzmann's constant and $n_m$ is the number of molecules per unit mass, so that $R$ is proportional to the number of molecules contained in a unit mass. Since $R^* = 8.314\,\mathrm{J\,mol^{-1}\,K^{-1}}$ and, for dry air, $\mu = 29.0 \times 10^{-3}\,\mathrm{kg\,mol^{-1}}$ we obtain $R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$. Air has virtually constant composition except for variations in water vapour; these variations make the gas constant, $R$, in the equation of state for air a weak function of the water vapour content but for now we regard $R$ as a constant.

For a liquid such as seawater no simple expression akin to (1.53) is easily derivable, and semi-empirical equations are usually resorted to. For water in a laboratory setting a reasonable approximation of the equation of state is $\rho = \rho_0[1-\beta_T(T-T_0)]$, where $\beta_T$ is a thermal expansion coefficient and $\rho_0$ and $T_0$ are constants. In the ocean the density is also significantly affected by pressure and dissolved salts: seawater is a solution of many ions in water — chloride ($\approx$ 1.9% by weight), sodium (1%), sulfate (0.26%), magnesium (0.13%) and so on, with a total average concentration of about 35‰ (ppt, or parts per thousand). The ratios of the fractions of these salts are almost constant throughout the ocean, and their total concentration may be parameterized by a single measure, the *salinity*, $S$.[2] Given this, the density of seawater is a function of three variables — pressure, temperature, and salinity — and we may write the conventional equation of state as

$$\rho = \rho(T, S, p) \qquad \text{or} \qquad \alpha = \alpha(T, S, p), \tag{1.54}$$

where $\alpha = 1/\rho$ is the specific volume, or inverse density. For small variations around a reference value we have

$$\mathrm{d}\alpha = \left(\frac{\partial \alpha}{\partial T}\right)_{S,p} \mathrm{d}T + \left(\frac{\partial \alpha}{\partial S}\right)_{T,p} \mathrm{d}S + \left(\frac{\partial \alpha}{\partial p}\right)_{T,S} \mathrm{d}p = \alpha\,(\beta_T\,\mathrm{d}T - \beta_S\,\mathrm{d}S - \beta_p\,\mathrm{d}p), \tag{1.55}$$

where the rightmost expression serves to define the thermal expansion coefficient $\beta_T$, the saline contraction coefficient $\beta_S$, and the compressibility coefficient (or inverse bulk modulus) $\beta_p$. In general these quantities are not constants, but for small variations around a reference state they may be treated as such and we have

$$\alpha = \alpha_0 \left[1 + \beta_T(T - T_0) - \beta_S(S - S_0) - \beta_p(p - p_0)\right]. \tag{1.56}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

Typical values of these parameters, with variations typically encountered through the ocean, are: $\beta_T \approx 2\,(\pm 1.5) \times 10^{-4}\,\mathrm{K}^{-1}$, $\beta_S \approx 7.6\,(\pm 0.2) \times 10^{-4}\,\mathrm{ppt}^{-1}$, $\beta_p \approx 4.4\,(\pm 0.5) \times 10^{-10}\,\mathrm{Pa}^{-1}$. The value of $\beta_p$ is also related to the speed of sound, $c_s$, by $\beta_p = \alpha_0 / c_s^2$. Since the variations around the mean density are small, (1.56) implies that

$$\rho = \rho_0 \left[ 1 - \beta_T (T - T_0) + \beta_S (S - S_0) + \beta_p (p - p_0) \right]. \tag{1.57}$$

In the ocean the pressure term leads to larger density changes than either the salinity or temperature terms but it is not normally as important for the dynamics, because pressure is largely determined by the hydrostatic pressure giving a large vertical density gradient. It is the lateral variations in density that are often more important for the dynamics and these are affected just as much by the saline and temperature terms.

A linear equation of state for seawater is emphatically *not* accurate enough for quantitative oceanography; as mentioned the $\beta$ parameters in (1.56) themselves vary with pressure, temperature and (more weakly) salinity so introducing nonlinearities to the equation. The most important of these are captured by an equation of state of the form

$$\alpha = \alpha_0 \left[ 1 + \beta_T (1 + \gamma^* p)(T - T_0) + \frac{\beta_T^*}{2} (T - T_0)^2 - \beta_S (S - S_0) - \beta_p (p - p_0) \right]. \tag{1.58}$$

The starred constants $\beta_T^*$ and $\gamma^*$ capture the leading nonlinearities: $\gamma^*$ is the *thermobaric parameter,* which determines the extent to which the thermal expansion depends on pressure, and $\beta_T^*$ is the second thermal expansion coefficient.[3] Even this equation of state has some quantitative deficiencies and more complicated empirical formulae are often used if very high accuracy is needed. The variation of density of seawater with temperature, salinity and pressure is illustrated in Fig. 1.3, with more discussion in Section 1.7.2.

Clearly, the equation of state introduces, in general, a sixth unknown, temperature, and we will have to introduce another physical principle — one coming from thermodynamics — to obtain a complete set of equations. However, if the equation of state were such that it linked only density and pressure, without introducing another variable, then the equations would be complete; the simplest case of all is a constant density fluid for which the equation of state is just $\rho =$ constant. A fluid for which the density is a function of pressure alone is called a *barotropic fluid* or a *homentropic fluid*; otherwise, it is a *baroclinic fluid*. (In this context, 'barotropic' is a shortening of the original phrase 'auto-barotropic'.) Equations of state of the form $p = C\rho^\gamma$, where $\gamma$ is a constant, are called 'polytropic'.

## 1.5 THERMODYNAMIC RELATIONS

In this section we review a few aspects of thermodynamics. We provide neither a complete nor an a priori development of the subject; rather, we focus on aspects that are particularly relevant to fluid dynamics, and that are needed to derive a 'thermodynamic equation' for fluids. Readers whose interest is solely in an ideal gas or a simple Boussinesq fluid may skim this section and then refer to it later as needed.

### 1.5.1 A Few Fundamentals

A fundamental postulate of thermodynamics is that the internal energy of a system in equilibrium is a function of its extensive properties: volume, entropy, and the mass of its various constituents. Extensive means that the property value is proportional to the amount of material present, in contrast to an intensive property such as temperature. For our purposes it is more convenient to divide all of these quantities by the mass of fluid present, so expressing the internal energy per unit mass (or the specific internal energy) $I$, as a function of the specific volume $\alpha = \rho^{-1}$, the specific entropy

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

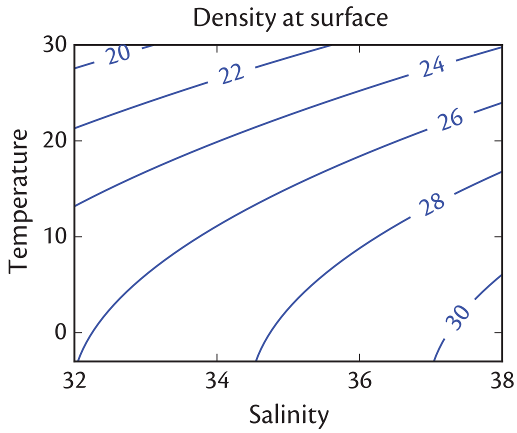


**Fig. 1.3** Contours of density as a function of temperature and salinity for seawater. Contour labels are (density – 1000) kg m$^{-3}$. Left panel: at sea-level ($p = 10^5$ Pa, or 1000 mb). Right panel: at $p = 4 \times 10^7$ Pa (about 4 km depth). In both cases the contours are slightly convex, so that if two parcels at the same density but different temperatures and salinities are mixed, the resulting parcel is of higher density. (The average temperature is not exactly conserved on mixing, but it very nearly is.)

$\eta$, and the mass fractions of its various components. (However, multiplying by the mass does not turn an intensive variable into a properly extensive one.) Our interest is in two-component fluids (dry air and water vapour, or water and salinity) so that we may parameterize the composition by a single parameter, $S$. (We follow conventional thermodynamical notation as much as possible, except that we use $I$ instead of $u$ for internal energy, since $u$ is a fluid velocity, and $\eta$ instead of $S$ for entropy, since $S$ is salinity.) We can also write entropy in terms of internal energy, density and salinity, and thus we have

$$I = I(\alpha, \eta, S) \qquad \text{or} \qquad \eta = \eta(I, \alpha, S). \tag{1.59a,b}$$

Given the functional forms on the right-hand sides, either of these expressions constitutes a complete description of the macroscopic state of a system in equilibrium, and we call either of them the *fundamental equation of state*. The thermal equation of state can be derived from (1.59), but not vice versa. The first differential of (1.59a) gives, formally,

$$\mathrm{d}I = \left(\frac{\partial I}{\partial \eta}\right)_{\alpha,S} \mathrm{d}\eta + \left(\frac{\partial I}{\partial \alpha}\right)_{\eta,S} \mathrm{d}\alpha + \left(\frac{\partial I}{\partial S}\right)_{\alpha,\eta} \mathrm{d}S. \tag{1.60}$$

We will now ascribe physical meaning to these differentials.

Conservation of energy states that the internal energy of a body may change because of work done by or on it, or because of a heat input, or because of a change in its chemical composition. We write this as

$$\mathrm{d}I = \text{đ}Q + \text{đ}W + \text{đ}C, \tag{1.61}$$

where đ$W$ is the work done on the body, đ$Q$ is the heat input to the body, and đ$C$ is the change in internal energy caused by a change in its chemical composition (e.g., its salinity, or water vapour content), sometimes called the 'chemical work'. The infinitesimal quantities on the right-hand side (denoted with a đ) are so-called imperfect or inexact differentials:[4] $Q$, $W$ and $C$ are not functions of the state of a body, and the internal energy cannot be regarded as the sum of a 'heat' and a 'work'. We should think of heat and work as having meaning only as fluxes of energy, or rates of energy input,

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

and not as amounts of energy; their sum changes the internal energy of a body, which *is* a function of its state. Equation (1.61) is sometimes called 'the first law of thermodynamics' (discussed more in Appendix B at the end of the chapter). We first consider the causes of variations of the quantities on the right-hand side, and then make a connection to (1.60).

*Heat input:* In an infinitesimal quasi-static or reversible process, if an amount of heat $đQ$ (per unit mass) is externally supplied then the specific entropy $\eta$ will change according to

$$T\,\mathrm{d}\eta = đQ. \tag{1.62}$$

The entropy is a function of the state of a body and is, by definition, an adiabatic invariant. As we are dealing with the amount of a quantity per unit mass, $\eta$ is the specific entropy, although we will often refer to it just as the entropy. We may regard (1.61) as defining the heat input, $đQ$, by way of a statement of conservation of energy, and (1.62) then says that there is a function of state, the entropy, that changes by an amount equal to the heat input divided by the temperature.

*Work done:* The work done on a body during a reversible process is equal to the pressure times its change in volume, and the work is positive if the volume change is negative. Thus if an infinitesimal amount of work $đW$ (per unit mass) is applied to a body then its thermodynamic state will change according to

$$-p\,\mathrm{d}\alpha = đW, \tag{1.63}$$

where $\alpha = 1/\rho$ is the specific volume of the fluid and $p$ is the pressure.

*Composition:* The chemical work, which produces a change in internal energy due to a small change in composition, $\mathrm{d}S$, is given by

$$đC = \mu\,\mathrm{d}S, \tag{1.64}$$

where $\mu$ is the *chemical potential* of the solution. In the ocean, salinity is the compositional variable and changes arise through precipitation and evaporation at the surface and molecular diffusion. When salinity changes the internal energy of a fluid parcel changes by (1.64), but this change is usually small compared to other changes in internal energy. In practice, the most important effect of salinity is that it changes the density of seawater. In the atmosphere the composition of a parcel of air varies mainly according to the amount of water vapour present. Since water vapour and dry air have different chemical potentials these variations cause changes in internal energy, but in the absence of phase-changes the changes are small. An important compositional effect does arise when condensation or evaporation occurs, for then energy is released (or required), as discussed in Chapter 18.

Collecting equations (1.61)–(1.64) together we have

$$\boxed{\mathrm{d}I = T\,\mathrm{d}\eta - p\,\mathrm{d}\alpha + \mu\,\mathrm{d}S.} \tag{1.65}$$

We refer to this as *the fundamental thermodynamic relation*. The fundamental equation of state, (1.59), describes the properties of a particular fluid, and the fundamental relation, (1.65), is associated with conservation of energy. Much of classical thermodynamics follows from these two expressions.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

### 1.5.2 Thermodynamic Potentials and Maxwell Relations

Given the fundamental thermodynamic relation, various other 'thermodynamic potentials' and relations between variables can be derived that prove extremely useful. The thermodynamic potentials are, like internal energy and entropy, functions of the state but they have different natural variables by which they are expressed.

If we begin with the internal energy itself, then from (1.60) and (1.65) it follows that

$$T = \left(\frac{\partial I}{\partial \eta}\right)_{\alpha,S}, \qquad p = -\left(\frac{\partial I}{\partial \alpha}\right)_{\eta,S}, \qquad \mu = \left(\frac{\partial I}{\partial S}\right)_{\eta,\alpha}. \tag{1.66a,b,c}$$

These may be regarded as the defining relations for these variables; because of the connection between (1.61) and (1.65) these are not just formal definitions, and the pressure and temperature so defined are indeed related to our intuitive concepts of these variables and to the motion of the fluid molecules. If we write

$$\mathrm{d}\eta = \frac{1}{T}\,\mathrm{d}I + \frac{p}{T}\,\mathrm{d}\alpha - \frac{\mu}{T}\,\mathrm{d}S, \tag{1.67}$$

it is also clear that

$$p = T\left(\frac{\partial \eta}{\partial \alpha}\right)_{I,S}, \qquad T^{-1} = \left(\frac{\partial \eta}{\partial I}\right)_{\alpha,S}, \qquad \mu = -T\left(\frac{\partial \eta}{\partial S}\right)_{I,\alpha}. \tag{1.68a,b,c}$$

We also see that $I$ and $\alpha$ (and $S$) are the natural variables for entropy.

Because the right-hand side of (1.65) is equal to an exact differential, the second derivatives are independent of the order of differentiation. That is,

$$\frac{\partial^2 I}{\partial \eta\,\partial \alpha} = \frac{\partial^2 I}{\partial \alpha\,\partial \eta}, \tag{1.69}$$

and therefore, using (1.66)

$$\left(\frac{\partial T}{\partial \alpha}\right)_{\eta} = -\left(\frac{\partial p}{\partial \eta}\right)_{\alpha}. \tag{1.70}$$

This is one of the *Maxwell relations,* which are a collection of four similar relations that follow directly from the fundamental thermodynamic relation (1.65) and simple relations between second derivatives. (Additional Maxwell-like relations exist if we consider chemical effects.) To derive the other Maxwell relations we will introduce thermodynamic potentials enthalpy, $h$, the Gibbs function, $g$, and the free energy, $f$. These are all closely related to the internal energy and they are all extensive functions (and then denoted with an uppercase letter), but for fluid-dynamical purposes it is convenient to divide them by the mass and use their specific forms, denoted with a lowercase letter (with the exception of $I$ itself, the specific internal energy).

Define the *enthalpy* of a fluid by

$$h \equiv I + p\alpha, \tag{1.71}$$

and then (1.65) becomes

$$\mathrm{d}h = T\,\mathrm{d}\eta + \alpha\,\mathrm{d}p + \mu\,\mathrm{d}S. \tag{1.72}$$

Evidently, the natural variables for enthalpy are entropy and pressure so that, in general,

$$\mathrm{d}h = \left(\frac{\partial h}{\partial \eta}\right)_{p,S}\mathrm{d}\eta + \left(\frac{\partial h}{\partial p}\right)_{\eta,S}\mathrm{d}p + \left(\frac{\partial h}{\partial S}\right)_{\eta,p}\mathrm{d}S. \tag{1.73}$$

Comparing the last two equations we have

$$T = \left(\frac{\partial h}{\partial \eta}\right)_{p,S}, \qquad \alpha = \left(\frac{\partial h}{\partial p}\right)_{\eta,S}, \qquad \mu = \left(\frac{\partial h}{\partial S}\right)_{\eta,p}. \tag{1.74}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

Noting that

$$\frac{\partial^2 h}{\partial \eta \, \partial p} = \frac{\partial^2 h}{\partial p \, \partial \eta} \tag{1.75}$$

we evidently must have

$$\left(\frac{\partial T}{\partial p}\right)_\eta = \left(\frac{\partial \alpha}{\partial \eta}\right)_p, \tag{1.76}$$

and this is our second Maxwell relation.

To obtain the third, we write

$$\mathrm{d}I = T\,\mathrm{d}\eta - p\,\mathrm{d}\alpha + \mu\,\mathrm{d}S = \mathrm{d}(T\eta) - \eta\,\mathrm{d}T - \mathrm{d}(p\alpha) + \alpha\,\mathrm{d}p + \mu\,\mathrm{d}S, \tag{1.77}$$

or

$$\mathrm{d}g = -\eta\,\mathrm{d}T + \alpha\,\mathrm{d}p + \mu\,\mathrm{d}S \quad \text{where} \quad g \equiv I - T\eta + p\alpha = h - TS. \tag{1.78a,b}$$

The quantity $g$ is the *Gibbs function,* also known as the 'Gibbs free energy' or 'Gibbs potential'. (We use $g$ for gravity and $g$ for the specific Gibbs function.) Now, formally, we have,

$$\mathrm{d}g = \left(\frac{\partial g}{\partial T}\right)_{p,S} \mathrm{d}T + \left(\frac{\partial g}{\partial p}\right)_{T,S} \mathrm{d}p + \left(\frac{\partial g}{\partial S}\right)_{T,p} \mathrm{d}S. \tag{1.79}$$

Comparing the last two equations we see that

$$\eta = -\left(\frac{\partial g}{\partial T}\right)_{p,S}, \qquad \alpha = \left(\frac{\partial g}{\partial p}\right)_{T,S}, \qquad \mu = \left(\frac{\partial g}{\partial S}\right)_{T,p}. \tag{1.80}$$

Furthermore, because

$$\frac{\partial^2 g}{\partial p \, \partial T} = \frac{\partial^2 g}{\partial T \, \partial p} \tag{1.81}$$

we have our third Maxwell equation,

$$\left(\frac{\partial \eta}{\partial p}\right)_T = -\left(\frac{\partial \alpha}{\partial T}\right)_p. \tag{1.82}$$

The Gibbs function is unique among the thermodynamic potentials in that its natural variables, $T$ and $p$, are intensive quantities.

The fourth Maxwell equation makes use of the specific free energy or Helmholtz function, $f$, where

$$f \equiv I - T\eta, \qquad \text{and} \qquad \mathrm{d}f = -\eta\,\mathrm{d}T - p\,\mathrm{d}\alpha + \mu\,\mathrm{d}S, \tag{1.83}$$

giving

$$\left(\frac{\partial \eta}{\partial \alpha}\right)_T = \left(\frac{\partial p}{\partial T}\right)_\alpha, \tag{1.84}$$

and all four of these Maxwell equations are summarized in the box on the next page. All of them follow from the fundamental thermodynamic relation, (1.65), which is the real silver hammer of thermodynamics. The fundamental relation can also be written in the following form connecting internal energy, enthalpy and entropy,

$$\boxed{\mathrm{d}I + p\,\mathrm{d}\alpha = \mathrm{d}h - \alpha\,\mathrm{d}p = T\mathrm{d}\eta + \mu\,\mathrm{d}S,} \tag{1.85}$$

which turns out to be useful for fluid dynamical applications.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

### Thermodynamic Functions and Maxwell Relations

The four canonical Maxwell relations, with their associated potentials, are

Internal energy:

$$I = I,$$
$$\mathrm{d}I = T\,\mathrm{d}\eta - p\,\mathrm{d}\alpha + \mu\,\mathrm{d}S,$$
$$\left(\frac{\partial T}{\partial \alpha}\right)_{\eta,S} = -\left(\frac{\partial p}{\partial \eta}\right)_{\alpha,S}.$$

Enthalpy:

$$h = I + p\alpha,$$
$$\mathrm{d}h = T\,\mathrm{d}\eta + \alpha\,\mathrm{d}p + \mu\,\mathrm{d}S,$$
$$\left(\frac{\partial T}{\partial p}\right)_{\eta,S} = \left(\frac{\partial \alpha}{\partial \eta}\right)_{p,S}.$$

Gibbs function:

$$g = I - T\eta + p\alpha,$$
$$\mathrm{d}g = -\eta\,\mathrm{d}T + \alpha\,\mathrm{d}p + \mu\,\mathrm{d}S,$$
$$\left(\frac{\partial \eta}{\partial p}\right)_{T,S} = -\left(\frac{\partial \alpha}{\partial T}\right)_{p,S}.$$

Helmholtz free energy:

$$f = I - T\eta,$$
$$\mathrm{d}f = -\eta\,\mathrm{d}T - p\,\mathrm{d}\alpha + \mu\,\mathrm{d}S,$$
$$\left(\frac{\partial \eta}{\partial \alpha}\right)_{T,S} = \left(\frac{\partial p}{\partial T}\right)_{\alpha,S}.$$

#### *Meaning of the state functions*

Internal energy, enthalpy, the Gibbs function, the free energy and the entropy are all state functions from which other thermodynamic quantities can be derived, but with different meanings and uses. The utility of these quantities will become apparent as we proceed, but here is a brief summary.

The internal energy of a body is the total energy within a body, excluding the kinetic energy and the potential energy due to external fields like gravity. It is an invariant if the volume is fixed and there is no heating or chemical change to the body (this is the first law). The other state functions are related to internal energy by Legendre transformations, but they are not necessarily equal to the energy that a body contains. The enthalpy is the Legendre transformation of the internal energy (a function of entropy, density and composition) to a function of entropy, pressure and composition, and it is important in fluids because the energy transfer between a fluid parcel and its environment is associated with a flux of enthalpy, not internal energy. If a parcel is adiabatically displaced its change in potential energy will be balanced by a change in enthalpy, for it is enthalpy that accounts for the work done by pressure forces. When two adjacent parcels at the same pressure mix, their total enthalpy is conserved.

The Gibbs function is useful because it is constant for systems at constant temperature, pressure and composition. Its natural variables, temperature, pressure and composition are all measurable and for that reason it finds use as the fundamental state function from which all other thermodynamic variables may be derived. The Helmholtz free energy is also sometimes used as a fundamental state variable, and is useful for systems at constant temperature, density and composition for then it is constant. For small isothermal and isohaline changes, the increase of free energy is equal to the work done on the system. Free energy is not in fact commonly used in the atmospheric or oceanic sciences. Finally, whole books have been written on entropy (which is not a Legendre transformation of the internal energy and is not a thermodynamic potential in the same sense as the others). Suffice it to say here that entropy is the state function that responds directly to heating and it is a measure of the disorder of a system.[5]

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

### Fundamental equation of state

The fundamental equation of state (1.59) gives complete information about a fluid in thermodynamic equilibrium, and given this we can obtain expressions for the temperature, pressure and chemical potential using (1.66). These expressions are also equations of state; however, each of them, taken individually, contains less information than the fundamental equation because a derivative has been taken. Equivalent to the fundamental equation of state are, using (1.72), an expression for the enthalpy as a function of its natural variables pressure, entropy and composition, or, using (1.78), the Gibbs function as a function of pressure, temperature and composition. Of these, the Gibbs function is particularly useful in practice because the pressure, temperature and composition may all be measured in the laboratory. Given the fundamental equation of state, the thermodynamic state of a body is fully specified by a knowledge of any two of $p, \rho, T, \eta$ and $I$, plus its composition. The thermal equation of state, (1.52), is obtained by using (1.59a) to eliminate entropy from (1.66a) and (1.66b).

One simple fundamental equation of state is to take the internal energy to be a function of density and not entropy; that is, $I = I(\alpha)$. Bodies with such a property are called *homentropic*. Using (1.66), temperature and chemical potential have no role in the fluid dynamics and the density is a function of pressure alone — the defining property of a barotropic fluid.[6] Neither water nor air is, in general, homentropic but under some circumstances the flow may be adiabatic and $p = p(\rho)$.

In an ideal gas the molecules do not interact except by elastic collisions, and the volume of the molecules is presumed to be negligible compared to the total volume they occupy. The internal energy of the gas then depends only on temperature, and not on density. A *simple* ideal gas, also called a *perfect* gas (although nomenclature in the literature varies), is an ideal gas for which the heat capacity is constant, so that

$$I = cT, \tag{1.86}$$

where $c$ is a constant. Using this and the conventional ideal gas equation, $p = \rho RT$ (where $R$ is also constant), along with the fundamental thermodynamic relation (1.65), we can infer the fundamental equation of state; however, we will defer that until we discuss potential temperature in Section 1.6.1 — if curious, look ahead to page 25. A general ideal gas also obeys $p = \rho RT$, but has heat capacities that may be a function of temperature (but only of temperature, if composition is fixed), but in this book we only deal with simple ideal gases.

### Internal energy and specific heats

We can obtain some useful relations between the internal energy and specific heat capacities, and some useful estimates of their values, by some simple manipulations of the fundamental thermodynamic relation. Assuming that the composition of the fluid is constant, (1.65) is

$$T\,\mathrm{d}\eta = \mathrm{d}I + p\,\mathrm{d}\alpha, \tag{1.87}$$

so that, taking $I$ to be a function of $\alpha$ and $T$,

$$T\,\mathrm{d}\eta = \left(\frac{\partial I}{\partial T}\right)_{\alpha,S}\mathrm{d}T + \left[\left(\frac{\partial I}{\partial \alpha}\right)_{T,S} + p\right]\mathrm{d}\alpha. \tag{1.88}$$

From this, we see that the heat capacity at constant volume (i.e., constant $\alpha$) $c_v$ is given by

$$c_v \equiv T\left(\frac{\partial \eta}{\partial T}\right)_{\alpha,S} = \left(\frac{\partial I}{\partial T}\right)_{\alpha,S}. \tag{1.89}$$

Thus, $c$ in (1.86) is equal to $c_v$.

Similarly, using (1.72) we have

$$T\,\mathrm{d}\eta = \mathrm{d}h - \alpha\,\mathrm{d}p = \left(\frac{\partial h}{\partial T}\right)_{p,S}\mathrm{d}T + \left[\left(\frac{\partial h}{\partial p}\right)_{T,S} - \alpha\right]\mathrm{d}p. \tag{1.90}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

The heat capacity at constant pressure, $c_p$, is then given by

$$c_p \equiv T\left(\frac{\partial \eta}{\partial T}\right)_{p,S} = \left(\frac{\partial h}{\partial T}\right)_{p,S}. \tag{1.91}$$

For an ideal gas $h = I + RT = T(c_v + R)$. But $c_p = (\partial h/\partial T)_p$, and hence $c_p = c_v + R$. For future use we define $\gamma \equiv c_p/c_v$ and $\kappa \equiv R/c_p$, and $(\gamma - 1)/\gamma = \kappa$. Statistical mechanics tells us that for a simple ideal gas the internal energy is equal to $kT/2$ per molecule, or $RT/2$ per unit mass, for each excited degree of freedom, where $k$ is the Boltzmann constant and $R$ the gas constant. The diatomic molecules $N_2$ and $O_2$ that form most of our atmosphere have two rotational and three translational degrees of freedom, so that $I \approx 5RT/2$, and so $c_v \approx 5R/2$ and $c_p \approx 7R/2$, both being constants. These are in fact very good approximations to the measured values for the Earth's atmosphere, and give $c_v \approx 714\,\mathrm{J\,kg^{-1}K^{-1}}$ and $c_p \approx 1000\,\mathrm{J\,kg^{-1}K^{-1}}$ (whereas $c_p$ is measured to be $1003\,\mathrm{J\,kg^{-1}K^{-1}}$). The internal energy is simply $c_vT$ and the enthalpy is $c_pT$. For a liquid, especially one like seawater that contains dissolved salts, no such simple relations are possible: the heat capacities are functions of the state of the fluid, and the internal energy is a function of pressure (or density) as well as temperature and composition.

## 1.6 THERMODYNAMIC EQUATIONS FOR FLUIDS

The thermodynamic relations — for example (1.65) — apply to identifiable bodies or systems; thus, the heat input affects the fluid parcel to which it is applied, and we can apply the material derivative to the above thermodynamic relations to obtain equations of motion for a moving fluid. In doing so we make two assumptions:

*(i)* That locally the fluid is in thermodynamic equilibrium. This means that, although the thermodynamic quantities like temperature, pressure and density vary in space and time, locally they are related by the thermodynamic relations such as the equation of state and Maxwell's relations.

*(ii)* That macroscopic fluid motions are reversible and so not entropy producing. Thus, such effects as the viscous dissipation of energy, radiation, and conduction may produce entropy whereas the macroscopic fluid motion itself does not.

The first point requires that the temperature variation on the macroscopic scales must be slow enough that there can exist a volume that is small compared to the scale of macroscopic variations, so that temperature is effectively constant within it, but that is also sufficiently large to contain enough molecules so that macroscopic variables such as temperature have a proper meaning.

Now from (1.61), conservation of energy for an infinitesimal fluid parcel may be written as

$$\mathrm{d}I = -p\,\mathrm{d}\alpha + đQ_E, \tag{1.92}$$

where $p\,\mathrm{d}\alpha$ is the work done by the parcel and $đQ_E$ is the total energy input to the parcel with contributions from heating and changes in composition. Given the first assumption above, we may form the material derivative of (1.92) to obtain

$$\frac{\mathrm{D}I}{\mathrm{D}t} + p\frac{\mathrm{D}\alpha}{\mathrm{D}t} = \dot{Q}_E, \tag{1.93}$$

where $\dot{Q}_E$ is the rate of total energy input, per unit mass, with (as for $đQ_E$) possible contributions from thermal fluxes (including radiative heating, thermal diffusion and heat generated by viscous damping) and fluxes of composition, and note that $\dot{Q}_E$ does not include any mechanical effects. (In general, both the diffusion of heat and composition depend on the gradients of both temperature and composition, although the thermal flux is largely determined by the temperature gradient, and the compositional flux by the gradient of composition.)

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

Using the mass continuity equation in the form $\mathrm{D}\alpha/\mathrm{D}t = \alpha\nabla\cdot\boldsymbol{v}$, we write (1.93) as

$$\boxed{\frac{\mathrm{D}I}{\mathrm{D}t} + p\alpha\nabla\cdot\boldsymbol{v} = \dot{Q}_E.} \tag{1.94}$$

This is the *internal energy equation* for a fluid. Internal energy is not a conservative variable because of the compression term involving $\nabla\cdot\boldsymbol{v}$. The internal energy equation may also be written in terms of enthalpy, and using (1.71) and (1.93) we obtain the equivalent equation,

$$\frac{\mathrm{D}h}{\mathrm{D}t} - \alpha\frac{\mathrm{D}p}{\mathrm{D}t} = \dot{Q}_E. \tag{1.95}$$

Because $\dot{Q}_E$ contains, in general, energy fluxes due to changes in composition, we need to know that composition. The composition of a fluid parcel is carried with it as it moves, and changes only if there are non-conservative sources and sinks, such as diffusive fluxes. Thus, and analogously to (1.41), the evolution of composition is determined by

$$\frac{\mathrm{D}S}{\mathrm{D}t} = \dot{S}, \tag{1.96}$$

where $\dot{S}$ represents all the non-conservative terms.

Rather than use an internal energy equation, we may use the fundamental thermodynamic relation to infer an evolution equation for entropy. Thus, forming the material derivative from (1.65) and using (1.94) and (1.96), we obtain the *entropy equation*

$$\boxed{\frac{\mathrm{D}\eta}{\mathrm{D}t} = \frac{1}{T}\dot{Q}_E - \frac{\mu}{T}\dot{S} \equiv \frac{1}{T}\dot{Q},} \tag{1.97}$$

where $\dot{Q}$ is the heating rate per unit mass. This equation is simply the material derivative of (1.62), along with assumption *(ii)* above. The heating of a fluid parcel generally needs to be *derived,* since it involves the viscous dissipation of energy as well as radiative and diffusive fluxes. The derivation is the topic of Appendix A at the chapter end, and from now on we assume the heating and energy input are known quantities.

The entropy equation is not independent of the internal energy equation, but is connected via the thermodynamic relations of Sections 1.5.1 and 1.5.2, and by the equation of state. If we use the internal energy equation we can in principle then calculate the entropy using the equation of state in the form $\eta = \eta(I, \alpha, S)$, or if we use the entropy equation we can then calculate the internal energy using $I = I(\eta, \alpha, S)$. Indeed, the internal energy equation and the entropy equation are both commonly referred to as 'the thermodynamic equation'. The heating term, $\dot{Q}$, is, however, not always easy to accurately determine in practice — it is affected by gradients of composition as well as viscosity and radiative fluxes — and the use of the internal energy equation may be more straightforward than the use of an entropy equation, especially in multi-component fluids. On the other hand, internal energy is affected by the $\nabla\cdot\boldsymbol{v}$ term in (1.94) and in a liquid this is small but non-zero and that may cause difficulties. See Section 1.7.3 for more discussion.

In any case, given evolution equations for composition and internal energy or entropy, and given the fundamental equation of state, we have, in principle, a complete set of equations for a fluid, as summarized in the shaded box on the facing page. Let us now look at the special, but very important, case of a dry ideal gas.

## Fundamental Equations of Motion of a Fluid

The following equations constitute, in principle, a complete set of equations for a fluid heated at a rate $\dot{Q}$ and whose composition, $S$, changes at a rate $\dot{S}$.

*Evolution equations for velocity, density and composition*

$$\frac{D\boldsymbol{v}}{Dt} = -\frac{1}{\rho}\nabla p + \nu\nabla^2\boldsymbol{v} + \boldsymbol{F}, \qquad \frac{D\rho}{Dt} + \rho\nabla\cdot\boldsymbol{v} = 0, \qquad \frac{DS}{Dt} = \dot{S}, \tag{F.1}$$

where $\boldsymbol{F}$ is some body force per unit mass, such as gravity.

*Internal energy equation or entropy equation (the 'thermodynamic' equation)*

$$\frac{DI}{Dt} + \frac{p}{\rho}\nabla\cdot\boldsymbol{v} = \dot{Q}_E \qquad \text{or} \qquad \frac{D\eta}{Dt} = \frac{1}{T}\dot{Q}, \tag{F.2}$$

where $\dot{Q}$ is the heating and $\dot{Q}_E = \dot{Q} + \mu\dot{S}$ is the total rate of energy input.

*Fundamental equation of state for internal energy, $I$, entropy, $\eta$, or Gibbs function, $g$*

$$I = I(\rho,\eta,S), \quad \eta = \eta(\rho,I,S) \quad \text{or} \quad g = g(T,p,S). \tag{F.3}$$

*Diagnostic equations for temperature and pressure*

$$T = \left(\frac{\partial I}{\partial \eta}\right)_{\alpha,S} \quad \text{or} \quad T^{-1} = \left(\frac{\partial \eta}{\partial I}\right)_{\alpha,S}. \tag{F.4}$$

$$p = -\left(\frac{\partial I}{\partial \alpha}\right)_{\eta,S} = \rho^2\left(\frac{\partial I}{\partial \rho}\right)_{\eta,S} \quad \text{or} \quad p = T\left(\frac{\partial \eta}{\partial \alpha}\right)_{I,S}. \tag{F.5}$$

The actual method of solution of these equations will depend on the equation of state. For example, for an ideal gas $T = I/c_v$ and the thermal equation of state, $p = \rho RT$, may be used to infer pressure.

The equations describing fluid motion are called the *Euler equations* if the viscous term is omitted, and the *Navier–Stokes equations* if viscosity is included.[7] These appellations are often taken to mean only the momentum and mass conservation equations, and the Euler equations are sometimes taken to be the equations for a fluid of constant density.

### 1.6.1 Thermodynamic Equation for an Ideal Gas

For a dry ideal gas the internal energy is a function of temperature only and $dI = c_v\,dT$. The first law of thermodynamics becomes

$$đQ = c_v\,dT + p\,d\alpha, \qquad \text{or} \qquad đQ = c_p\,dT - \alpha\,dp, \tag{1.98a,b}$$

where the second expression is derived using $\alpha = RT/p$ and $c_p - c_v = R$. Forming the material derivative of the above gives two forms of the internal energy equation:

$$c_v\frac{DT}{Dt} + p\frac{D\alpha}{Dt} = \dot{Q} \qquad \text{or} \qquad c_p\frac{DT}{Dt} - \frac{RT}{p}\frac{Dp}{Dt} = \dot{Q}. \tag{1.99a,b}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

Using the mass continuity equation, (1.99a) may be written as

$$c_v \frac{\mathrm{D}T}{\mathrm{D}t} + p\alpha\nabla\cdot\boldsymbol{v} = \dot{Q}. \tag{1.100}$$

This is one of the most common and useful forms of the thermodynamic equation for the atmosphere. A less common but equivalent form arises if we use the ideal gas equation to eliminate $T$ in favour of $p$, giving

$$\frac{\mathrm{D}p}{\mathrm{D}t} + \gamma p\nabla\cdot\boldsymbol{v} = \dot{Q}\frac{\rho R}{c_v}. \tag{1.101}$$

The Earth's atmosphere also contains water vapour with mixing ratio $q$ (commonly referred to as specific humidity), and an evolution equation for it takes the form

$$\frac{\mathrm{D}q}{\mathrm{D}t} = \dot{q}, \tag{1.102}$$

where $\dot{q}$ represents the effects of condensation and evaporation. The main thermodynamic effects of water vapour occur when it condenses and latent heat is released; this heating appears in the thermodynamic equation, with $\dot{Q} = -L\dot{q}$, where $L$ is the latent heat of condensation, as discussed in Chapter 18.

### *Potential temperature, potential density and entropy*

We can use entropy instead of temperature for our thermodynamic equation, and this corresponds to using (1.97) instead of (1.94). However, it is common in meteorology to express entropy in terms of a temperature-like quantity, potential temperature, which has a more intuitive appeal to some.[8] Seawater also has a potential temperature variable, but with a different form.

We begin with the observation that when a fluid parcel changes pressure adiabatically, it will expand or contract and, using (1.98b) with $đQ = 0$, its temperature change is determined by $c_p\,\mathrm{d}T = \alpha\,\mathrm{d}p$. This temperature change is plainly not caused by heating, but we may construct a temperature-like quantity that changes *only* if diabatic effects are present; specifically *potential temperature,* $\theta$, is defined to be the temperature that a fluid would have if moved adiabatically and at constant composition to some reference pressure (usually taken to be 1000 hPa, which is close to the pressure at the Earth's surface). Thus, in adiabatic flow the potential temperature obeys $\mathrm{D}\theta/\mathrm{D}t = 0$.

In order to relate $\theta$ to the other thermodynamic variables we use (1.98b) and the equation of state for an ideal gas to write the fundamental thermodynamic relation as

$$\mathrm{d}\eta = c_p\,\mathrm{d}\ln T - R\,\mathrm{d}\ln p. \tag{1.103}$$

If we move a parcel adiabatically ($\mathrm{d}\eta = 0$) from $p$ to $p_R$ the temperature changes, by definition, from $T$ to $\theta$, and (1.103) gives

$$\int_T^\theta c_p\,\mathrm{d}\ln T - \int_p^{p_R} R\,\mathrm{d}\ln p = 0. \tag{1.104}$$

For constant $c_p$ and $R$ this equation may be solved to give

$$\theta = T\left(\frac{p_R}{p}\right)^\kappa, \tag{1.105}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

where $p_R$ is the reference pressure and $\kappa \equiv R/c_p$. (Another derivation of this result is given in Appendix A.) It follows from (1.103) and (1.105) that potential temperature is related to entropy by

$$\mathrm{d}\eta = c_p \,\mathrm{d}\ln\theta, \tag{1.106}$$

and, if $c_p$ is constant, which it nearly is for Earth's atmosphere,

$$\eta = c_p \ln\theta. \tag{1.107}$$

Forming the material derivative of (1.107) and using (1.97) we obtain

$$\boxed{c_p \frac{\mathrm{D}\theta}{\mathrm{D}t} = \frac{\theta}{T}\dot{Q},} \tag{1.108}$$

with $\theta$ given by (1.105). Equations (1.100), (1.101) and (1.108) are all equivalent forms of the thermodynamic equation for an ideal gas.

The *potential density*, $\rho_\theta$, is the density that a fluid parcel would have if moved adiabatically and at constant composition to a reference pressure, $p_R$. If the equation of state is written as $\rho = f(p,T)$ then the potential density is just $\rho_\theta = f(p_R,\theta)$ and for an ideal gas we therefore have

$$\rho_\theta = \frac{p_R}{R\theta} = \rho\left(\frac{p_R}{p}\right)^{1/\gamma}. \tag{1.109}$$

Finally, for later use, a little manipulation of the equation of state for an ideal gas, for small variations around a reference state, reveals that

$$\frac{\delta\theta}{\theta} = \frac{\delta T}{T} - \kappa\frac{\delta p}{p} = \frac{1}{\gamma}\frac{\delta p}{p} - \frac{\delta\rho}{\rho}. \tag{1.110}$$

*The fundamental equation of state for an ideal gas*

Equations (1.107) and (1.105) are closely related to the fundamental equation of state, and using $I = c_v T$ and the equation of state $p = \rho RT$, we can express the entropy explicitly in terms of the density and the internal energy. Similarly, we may derive an expression for Gibbs function as a function of pressure and temperature, and we find

$$\boxed{\eta = c_v \ln I - R\ln\rho + A, \qquad g = c_p T(1-\ln T) + RT\ln p + BT + C,} \tag{1.111a,b}$$

where $A$, $B$ and $C$ are constants. Either of these two expressions may be regarded as *the fundamental equation of state for a simple ideal gas.* They could in fact be used to define a simple ideal gas (although we have not motivated that approach) and if we were to *begin* with either of them we could derive all the other thermodynamic quantities of interest. For example, using (1.68a) and (1.111a) we immediately recover $p = \rho RT$, and using (1.68b) we obtain $T = c_v/I$. Similarly, from the Gibbs function we obtain $\alpha = (\partial g/\partial p)_T = RT/p$, and the entropy satisfies $\eta = -(\partial g/\partial T)_p$. We provide a more complete set of derivations from the Gibbs function in Appendix A on page 47 and, for moist air, the appendix on page 720.

### 1.6.2 ✦ Other Forms of the Thermodynamic Equation

For a liquid such as seawater no simple exact equation of state exists and writing down a useful thermodynamic equation is not easy. Thus, although (1.97) holds in general, we need to be able to evaluate the heating and we need an expression relating entropy to the other thermodynamic variables — that is, an equation of state. For quantitative modelling and observational work such an equation of state must be quite accurate and we come back to this in Section 1.7. However, we can gain understanding — for both liquids and gases — by beginning with the entropy equation and making simplifications to it, as follows.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

### Forms of the Thermodynamic Equation

*General form*

$$\frac{\mathrm{D}I}{\mathrm{D}t} + p\frac{\mathrm{D}\alpha}{\mathrm{D}t} = \dot{Q}_E \qquad \text{or} \qquad \frac{\mathrm{D}I}{\mathrm{D}t} + p\alpha\nabla\cdot\boldsymbol{v} = \dot{Q}_E, \tag{T.1a,b}$$

where $I$ is the internal energy and $\dot{Q}_E$ is the rate of energy input, per unit mass. This may be written in terms of enthalpy, $h$, or entropy, $\eta$

$$\frac{\mathrm{D}h}{\mathrm{D}t} - \alpha\frac{\mathrm{D}p}{\mathrm{D}t} = \dot{Q}_E, \qquad T\frac{\mathrm{D}\eta}{\mathrm{D}t} = \dot{Q}_E - \mu\dot{S} = \dot{Q}, \tag{T.2a,b}$$

where $\dot{Q}$ is the heating rate and $\dot{S}$ the rate of change of composition. For a fluid parcel of constant composition, $c_p \,\mathrm{d}\ln\theta = \mathrm{d}\eta$ and (T.2b) may be written as a potential temperature equation.

*Ideal gas*

For an ideal gas $\mathrm{d}I = c_v\,\mathrm{d}T$, $\mathrm{d}h = c_p\mathrm{d}T$, $\mathrm{d}\eta = c_p\mathrm{d}\ln\theta$ and the adiabatic thermodynamic equation may be written in the following equivalent, exact, forms:

$$\begin{aligned} c_p\frac{\mathrm{D}T}{\mathrm{D}t} - \alpha\frac{\mathrm{D}p}{\mathrm{D}t} &= 0, &\qquad \frac{\mathrm{D}p}{\mathrm{D}t} + \gamma p\nabla\cdot\boldsymbol{v} &= 0, \\ c_v\frac{\mathrm{D}T}{\mathrm{D}t} + p\alpha\nabla\cdot\boldsymbol{v} &= 0, &\qquad \frac{\mathrm{D}\theta}{\mathrm{D}t} &= 0, \end{aligned} \tag{T.3}$$

where $\theta = T(p_R/p)^\kappa$, and energy or heating terms in various forms appear on the right-hand sides as needed. The two expressions on the second line are usually the most useful in modelling and theoretical work in meteorology.

*I. Thermodynamic equation using pressure and density*

If we regard $\eta$ as a function of pressure and density, and salinity $S$ where appropriate, we obtain

$$\begin{aligned} T\,\mathrm{d}\eta &= T\left(\frac{\partial\eta}{\partial\rho}\right)_{p,S}\mathrm{d}\rho + T\left(\frac{\partial\eta}{\partial p}\right)_{\rho,S}\mathrm{d}p + T\left(\frac{\partial\eta}{\partial S}\right)_{\rho,p}\mathrm{d}S \\ &= T\left(\frac{\partial\eta}{\partial\rho}\right)_{p,S}\mathrm{d}\rho - T\left(\frac{\partial\eta}{\partial\rho}\right)_{p,S}\left(\frac{\partial\rho}{\partial p}\right)_{\eta,S}\mathrm{d}p + T\left(\frac{\partial\eta}{\partial S}\right)_{\rho,p}\mathrm{d}S. \end{aligned} \tag{1.112}$$

Forming the material derivative, and using (1.97) and (1.96), we obtain for a moving fluid

$$T\left(\frac{\partial\eta}{\partial\rho}\right)_{p,S}\frac{\mathrm{D}\rho}{\mathrm{D}t} - T\left(\frac{\partial\eta}{\partial\rho}\right)_{p,S}\left(\frac{\partial\rho}{\partial p}\right)_{\eta,S}\frac{\mathrm{D}p}{\mathrm{D}t} = \dot{Q} - T\left(\frac{\partial\eta}{\partial S}\right)_{\rho,p}\dot{S}. \tag{1.113}$$

But $(\partial p/\partial\rho)_{\eta,S} = c_s^2$ where $c_s$ is the speed of sound (see Section 1.8). This is a measurable quantity in a fluid, and often nearly constant, and so useful to keep in an equation. The thermodynamic equation may then be written in the form

$$\boxed{\frac{\mathrm{D}\rho}{\mathrm{D}t} - \frac{1}{c_s^2}\frac{\mathrm{D}p}{\mathrm{D}t} = Q_{[\rho,p]},} \tag{1.114}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

where $Q_{[\rho,p]} \equiv (\partial\rho/\partial\eta)_{p,S}\dot{Q}/T + (\partial\rho/\partial S)_{\eta,p}\dot{S}$ represents the effects of entropy and salinity source terms. This form of the thermodynamic equation is valid for both liquids and gases.

### Approximations using pressure and density

The speed of sound in a fluid is related to its compressibility — the less compressible the fluid, the greater the sound speed. In a liquid, sound speed is often sufficiently high that the second term in (1.114) can be neglected, and the thermodynamic equation takes the simple form:

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = Q_{[\rho,p]}. \tag{1.115}$$

The above equation is a very good approximation for many laboratory fluids. It is a *thermodynamic* equation, arising from the principle of conservation of energy for a liquid; it is a very different equation from the mass conservation equation, which for compressible fluids is also an evolution equation for density.

In the ocean the enormous pressures resulting from columns of seawater kilometres deep mean that although the second term in (1.114) may be small, it is not negligible, and a better approximation results if we suppose that the pressure is given by the weight of the fluid above it — the hydrostatic approximation. In this case $\mathrm{d}p = -\rho g\,\mathrm{d}z$ and (1.114) becomes

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \frac{\rho g}{c_s^2}\frac{\mathrm{D}z}{\mathrm{D}t} = Q_{[\rho,p]}. \tag{1.116}$$

In the second term the height field varies much more than the density field, so a good approximation is to replace $\rho$ by a constant, $\rho_0$, in this term only. Taking the speed of sound also to be constant gives

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\rho + \frac{\rho_0 z}{H_\rho}\right) = Q_{[\rho,p]}, \qquad \text{where} \qquad H_\rho = c_s^2/g. \tag{1.117a,b}$$

$H_\rho$ is the *density scale height* of the ocean. In water, $c_s \approx 1500\,\mathrm{m\,s^{-1}}$ so that $H_\rho \approx 200\,\mathrm{km}$. The quantity in brackets on the left-hand side of (1.117a) is (in this approximation) the *potential density*, this being the density that a parcel would have if moved adiabatically and with constant composition to the reference height $z = 0$. The adiabatic lapse rate of density is the rate at which the density of a parcel changes when undergoing an adiabatic displacement. From (1.117) it is approximately

$$-\left(\frac{\partial\rho}{\partial z}\right)_\eta \approx \frac{\rho_0 g}{c_s^2} \approx 5\,(\mathrm{kg\,m^{-3}})/\mathrm{km}, \tag{1.118}$$

so that if a parcel is moved adiabatically from the surface to the deep ocean (5 km depth, say) its density will increase by about $25\,\mathrm{kg\,m^{-3}}$, a fractional change of about 1/40 or 2.5%.

### II. Thermodynamic equation using pressure and temperature

Taking entropy to be a function of pressure and temperature (and salinity if appropriate), we have

$$\begin{aligned} T\,\mathrm{d}\eta &= T\left(\frac{\partial\eta}{\partial T}\right)_{p,S}\mathrm{d}T + T\left(\frac{\partial\eta}{\partial p}\right)_{T,S}\mathrm{d}p + T\left(\frac{\partial\eta}{\partial S}\right)_{T,p}\mathrm{d}S \\ &= c_p\,\mathrm{d}T + T\left(\frac{\partial\eta}{\partial p}\right)_{T,S}\mathrm{d}p + T\left(\frac{\partial\eta}{\partial S}\right)_{T,p}\mathrm{d}S. \end{aligned} \tag{1.119}$$

For a moving fluid, and using (1.97) and (1.96), this implies,

$$\frac{\mathrm{D}T}{\mathrm{D}t} + \frac{T}{c_p}\left(\frac{\partial\eta}{\partial p}\right)_{T,S}\frac{\mathrm{D}p}{\mathrm{D}t} = Q_{[T,p]}, \tag{1.120}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

where $Q_{[T,p]} \equiv \dot{Q}/c_p - Tc_p^{-1}\dot{S}(\partial\eta/\partial S)$. Now substitute the Maxwell relation (1.82) in the form

$$\left(\frac{\partial\eta}{\partial p}\right)_T = \frac{1}{\rho^2}\left(\frac{\partial\rho}{\partial T}\right)_p \tag{1.121}$$

to give

$$\frac{\mathrm{D}T}{\mathrm{D}t} + \frac{T}{c_p\rho^2}\left(\frac{\partial\rho}{\partial T}\right)_p \frac{\mathrm{D}p}{\mathrm{D}t} = Q_{[T,p]}, \qquad \text{or} \qquad \frac{\mathrm{D}T}{\mathrm{D}t} - \frac{T}{c_p}\left(\frac{\partial\alpha}{\partial T}\right)_p \frac{\mathrm{D}p}{\mathrm{D}t} = Q_{[T,p]}. \tag{1.122a,b}$$

The density and temperature are related through a coefficient of thermal expansion $\beta_T$ where

$$\left(\frac{\partial\rho}{\partial T}\right)_p = -\beta_T\rho. \tag{1.123}$$

Equation (1.122) then becomes

$$\boxed{\frac{\mathrm{D}T}{\mathrm{D}t} - \frac{\beta_T T}{c_p\rho}\frac{\mathrm{D}p}{\mathrm{D}t} = Q_{[T,p]}\,.} \tag{1.124}$$

This form of the thermodynamic equation is valid for both liquids and gases. In an ideal gas we have $\beta_T = 1/T$, whereas in a liquid $\beta_T$ is usually quite small.

### *Approximations using pressure and temperature*

In the hydrostatic approximation we suppose that the pressure in (1.124) varies according only to the weight of the fluid above it. Then $\mathrm{d}p = -\rho g\,\mathrm{d}z$ and (1.124) becomes

$$\frac{1}{T}\frac{\mathrm{D}T}{\mathrm{D}t} + \frac{\beta_T g}{c_p}\frac{\mathrm{D}z}{\mathrm{D}t} = \frac{Q_{[T,p]}}{T}. \tag{1.125}$$

For an ideal gas we have $\beta_T = 1/T$, whence, if $c_p$ is constant,

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(c_pT + gz\right) = c_pQ_{[T,p]}. \tag{1.126}$$

The quantity $c_pT + gz$ is known as the dry static energy and we will encounter it again throughout the book. The above equation is closely related to the potential temperature form of the thermodynamic equation, since in hydrostatic balance a little manipulation reveals that

$$\frac{T}{\theta}\frac{\partial\theta}{\partial z} = \frac{\partial T}{\partial z} + \frac{g}{c_p}. \tag{1.127}$$

The quantity $T + gz/c_p$ is a height form of potential temperature for an ideal gas in hydrostatic balance, being the temperature that a fluid at a level $z$ and temperature $T$ would have if moved adiabatically to a reference level of $z = 0$.

If $\beta_T$ is constant, which is a fair approximation for many liquids, then for small variations of temperature around the value $T_0$, (1.125) simplifies to

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(T + \frac{T_0 z}{H_T}\right) = Q_{[T,p]}, \qquad \text{where} \qquad H_T = \frac{c_p}{\beta_T g}. \tag{1.128a,b}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

The quantity $H_T$ is the *temperature scale height* of the fluid, and with the oceanic values $\beta_T \approx 2 \times 10^{-4}\,\mathrm{K}^{-1}$ and $c_p \approx 4 \times 10^3\,\mathrm{J\,kg^{-1}\,K^{-1}}$ we obtain $H_T \approx 2000\,\mathrm{km}$. The field $T + T_0 z/H_T$ is a height form of potential temperature for liquids; that is,

$$\theta \approx T + \frac{\beta_T g T_0}{c_p} z. \tag{1.129}$$

The temperature changes because of the work done by or on the fluid parcel as it expands or is compressed. In seawater the expansion coefficient $\beta_T$ and $c_p$ are functions of pressure and (1.129) is not good enough if high accuracy is required, whereas in a laboratory setting we can often simply neglect the term involving $\beta_T$.

The adiabatic lapse rate of temperature is the rate at which the temperature of a parcel changes in the vertical when undergoing an adiabatic displacement. From (1.125) it is

$$\Gamma_z \equiv -\left(\frac{\partial T}{\partial z}\right)_\eta = \frac{T g \beta_T}{c_p}. \tag{1.130}$$

In general $\Gamma_z$ is a function of temperature, salinity and pressure, but it is a calculable quantity if $\beta_T$ is known and, with the oceanic values above, it is approximately $0.15\,\mathrm{K\,km^{-1}}$. Equation (1.130) is not accurate enough for quantitative oceanography because the expansion coefficient is a function of pressure; nor is it a good measure of stability, because of the effects of salt. In a dry atmosphere the ideal gas relationship gives $\beta_T = 1/T$ and so

$$\Gamma_z = \frac{g}{c_p}, \tag{1.131}$$

which is approximately $10\,\mathrm{K\,km^{-1}}$. The only approximation involved in deriving this is the use of the hydrostatic relationship.

It is noteworthy that the scale heights given by (1.117b) and (1.128b) differ so much. The first is due to the pressure compressibility of seawater [and so related to $c_s^2$, or $\beta_p$ in (1.57)] whereas the second is due to the change of density with temperature [$\beta_T$ in (1.57)], and is the distance over which the difference between temperature and potential temperature changes by an amount equal to the temperature itself (i.e., by about 273 K). The two heights differ so much because the value of the thermal expansion coefficient is not directly related to the pressure compressibility — for example, fresh water at 4° C has zero thermal expansivity, and so would have an infinite temperature scale height, but its pressure compressibility differs little from water at 20° C.

### *III. Thermodynamic equation using density and temperature*

Taking entropy to be a function of density and temperature (and salinity if appropriate) we have

$$\begin{aligned} T\,\mathrm{d}\eta &= T\left(\frac{\partial \eta}{\partial T}\right)_{\alpha,S} \mathrm{d}T + T\left(\frac{\partial \eta}{\partial \alpha}\right)_{T,S} \mathrm{d}\alpha + T\left(\frac{\partial \eta}{\partial S}\right)_{T,\alpha} \mathrm{d}S \\ &= c_v\,\mathrm{d}T + T\left(\frac{\partial \eta}{\partial \alpha}\right)_{T,S} \mathrm{d}\alpha + T\left(\frac{\partial \eta}{\partial S}\right)_{T,\alpha} \mathrm{d}S. \end{aligned} \tag{1.132}$$

For a moving fluid this implies,

$$\frac{\mathrm{D}T}{\mathrm{D}t} + \frac{T}{c_v}\left(\frac{\partial \eta}{\partial \alpha}\right)_{T,S} \frac{\mathrm{D}\alpha}{\mathrm{D}t} = Q_{[T,\alpha]}, \tag{1.133}$$

where $Q_{[T,\alpha]} \equiv \dot{Q}/c_v - T c_v^{-1} \dot{S}(\partial\eta/\partial S)$.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

For an ideal gas, and using (1.84), which is one of Maxwell's relations, (1.133) may be written as

$$c_v \frac{\mathrm{D}T}{\mathrm{D}t} + p\frac{\mathrm{D}\alpha}{\mathrm{D}t} = \dot{Q}, \tag{1.134}$$

so recovering the internal energy equation (1.99a). On the other hand, for a liquid of nearly constant density, the second term on the left-hand side of (1.133) is small, and $c_p \approx c_v$, and we have to a first approximation $\mathrm{D}T/\mathrm{D}t = Q_{[T,\alpha]}$.

Various forms of the thermodynamic equation are summarized in the box on page 26. For an ideal gas, (1.114) and (1.124) are exactly equivalent to (1.100) or (1.101), and numerical models of an ideal gas usually use either a prognostic equation for internal energy $c_vT$ or potential temperature. A discussion of an accurate thermodynamic equation for seawater is given below, with some summary remarks in the box on the next page.

## 1.7 ✦ THERMODYNAMICS OF SEAWATER

We now discuss the thermodynamics of liquids such as seawater in a little more detail. Readers whose interest is mainly in an ideal gas may skim this section. We begin with a phenomenological discussion of potential temperature and potential density followed by a more accurate treatment of the equation of state.

### 1.7.1 Potential Temperature, Potential Density and Entropy

*Potential temperature and entropy*

The potential temperature is defined as the temperature that a parcel would have if moved adiabatically and at constant composition to a given reference pressure $p_R$, often taken as $10^5$ Pa (or 1000 hPa, or 1000 mb, approximately the pressure at the sea-surface). Thus it may be calculated, at least in principle, by the integral

$$\theta(S,T,p,p_R) = T + \int_p^{p_R} \Gamma'_{ad}(S,T,p')\,\mathrm{d}p', \tag{1.135}$$

where $\Gamma'_{ad} = (\partial T/\partial p)_{\eta,S}$. Such integrals may be hard to calculate, and if we know the equation of state in the form $\eta = \eta(S,T,p)$ then we can calculate potential temperature more directly because potential temperature must satisfy

$$\eta(S,T,p) = \eta(S,\theta,p_R). \tag{1.136}$$

Solving this equation for $\theta$ gives, in principle, $\theta = \theta(\eta, S, p_R) = \theta(T,S,p)$, and examples will be given in (1.152) and in Appendix A. Potential temperature is not a materially conserved variable in the presence of salinity changes.

For a parcel of constant composition, changes in entropy are directly related to changes in potential temperature because, from the right-hand side of (1.136),

$$\mathrm{d}\eta = \left(\frac{\partial\eta(S,\theta,p_R)}{\partial\theta}\right)_S \mathrm{d}\theta. \tag{1.137}$$

Thus, if a fluid parcel moves adiabatically and at constant composition then $\mathrm{d}\eta = 0$ and $\mathrm{d}\theta = 0$. Furthermore, if we express entropy as a function of temperature and pressure then

$$T\,\mathrm{d}\eta = T\left(\frac{\partial\eta}{\partial T}\right)_{p,S}\mathrm{d}T + T\left(\frac{\partial\eta}{\partial p}\right)_{T,S}\mathrm{d}p + T\left(\frac{\partial\eta}{\partial S}\right)_{p,T}\mathrm{d}S = c_p\,\mathrm{d}T + T\left(\frac{\partial\eta}{\partial p}\right)_{T,S}\mathrm{d}p + T\left(\frac{\partial\eta}{\partial S}\right)_{p,T}\mathrm{d}S, \tag{1.138}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

## Thermodynamics of Liquids

Liquids, unlike ideal gases, do not have a simple equation of state and this has ramifications for the thermodynamic equation. For seawater, a very accurate, albeit equally complex, equation of state is given by TEOS-10. The simpler expression for the Gibbs function, (1.146), is a good approximation in many circumstances and using it we can derive the thermal equation of state and some useful forms of the thermodynamic equation.

*Thermal equation of state*

A very useful expression for many purposes is given by

$$\alpha = \alpha_0 \left[ 1 + \beta_T(1 + \gamma^* p)(T - T_0) + \frac{\beta_T^*}{2}(T - T_0)^2 - \beta_S(S - S_0) - \beta_p(p - p_0) \right]. \tag{TL.1}$$

For laboratory situations where pressure variations are not large a useful approximation is

$$\alpha = \alpha_0 \left[ 1 + \beta_T(T - T_0) - \beta_S(S - S_0) \right]. \tag{TL.2}$$

*Thermodynamic equation*

Using entropy or potential enthalpy, as discussed in Section 1.7.3, as a primary thermodynamic variable is the best way to proceed if high accuracy is required. For idealized or laboratory work we can make further approximations, as follows.

*Entropy evolution:*

Using the hydrostatic approximation and simplifying (1.148) gives

$$\frac{\mathrm{D}\eta}{\mathrm{D}t} = 0, \qquad \eta = c_{p0} \ln \frac{T}{T_0} \left[ 1 + \beta_S^*(S - S_0) \right] + gz\beta_T. \tag{TL.3}$$

Given entropy and salinity we can infer temperature and, using (TL.1), density.

*Potential temperature and potential density:*

Potential temperature or potential density are could also be used as thermodynamic variables. Accurate expressions can be derived, but approximate expressions are

$$\frac{\mathrm{D}\theta}{\mathrm{D}t} = 0, \qquad \theta = T + \frac{\beta_T g T_0 z}{c_{p0}}, \tag{TL.4}$$

$$\text{or} \qquad \frac{\mathrm{D}\rho_\theta}{\mathrm{D}t} = 0, \qquad \rho_\theta = \rho + \frac{\rho_0 g z}{c_s^2}. \tag{TL.5}$$

Given $\theta$ or $\rho_\theta$, as well as salinity, we then infer density using an appropriate equation of state as needed.

using the definition of $c_p$. If we evaluate this expression at the reference pressure, where $T = \theta$ and $\mathrm{d}p = 0$, and consider isohaline changes with $\mathrm{d}S = 0$, then we have $\theta\,\mathrm{d}\eta = c_p(p_R, \theta, S)\,\mathrm{d}\theta$, and therefore

$$\mathrm{d}\eta = c_p(p_R, \theta, S)\frac{\mathrm{d}\theta}{\theta}, \tag{1.139}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

and $d\eta/d\theta = c_p(p_R, \theta, S)/\theta$. In the special case of constant $c_p$ integration yields

$$\eta = c_p \ln\theta + \text{constant}, \tag{1.140}$$

as for a simple ideal gas — see (1.107). Given (1.139), the thermodynamic equation can be written

$$c_p \frac{D\theta}{Dt} = \frac{\theta}{T}\dot{Q}, \tag{1.141}$$

where the right-hand side represents heating. We can use (1.141) as a thermodynamic equation instead of (1.97), although we must now relate $\theta$ instead of $\eta$ to the other state variables via an equation of state.

The notion of potential temperature is useful because it is connected to the actual temperature, with which we are familiar; roughly speaking, potential temperature is temperature plus a correction for the effects of thermal expansion. Entropy, on the other hand, may seem alien and unnecessarily exotic. However, the use of potential temperature brings no true simplifications to the equations of motion beyond those already afforded by the use of entropy as a thermodynamic variable.

### *Potential density*

Potential density, $\rho_\theta$, is the density that a parcel would have if it were moved adiabatically and with fixed composition to a given reference pressure, $p_R$, that is often, but not always, taken as $10^5$ Pa, or 1 bar. If the equation of state is of the form $\rho = \rho(S, T, p)$ then by definition we have

$$\rho_\theta = \rho(S, \theta, p_R). \tag{1.142}$$

For a parcel moving adiabatically and at fixed salinity its potential density is therefore conserved, and it is the vertical gradient of potential density that provides the appropriate measure of stability (as we find in Section 2.10.1). Because density of seawater is nearly constant we can obtain an approximate expression for potential density by Taylor-expanding the density around the potential density at the reference level at which $T = \theta$ and $p = p_R$. At first order we then have

$$\rho(S, \theta, p) \approx \rho(S, \theta, p_R) + (p - p_R)\left(\frac{\partial p}{\partial \rho}\right)_{S,\theta}. \tag{1.143}$$

The first term on the right hand side is, by definition, the potential density and the derivative in the second term is the inverse of the square of speed of sound, evaluated at the reference level, and so

$$\boxed{\rho_\theta \approx \rho - \frac{1}{c_s^2}(p - p_R) \approx \rho + \frac{\rho_0 g z}{c_s^2}.} \tag{1.144}$$

To obtain the right-most expression we use hydrostatic balance and take $p = -\rho_0 g z$ and $p_R = 0$ (at $z = 0$), giving the same expression as occurs in (1.117).

Because the density of seawater is nearly constant, it is common in oceanography to subtract the amount $1000\,\mathrm{kg\,m^{-3}}$ before quoting its value; then, depending on whether we are referring to *in situ* density or the potential density the results are called $\sigma_T$ ('sigma-tee') or $\sigma_\theta$ ('sigma-theta') respectively. Thus,

$$\sigma_T = \rho(p, T, S) - 1000, \qquad \sigma_\theta = \rho(p_R, \theta, S) - 1000. \tag{1.145a,b}$$

Instead of the subscript $\theta$, a number can be used to denote the level to which potential density is referenced. Thus, $\sigma_0$ is the potential density referenced to the surface and $\sigma_2$ is the potential density referenced to 200 bars of pressure, or about 2 kilometres depth.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

| Parameter | Description | Value |
|---|---|---|
| $\rho_0$ | Reference density | $1.027 \times 10^3$ kg m$^{-3}$ |
| $\alpha_0$ | Reference specific volume | $9.738 \times 10^{-4}$ m$^3$ kg$^{-1}$ |
| $T_0$ | Reference temperature | 283K |
| $S_0$ | Reference salinity | 35 ppt = 35 g kg$^{-1}$ |
| $c_{s0}$ | Reference sound speed | 1490 m s$^{-1}$ |
| $\beta_T$ | First thermal expansion coefficient | $1.67 \times 10^{-4}$ K$^{-1}$ |
| $\beta_T^*$ | Second thermal expansion coefficient | $1.00 \times 10^{-5}$ K$^{-2}$ |
| $\beta_S$ | Haline contraction coefficient | $0.78 \times 10^{-3}$ ppt$^{-1}$ |
| $\beta_p$ | Compressibility coefficient ($= \alpha_0/c_{s0}^2$) | $4.39 \times 10^{-10}$ m s$^2$ kg$^{-1}$ |
| $\gamma^*$ | Thermobaric parameter ($\approx \gamma'^*$) | $1.1 \times 10^{-8}$ Pa$^{-1}$ |
| $c_{p0}$ | Specific heat capacity at constant pressure | 3986 J kg$^{-1}$ K$^{-1}$ |
| $\beta_S^*$ | Haline heat capacity coefficient | $1.5 \times 10^{-3}$ ppt$^{-1}$ |

**Table 1.2** Various thermodynamic and equation-of-state parameters appropriate for the seawater equations of state (1.58) and (1.146). The unit ppt (or ‰) is parts per thousand by weight, or g/kg.

We cannot use $p_R = 0$ everywhere and maintain accuracy. Thus, for a parcel near 2 km, $\sigma_2$ is more relevant than $\sigma_0$. The 'neutral density' (or quasi-neutral density) is a semi-empirical way to avoid this reference-level difficulty. Neutral density is, by construction, a quantity such that the buoyancy force is locally perpendicular to its iso-surfaces, so that a parcel that is displaced adiabatically along a neutral density iso-surface will remain neutrally buoyant.[9] Neutral density is not a thermodynamic state variable and because of form of the seawater equation of state there is no continuous, unique field of neutral density extending through the ocean. Wherever it appears in figures in this book it can be assumed that potential density would look similar.

### 1.7.2 Equation of State for Seawater

Oceanographers go to great lengths to obtain an accurate equation of state and other physical properties of seawater, and we noted in Section 1.4 that seawater has some nonlinear properties that, although small, are nevertheless important. We need to be able to calculate these properties, and in this section we illustrate how the thermodynamic variables for seawater — the conventional equation of state, an expression for potential temperature, and so on — can be obtained directly from the fundamental equation of state. Writing the fundamental equation in the form $I = I(\eta, S, \alpha)$ is not practically useful, because the variables are not easily measured in the laboratory. However, if we cast the fundamental equation in terms of the Gibbs function, $g = I - T\eta + p\alpha$, then $\mathrm{d}g = -\eta\, \mathrm{d}T + \alpha\, \mathrm{d}p + \mu\, \mathrm{d}S$ and the independent variables are the familiar and measurable $(T, S, p)$. A similar, but simpler, procedure is carried out for an ideal gas in Appendix A.

A Gibbs function that reproduces the properties of seawater with high accuracy is very complicated, but we can write down a Gibbs function that, although slightly less accurate, captures the most important properties with a certain degree of economy and transparency. The expression is[10]

$$
\begin{aligned}
g = g_0 - \eta_0(T - T_0) + \mu_0(S - S_0) - c_{p0}T\left[\ln(T/T_0) - 1\right]\left[1 + \beta_S^*(S - S_0)\right] \\
+ \alpha_0(p - p_0)\Big[1 + \beta_T(T - T_0) - \beta_S(S - S_0) - \frac{\beta_p}{2}(p - p_0) \\
+ \frac{\beta_T\gamma^*}{2}(p - p_0)(T - T_0) + \frac{\beta_T^*}{2}(T - T_0)^2\Big].
\end{aligned}
\tag{1.146}
$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

In this equation the variables are $g, T, S$ and $p$. The parameters (which, as in (1.58), all have subscripts or stars, with the starred parameters giving rise to nonlinear effects) are all constants that could in principle be determined in the laboratory with the help of the derived quantities like heat capacity, and their approximate values are given in Table 1.2. We will take $p_0 = 0$ and $\beta_p = \alpha_0/c_{s0}^2$, where $c_{s0}$ is a reference sound speed. Equation (1.146) is in fact quite accurate for most oceanographic situations, and from it we may derive the following quantities of interest:

– The conventional or thermal equation of state, $\alpha = (\partial g/\partial p)_{T,S}$:

$$\alpha = \alpha_0 \left[ 1 + \beta_T (1 + \gamma^* p)(T - T_0) + \frac{\beta_T^*}{2}(T - T_0)^2 - \beta_S (S - S_0) - \beta_p p \right]. \tag{1.147}$$

– The entropy, $\eta = -(\partial g/\partial T)_{p,S}$:

$$\eta = \eta_0 + c_{p0} \ln \frac{T}{T_0} \left[ 1 + \beta_S^* (S - S_0) \right] - \alpha_0 p \left[ \beta_T + \beta_T \gamma^* \frac{p}{2} + \beta_T^* (T - T_0) \right]. \tag{1.148}$$

For temperatures in the range 0°–30° Celsius, entropy increases linearly with temperature to within a few percent.

– The heat capacity, $c_p = T(\partial \eta/\partial T)_{p,S}$:

$$c_p = c_{p0} \left[ 1 + \beta_S^* (S - S_0) \right] - \alpha_0 p \beta_T^* T. \tag{1.149}$$

This is to a first approximation constant, varying mildly with salinity and more weakly with temperature and pressure.

– The thermal expansion coefficient, $\widehat{\beta}_T = \alpha^{-1}(\partial \alpha/\partial T)_{S,p}$:

$$\widehat{\beta}_T = (\alpha_0/\alpha) \left[ \beta_T + \beta_T \gamma^* p + \beta_T^* (T - T_0) \right], \tag{1.150}$$

where $\alpha$ is given by (1.147).

– The adiabatic lapse rate, $\Gamma = (\partial T/\partial p)_{\eta,S}$. Using (1.138) gives

$$\Gamma = \left( \frac{\partial T}{\partial p} \right)_{\eta,S} = -\frac{T}{c_p} \left( \frac{\partial \eta}{\partial p} \right)_{T,S} = \frac{T}{c_p} \alpha_0 [\beta_T (1 + \gamma^* p) + \beta_T^* (T - T_0)]. \tag{1.151}$$

where $c_p$ is given by (1.149).

### *Potential temperature and potential density, revisited*

An expression for the potential temperature, $\theta$, may be obtained by solving (1.136) for $\theta$. In general, such an equation must be solved numerically, but for our equation of state, using (1.148) and taking $p_R = 0$, we find

$$\theta = T \exp \left\{ -\frac{\alpha_0 \beta_T p}{c_p'} \left[ 1 + \frac{1}{2} \gamma^* p + \frac{\beta_T^*}{\beta_T} (T - T_0) \right] \right\}, \tag{1.152}$$

where $c_p' = c_{p0} \left[ 1 + \beta_S^* (S - S_0) \right]$. Equation (1.152) is a relationship between $T, \theta$ and $p$ analogous to (1.105) for an ideal gas. The exponent itself is small, the second and third terms in square brackets are small compared to unity, the deviations of both $T$ and $\theta$ from $T_0$ are also presumed to be small,

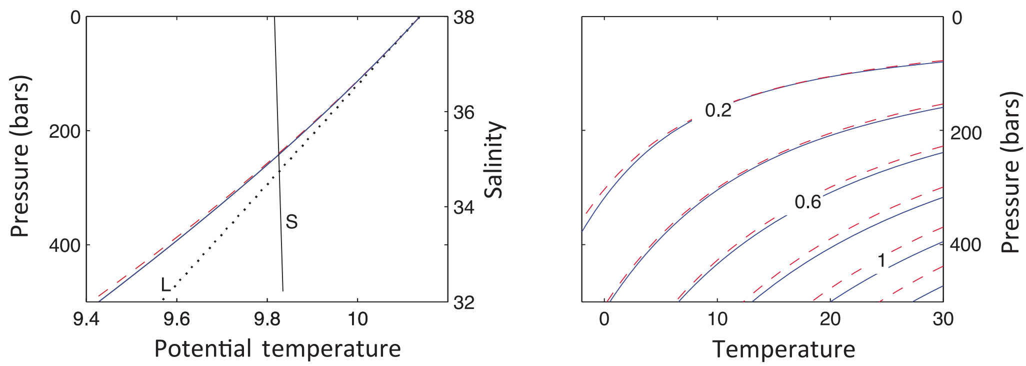


**Fig. 1.4** Examples of the variation of potential temperature of seawater with pressure, temperature and salinity. Left panel: the sloping lines show potential temperature as a function of pressure at fixed salinity ($S = 35$ psu) and temperature (10° C). The solid line is computed using an accurate, empirical equation of state, the almost-coincident dashed line uses the simpler expression (1.152) and the dotted line (labelled L) uses the linear expression (1.153c). The near vertical solid line, labelled S, shows the variation of potential temperature with salinity at fixed temperature and pressure. Right panel: Contours of the difference between temperature and potential temperature, $(T - \theta)$ in the pressure–temperature plane, for $S = 35$ psu. The dashed lines use (1.152) and the solid lines use an accurate empirical formula. (Note: 100 bars of pressure ($10^7$ Pa or 10 MPa) is approximately 1 km depth.)

and $c'_p$ is nearly constant. Taking advantage of all of this enables (1.152) to be rewritten, with increasing levels of approximation, as

$$T' \approx \frac{T_0\alpha_0\beta_T}{c_{p0}} p\left(1 + \frac{1}{2}\gamma^* p + T_0\frac{\alpha_0\beta_T^*}{c_{p0}}p\right) + \theta'\left(1 + T_0\frac{\alpha_0\beta_T^*}{c_{p0}}p\right) \tag{1.153a}$$

$$\approx \frac{T_0\alpha_0\beta_T}{c_{p0}} p\left(1 + \frac{1}{2}\gamma^* p\right) + \theta'\left(1 + T_0\frac{\alpha_0\beta_T^*}{c_{p0}}p\right) \tag{1.153b}$$

$$\approx \frac{T_0\alpha_0\beta_T}{c_{p0}} p + \theta', \tag{1.153c}$$

where $T' = T - T_0$ and $\theta' = \theta - T_0$. The last of the three, (1.153c), holds for a linear equation of state, and is useful for calculating approximate differences between temperature and potential temperature; making use of the hydrostatic approximation reveals that it is essentially the same as (1.129). Plots of the difference between temperature and potential temperature, using both a highly accurate empirical equation of state and using the simplified equation, (1.152), are given in Fig. 1.4, and some examples of the density variation of seawater are given in Fig. 1.5.

To obtain an equation of state that gives density in terms of potential temperature, pressure and salinity we use (1.153b) in the equation of state, (1.147), to give

$$\alpha \approx \alpha_0\left[1 - \frac{\alpha_0 p}{c'^2_{s0}} + \beta_T(1 + \gamma'^* p)\theta' + \frac{1}{2}\beta_T^*\theta'^2 - \beta_S(S - S_0)\right], \tag{1.154}$$

where $\gamma'^* = \gamma^* + T_0\beta_T^*\alpha_0/c_{p0}$ and $c_{s0}^{-2} = c'^{-2}_{s0} - \beta_T^2 T_0/c_p$. The parameters $\gamma^*$ and $\gamma'^*$ differ by a few percent, and $c_s^2$ and $c'^2_s$ differ by a few parts in a thousand, and we may neglect the differences. We may further approximate (1.154) by using the hydrostatic pressure instead of the actual pressure; thus, letting $p = -g(z - z_0)/\alpha_0$ where $z_0$ is the nominal value of $z$ at which $p = 0$, we obtain

$$\alpha \approx \alpha_0\left[1 + \frac{g(z - z_0)}{c_{s0}^2} + \beta_T\left(1 - \gamma^*\frac{g(z - z_0)}{\alpha_0}\right)\theta' + \frac{\beta_T^*}{2}\theta'^2 - \beta_S(S - S_0)\right]. \tag{1.155}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

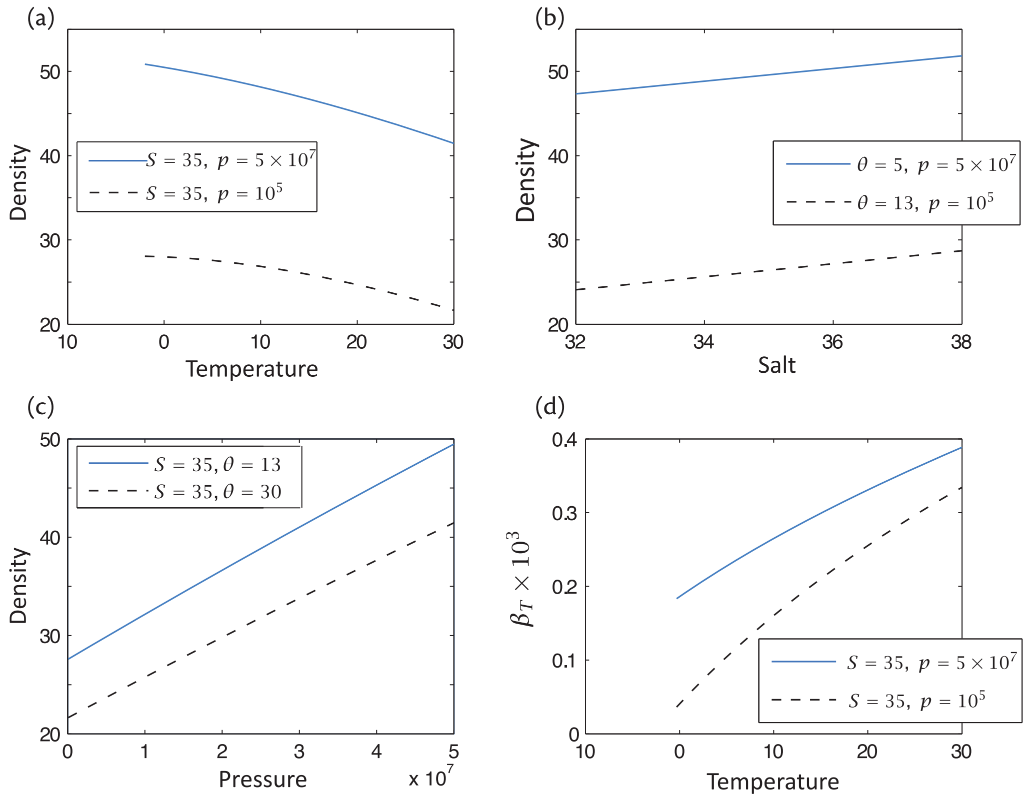


**Fig. 1.5** Examples of the variation of density of seawater, $(\rho - 1000)\,\mathrm{kg\,m^{-3}}$. (a) With potential temperature (° C); (b) with salt (g/kg); and (c) with pressure (Pa), for seawater. Panel (d), shows the thermal expansion coefficient, $\widehat{\beta}_T = -\rho^{-1}(\partial\rho/\partial T)_{p,S}\,\mathrm{K}^{-1}$, for each of the two curves in panel (a).

Using $z$ instead of $p$ in the equation of state entails a slight loss of accuracy, but is necessary to ensure that the Boussinesq equations maintain good conservation properties, as will be discussed in Sections 2.4 and 4.7.1.

Given an expression for density in terms of potential temperature we can write down an expression for potential density, $\rho_\theta$, since by definition it is just the density at a reference pressure, $p_R$. From (1.147) we have, assuming density variations are small,

$$\rho_\theta \approx \rho_0\left[1 + \frac{\alpha_0 p_R}{c_{s0}^2} - \beta_T(1+\gamma^* p_R)\theta' - \frac{1}{2}\beta_T^*\theta'^2 + \beta_S(S - S_0)\right], \tag{1.156a}$$

$$\approx \rho + (p_R - p)\left(\frac{1}{c_{s0}^2} - \beta_T\gamma^*\rho_0\theta'\right). \tag{1.156b}$$

This expression may be compared to the more approximate one, (1.144). The second term in large brackets in (1.156b) is quite small and is a manifestation of the thermobaric effect, that the compressibility of seawater is a weak function of temperature.

### 1.7.3 Potential Enthalpy as a Thermodynamic Variable

As discussed in Section 1.6, there are various forms of thermodynamic equation that are, in principle, equivalent. However, they are not equivalent in practice, even if the fundamental equation of state of is known. Thus, for example, for a nearly incompressible fluid like seawater the velocity

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

divergence is small but non-zero and numerically integrating the internal energy equation, (1.94), may be awkward and inaccurate. An ideal choice might be some conservative variable that obeys

$$\rho \frac{D\chi}{Dt} = \nabla \cdot \boldsymbol{F}_\chi \qquad \text{or equivalently} \qquad \frac{\partial}{\partial t}(\rho\chi) + \nabla \cdot (\rho \boldsymbol{v} \chi) = \nabla \cdot \boldsymbol{F}_\chi, \tag{1.157}$$

where $\chi$ is a thermodynamical variable from which all the other variables can be inferred using the fundamental equation of state, and $\boldsymbol{F}_\chi$ represents molecular and radiative fluxes. Unfortunately, there is no variable that exactly has an advective left-hand side and a right-hand side that is the divergence of a flux — but some are nearly so, as we discuss.

### *Entropy*

An obvious choice for a primary thermodynamic variable is entropy, an advected variable satisfying

$$\frac{D\eta}{Dt} = \frac{\dot{Q}}{T}, \tag{1.158}$$

where $\dot{Q}$ is the heating. The main difficulties with an entropic approach are the determination of the heating and the accurate treatment of that heating once determined. The right-hand side does not have a conservative form even in the case which the heating is due solely to diffusive molecular fluxes and radiation. A further, albeit minor, complication is that heating is affected by irreversible molecular fluxes of salt in the interior. Further difficulties arise in using (1.158) in a turbulent ocean, because heating is affected not only by thermal effects but also by freshwater fluxes at the ocean surface and unresolved fluxes of salinity in the ocean interior. Care must taken to ensure that parameterized fluxes of temperature and salinity do produce (and not reduce) entropy.

Potential temperature, $\theta$, provides no fundamental advantage over entropy, because $\theta$ is a function of both entropy and salinity and still requires a computation of heating. Indeed, because $d\eta \approx c_p d\ln\theta$ the evolution equation for potential temperature properly involves the heat capacity, $c_p$, which is a function (albeit a weak one) of both salinity and pressure. Having said all this, the above problems are not in practice large ones, and entropy and (more commonly) potential temperature have been used very successfully in quantitative ocean models.

### *Potential enthalpy*

An alternative to entropy is to construct a near-conservative variable related to enthalpy. Consider first the fundamental thermodynamic relation, (1.85), in the form

$$\frac{DI}{Dt} + p\frac{D\alpha}{Dt} = T\frac{D\eta}{Dt} + \mu\frac{DS}{Dt} = \frac{Dh}{Dt} - \alpha\frac{Dp}{Dt} = \dot{Q}_E, \tag{1.159}$$

where $\dot{Q}_E = \dot{Q} + \mu\dot{S}$ is the total rate of non-mechanical energy input per unit mass, $\dot{Q}$ is the heating and $\dot{S}$ represents saline sources and sinks. Now, it is usually easier and more accurate to determine energy input, $\dot{Q}_E$, than the heating, $\dot{Q}$, because $\dot{Q}_E$ is very nearly equal to the divergence of an energy flux. That is to say, $\rho\dot{Q}_E \approx \nabla \cdot \boldsymbol{F}_E$ where $\boldsymbol{F}_E$ is an energy flux, with the small difference arising from the heating due to viscous dissipation of kinetic energy — see Appendix B. This property suggests the use of internal energy or enthalpy as a primary thermodynamical variable, but neither are conservative quantities because of the $D\alpha/Dt$ and $Dp/Dt$ terms in (1.159). However we can form a quantity, *potential enthalpy,* that very nearly is conservative, as follows.[11]

The potential enthalpy, $h^0$, is defined to be the enthalpy that a fluid parcel has if taken at constant composition and entropy from its current location to a fixed reference pressure, $p_R$. Thus,

$$h^0(S, \eta, p_R) \equiv h(S, \eta, p) + \int_p^{p_R} \left(\frac{\partial h}{\partial p}\right)_{S,\eta} dp' = h(S, \eta, p) + \int_p^{p_R} \alpha \, dp', \tag{1.160}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

using (1.74). That is, $h = h^0 + h^d$ where

$$h^d(\eta, p, S) = -\int_p^{p_R} \left(\frac{\partial h}{\partial p}\right)_{S,\eta} \mathrm{d}p' = -\int_p^{p_R} \alpha \, \mathrm{d}p' \tag{1.161}$$

is the 'dynamic enthalpy'. The material derivative of the dynamic enthalpy is given by

$$\frac{\mathrm{D}h^d}{\mathrm{D}t} = \frac{\mathrm{D}\eta}{\mathrm{D}t}\frac{\partial h^d}{\partial \eta} + \frac{\mathrm{D}S}{\mathrm{D}t}\frac{\partial h^d}{\partial S} + \alpha\frac{\mathrm{D}p}{\mathrm{D}t}. \tag{1.162}$$

To obtain this expression we have used $(\partial h^d/\partial p)_{\eta,S} = (\partial h/\partial p)_{\eta,S} = \alpha$, because $h^0$ is not a function of pressure. Using (1.162) and (1.160) we obtain

$$\frac{\mathrm{D}h^0}{\mathrm{D}t} = \frac{\mathrm{D}h}{\mathrm{D}t} - \frac{\mathrm{D}h^d}{\mathrm{D}t} = \frac{\mathrm{D}h}{\mathrm{D}t} - \left[\alpha\frac{\mathrm{D}p}{\mathrm{D}t} + \frac{\mathrm{D}\eta}{\mathrm{D}t}\frac{\partial h^d}{\partial \eta} + \frac{\mathrm{D}S}{\mathrm{D}t}\frac{\partial h^d}{\partial S}\right], \tag{1.163}$$

and using the fundamental thermodynamic relation, (1.159), we can write this equation as

$$\frac{\mathrm{D}h^0}{\mathrm{D}t} = \dot{Q}_E - \frac{\partial h^d}{\partial \eta}\dot{\eta} - \frac{\partial h^d}{\partial S}\dot{S}. \tag{1.164}$$

The second and third terms on the right-hand side are much smaller than the other terms and can be neglected in most oceanographic applications. To see this, realize that an approximate size of the first term on the right-hand side is $c_p \mathrm{d}\theta/\mathrm{d}t$, whereas the approximate size of the second term is $(\partial h^d/\partial \eta) \times (c_p \theta^{-1} \mathrm{d}\theta/\mathrm{d}t)$. That is, the second term is smaller than the first by the factor

$$\gamma = \frac{1}{\theta}\frac{\partial h^d}{\partial \eta} = \frac{1}{c_p}\frac{\partial h_d}{\partial \theta} \sim \frac{1}{c_p}\frac{\partial \alpha}{\partial \theta}\Delta p, \tag{1.165}$$

using (1.161), where $\Delta p = p - p_R \le 4 \times 10^7$ Pa. Putting in values from the seawater equation of state, (1.154), we find

$$\gamma \sim \frac{1}{c_p}\frac{\partial \alpha}{\partial \theta}\Delta p \approx \frac{\beta_T \Delta p}{\rho_0 c_p} \approx \frac{1.7 \times 10^{-4} \;\; 4 \times 10^7}{10^3 \;\; 4 \times 10^3} \approx 1.7 \times 10^{-3}, \tag{1.166}$$

with a similarly small value for the saline term, where the smallness of these terms ultimately stems from the near incompressibility of seawater. These terms may actually be smaller than (1.166) suggests if we choose $p_R$ appropriately. Most of the energy flux into the ocean occurs at the ocean surface and if we choose $p_R = 0$ the fluxes affect enthalpy and potential enthalpy in the same way.

Given the above arguments, an accurate and computable thermodynamic equation for seawater is (1.164) with the last two terms neglected and the source term in flux form, namely

$$\rho\frac{\mathrm{D}h^0}{\mathrm{D}t} = \nabla \cdot \boldsymbol{F}_E \qquad \text{or} \qquad \frac{\partial}{\partial t}(\rho h^0) + \nabla \cdot (\rho \boldsymbol{v} h^0) = \nabla \cdot \boldsymbol{F}_E. \tag{1.167}$$

Given that the right-hand side is a flux divergence, in steady state the interior fluxes of potential enthalpy are (in this approximation) in exact balance with the energy fluxes at the ocean boundaries. The integral of $\rho h^0$ is thus a sensible measure of the total non-kinetic energy, or 'heat content', of the ocean because it responds almost exactly to energy fluxes at the ocean surface.[12]

Why does internal energy not have these same advantages since, from (1.94), it too is conservative apart from a small compression term? The reason is that internal energy can be changed by processes that are entirely internal to the ocean, whereas enthalpy cannot. When two adjacent

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

fluid parcels mix, the total enthalpy is conserved because of the form of the thermodynamic equation, $\rho \mathrm{D}h/\mathrm{D}t - \mathrm{D}p/\mathrm{D}t = \rho Q_E$. If $\rho Q_E$ is the divergence of a flux then the total enthalpy ($\int \rho h \, \mathrm{d}V$) is conserved on mixing (because the parcels are at the same pressure), but this conservative property is not shared by internal energy, because of the compression term, $p\mathrm{D}\alpha/\mathrm{D}t$, in the internal energy equation. Indeed, when two parcels at different temperatures and salinities are mixed the density of the resulting parcel is higher than the average of the two (see Fig. 1.3); this effect is called cabbeling and the contraction leads to a small increase in internal energy. (The differences in the conservation of enthalpy and internal energy arise from the differences in the equation of motion for enthalpy and internal energy and not directly from the equation of state.) Thus, internal energy is not a good measure of the 'heat content' of the ocean and, furthermore, in practice the very small compression term would be hard to treat accurately. In contrast, internal energy often *is* used as a primary thermodynamic variable in ideal-gas atmospheric models, where $I = c_v T$.

Thus, in short, the practical advantages of potential enthalpy are three-fold: (i) Potential enthalpy is nearly conservative, as in (1.167); (ii) As a consequence, potential enthalpy is itself a useful measure of the non-kinetic energy of the ocean; (iii) Energy flux is a little more easily and more accurately computed than heating. Property (ii) is not shared by internal energy, since that is not conserved when parcels mix in the interior. In practice, potential enthalpy is almost proportional to potential temperature, as we now see.

### *Conservative temperature and an expression for potential enthalpy*

To make a connection to the perhaps more familiar potential temperature it is convenient to define *conservative temperature,* $\Theta$, by

$$\Theta \equiv \frac{h^0(\eta, p_R, S)}{c_p^0}, \tag{1.168}$$

where $c_p^0 = 3991.87\,\mathrm{J\,kg^{-1}\,K^{-1}}$ is the average heat capacity at the ocean surface and $p_R = 0$. (Potential enthalpy can have any reference pressure; conservative temperature has, by definition $p_R = 0$.) As we see below, conservative temperature is then very similar to potential temperature.

The enthalpy is given in terms of the Gibbs function by

$$h = g + T\eta = g - T\left(\frac{\partial g}{\partial T}\right)_{p,S}. \tag{1.169}$$

The potential enthalpy is therefore given by

$$h^0(S, T, p) = h(S, \theta, p_R) = g(S, \theta, p_R) - T\frac{\partial}{\partial T} g(S, \theta, p_R), \tag{1.170}$$

where $\theta$ is the potential temperature referenced to $p = p_R$ (and $p_R = 0$) and the last term is the derivative of the Gibbs function evaluated at $T = \theta$ and $p = p_R$. We can evaluate the derivative using the seawater equation of state, (1.146), giving, to within a constant factor,

$$\begin{aligned} h(T, p, S) = {}& \mu_0 S + c_{p0} T\left[1 + \beta_S^*(S - S_0)\right] \\ & + \alpha_0 p\left[1 - \beta_T T_0 - \beta_S(S - S_0) - \frac{\beta_P p}{2} - \frac{\beta_T \gamma^*}{2} p T_0 + \frac{\beta_T^*}{2}(T_0^2 - T^2)\right], \end{aligned} \tag{1.171}$$

and

$$h^0(\theta, 0, S) = h(\theta, 0, S) = \theta c_{p0}\left[1 + \beta_S^*(S - S_0)\right] + \mu_0 S. \tag{1.172}$$

The last term on the right-hand side of the above equation, $\mu_0 S$, is over two orders of magnitude smaller than the first. Also, from (1.149), the factor multiplying $\theta$ in (1.172) is just the heat capacity at $p = 0$, namely $c_p^0$, which varies only by few percent over the ocean. Thus, using the definition of conservative temperature given in (1.168), we have to a good approximation

$$c_p^0 \Theta \equiv h^0(\theta, 0, S) \approx c_p^0 \theta. \tag{1.173}$$

That is, to this approximation, conservative temperature equals potential temperature. Finally, if we are to use potential enthalpy (or entropy) as a thermodynamic variable we must manipulate the equation of state to obtain variables such as pressure or density from it, analogous to (1.154).

## 1.8 SOUND WAVES

*Full of sound and fury, signifying nothing.*
William Shakespeare, *Macbeth*, c. 1606.

We now consider, rather briefly, one of the most common phenomena in fluid dynamics, yet one which is in most circumstances relatively unimportant for geophysical fluid dynamics — sound waves. Their unimportance stems from the fact that the pressure disturbances produced by sound waves are a tiny fraction of the ambient pressure and are too small to affect the circulation. For example, the ambient surface pressure in the atmosphere is about $10^5$ Pa and variations due to large-scale weather phenomena are about $10^3$ Pa or larger, whereas sound waves of 70 dB (i.e., a loud conversation) produce pressure variations of about 0.06 Pa. (To convert, dBs $= 20 \log_{10}(\Delta p / p_r)$ where $\Delta p$ is the pressure change in Pascals and $p_r = 2 \times 10^{-5}$.)

The smallness of the disturbance produced by sound waves justifies a linearization of the equations of motion and we do so about a spatially uniform basic state that is a time-independent solution to the equations of motion. Thus, we write $\boldsymbol{v} = \boldsymbol{v}_0 + \boldsymbol{v}'$, $\rho = \rho_0 + \rho'$ (where a subscript 0 denotes a basic state and a prime denotes a perturbation) and so on, substitute in the equations of motion, and neglect terms involving products of primed quantities. By choice of our reference frame we will simplify matters further by setting $\boldsymbol{v}_0 = 0$. The linearized momentum and mass conservation equations are then, respectively,

$$\rho_0 \frac{\partial \boldsymbol{v}'}{\partial t} = -\nabla p', \qquad \frac{\partial \rho'}{\partial t} = -\rho_0 \nabla \cdot \boldsymbol{v}'. \tag{1.174a,b}$$

(These linear equations do not in themselves determine the magnitude of the disturbance.) Now, sound waves are largely adiabatic. Thus,

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \left(\frac{\partial p}{\partial \rho}\right)_\eta \frac{\mathrm{d}\rho}{\mathrm{d}t}, \tag{1.175}$$

where $(\partial p/\partial \rho)_\eta$ is the derivative at constant entropy, whose particular form is given by the equation of state for the fluid at hand. Then, from (1.174) and (1.175) we obtain a single equation for pressure,

$$\frac{\partial^2 p'}{\partial t^2} = c_s^2 \nabla^2 p', \tag{1.176}$$

where $c_s^2 = (\partial p/\partial \rho)_\eta$. Equation (1.176) is the classical wave equation; solutions propagate at a speed $c_s$, which may be identified as the speed of sound. For adiabatic flow in an ideal gas, manipulation of the equation of state leads to $p = C\rho^\gamma$, where $\gamma = c_p/c_v$, whence $c_s^2 = \gamma p/\rho = \gamma RT$. Values of $\gamma$ typically range from 5/3 for a monatomic gas to 7/5 for a diatomic gas, and so for air, which is almost entirely diatomic, we find $c_s \approx 350\,\mathrm{m\,s^{-1}}$ at 300 K. In seawater no such theoretical approximation is easily available, but measurements show that $c_s \approx 1500\,\mathrm{m\,s^{-1}}$.

## 1.9 COMPRESSIBLE AND INCOMPRESSIBLE FLOW

Although there may be no fluids of truly constant density, in many cases the density of a fluid will vary so little that it is a very good approximation to consider the density effectively constant. The fluid is then said to be *incompressible*. (Some sources take incompressible to mean that density is unaffected by pressure. We take it to mean that density is also unaffected by temperature and composition.) For example, in the Earth's oceans the density varies by less than 5% (usually much less) even though the pressure at the ocean bottom is several hundred times that at the surface. We first consider how the mass continuity equation simplifies when density is truly constant, and then consider conditions under which treating density as constant is a good approximation.

### 1.9.1 Constant Density Fluids

If a fluid is strictly of constant density then the mass continuity equation, $\mathrm{D}\rho/\mathrm{D}t + \rho\nabla\cdot\boldsymbol{v} = 0$, simplifies to

$$\boxed{\nabla\cdot\boldsymbol{v} = 0.} \tag{1.177}$$

The *prognostic* equation (1.36) has become a *diagnostic* equation (1.177), or a constraint to be satisfied by the velocity. The volume of each material fluid element is therefore constant; to see this recall that conservation of mass is $\mathrm{D}(\rho\Delta V)/\mathrm{D}t = 0$ and if $\rho$ is constant this becomes $\mathrm{D}\Delta V/\mathrm{D}t = 0$, whence (1.177) is recovered because $\mathrm{D}\Delta V/\mathrm{D}t = \Delta V\nabla\cdot\boldsymbol{v}$ .

### 1.9.2 Incompressible Flows

In reality no fluid is truly incompressible and for (1.177) to approximately hold we just require that

$$\left|\frac{\mathrm{D}\rho}{\mathrm{D}t}\right| \ll \rho\left(\left|\frac{\partial u}{\partial x}\right| + \left|\frac{\partial v}{\partial y}\right| + \left|\frac{\partial w}{\partial z}\right|\right); \tag{1.178}$$

that is, the material derivative of density is much smaller than the individual terms constituting the divergence. As a working definition we say that *in an incompressible fluid, density changes are so small that they have a negligible effect on the mass balance.* We do not need to assume that the densities of differing fluid elements are similar to each other, but in the ocean (and in most liquids) variations in density, $\delta\rho$, are in fact everywhere small compared to the mean density, $\rho_0$. A sufficient condition for incompressibility, then, is that

$$\frac{\delta\rho}{\rho_0} \ll 1. \tag{1.179}$$

The fact that $\nabla\cdot\boldsymbol{v} = 0$ does *not* imply that we may independently use $\mathrm{D}\rho/\mathrm{D}t = 0$. Indeed for a liquid with an equation of state $\rho = \rho_0(1-\beta_T(T-T_0))$ and a thermodynamic equation $c_p DT/Dt = \dot{Q}$ we have

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\frac{\beta_T\rho_0}{c_p}\dot{Q}. \tag{1.180}$$

Furthermore, incompressibility does not necessarily imply the neglect of density variations in the momentum equation — it is only in the mass continuity equation that density variations are neglected, as will become apparent in our discussion of the Boussinesq equations in chapter 2.

#### *Conditions for incompressibility*

The conditions under which incompressibility is a good approximation to the full mass continuity equation depend not only on the physical nature of the fluid but also on the flow itself. The condition that density is largely unaffected by pressure gives one necessary condition for the legitimate

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

use of (1.177), as follows. First assume adiabatic flow, and omit the gravitational term. Then

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \left(\frac{\partial p}{\partial \rho}\right)_\eta \frac{\mathrm{d}\rho}{\mathrm{d}t} = c_s^2 \frac{\mathrm{d}\rho}{\mathrm{d}t}, \tag{1.181}$$

so that the density and pressure variations of a fluid parcel are related by

$$\delta p \sim c_s^2 \delta \rho. \tag{1.182}$$

From the momentum equation we estimate

$$\frac{U^2}{L} \sim \frac{1}{L}\frac{\delta p}{\rho_0}, \tag{1.183}$$

where $U$ and $L$ are typical velocities and lengths and where $\rho_0$ is a representative value of the density. Using (1.182) and (1.183) gives $U^2 \sim c_s^2 \delta\rho/\rho_0$. The incompressibility condition (1.179) then becomes

$$\frac{U^2}{c_s^2} \ll 1. \tag{1.184}$$

Thus, for a flow to be incompressible the fluid velocities must be less than the speed of sound; that is, the Mach number, $M \equiv U/c_s$, must be small.

In the Earth's atmosphere it is apparent that density changes with height. To estimate how much density does change, let us first assume hydrostatic balance and an ideal gas, so that $\partial p/\partial z = -\rho g$. If we also assume that atmosphere is isothermal then

$$\frac{\partial p}{\partial z} = \left(\frac{\partial p}{\partial \rho}\right)_T \frac{\partial \rho}{\partial z} = RT_0 \frac{\partial \rho}{\partial z}. \tag{1.185}$$

Using hydrostasy and (1.185) gives

$$\rho = \rho_0 \exp(-z/H_\rho), \tag{1.186}$$

where $H_\rho = RT_0/g$ is the (density) *scale height* of the atmosphere. (It is also the pressure scale height here.) It is easy to see that density changes are negligible only if we concern ourselves with motion less than the scale height, so this is another necessary condition for incompressibility.

In the atmosphere, although the Mach number is small for most flows, vertical displacements often exceed the scale height and the flow cannot then be considered incompressible. In the ocean, density changes from all causes are small and in most circumstances the ocean may be considered to contain an incompressible fluid.

## 1.10 THE ENERGY BUDGET

The total energy of a fluid includes the kinetic, potential and internal energies. Both fluid flow and pressure forces will, in general, move energy from place to place, but we nevertheless expect, even demand, energy to be conserved in an enclosed volume. Is it?

### 1.10.1 Constant Density Fluid

For a constant density fluid the momentum equation and the mass continuity equation $\nabla \cdot \boldsymbol{v} = 0$, are sufficient to completely determine the evolution of a system. The momentum equation is

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla(\phi + \Phi) + \nu \nabla^2 \boldsymbol{v}, \tag{1.187}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

where $\phi = p/\rho_0$ and $\Phi$ is the potential for any conservative force per unit mass (e.g., $gz$ for a uniform gravitational field). We can rewrite the advective term on the left-hand side using the identity,

$$(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\boldsymbol{v}\times\boldsymbol{\omega} + \nabla(\boldsymbol{v}^2/2), \tag{1.188}$$

where $\boldsymbol{\omega} \equiv \nabla\times\boldsymbol{v}$ is the *vorticity*, discussed more in later chapters. Then, omitting viscosity, we have

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{\omega}\times\boldsymbol{v} = -\nabla B, \tag{1.189}$$

where $B = (\phi+\Phi+\boldsymbol{v}^2/2)$ is the *Bernoulli function* for constant density flow. Consider for a moment steady flows ($\partial\boldsymbol{v}/\partial t = 0$). Streamlines are, by definition, parallel to $\boldsymbol{v}$ everywhere, and the vector $\boldsymbol{v}\times\boldsymbol{\omega}$ is everywhere orthogonal to the streamlines, so that taking the dot product of the steady version of (1.189) with $\boldsymbol{v}$ gives $\boldsymbol{v}\cdot\nabla B = 0$. That is, for steady flows the Bernoulli function is constant along a streamline, and $\mathrm{D}B/\mathrm{D}t = 0$.

Reverting to the time-varying case, take the dot product of (1.189) with $\boldsymbol{v}$ and include the density to yield

$$\frac{1}{2}\frac{\partial \rho_0\boldsymbol{v}^2}{\partial t} + \rho_0\boldsymbol{v}\cdot(\boldsymbol{\omega}\times\boldsymbol{v}) = -\rho_0\boldsymbol{v}\cdot\nabla B. \tag{1.190}$$

The second term on the left-hand side vanishes identically. Defining the kinetic energy density $K$, or energy per unit volume, by $K = \rho_0\boldsymbol{v}^2/2$, (1.190) becomes an expression for the rate of change of $K$,

$$\frac{\partial K}{\partial t} + \nabla\cdot(\rho_0\boldsymbol{v}B) = 0. \tag{1.191}$$

Because $\Phi$ is time-independent this may be written

$$\frac{\partial E}{\partial t} + \nabla\cdot(\rho_0\boldsymbol{v}B) = 0, \tag{1.192}$$

where $E = K + \rho_0\Phi$ is the total energy density (i.e, the total energy per unit volume). This has the form of a general conservation equation in which a local change in a quantity is balanced by the divergence of its flux. However, the energy flux, $\rho_0\boldsymbol{v}B = \rho_0\boldsymbol{v}(\boldsymbol{v}^2/2 + \Phi + \phi)$, is *not* simply the velocity times the energy density $\rho_0(\boldsymbol{v}^2/2+\Phi)$; there is an additional term, $\boldsymbol{v}p$, that represents the energy transfer occurring when work is done by the fluid against the pressure force.

Now consider a volume through which there is no mass flux, for example a domain bounded by rigid walls. The rate of change of energy within that volume is then given by the integral of (1.192),

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V E\,\mathrm{d}V = -\int_V \nabla\cdot(\rho_0\boldsymbol{v}B)\,\mathrm{d}V = -\int_S \rho_0 B\boldsymbol{v}\cdot\,\mathrm{d}\boldsymbol{S} = 0, \tag{1.193}$$

using the divergence theorem. Thus, the total energy within the volume is conserved. The total kinetic energy is also conserved, total gravitational potential energy is equal to $\int_V \rho_0 gz\,\mathrm{d}V$, and this is a constant, unaffected by a rearrangement of the fluid. Thus, in a constant density fluid there is no exchange between kinetic energy and potential energy.

### 1.10.2 Variable Density Fluids

We start with the inviscid momentum equation with a time-independent potential $\Phi$,

$$\rho\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla p - \rho\nabla\Phi, \tag{1.194}$$

and take its dot product with $\boldsymbol{v}$ to obtain an equation for the evolution of kinetic energy,

$$\frac{1}{2}\rho\frac{\mathrm{D}\boldsymbol{v}^2}{\mathrm{D}t} = -\boldsymbol{v}\cdot\nabla p - \rho\boldsymbol{v}\cdot\nabla\Phi = -\nabla\cdot(p\boldsymbol{v}) + p\nabla\cdot\boldsymbol{v} - \rho\boldsymbol{v}\cdot\nabla\Phi. \tag{1.195}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

The internal energy equation for adiabatic flow is

$$\rho \frac{\mathrm{D}I}{\mathrm{D}t} = -p\nabla \cdot \boldsymbol{v}. \tag{1.196}$$

Finally, and somewhat trivially, the potential energy density obeys

$$\rho \frac{\mathrm{D}\Phi}{\mathrm{D}t} = \rho \boldsymbol{v} \cdot \nabla \Phi. \tag{1.197}$$

Adding (1.195), (1.196) and (1.197) we obtain

$$\rho \frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{1}{2}\boldsymbol{v}^2 + I + \Phi\right) = -\nabla \cdot (p\boldsymbol{v}), \tag{1.198}$$

which, on expanding the material derivative and using the mass conservation equation, becomes

$$\frac{\partial}{\partial t}\left[\rho\left(\frac{1}{2}\boldsymbol{v}^2 + I + \Phi\right)\right] + \nabla \cdot \left[\rho \boldsymbol{v}\left(\frac{1}{2}\boldsymbol{v}^2 + I + \Phi + p/\rho\right)\right] = 0. \tag{1.199}$$

This may be written

$$\boxed{\frac{\partial E}{\partial t} + \nabla \cdot [\boldsymbol{v}(E + p)] = 0,} \tag{1.200}$$

where $E = \rho(\boldsymbol{v}^2/2 + I + \Phi)$ is the total energy per unit volume of the fluid. This is the energy equation for an unforced, inviscid and adiabatic, compressible fluid. The energy flux term vanishes when integrated over a closed domain with rigid boundaries, implying that the total energy is conserved. However, there can be an exchange of energy between kinetic, potential and internal components. It is the divergent term, $\nabla \cdot \boldsymbol{v}$, that connects the kinetic energy equation, (1.195), and the internal energy equation, (1.196). In an incompressible fluid this term is absent, and the internal energy is divorced from the other components of energy. This consideration will be important when we consider the Boussinesq equations in Section 2.4. Note finally that the flux of energy, $\boldsymbol{F}_E = \boldsymbol{v}(E + p)$ is not equal to the velocity times the energy; rather, energy is also transferred by pressure. We may write the energy flux as

$$\boxed{\boldsymbol{F}_E = \rho \boldsymbol{v}\left(\frac{\boldsymbol{v}^2}{2} + \Phi + h\right),} \tag{1.201}$$

where $h = I + p/\rho$ is the enthalpy. That is, the local rate of change of energy is effected by the fluxes of kinetic and potential energy and *enthalpy*, not internal energy.

### *Bernoulli's theorem*

The quantity

$$B = \left(E + \frac{p}{\rho}\right) = \left(\frac{1}{2}\boldsymbol{v}^2 + I + \Phi + p\alpha\right) = \left(\frac{1}{2}\boldsymbol{v}^2 + h + \Phi\right), \tag{1.202}$$

is the general form of the Bernoulli function, equal to the sum of the kinetic energy, the potential energy and the enthalpy. Equation (1.200) may be written as

$$\boxed{\frac{\partial E}{\partial t} + \nabla \cdot (\rho \boldsymbol{v} B) = 0.} \tag{1.203}$$

This equation may also be written $\partial(\rho B)/\partial t + \nabla \cdot (\rho \boldsymbol{v} B) = \partial p/\partial t$, so obviously the Bernoulli function itself is not conserved, even for adiabatic flow. For steady flow $\nabla \cdot (\rho \boldsymbol{v}) = 0$, and the $\partial/\partial t$ terms

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

vanish so that (1.203) may be written $\boldsymbol{v}\cdot\nabla B = 0$, or even $\mathrm{D}B/\mathrm{D}t = 0$. The Bernoulli function is then a constant along streamlines, a result commonly known as Bernoulli's theorem.[13] For adiabatic flow at constant composition we also have $\mathrm{D}\theta/\mathrm{D}t = 0$. Thus, steady flow is both along surfaces of constant $\theta$ and along surfaces of constant $B$, and the vector

$$\boldsymbol{l} = \nabla\theta \times \nabla B \tag{1.204}$$

is parallel to streamlines. A related result for unsteady flow is given in Section 4.8.

*Viscous effects*

We might expect that viscosity will always act to reduce the kinetic energy of a flow, and we will demonstrate this for a constant density fluid. Retaining the viscous term in (1.187), the energy equation becomes

$$\frac{\mathrm{d}\widehat{E}}{\mathrm{d}t} \equiv \frac{\mathrm{d}}{\mathrm{d}t}\int_V E\,\mathrm{d}V = \mu\int_V \boldsymbol{v}\cdot\nabla^2\boldsymbol{v}\,\mathrm{d}V. \tag{1.205}$$

The right-hand side is negative definite. To see this we use the vector identity

$$\nabla\times(\nabla\times\boldsymbol{v}) = \nabla(\nabla\cdot\boldsymbol{v}) - \nabla^2\boldsymbol{v}, \tag{1.206}$$

and because $\nabla\cdot\boldsymbol{v} = 0$ we have $\nabla^2\boldsymbol{v} = -\nabla\times\boldsymbol{\omega}$, where $\boldsymbol{\omega}\equiv\nabla\times\boldsymbol{v}$. Thus,

$$\frac{\mathrm{d}\widehat{E}}{\mathrm{d}t} = -\mu\int_V \boldsymbol{v}\cdot(\nabla\times\boldsymbol{\omega})\,\mathrm{d}V = -\mu\int_V \boldsymbol{\omega}\cdot(\nabla\times\boldsymbol{v})\,\mathrm{d}V = -\mu\int_V \boldsymbol{\omega}^2\,\mathrm{d}V, \tag{1.207}$$

after integrating by parts, providing $\boldsymbol{v}\times\boldsymbol{\omega}$ vanishes at the boundary. Thus, viscosity acts to extract kinetic energy from the flow. The loss of kinetic energy reappears as an irreversible warming of the fluid (called 'Joule heating'), and the total energy of the fluid is conserved, but this effect plays no role in a constant density fluid. The heating is normally locally small, at least in the Earth's ocean and atmosphere, but it is needed to preserve total energy.

### 1.10.3 Enthalpy, Static Energy and Energy Flux

We saw above that it is the quantity

$$B = \text{kinetic energy} + \text{potential energy} + \text{enthalpy}, \tag{1.208}$$

that, when multiplied by $\rho\boldsymbol{v}$, fluxes the energy. However, the conserved energy contains only the kinetic and potential energies plus the internal energy. The difference between $B$ and energy is the fluid dynamical analogue of the thermodynamical principle that it is the flux of *enthalpy* that changes the energy of a system, because this accounts for work done by the pressure.

For an example, consider an ideal gas in a uniform gravitational field for which $I = c_v T$ and $p\alpha = RT$ so that $B = c_v T + RT + \mathrm{KE} + gz = c_p T + \mathrm{KE} + gz$. The sum of the enthalpy and the potential energy,

$$h_d^* \equiv c_p T + gz, \tag{1.209}$$

is known as the *dry static energy*. Dry static energy is not an integral conserved quantity nor is it a measure of the total energy itself. To see the importance of enthalpy fluxes, consider an adiabatic rearrangement of a fluid with $\mathrm{d}\eta = 0$. From (1.87) we then have $\mathrm{d}I = -p\,\mathrm{d}\alpha$, meaning that the internal energy changes because of work done. At the same time, from (1.72) with $\mathrm{d}\eta = 0$ we also have $\mathrm{d}h = \alpha\,\mathrm{d}p$. For a fluid in hydrostatic balance, (1.51), we have that $\mathrm{d}p = -g\,\mathrm{d}z/\alpha$ and thus

$$\mathrm{d}(h + gz) = 0. \tag{1.210}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

The change in potential energy of a parcel arises from a change in the enthalpy, not the internal energy, because of the work that must be done by the pressure force to move the parcel.

The form of the conservation law (1.210) is different from the total energy conservation previously derived. Equation (1.200) is an *integral* conservation law, whereas (1.210) is a *parcel* conservation law, which we might write in fluid dynamical form as

$$\frac{\mathrm{D}}{\mathrm{D}t}(h + gz) = 0. \tag{1.211}$$

Equation (1.211) is a form of the thermodynamic equation for adiabatic changes to an ideal gas in hydrostatic balance, similar to (1.126) or (1.141). The equivalence arises because changes in $h_d^*$ and $\theta$ are related by $\mathrm{d}h_d^* = c_p\,\mathrm{d}T + g\,\mathrm{d}z = c_p(T/\theta)\,\mathrm{d}\theta$, as in (1.127).

*Potential enthalpy* (see Section 1.7.3 for an oceanographic discussion) is the enthalpy that a parcel would have if adiabatically moved to a reference pressure, and since the enthalpy for an ideal gas is $c_pT$, the potential enthalpy is just $c_p\theta$. The dry static energy, $c_pT + gz$, is the enthalpy that a parcel at a height $z$ and temperature $T$ would have if moved adiabatically and hydrostatically to $z = 0$, and is thus a special form of potential enthalpy (and sometimes called 'generalized enthalpy'). Energy and enthalpy both play major roles in the chapters ahead, but let's round off this chapter with an introduction to scaling in a pure fluid-dynamical setting.

## 1.11 AN INTRODUCTION TO NONDIMENSIONALIZATION AND SCALING

The units we use to measure length, velocity and so on are irrelevant to the dynamics and it is useful to express the equations of motion in terms of 'nondimensional' variables, by which we mean expressing every variable as the ratio of its value to some reference value. We choose the reference as a natural one for a given flow so that, as far as possible, the nondimensional variables are order-unity quantities, and doing this is called *scaling the equations*. There is no reference that is universally appropriate, and much of the art of fluid dynamics lies in choosing sensible scaling factors for the problem at hand. We introduce the methodology here with a simple example.

### 1.11.1 The Reynolds Number

Consider the constant-density momentum equation in Cartesian coordinates. If a typical velocity is $U$, a typical length is $L$, a typical time scale is $T$, and a typical value of the pressure deviation is $\Phi$, then the approximate sizes of the various terms in the momentum equation are given by

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\nabla\phi + \nu\nabla^2\boldsymbol{v}, \tag{1.212a}$$

$$\frac{U}{T} \qquad \frac{U^2}{L} \sim \frac{\Phi}{L} \qquad \nu\frac{U}{L^2}. \tag{1.212b}$$

The ratio of the inertial terms to the viscous terms is $(U^2/L)/(\nu U/L^2) = UL/\nu$, and this is the *Reynolds number*.[14] More formally, we can nondimensionalize the momentum equation by writing

$$\widehat{\boldsymbol{v}} = \frac{\boldsymbol{v}}{U}, \qquad \widehat{\boldsymbol{x}} = \frac{\boldsymbol{x}}{L}, \qquad \hat{t} = \frac{t}{T}, \qquad \widehat{\phi} = \frac{\phi}{\Phi}, \tag{1.213}$$

where the terms with hats on are *nondimensional* values of the variables and the capitalized quantities are known as *scaling values*, and these are the approximate magnitudes of the variables. We now choose the scaling values so that the nondimensional variables are of order unity, or $\widehat{u} = \mathcal{O}(1)$. Thus, for example, we choose $U$ so that $u = \mathcal{O}(U)$, where this notation should be taken to mean that the magnitude of the variable $u$ is of order $U$, or that $u \sim U$, and we say that '$u$ scales like $U$'.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

Because there are no external forces in this problem, appropriate scaling values for time and pressure are

$$T = \frac{L}{U}, \qquad \Phi = U^2. \tag{1.214}$$

Substituting (1.213) and (1.214) into the momentum equation gives

$$\frac{U^2}{L}\left[\frac{\partial \widehat{\boldsymbol{v}}}{\partial \hat{t}} + (\widehat{\boldsymbol{v}} \cdot \nabla)\widehat{\boldsymbol{v}}\right] = -\frac{U^2}{L}\nabla\widehat{\phi} + \frac{\nu U}{L^2}\nabla^2\widehat{\boldsymbol{v}}, \tag{1.215}$$

where we use the convention that when $\nabla$ operates on a nondimensional variable it is a nondimensional operator. Equation (1.215) then simplifies to

$$\frac{\partial \widehat{\boldsymbol{v}}}{\partial \hat{t}} + (\widehat{\boldsymbol{v}} \cdot \nabla)\widehat{\boldsymbol{v}} = -\nabla\widehat{\phi} + \frac{1}{Re}\nabla^2\widehat{\boldsymbol{v}}, \qquad \text{where} \qquad Re \equiv \frac{UL}{\nu}. \tag{1.216a,b}$$

The parameter $Re$ is, as before, the Reynolds number. If we have chosen our length and velocity scales sensibly — that is, if we have scaled them properly — each variable in (1.216a) is order unity, with the viscous term being multiplied by $1/Re$. There are two important conclusions:

(i) The ratio of the importance of the inertial terms to the viscous terms is given by the Reynolds number, defined above. In the absence of other forces, such as those due to gravity and rotation, the Reynolds number is the only nondimensional parameter explicitly appearing in the momentum equation. Hence its value, along with the boundary conditions and geometry, controls the behaviour of the system.

(ii) More generally, by scaling the equations of motion appropriately the parameters determining the behaviour of the system become explicit. *Scaling the equations is intelligent nondimensionalization.*

Nondimensionalizing the equations does not, however, absolve the investigator from the responsibility of producing dimensionally correct equations. One should regard 'nondimensional' equations as dimensional equations in units appropriate for the problem at hand.

## APPENDIX A: THERMODYNAMICS OF AN IDEAL GAS FROM THE GIBBS FUNCTION

All the thermodynamic quantities of interest for a simple ideal gas, sometimes called a perfect gas, may be derived in a straightforward way from the fundamental equation of state. (An analogous treatment for moist air is given in Appendix A of Chapter 18.) To show this we begin with the specific Gibbs function, $g = I - T\eta + p\alpha$, which for an ideal gas is given by

$$\boxed{g = \mathcal{C}_p T(1 - \ln T) + \mathcal{R}T \ln p + BT + C,} \tag{1.217}$$

where $\mathcal{C}_p$, $\mathcal{R}$, $B$ and $C$ are constants (and our notation anticipates what $\mathcal{C}_p$ and $\mathcal{R}$ really are). The procedure below is especially useful when the Gibbs function is more complex than (1.217), with the main difficulty then lying in obtaining the Gibbs function in the first instance.

### *Density and the thermal equation of state*

From the fundamental relation involving the Gibbs function, (1.78a), specific volume is given by

$$\alpha = \left(\frac{\partial g}{\partial p}\right)_T = \frac{\mathcal{R}T}{p}, \qquad \text{or} \qquad p = \rho\mathcal{R}T. \tag{1.218}$$

Thus, $R$, the gas constant used elsewhere in this chapter, is equal to $\mathcal{R}$.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

### Entropy

$$\eta = -\left(\frac{\partial g}{\partial T}\right)_p = \mathcal{C}_p \ln T - \mathcal{R} \ln p - B \tag{1.219}$$

Using (1.218) and (1.221) in (1.219), the entropy may be expressed in terms of $I$ and $\alpha$, giving $\eta = (\mathcal{C}_p - R) \ln I - \mathcal{R} \ln \rho$ + constant, as in (1.111).

### The internal energy

Using (1.218) and (1.219) in the definition of the Gibbs function, $g = I - T\eta + p\alpha$, gives

$$I = g + T\eta - p\alpha = g - T\left(\frac{\partial g}{\partial T}\right)_p - p\left(\frac{\partial g}{\partial p}\right)_T, \tag{1.220}$$

and so

$$I = (\mathcal{C}_p - \mathcal{R})T + C = \mathcal{C}_v T + C, \tag{1.221}$$

where $\mathcal{C}_v \equiv \mathcal{C}_p - \mathcal{R}$. Equation (1.221) also suggests we may sensibly take $C = 0$, but no physical result in classical mechanics depends on this choice.

### Heat capacities

Let $c_p$ be the heat capacity at constant pressure. We have

$$c_p \equiv T\left(\frac{\partial \eta}{\partial T}\right)_p = \mathcal{C}_p. \tag{1.222}$$

To obtain the heat capacity at constant volume, $c_v$, first rewrite the entropy using the thermal equation of state as

$$\eta = (\mathcal{C}_p \ln T - \mathcal{R} \ln T + \mathcal{R} \ln \alpha) + \text{constant}. \tag{1.223}$$

We then have

$$c_v \equiv T\left(\frac{\partial \eta}{\partial T}\right)_\alpha = \mathcal{C}_p - \mathcal{R}. \tag{1.224}$$

### Adiabatic lapse rate

The adiabatic lapse rate, $\Gamma_p$, is the rate of change of temperature with pressure at constant entropy. Thus

$$\Gamma_p \equiv \left(\frac{\partial T}{\partial p}\right)_\eta, \tag{1.225}$$

Now, from (1.138) we have

$$\left(\frac{\partial T}{\partial p}\right)_\eta = -\frac{T}{\mathcal{C}_p}\left(\frac{\partial \eta}{\partial p}\right)_T, \tag{1.226}$$

and using (1.219) to evaluate the right-hand side gives

$$\Gamma_p = -\frac{T}{\mathcal{C}_p}\left(\frac{\partial \eta}{\partial p}\right)_T = \frac{T}{\mathcal{C}_p}\frac{\mathcal{R}}{p} = \frac{1}{\mathcal{C}_p \rho}, \tag{1.227}$$

using the thermal equation of state. It is common to write this expression in terms of a rate of change of temperature with respect to height by using the hydrostatic approximation, $\mathrm{d}p = -\rho g\, \mathrm{d}z$, whence

$$\Gamma_z \equiv -\left(\frac{\partial T}{\partial z}\right)_\eta = \frac{g}{\mathcal{C}_p}. \tag{1.228}$$

The physical significance of this quantity is explored in Chapter 2.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

*Potential temperature*

As in (1.136), potential temperature, $\theta$, satisfies

$$\eta(T, p) = \eta(\theta, p_R). \tag{1.229}$$

That is to say, $\theta$ is the temperature that a parcel will have if moved at constant entropy from a pressure $p$ to a reference pressure $p_R$. Using (1.219) gives

$$\mathcal{C}_p \ln T - \mathcal{R} \ln p = \mathcal{C}_p \ln \theta - \mathcal{R} \ln p_R, \qquad \text{or} \qquad \ln(T/\theta)^{\mathcal{C}_p} = \ln(p/p_R)^{\mathcal{R}}. \tag{1.230}$$

Re-arranging gives

$$\theta = T \left( \frac{p_R}{p} \right)^{\mathcal{R}/\mathcal{C}_p}. \tag{1.231}$$

We can derive the same result by noting that, by definition, potential temperature satisfies

$$\theta \equiv T(p_R) = T(p) + \int_p^{p_R} \left( \frac{\partial T}{\partial p'} \right)_\eta \mathrm{d}p' = T(p) + \int_p^{p_R} \frac{\mathcal{R}T}{\mathcal{C}_p p'} \mathrm{d}p', \tag{1.232}$$

where the rightmost expression uses (1.227). It is easy to verify that the solution to this integral equation is $T = \theta(p/p_R)^{\mathcal{R}/\mathcal{C}_p}$, although solving the equation *ab initio* is a little more difficult. Finally, in an ideal gas potential temperature can be related to entropy using (1.219) and (1.231), giving $\eta = \mathcal{C}_p \ln \theta + \text{constant}$.

*Enthalpy and potential enthalpy*

Enthalpy is related to the Gibbs function by

$$h = g + T\eta = g - T \left( \frac{\partial g}{\partial T} \right)_p = \mathcal{C}_p T. \tag{1.233}$$

Potential enthalpy, $h^0$, is the enthalpy that a parcel would have if moved adiabatically to a reference pressure. It therefore satisfies an equation similar to (1.229), namely $\eta(h, p) = \eta(h^0, p_R)$. Since $h = \mathcal{C}_p T$ we immediately obtain

$$h^0 = \mathcal{C}_p \theta. \tag{1.234}$$

That is, the potential enthalpy is equal to the potential temperature times the heat capacity at constant pressure.

## APPENDIX B: THE FIRST LAW OF THERMODYNAMICS FOR FLUIDS

In its usual form the first law states that changes in the internal energy of a body are equal to the sum of heat supplied, the work done, and changes due to composition (the chemical work), and in a reversible process the entropy of a body then changes according to the heat supplied. However, the heating is not known a priori and the first law may be regarded as a definition of heating by way of energy conservation; heating should then be considered a derived quantity, as we noted in our discussion of (1.61) and (1.62). The real problem is to determine what the heating actually is, and this is not wholly trivial for fluids that also have mechanical energy and viscosity, as well as thermal and compositional diffusion. The first law is not a statement of total energy conservation, since it does not involve kinetic energy.

To obtain an unambiguous prescription for the heating — and hence a useful thermodynamic equation — we *begin* with total energy conservation and work backwards to obtain the heating, and since energy conservation is the more fundamental physical law this procedure is natural.[15]

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

Effectively, we subtract off the evolution of mechanical energy from an equation for total energy conservation, and the remainder is the thermodynamic equation. Equivalently (and this is how we proceed below) we demand consistency between total energy conservation and an energy equation derived in a forward fashion using a thermodynamic equation representing the first law but with the heating unspecified, and thereby deduce an explicit expression for that heating. Underlying this procedure is the fundamental notion that the work done and the generation of mechanical energy are less ambiguous than heating because they can be measured and/or arise through well-defined forces in the momentum equation. We will assume that energy and energy fluxes are knowable or calculable, but we will not discuss the questions of what energy fundamentally is or why it is conserved.

For reference, we first write down the fundamental thermodynamic relation, (1.85), and its inputs,

$$\frac{\mathrm{D}I}{\mathrm{D}t} + p\frac{\mathrm{D}\alpha}{\mathrm{D}t} = T\frac{\mathrm{D}\eta}{\mathrm{D}t} + \mu\frac{\mathrm{D}S}{\mathrm{D}t} = \frac{\mathrm{D}h}{\mathrm{D}t} - \alpha\frac{\mathrm{D}p}{\mathrm{D}t} = \dot{Q}_E, \tag{1.235}$$

where $\dot{Q}_E = \dot{Q} + \mu\dot{S}$ accounts for the total energy input to a fluid parcel from both heating, $\dot{Q}$, and compositional changes, $\mu\dot{S}$. The inclusion of the term $\dot{Q}_E$ connects the above equation to the first law, but the first law is not useful until we know what the heating is.

## B.1 Single Component Fluid

Consider the energetics of a fluid as in Section 1.10.2, with two additional effects: the fluid is viscous, and there is an additional energy source in the fluid, for example radiation or thermal conduction, that does not appear in the momentum equation. We write the momentum equation, (1.194), in Cartesian tensor notation (where repeated indices are summed) as

$$\rho\frac{\mathrm{D}v_i}{\mathrm{D}t} = -\frac{\partial p}{\partial x_i} - \rho\frac{\partial \Phi}{\partial x_i} + \mu_v\frac{\partial^2 v_i}{\partial x_j \partial x_j}, \tag{1.236}$$

where $\mu_v$ is the coefficient of viscosity and $\Phi$ is a time-independent external potential field, such as $gz$. The form of the viscous term is exact only for incompressible fluids with constant viscosity but it is usually an excellent approximation in the atmosphere and ocean, because the Mach number is small and the scale on which dissipation occurs is very much smaller than the density scale height. (In any case those restrictions can be relaxed.) We also write the entropy equation as

$$T\frac{\mathrm{D}\eta}{\mathrm{D}t} = \dot{Q}, \tag{1.237}$$

where $\dot{Q}$ is the heating, whose form we do not yet know. In a single component fluid we can use the fundamental thermodynamic relation, (1.235), to write (1.237) as an internal energy equation,

$$\rho\frac{\mathrm{D}I}{\mathrm{D}t} + p\nabla\cdot\boldsymbol{v} = \rho\dot{Q}, \tag{1.238}$$

also having used $\mathrm{D}\alpha/\mathrm{D}t = \alpha\nabla\cdot\boldsymbol{v}$.

We obtain a kinetic energy equation by multiplying (1.236) by $v_i$ to give

$$\frac{1}{2}\rho\frac{\mathrm{D}v_i^2}{\mathrm{D}t} = -\partial_i(pv_i) + p\partial_i v_i - \rho v_i\partial_i\Phi + \mu_v v_i\partial_j(\partial_j v_i), \tag{1.239}$$

where $\partial_i \equiv \partial/\partial x_i$. The viscous term may be written as

$$\mu_v v_i\partial_j(\partial_j v_i) = \mu_v\left[\partial_j(v_i\partial_j v_i) - (\partial_j v_i)^2\right]. \tag{1.240}$$

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

The first term on the right-hand side is the divergence of a flux, and so is energy conserving, and the second term is negative definite, representing kinetic energy dissipation.

We now proceed, just as in Section 1.10.2, to obtain a total energy equation from (1.238) and (1.239) and the result is

$$\frac{\partial}{\partial t}\left[\rho\left(\frac{1}{2}v_i^2 + I + \Phi\right)\right] + \partial_i\left[\rho v_i\left(\frac{1}{2}v_j^2 + I + \Phi + p/\rho\right)\right] = \mu_\nu\left[\partial_j(v_i\partial_j v_i) - (\partial_j v_i)^2\right] + \rho\dot{Q}. \quad (1.241)$$

Now, the general form of the energy conservation law for a fluid takes the form

$$\frac{\partial}{\partial t}\left[\rho\left(\frac{1}{2}v_i^2 + I + \Phi\right)\right] + \partial_i\left[\rho v_i\left(\frac{1}{2}v_j^2 + I + \Phi + p/\rho\right)\right] = \partial_j\left[\mu_\nu v_i\partial_j v_i + F_{Ej}\right], \quad (1.242)$$

where $F_{Ej}$ (or $\boldsymbol{F}_E$, and $\partial_j F_{Ej} = \nabla\cdot\boldsymbol{F}_E$) is the total energy flux due to radiation, conduction and any other effects — the important point being that the right-hand side of (1.242) must be the divergence of a flux in order to guarantee energy conservation. We have not used the first law to obtain (1.242), just the fundamental thermodynamic relation, the momentum equation and energy conservation. The above two equations are consistent only if the heating term has the form $\rho\dot{Q} = \partial_j F_{Ej} + \mu_\nu(\partial_j v_i)^2$, and the internal energy and entropy equations are then

$$\rho\frac{\mathrm{D}I}{\mathrm{D}t} + p\nabla\cdot\boldsymbol{v} = \nabla\cdot\boldsymbol{F}_E + \mu_\nu(\partial_j v_i)^2, \qquad \rho T\frac{\mathrm{D}\eta}{\mathrm{D}t} = \nabla\cdot\boldsymbol{F}_E + \mu_\nu(\partial_j v_i)^2. \quad (1.243a,b)$$

Evidently, the 'heating' of a fluid is given by the sum of the energy fluxes and a positive definite term due to viscous dissipation. Either of the above equivalent equations may be considered to be statements of the first law with explicit expressions for energy input and heating, and either of them then provides a useful predictive thermodynamic equation for a fluid.

## B.2 Multi-Component Fluids

Consider now a two component fluid, such as dry air and water vapour or water and salinity. We refer to the second component as concentration and we assume it obeys

$$\rho\frac{\mathrm{D}S}{\mathrm{D}t} = \nabla\cdot\boldsymbol{F}_S. \quad (1.244)$$

The only difference from the previous derivation is that, using the fundamental relation, the internal energy equation is now

$$\rho\frac{\mathrm{D}I}{\mathrm{D}t} + p\nabla\cdot\boldsymbol{v} = \dot{Q} + \mu\nabla\cdot\boldsymbol{F}_S, \quad (1.245)$$

where $\mu$ is the chemical potential and the second term on the right-hand side accounts for the effects of concentration fluxes on internal energy (i.e., the chemical work). Proceeding to calculate the energy equation as before, we find that (1.241) then has an additional term $\mu\nabla\cdot\boldsymbol{F}_S$ on the right-hand side. This is consistent with (1.242) only if the internal energy and entropy equations obey

$$\rho\frac{\mathrm{D}I}{\mathrm{D}t} + p\nabla\cdot\boldsymbol{v} = \nabla\cdot\boldsymbol{F}_E + \mu_\nu(\partial_j v_i)^2, \qquad \rho T\frac{\mathrm{D}\eta}{\mathrm{D}t} = \nabla\cdot\boldsymbol{F}_E - \mu\nabla\cdot\boldsymbol{F}_S + \mu_\nu(\partial_j v_i)^2, \quad (1.246a,b)$$

where the energy flux, $\boldsymbol{F}_E$, now includes the effects of any fluxes of composition. Again, either of the above two equations is a statement of the first law, and the right-hand side of (1.246b) is the heating. Additional terms may appear on the right-hand sides of the above equations if there are additional source or sink terms, for example a source of concentration.

Equation (1.246) differs from the single component case by the addition of a concentration flux. However, in a two component fluid the diffusion of temperature and of concentration are affected by gradients of both temperature and concentration, so that the heat flux itself differs from the single component case. Thermodynamics provides constraints on these fluxes, but the reader must look elsewhere to learn about them.[16] Finally, in both oceanography and meteorology the viscous heating term is small, at least on Earth, but it must be included if energy balance is desired — for example if incoming solar radiation is to balance outgoing infrared radiation.

## Notes

1 Joseph-Louis Lagrange (1736–1813) was a Franco-Italian, born and raised in Turin who then lived and worked mainly in Germany and France. He made notable contributions in analysis, number theory and mechanics and was recognized as one of the greatest mathematicians of the eighteenth century. He laid the foundations of the calculus of variations (to wit, the 'Lagrange multiplier') and first formulated the principle of least action, and his treatise *Mécanique Analytique* (1788) provides a unified analytic framework (it contains no diagrams, a feature emulated in Whittaker's *Treatise on Analytical Dynamics*, 1927) for all Newtonian mechanics.

Leonard Euler (1707–1783), a Swiss mathematician who lived and worked for extended periods in Berlin and St. Petersburg, made important contributions in many areas of mathematics and mechanics, including the analytical treatment of algebra, the theory of equations, calculus, number theory and classical mechanics. He was the first to establish the form of the equations of motion of fluid mechanics, writing down both the field description of fluids *and* what we now call the material or advective derivative.

Truesdell (1954) points out that 'Eulerian' and 'Lagrangian', especially the latter, are inappropriate eponyms. The Eulerian description was introduced by d'Alembert in 1749 and generalized by Euler in 1752, and the so-called Lagrangian description was introduced by Euler in 1759. (It is sometimes said that advances in mathematics are named after the next person to discover them after Euler — the Coriolis effect is another example.) The modern confusion evidently stems from a monograph by Dirichlet in 1860 that credits Euler in 1757 and Lagrange in 1788 for the respective methods.

Clifford Truesdell (1919–2000) was a remarkable figure himself, known both for his own contributions to many areas of continuum mechanics and for his scholarly investigations on the history of mathematics and science. He also had a trenchant and at times pungent writing style. Ball & James (2002) provide a biography.

2 Salinity is a mass fraction and thus is nondimensional, but it is commonly referred to in units of g/kg. For many years the measure of salinity of seawater that was used in oceanography was based on electrical conductivity and referred to as 'practical salinity', $S_P$, since this was (and still is) more easily measured. In thermodynamical calculations practical salinity is now largely dropped in favour of the true salinity, generally referred to as absolute salinity and denoted $S_A$. Differences between practical and absolute salinity are small but not negligible (Millero *et al.* 2008, IOC *et al.* 2010).

3 See also de Szoeke (2004). Nycander & Roquet (2015) and Roquet *et al.* (2015) show that this equation of state can, in fact, be used to give a quantitatively accurate simulation of the ocean.

4 The use of inexact differentials in thermodynamics is questionable for they do not have a straightforward mathematical foundation. Their use can be avoided and, were this a rigorous treatise on thermodynamics, probably should be avoided, but here they are useful artifacts. Reif (1965) and Callen (1985) both make use of them but Truesdell (1969) is particularly scathing on the matter.

5 It is said that the early students of ideas related to entropy were unusually prone to suicide, Ludwig Boltzmann being a tragic example. Thankfully there are many counter-examples, such as William Thompson (Lord Kelvin). He did foundational work early in his career on thermodynamics and among other achievements put forward a formulation of the second law. Neither this nor his much less successful later work seems to have caused him too much distress, and he lived for 83 years. Perhaps Truesdell (1969) gets it right when he says that entropy gives 'intense headaches to those who have studied thermodynamics'.

6 Because the word barotropic has other meanings — sometimes it is just taken to mean the vertical average — it might be better to always refer to fluids for which density is a function only of pressure as homentropic. Unfortunately the current usage is deeply ingrained and to insist on homentropic would be tilting at windmills.

7 Claude-Louis-Marie-Henri Navier (1785–1836) was a French civil engineer, professor at the École Polytechnique and later at the École des Ponts et Chaussée. He was an expert in road and bridge building (he developed the theory of suspension bridges) and, relatedly, made lasting theoretical contributions to the theory of elasticity, being the first to publish a set of general equations for the dynamics of an elastic solid. In fluid mechanics, he laid down the now-called *Navier–Stokes equations*, including the viscous terms, in 1822.

George Gabriel Stokes (1819–1903). Irish born (in Skreen, County Sligo), he was a professor of mathematics at Cambridge from 1849 until his retirement. As well as having a role in the development of fluid mechanics, especially through his considerations of viscous effects, Stokes worked on the dynamics of elasticity, fluorescence, the wave theory of light, and was (perhaps rather ill-advisedly in hindsight) a proponent of the idea of an ether permeating all space.

8 Potential temperature was known to William Thomson in 1857.

9 Jackett & McDougall (1997), extending McDougall (1987). A similar quantity was described by Eden & Willebrand (1999). de Szoeke (2000), Nycander (2011) and Tailleux (2016) provide more discussion.

10 Building from de Szoeke (2004) and with input from W. R. Young. See also Fofonoff (1959) and Warren (2006) for some historical background. A very accurate semi-empirical formula for the Gibbs function is given by Feistel (2008). Using this as a basis, seawater equations of state are now available in the form of the TEOS-10 standard (IOC *et al.* 2010) and from Roquet *et al.* (2015), and these fit laboratory measurements close to the accuracy of the measurements themselves.

11 Potential enthalpy was introduced to oceanography by McDougall (2003) and its use is advocated in IOC *et al.* (2010). The advantages and disadvantages of various thermodynamic variables, including entropy and potential enthalpy, are discussed there and in Graham & McDougall (2013). Useful discussion is also to be found in Warren (1999), Young (2010) and Nycander (2011). I am very grateful to T. McDougall for discussions on these and other thermodynamic matters.

12 Referring to the 'heat content' of a fluid borders on dangerous language, because heat itself is a type of energy transfer, like work, and not a state variable. On the other hand, heat content has an intuitive appeal and, provided it is properly understood, conveys a useful meaning in oceanography, since the compression work done on the ocean is small (as water is almost incompressible) and kinetic energy is small compared to internal and potential energy.

13 Bernoulli's theorem was developed mainly by Daniel Bernoulli (1700–1782). It was based on earlier work on the conservation of energy that Daniel had done with his father, Johann Bernoulli (1667–1748), and so perhaps should be known as Bernoullis' theorem. The two men fell out when Daniel was a young man, reputedly because of Johann's jealousy of Daniel's abilities, and subsequently had a very strained relationship. The Bernoulli family produced several (at least eight) talented mathematicians over three generations in the seventeenth and eighteenth centuries, and is often regarded as the most mathematically distinguished family of all time.

14 Osborne Reynolds (1842–1912) was an Irish born (Belfast) physicist who was professor of engineering at Manchester University from 1868–1905. His early work was in electricity and magnetism, but he is now most famous for his work in hydrodynamics. The 'Reynolds number', which determines the ratio of inertial to viscous forces, and the 'Reynolds stress', which is the stress on the mean flow due to the fluctuating components, are both named after him. He was also one of the first scientists to think about the concept of group velocity.

15 See also Landau & Lifshitz (1987) and IOC *et al.* (2010). For an interesting and somewhat idiosyncratic view of the first law applied to the ocean, read Warren (2006).

16 Onsager (1931), Salmon (1998). For example, in a single component fluid, total entropy increases if the heat flux is proportional to a downgradient temperature flux.

https://doi.org/10.1017/9781107588417.002 Published online by Cambridge University Press

### *Further Reading*

#### General fluid dynamics

There are numerous books on hydrodynamics, an early one being

Lamb, H., 1932. *Hydrodynamics.*
Lamb's book is a classic in the field, although now too dated to make it useful as an introduction.

Two somewhat more modern references, at a fairly advanced level, are

Batchelor, G. K., 1967. *An Introduction to Fluid Dynamics.*

Landau, L. D. & Lifshitz, E. M., 1987. *Fluid Mechanics.*
These two books both contain a detailed derivation of the equations of motion, including viscous and pressure forces.

At a more elementary level we have

Kundu, P., Cohen, I. &. Dowling, D., 2015. *Fluid Mechanics.*
This book is written at the advanced undergraduate/beginning graduate level, is easier-going than Batchelor or Landau & Lifshitz, and contains material on geophysical fluid dynamics.

For the connoisseur, a more specialized treatment is

Truesdell, C., 1954. *The Kinematics of Vorticity.*
Written in Truesdell's inimitable style, this book discusses many aspects of vorticity with numerous historical references. Truesdell's books are all gems in their own way.

#### Thermodynamics

There are many books on thermodynamics, and two that I have found particularly useful are

Reif, F., 1965. *Fundamentals of Statistical and Thermal Physics.*

Callen, H. B., 1985. *Thermodynamics and an Introduction to Thermostatistics.*
Reif's book has become something of a classic, and Callen provides an axiomatic approach that will be an antidote for those who feel that thermodynamic reasoning is mysterious or even circular.

For the subtopic of atmospheric thermodynamics see the further reading section at the end of Chapter 14.

#### Geophysical fluid dynamics

Gill, A. E., 1982. *Atmosphere–Ocean Dynamics.*
A richly textured book, especially strong on equatorial dynamics and gravity wave motion.

Pedlosky, J., 1987. *Geophysical Fluid Dynamics.*
A primary reference for flow at low Rossby number. Although the book requires some effort, there is a handsome pay-off for those who study it closely.

Holton, J. R. & Hakim, G., 2012. *An Introduction to Dynamical Meteorology.*
A very well-known textbook at the undergraduate/beginning graduate level.

Salmon, R., 1998. *Lectures on Geophysical Fluid Dynamics.*
Covers the fundamentals as well as Hamiltonian fluid dynamics, geostrophic turbulence and oceanic circulation.

*The heavens themselves, the planets, and this centre*
*Observe degree, priority, and place,*
*Insisture, course, proportion, season, form,*
*Office, and custom, in all line of order.*
*And therefore is the glorious planet Sol*
*In noble eminence enthroned and sphered.*
William Shakespeare, *Troilus and Cressida*, c. 1602.

*Eppur si muove. (And yet it does move.)*
Galileo Galilei, apocryphal, 1633.

# CHAPTER 2

# Effects of Rotation and Stratification

THE ATMOSPHERE AND OCEAN are shallow layers of fluid on a sphere, 'shallow' because their thickness is much less than their horizontal extent. Their motion is strongly influenced by two effects: rotation and stratification, the latter meaning that there is a mean vertical gradient of (potential) density that is often large compared with the horizontal gradient. Here we consider how the equations of motion are affected by these effects. First, we consider some elementary effects of rotation on a fluid and derive the Coriolis and centrifugal forces, and write down the equations of motion appropriate for motion on a sphere. Then we discuss some approximations to the equations of motion that are appropriate for large-scale flow in the ocean and atmosphere, in particular the hydrostatic and geostrophic approximations, and finally we look at the possible static instability of stratified flows.

## 2.1 EQUATIONS OF MOTION IN A ROTATING FRAME

Newton's second law of motion, that the acceleration of a body is proportional to the imposed force divided by the body's mass, applies in so-called inertial frames of reference; that is, frames that are stationary or moving only with a constant rectilinear velocity relative to the distant galaxies. Now Earth spins round its own axis with a period of almost 24 hours (23h 56m, the difference due to Earth's rotation around the Sun) and so the surface of the Earth manifestly is not an inertial frame. Nevertheless, it is very convenient to describe the flow relative to Earth's surface (which in fact is moving at speeds of up to a few hundreds of metres per second), rather than in some inertial frame.[1] This necessitates recasting the equations into a form appropriate in a rotating frame of reference, and that is the subject of this section.

### 2.1.1 Rate of Change of a Vector

Consider first a vector $\boldsymbol{C}$ of constant length rotating relative to an inertial frame at a constant angular velocity $\boldsymbol{\Omega}$. Then, in a frame rotating with that same angular velocity it appears stationary and constant. If in a small interval of time $\delta t$ the vector $\boldsymbol{C}$ rotates through a small angle $\delta\lambda$ then the change in $\boldsymbol{C}$, as perceived in the inertial frame, is given by (see Fig. 2.1)

$$\delta \boldsymbol{C} = |\boldsymbol{C}| \cos\vartheta \, \delta\lambda \, \boldsymbol{m}, \tag{2.1}$$



https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

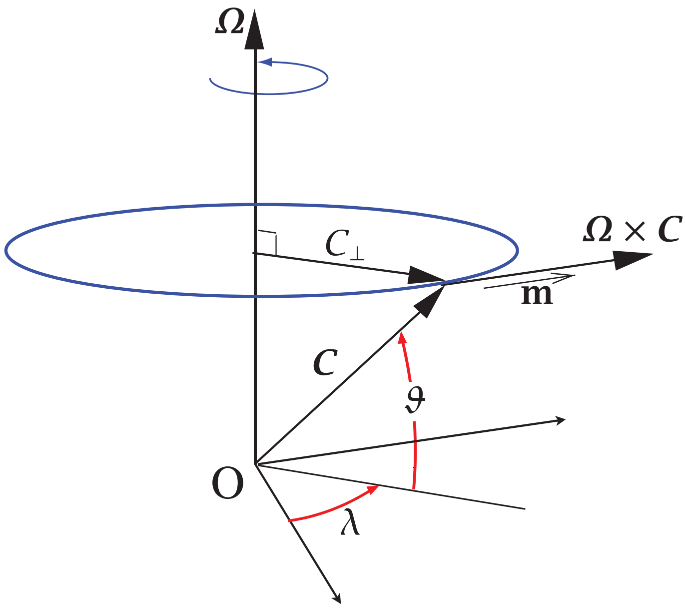


**Fig. 2.1** A vector $\boldsymbol{C}$ rotating at an angular velocity $\boldsymbol{\Omega}$. It appears to be a constant vector in the rotating frame, whereas in the inertial frame it evolves according to $(\mathrm{d}\boldsymbol{C}/\mathrm{d}t)_I = \boldsymbol{\Omega}\times\boldsymbol{C}$.

where the vector $\boldsymbol{m}$ is the unit vector in the direction of change of $\boldsymbol{C}$, which is perpendicular to both $\boldsymbol{C}$ and $\boldsymbol{\Omega}$. But the rate of change of the angle $\lambda$ is just, by definition, the angular velocity so that $\delta\lambda = |\boldsymbol{\Omega}|\delta t$ and

$$\delta\boldsymbol{C} = |\boldsymbol{C}||\boldsymbol{\Omega}| \sin\widehat{\vartheta}\, \boldsymbol{m}\, \delta t = \boldsymbol{\Omega}\times\boldsymbol{C}\,\delta t, \tag{2.2}$$

using the definition of the vector cross product, where $\widehat{\vartheta} = (\pi/2 - \vartheta)$ is the angle between $\boldsymbol{\Omega}$ and $\boldsymbol{C}$. Thus

$$\left(\frac{\mathrm{d}\boldsymbol{C}}{\mathrm{d}t}\right)_I = \boldsymbol{\Omega}\times\boldsymbol{C}, \tag{2.3}$$

where the left-hand side is the rate of change of $\boldsymbol{C}$ as perceived in the inertial frame.

Now consider a vector $\boldsymbol{B}$ that changes in the inertial frame. In a small time $\delta t$ the change in $\boldsymbol{B}$ as seen in the rotating frame is related to the change seen in the inertial frame by

$$(\delta\boldsymbol{B})_I = (\delta\boldsymbol{B})_R + (\delta\boldsymbol{B})_{rot}, \tag{2.4}$$

where the terms are, respectively, the change seen in the inertial frame, the change due to the vector itself changing as measured in the rotating frame, and the change due to the rotation. Using (2.2) $(\delta\boldsymbol{B})_{rot} = \boldsymbol{\Omega}\times\boldsymbol{B}\,\delta t$, and so the rates of change of the vector $\boldsymbol{B}$ in the inertial and rotating frames are related by

$$\boxed{\left(\frac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t}\right)_I = \left(\frac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t}\right)_R + \boldsymbol{\Omega}\times\boldsymbol{B}.} \tag{2.5}$$

This relation applies to a vector $\boldsymbol{B}$ that, as measured at any one time, is the same in both inertial and rotating frames.

### 2.1.2 Velocity and Acceleration in a Rotating Frame

The velocity of a body is not measured to be the same in the inertial and rotating frames, so care must be taken when applying (2.5) to velocity. First apply (2.5) to $\boldsymbol{r}$, the position of a particle to obtain

$$\left(\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right)_I = \left(\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right)_R + \boldsymbol{\Omega}\times\boldsymbol{r} \tag{2.6}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

or

$$\boldsymbol{v}_I = \boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r}. \tag{2.7}$$

We refer to $\boldsymbol{v}_R$ and $\boldsymbol{v}_I$ as the relative and inertial velocity, respectively, and (2.7) relates the two. Apply (2.5) again, this time to the velocity $\boldsymbol{v}_R$ to give

$$\left(\frac{\mathrm{d}\boldsymbol{v}_R}{\mathrm{d}t}\right)_I = \left(\frac{\mathrm{d}\boldsymbol{v}_R}{\mathrm{d}t}\right)_R + \boldsymbol{\Omega} \times \boldsymbol{v}_R, \tag{2.8}$$

or, using (2.7)

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{v}_I - \boldsymbol{\Omega} \times \boldsymbol{r})\right)_I = \left(\frac{\mathrm{d}\boldsymbol{v}_R}{\mathrm{d}t}\right)_R + \boldsymbol{\Omega} \times \boldsymbol{v}_R, \tag{2.9}$$

or

$$\left(\frac{\mathrm{d}\boldsymbol{v}_I}{\mathrm{d}t}\right)_I = \left(\frac{\mathrm{d}\boldsymbol{v}_R}{\mathrm{d}t}\right)_R + \boldsymbol{\Omega} \times \boldsymbol{v}_R + \frac{\mathrm{d}\boldsymbol{\Omega}}{\mathrm{d}t} \times \boldsymbol{r} + \boldsymbol{\Omega} \times \left(\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right)_I. \tag{2.10}$$

Then, noting that

$$\left(\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right)_I = \left(\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right)_R + \boldsymbol{\Omega} \times \boldsymbol{r} = (\boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r}), \tag{2.11}$$

and assuming that the rate of rotation is constant, (2.10) becomes

$$\left(\frac{\mathrm{d}\boldsymbol{v}_R}{\mathrm{d}t}\right)_R = \left(\frac{\mathrm{d}\boldsymbol{v}_I}{\mathrm{d}t}\right)_I - 2\boldsymbol{\Omega} \times \boldsymbol{v}_R - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \boldsymbol{r}). \tag{2.12}$$

This equation may be interpreted as follows. The term on the left-hand side is the rate of change of the relative velocity as measured in the rotating frame. The first term on the right-hand side is the rate of change of the inertial velocity as measured in the inertial frame (the inertial acceleration, which is, by Newton's second law, equal to the force on a fluid parcel divided by its mass). The second and third terms on the right-hand side (including the minus signs) are the *Coriolis force* and the *centrifugal force* per unit mass. Neither of these is a true force — they may be thought of as quasi-forces (i.e., 'as if' forces); that is, when a body is observed from a rotating frame it behaves as if unseen forces are present that affect its motion. If (2.12) is written, as is common, with the terms $+2\boldsymbol{\Omega} \times \boldsymbol{v}_r$ and $+\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \boldsymbol{r})$ on the left-hand side then these terms should be referred to as the Coriolis and centrifugal *accelerations.*[2]

### Centrifugal force

If $\boldsymbol{r}_\perp$ is the perpendicular distance from the axis of rotation (see Fig. 2.1 and substitute $\boldsymbol{r}$ for $\boldsymbol{C}$), then, because $\boldsymbol{\Omega}$ is perpendicular to $\boldsymbol{r}_\perp$, $\boldsymbol{\Omega} \times \boldsymbol{r} = \boldsymbol{\Omega} \times \boldsymbol{r}_\perp$. Then, using the vector identity $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \boldsymbol{r}_\perp) = (\boldsymbol{\Omega} \cdot \boldsymbol{r}_\perp)\boldsymbol{\Omega} - (\boldsymbol{\Omega} \cdot \boldsymbol{\Omega})\boldsymbol{r}_\perp$ and noting that the first term is zero, we see that the centrifugal force per unit mass is just given by

$$\boldsymbol{F}_{ce} = -\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \boldsymbol{r}) = \Omega^2 \boldsymbol{r}_\perp. \tag{2.13}$$

This may usefully be written as the gradient of a scalar potential,

$$\boldsymbol{F}_{ce} = -\nabla \Phi_{ce}, \tag{2.14}$$

where $\Phi_{ce} = -(\Omega^2 r_\perp^2)/2 = -(\boldsymbol{\Omega} \times \boldsymbol{r}_\perp)^2/2$.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

*Coriolis force*

The Coriolis force per unit mass is given by

$$\boldsymbol{F}_{Co} = -2\boldsymbol{\Omega} \times \boldsymbol{v}_R. \tag{2.15}$$

It plays a central role in much of geophysical fluid dynamics and will be considered extensively later on. For now, we just note three basic properties:

*(i)* There is no Coriolis force on bodies that are stationary in the rotating frame.
*(ii)* The Coriolis force acts to deflect moving bodies at right angles to their direction of travel.
*(iii)* The Coriolis force does no work on a body because it is perpendicular to the velocity, and so $\boldsymbol{v}_R \cdot (\boldsymbol{\Omega} \times \boldsymbol{v}_R) = 0$.

### 2.1.3 Momentum Equation in a Rotating Frame

Since (2.12) simply relates the accelerations of a particle in the inertial and rotating frames, then in the rotating frame of reference the momentum equation may be written

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\boldsymbol{\Omega} \times \boldsymbol{v} = -\frac{1}{\rho}\nabla p - \nabla\Phi, \tag{2.16}$$

incorporating the centrifugal term into the potential, $\Phi$. We have dropped the subscript $R$; henceforth, unless we need to be explicit (as in the next section), all velocities without a subscript will be considered to be relative to the rotating frame.

### 2.1.4 Mass and Tracer Conservation in a Rotating frame

Let $\varphi$ be a scalar field that, in the inertial frame, obeys

$$\frac{\mathrm{D}\varphi}{\mathrm{D}t} + \varphi\nabla \cdot \boldsymbol{v}_I = 0. \tag{2.17}$$

Now, observers in both the rotating and inertial frame measure the same value of $\varphi$. Further, $\mathrm{D}\varphi/\mathrm{D}t$ is simply the rate of change of $\varphi$ associated with a material parcel, and therefore is reference frame invariant. Thus, without further ado, we write

$$\left(\frac{\mathrm{D}\varphi}{\mathrm{D}t}\right)_R = \left(\frac{\mathrm{D}\varphi}{\mathrm{D}t}\right)_I, \tag{2.18}$$

where $(\mathrm{D}\varphi/\mathrm{D}t)_R = (\partial\varphi/\partial t)_R + \boldsymbol{v}_R \cdot \nabla\varphi$ and $(\mathrm{D}\varphi/\mathrm{D}t)_I = (\partial\varphi/\partial t)_I + \boldsymbol{v}_I \cdot \nabla\varphi$, and the local temporal derivatives $(\partial\varphi/\partial t)_R$ and $(\partial\varphi/\partial t)_I$ are evaluated at fixed locations in the rotating and inertial frames, respectively.

Further, using (2.7), we have that

$$\nabla \cdot \boldsymbol{v}_I = \nabla \cdot (\boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r}) = \nabla \cdot \boldsymbol{v}_R, \tag{2.19}$$

since $\nabla \cdot (\boldsymbol{\Omega} \times \boldsymbol{r}) = 0$. Thus, using (2.18) and (2.19), (2.17) is equivalent to

$$\frac{\mathrm{D}\varphi}{\mathrm{D}t} + \varphi\nabla \cdot \boldsymbol{v}_R = 0, \tag{2.20}$$

where all observables are measured in the *rotating* frame. Thus, the equation for the evolution of a scalar whose measured value is the same in rotating and inertial frames is unaltered by the presence of rotation. In particular, the mass conservation equation is unaltered by the presence of rotation.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

Although we have taken (2.18) as true a priori, the individual components of the material derivative differ in the rotating and inertial frames. In particular

$$\left(\frac{\partial \varphi}{\partial t}\right)_I = \left(\frac{\partial \varphi}{\partial t}\right)_R - (\boldsymbol{\Omega} \times \boldsymbol{r}) \cdot \nabla\varphi, \tag{2.21}$$

because $\boldsymbol{\Omega} \times \boldsymbol{r}$ is the velocity, in the inertial frame, of a uniformly rotating body. Similarly,

$$\boldsymbol{v}_I \cdot \nabla\varphi = (\boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r}) \cdot \nabla\varphi. \tag{2.22}$$

Adding the last two equations reprises and confirms (2.18).

## 2.2 EQUATIONS OF MOTION IN SPHERICAL COORDINATES

The Earth is very nearly spherical and it might appear obvious that we should cast our equations in spherical coordinates. Although this does turn out to be true, the presence of a centrifugal force causes some complications that we should first discuss. The reader who is willing ab initio to treat the Earth as a perfect sphere and to neglect the horizontal component of the centrifugal force may skip the next section.

### 2.2.1 ✦ The Centrifugal Force and Spherical Coordinates

The centrifugal force is a potential force, like gravity, and so we may therefore define an 'effective gravity' equal to the sum of the true, or Newtonian, gravity and the centrifugal force. The Newtonian gravitational force is directed approximately toward the centre of the Earth, with small deviations due mainly to the Earth's oblateness. The line of action of the effective gravity will in general differ slightly from this, and therefore have a component in the 'horizontal' plane, that is the plane perpendicular to the radial direction. The magnitude of the centrifugal force is $\Omega^2 r_\perp$, and so the effective gravity is given by

$$\boldsymbol{g} \equiv \boldsymbol{g}_{eff} = \boldsymbol{g}_{grav} + \Omega^2 \boldsymbol{r}_\perp, \tag{2.23}$$

where $\boldsymbol{g}_{grav}$ is the Newtonian gravitational force due to the gravitational attraction of the Earth and $\boldsymbol{r}_\perp$ is normal to the rotation vector (in the direction $\boldsymbol{C}$ in Fig. 2.2), with $r_\perp = r\cos\vartheta$. Both gravity and centrifugal force are potential forces and therefore we may define the *geopotential*, $\Phi$, such that

$$\boldsymbol{g} = -\nabla\Phi. \tag{2.24}$$

Surfaces of constant $\Phi$ are not quite spherical because $r_\perp$, and hence the centrifugal force, vary with latitude (Fig. 2.2); this has certain ramifications, as we now discuss.

The components of the centrifugal force parallel and perpendicular to the radial direction are $\Omega^2 r\cos^2\vartheta$ and $\Omega^2 r\cos\vartheta\sin\vartheta$. Newtonian gravity is much larger than either of these, and at the Earth's surface the ratio of centrifugal to gravitational terms is approximately, and no more than,

$$\alpha \approx \frac{\Omega^2 a}{g} \approx \frac{(7.27\times10^{-5})^2 \times 6.4\times10^6}{9.8} \approx 3\times10^{-3}. \tag{2.25}$$

(At the equator and pole the horizontal component of the centrifugal force is zero and the effective gravity is aligned with Newtonian gravity.) The angle between $\boldsymbol{g}$ and the line to the centre of the Earth is given by a similar expression and so is also small, typically around $3\times10^{-3}$ radians. However, the horizontal component of the centrifugal force is still large compared to the Coriolis force, the ratio of their magnitudes in mid-latitudes being given by

$$\frac{\text{horizontal centrifugal force}}{\text{Coriolis force}} \approx \frac{\Omega^2 a\cos\vartheta\sin\vartheta}{2\Omega|u|} \approx \frac{\Omega a}{4|u|} \approx 10, \tag{2.26}$$

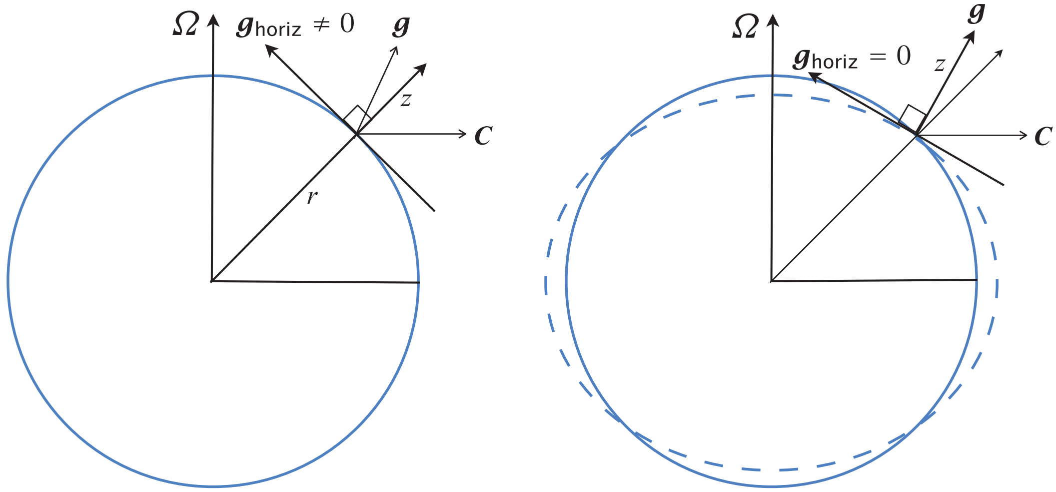


**Fig. 2.2** Left: directions of forces and coordinates in true spherical geometry. $\boldsymbol{g}$ is the effective gravity (including the centrifugal force, $\boldsymbol{C}$) and its horizontal component is evidently non-zero. Right: a modified coordinate system, in which the vertical direction is defined by the direction of $\boldsymbol{g}$, and so the horizontal component of $\boldsymbol{g}$ is identically zero. The dashed line schematically indicates a surface of constant geopotential. The differences between the direction of $\boldsymbol{g}$ and the direction of the radial coordinate, and between the sphere and the geopotential surface, are much exaggerated and in reality are similar to the thickness of the lines themselves.

using $u = 10\,\mathrm{m\,s^{-1}}$. The centrifugal term therefore dominates over the Coriolis term, and is largely balanced by a pressure gradient force. Thus, if we adhered to true spherical coordinates, both the horizontal and radial components of the momentum equation would be dominated by a static balance between a pressure gradient and gravity or centrifugal terms. Although in principle there is nothing wrong with writing the equations this way, it obscures the dynamical balances involving the Coriolis force and pressure that determine the large-scale horizontal flow.

A way around this problem is to use the direction of the geopotential force to *define* the vertical direction, and then for all geometric purposes to regard the surfaces of constant $\Phi$ as if they were true spheres.[3] The horizontal component of effective gravity is then identically zero, and we have traded a potentially large dynamical error for a very small geometric error. In fact, over time, the Earth has developed an equatorial bulge to compensate for and neutralize the centrifugal force, so that the effective gravity does act in a direction virtually normal to the Earth's surface; that is, the surface of the Earth is an oblate spheroid of nearly constant geopotential. The geopotential $\Phi$ is then a function of the vertical coordinate alone, and for many purposes we can just take $\Phi = gz$; that is, the direction normal to geopotential surfaces, the local vertical, is, in this approximation, taken to be the direction of increasing $r$ in spherical coordinates. It is because the oblateness is very small (the polar diameter is about 12 714 km, whereas the equatorial diameter is about 12 756 km) that using spherical coordinates is a very accurate way to map the spheroid. If the angle between effective gravity and a natural direction of the coordinate system were not small then more heroic measures would be called for.

If the solid Earth did not bulge at the equator, the *behaviour* of the atmosphere and ocean would differ significantly from that of the present system. For example, the surface of the ocean is, necessarily, very nearly a geopotential surface; if the solid Earth were exactly spherical then the ocean would perforce become much deeper at low latitudes and the ocean basins would dry out completely at high latitudes. We could still choose to use the spherical coordinate system discussed above to describe the dynamics, but the shape of the surface of the solid Earth would have to

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

be represented by a topography, with the topographic height increasing monotonically polewards nearly everywhere.

### 2.2.2 Some Identities in Spherical Coordinates

The location of a point is given by the coordinates $(\lambda, \vartheta, r)$ where $\lambda$ is the angular distance eastwards (i.e., longitude), $\vartheta$ is angular distance polewards (i.e., latitude) and $r$ is the radial distance from the centre of the Earth — see Fig. 2.3. (In some other fields of study co-latitude is used as a spherical coordinate.) If $a$ is the radius of the Earth, then we also define $z = r - a$. At a given location we may also define the Cartesian increments $(\delta x, \delta y, \delta z) = (r\cos\vartheta\,\delta\lambda, r\delta\vartheta, \delta r)$.

For a scalar quantity $\phi$ the material derivative in spherical coordinates is

$$\frac{\mathrm{D}\phi}{\mathrm{D}t} = \frac{\partial \phi}{\partial t} + \frac{u}{r\cos\vartheta}\frac{\partial \phi}{\partial \lambda} + \frac{v}{r}\frac{\partial \phi}{\partial \vartheta} + w\frac{\partial \phi}{\partial r}, \tag{2.27}$$

where the velocity components corresponding to the coordinates $(\lambda, \vartheta, r)$ are

$$(u, v, w) \equiv \left( r\cos\vartheta \frac{\mathrm{D}\lambda}{\mathrm{D}t}, r\frac{\mathrm{D}\vartheta}{\mathrm{D}t}, \frac{\mathrm{D}r}{\mathrm{D}t} \right). \tag{2.28}$$

That is, $u$ is the zonal velocity, $v$ is the meridional velocity and $w$ is the vertical velocity. If we define $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ to be the unit vectors in the direction of increasing $(\lambda, \vartheta, r)$ then

$$\boldsymbol{v} = \mathbf{i}u + \mathbf{j}v + \mathbf{k}w. \tag{2.29}$$

Note also that $\mathrm{D}r/\mathrm{D}t = \mathrm{D}z/\mathrm{D}t$.

The divergence of a vector $\boldsymbol{B} = \mathbf{i}\,B^\lambda + \mathbf{j}\,B^\vartheta + \mathbf{k}\,B^r$ is

$$\nabla \cdot \boldsymbol{B} = \frac{1}{\cos\vartheta}\left[ \frac{1}{r}\frac{\partial B^\lambda}{\partial \lambda} + \frac{1}{r}\frac{\partial}{\partial \vartheta}(B^\vartheta \cos\vartheta) + \frac{\cos\vartheta}{r^2}\frac{\partial}{\partial r}(r^2 B^r) \right]. \tag{2.30}$$

The vector gradient of a scalar is:

$$\nabla\phi = \mathbf{i}\frac{1}{r\cos\vartheta}\frac{\partial \phi}{\partial \lambda} + \mathbf{j}\frac{1}{r}\frac{\partial \phi}{\partial \vartheta} + \mathbf{k}\frac{\partial \phi}{\partial r}. \tag{2.31}$$

The Laplacian of a scalar is:

$$\nabla^2\phi \equiv \nabla\cdot\nabla\phi = \frac{1}{r^2\cos\vartheta}\left[ \frac{1}{\cos\vartheta}\frac{\partial^2\phi}{\partial \lambda^2} + \frac{\partial}{\partial \vartheta}\left(\cos\vartheta\frac{\partial \phi}{\partial \vartheta}\right) + \cos\vartheta\frac{\partial}{\partial r}\left(r^2\frac{\partial \phi}{\partial r}\right) \right]. \tag{2.32}$$

The curl of a vector is:

$$\operatorname{curl}\boldsymbol{B} = \nabla\times\boldsymbol{B} = \frac{1}{r^2\cos\vartheta}\begin{vmatrix} \mathbf{i}\,r\cos\vartheta & \mathbf{j}\,r & \mathbf{k} \\ \partial/\partial\lambda & \partial/\partial\vartheta & \partial/\partial r \\ B^\lambda r\cos\vartheta & B^\vartheta r & B^r \end{vmatrix}. \tag{2.33}$$

The vector Laplacian $\nabla^2\boldsymbol{B}$ (used for example when calculating viscous terms in the momentum equation) may be obtained from the vector identity:

$$\nabla^2\boldsymbol{B} = \nabla(\nabla\cdot\boldsymbol{B}) - \nabla\times(\nabla\times\boldsymbol{B}). \tag{2.34}$$

Only in Cartesian coordinates does this take the simple form:

$$\nabla^2\boldsymbol{B} = \frac{\partial^2\boldsymbol{B}}{\partial x^2} + \frac{\partial^2\boldsymbol{B}}{\partial y^2} + \frac{\partial^2\boldsymbol{B}}{\partial z^2}. \tag{2.35}$$

The expansion in spherical coordinates is of itself, to most eyes, rather uninformative.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

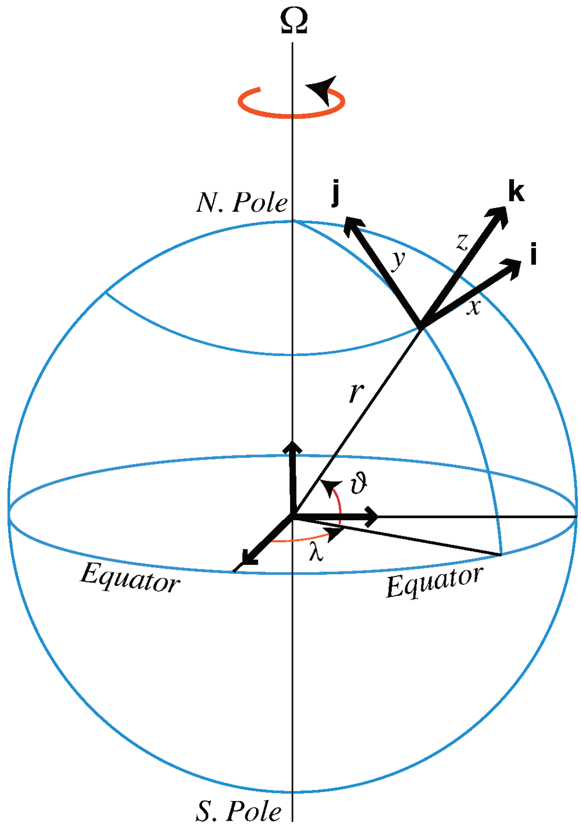


**Fig. 2.3** The spherical coordinate system. The orthogonal unit vectors **i**, **j** and **k** point in the direction of increasing longitude $\lambda$, latitude $\vartheta$, and altitude $z$. Locally, one may apply a Cartesian system with variables $x$, $y$ and $z$ measuring distances along **i**, **j** and **k**.

### *Rate of change of unit vectors*

In spherical coordinates the defining unit vectors are **i**, the unit vector pointing eastwards, parallel to a line of latitude; **j** is the unit vector pointing polewards, parallel to a meridian; and **k**, the unit vector pointing radially outward. The directions of these vectors change with location, and in fact this is the case in nearly all coordinate systems, with the notable exception of the Cartesian one, and thus their material derivative is not zero. One way to evaluate this is to consider geometrically how the coordinate axes change with position. Another way, and the way that we shall proceed, is to first obtain the effective rotation rate $\boldsymbol{\Omega}_{flow}$, relative to the Earth, of a unit vector as it moves with the flow, and then apply (2.3). Specifically, let the fluid velocity be $\boldsymbol{v} = (u, v, w)$. The meridional component, $v$, produces a displacement $r\delta\vartheta = v\delta t$, and this gives rise to a local effective vector rotation rate around the local zonal axis of $-(v/r)\mathbf{i}$, the minus sign arising because a displacement in the direction of the north pole is produced by negative rotational displacement around the **i** axis. Similarly, the zonal component, $u$, produces a displacement $\delta\lambda r\cos\vartheta = u\delta t$ and so an effective rotation rate, about the Earth's rotation axis, of $u/(r\cos\vartheta)$. Now, a rotation around the Earth's rotation axis may be written as (see Fig. 2.4)

$$\boldsymbol{\Omega} = \Omega(\mathbf{j}\cos\vartheta + \mathbf{k}\sin\vartheta). \tag{2.36}$$

If the scalar rotation rate is not $\Omega$ but is $u/(r\cos\vartheta)$, then the vector rotation rate is

$$\frac{u}{r\cos\vartheta}(\mathbf{j}\cos\vartheta + \mathbf{k}\sin\vartheta) = \mathbf{j}\frac{u}{r} + \mathbf{k}\frac{u\tan\vartheta}{r}. \tag{2.37}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

Thus, the total rotation rate of a vector that moves with the flow is

$$\boldsymbol{\Omega}_{flow} = -\mathbf{i}\frac{v}{r} + \mathbf{j}\frac{u}{r} + \mathbf{k}\frac{u\tan\vartheta}{r}. \tag{2.38}$$

Applying (2.3) to (2.38), we find

$$\frac{\mathrm{D}\mathbf{i}}{\mathrm{D}t} = \boldsymbol{\Omega}_{flow} \times \mathbf{i} = \frac{u}{r\cos\vartheta}(\mathbf{j}\sin\vartheta - \mathbf{k}\cos\vartheta), \tag{2.39a}$$

$$\frac{\mathrm{D}\mathbf{j}}{\mathrm{D}t} = \boldsymbol{\Omega}_{flow} \times \mathbf{j} = -\mathbf{i}\frac{u}{r}\tan\vartheta - \mathbf{k}\frac{v}{r}, \tag{2.39b}$$

$$\frac{\mathrm{D}\mathbf{k}}{\mathrm{D}t} = \boldsymbol{\Omega}_{flow} \times \mathbf{k} = \mathbf{i}\frac{u}{r} + \mathbf{j}\frac{v}{r}. \tag{2.39c}$$

### 2.2.3 Equations of Motion

*Mass conservation and thermodynamic equation*

The mass conservation equation, (1.36a), expanded in spherical co-ordinates, is

$$\frac{\partial\rho}{\partial t} + \frac{u}{r\cos\vartheta}\frac{\partial\rho}{\partial\lambda} + \frac{v}{r}\frac{\partial\rho}{\partial\vartheta} + w\frac{\partial\rho}{\partial r} + \frac{\rho}{r\cos\vartheta}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial}{\partial\vartheta}(v\cos\vartheta) + \frac{1}{r}\frac{\partial}{\partial r}(wr^2\cos\vartheta)\right] = 0. \tag{2.40}$$

Equivalently, using the form (1.36b), this is

$$\frac{\partial\rho}{\partial t} + \frac{1}{r\cos\vartheta}\frac{\partial(u\rho)}{\partial\lambda} + \frac{1}{r\cos\vartheta}\frac{\partial}{\partial\vartheta}(v\rho\cos\vartheta) + \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 w\rho) = 0. \tag{2.41}$$

The thermodynamic equation, (1.108), is a tracer advection equation. Thus, using (2.27), its (adiabatic) spherical coordinate form is

$$\frac{\mathrm{D}\theta}{\mathrm{D}t} = \frac{\partial\theta}{\partial t} + \frac{u}{r\cos\vartheta}\frac{\partial\theta}{\partial\lambda} + \frac{v}{r}\frac{\partial\theta}{\partial\vartheta} + w\frac{\partial\theta}{\partial r} = 0, \tag{2.42}$$

and similarly for tracers such as water vapour or salt.

*Momentum equation*

Recall that the inviscid momentum equation is:

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\boldsymbol{\Omega}\times\boldsymbol{v} = -\frac{1}{\rho}\nabla p - \nabla\Phi, \tag{2.43}$$

where $\Phi$ is the geopotential. In spherical coordinates the directions of the coordinate axes change with position and so the component expansion of (2.43) is

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \frac{\mathrm{D}u}{\mathrm{D}t}\mathbf{i} + \frac{\mathrm{D}v}{\mathrm{D}t}\mathbf{j} + \frac{\mathrm{D}w}{\mathrm{D}t}\mathbf{k} + u\frac{\mathrm{D}\mathbf{i}}{\mathrm{D}t} + v\frac{\mathrm{D}\mathbf{j}}{\mathrm{D}t} + w\frac{\mathrm{D}\mathbf{k}}{\mathrm{D}t} \tag{2.44a}$$

$$= \frac{\mathrm{D}u}{\mathrm{D}t}\mathbf{i} + \frac{\mathrm{D}v}{\mathrm{D}t}\mathbf{j} + \frac{\mathrm{D}w}{\mathrm{D}t}\mathbf{k} + \boldsymbol{\Omega}_{flow}\times\boldsymbol{v}, \tag{2.44b}$$

using (2.39). Using either (2.44a) and the expressions for the rates of change of the unit vectors given in (2.39), or (2.44b) and the expression for $\boldsymbol{\Omega}_{flow}$ given in (2.38), (2.44) becomes

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \mathbf{i}\left(\frac{\mathrm{D}u}{\mathrm{D}t} - \frac{uv\tan\vartheta}{r} + \frac{uw}{r}\right) + \mathbf{j}\left(\frac{\mathrm{D}v}{\mathrm{D}t} + \frac{u^2\tan\vartheta}{r} + \frac{vw}{r}\right) + \mathbf{k}\left(\frac{\mathrm{D}w}{\mathrm{D}t} - \frac{u^2+v^2}{r}\right). \tag{2.45}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

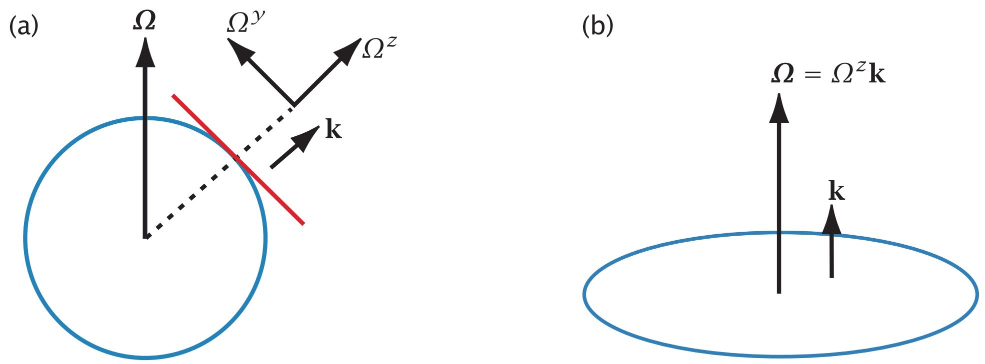


**Fig. 2.4** (a) On the sphere the rotation vector $\boldsymbol{\Omega}$ can be decomposed into two components, one in the local vertical and one in the local horizontal, pointing toward the pole. That is, $\boldsymbol{\Omega} = \Omega_y\mathbf{j}+\Omega_z\mathbf{k}$ where $\Omega_y = \Omega\cos\vartheta$ and $\Omega_z = \Omega\sin\vartheta$. In geophysical fluid dynamics, the rotation vector in the local vertical is often the more important component in the horizontal momentum equations. On a rotating disk, (b), the rotation vector $\boldsymbol{\Omega}$ is parallel to the local vertical $\mathbf{k}$.

Using the definition of a vector cross product the Coriolis term is:

$$2\boldsymbol{\Omega}\times\boldsymbol{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 2\Omega\cos\vartheta & 2\Omega\sin\vartheta \\ u & v & w \end{vmatrix}$$
$$= \mathbf{i}\,(2\Omega w\cos\vartheta - 2\Omega v\sin\vartheta) + \mathbf{j}\,2\Omega u\sin\vartheta - \mathbf{k}\,2\Omega u\cos\vartheta. \tag{2.46}$$

Using (2.45) and (2.46), and the gradient operator given by (2.31), the momentum equation (2.43) becomes:

$$\frac{\mathrm{D}u}{\mathrm{D}t} - \left(2\Omega + \frac{u}{r\cos\vartheta}\right)(v\sin\vartheta - w\cos\vartheta) = -\frac{1}{\rho r\cos\vartheta}\frac{\partial p}{\partial\lambda}, \tag{2.47a}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t} + \frac{wv}{r} + \left(2\Omega + \frac{u}{r\cos\vartheta}\right)u\sin\vartheta = -\frac{1}{\rho r}\frac{\partial p}{\partial\vartheta}, \tag{2.47b}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} - \frac{u^2+v^2}{r} - 2\Omega u\cos\vartheta = -\frac{1}{\rho}\frac{\partial p}{\partial r} - g. \tag{2.47c}$$

The terms involving $\Omega$ are called Coriolis terms, and the quadratic terms on the left-hand sides involving $1/r$ are often called metric terms.

### 2.2.4 The Primitive Equations

The so-called *primitive equations* of motion are simplifications of the above equations frequently used in atmospheric and oceanic modelling.[4] Three related approximations are involved:

(i) *The hydrostatic approximation.* In the vertical momentum equation the gravitational term is assumed to be balanced by the pressure gradient term, so that

$$\frac{\partial p}{\partial z} = -\rho g. \tag{2.48}$$

The advection of vertical velocity, the Coriolis terms, and the metric term $(u^2+v^2)/r$ are all neglected.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

(ii) *The shallow-fluid approximation.* We write $r = a + z$ where the constant $a$ is the radius of the Earth and $z$ increases in the radial direction. The coordinate $r$ is then replaced by $a$ except where it is used as the differentiating argument. Thus, for example,

$$\frac{1}{r^2}\frac{\partial(r^2 w)}{\partial r} \to \frac{\partial w}{\partial z}. \tag{2.49}$$

(iii) *The traditional approximation.* Coriolis terms in the horizontal momentum equations involving the vertical velocity, and the still smaller metric terms $uw/r$ and $vw/r$, are neglected.

The second and third of these approximations should be taken, or not, together, the underlying reason being that they both relate to the presumed small aspect ratio of the motion, so the approximations succeed or fail together. If we make one approximation but not the other then we are being asymptotically inconsistent, and angular momentum and energy conservation are not assured.[5] The hydrostatic approximation also depends on the small aspect ratio of the flow, but in a slightly different way. For large-scale flow in the terrestrial atmosphere and ocean all three approximations are in fact very accurate approximations. We defer a more complete treatment until Section 2.7, in part because a treatment of the hydrostatic approximation is done most easily in the context of the Boussinesq equations, derived in Section 2.4.

Making these approximations, the momentum equations for a shallow layer are

$$\frac{\mathrm{D}u}{\mathrm{D}t} - 2\Omega\sin\vartheta v - \frac{uv}{a}\tan\vartheta = -\frac{1}{a\rho\cos\vartheta}\frac{\partial p}{\partial\lambda}, \tag{2.50a}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t} + 2\Omega\sin\vartheta u + \frac{u^2\tan\vartheta}{a} = -\frac{1}{\rho a}\frac{\partial p}{\partial\vartheta}, \tag{2.50b}$$

$$0 = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g, \tag{2.50c}$$

where

$$\frac{\mathrm{D}}{\mathrm{D}t} = \left(\frac{\partial}{\partial t} + \frac{u}{a\cos\vartheta}\frac{\partial}{\partial\lambda} + \frac{v}{a}\frac{\partial}{\partial\vartheta} + w\frac{\partial}{\partial z}\right). \tag{2.51}$$

We note the ubiquity of the factor $2\Omega\sin\vartheta$, and take the opportunity to define the *Coriolis parameter*, $f \equiv 2\Omega\sin\vartheta$. The associated mass conservation equation for a shallow fluid layer is:

$$\frac{\partial\rho}{\partial t} + \frac{u}{a\cos\vartheta}\frac{\partial\rho}{\partial\lambda} + \frac{v}{a}\frac{\partial\rho}{\partial\vartheta} + w\frac{\partial\rho}{\partial z} + \rho\left[\frac{1}{a\cos\vartheta}\frac{\partial u}{\partial\lambda} + \frac{1}{a\cos\vartheta}\frac{\partial}{\partial\vartheta}(v\cos\vartheta) + \frac{\partial w}{\partial z}\right] = 0, \tag{2.52}$$

or equivalently,

$$\frac{\partial\rho}{\partial t} + \frac{1}{a\cos\vartheta}\frac{\partial(u\rho)}{\partial\lambda} + \frac{1}{a\cos\vartheta}\frac{\partial}{\partial\vartheta}(v\rho\cos\vartheta) + \frac{\partial(w\rho)}{\partial z} = 0. \tag{2.53}$$

### 2.2.5 Primitive Equations in Vector Form

The primitive equations on a sphere may be written in a compact vector form provided we make a slight reinterpretation of the material derivative of the coordinate axes. Instead of (2.39) we take the material derivative of the unit vectors to be

$$\frac{\mathrm{D}\mathbf{i}}{\mathrm{D}t} = \widetilde{\boldsymbol{\Omega}}_{flow} \times \mathbf{i} = \mathbf{j}\frac{u\tan\vartheta}{a}, \tag{2.54a}$$

$$\frac{\mathrm{D}\mathbf{j}}{\mathrm{D}t} = \widetilde{\boldsymbol{\Omega}}_{flow} \times \mathbf{j} = -\mathbf{i}\frac{u\tan\vartheta}{a}, \tag{2.54b}$$

where $\widetilde{\boldsymbol{\Omega}}_{flow} = \mathbf{k}u\tan\vartheta/a$, which is the vertical component of (2.38) with $r$ replaced by $a$. Given (2.54), the primitive equations (2.50a) and (2.50b) may be written as

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} = -\frac{1}{\rho}\nabla_z p, \tag{2.55}$$

where $\boldsymbol{u} = u\mathbf{i} + v\mathbf{j} + 0\,\mathbf{k}$ is the horizontal velocity, $\nabla_z p = [(a\cos\vartheta)^{-1}\partial p/\partial\lambda,\ a^{-1}\partial p/\partial\vartheta]$ is the gradient operator at constant $z$, and $\boldsymbol{f} = f\mathbf{k} = 2\Omega\sin\vartheta\mathbf{k}$. In (2.55) the material derivative of the horizontal velocity is given by

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \mathbf{i}\frac{\mathrm{D}u}{\mathrm{D}t} + \mathbf{j}\frac{\mathrm{D}v}{\mathrm{D}t} + u\frac{\mathrm{D}\mathbf{i}}{\mathrm{D}t} + v\frac{\mathrm{D}\mathbf{j}}{\mathrm{D}t}. \tag{2.56}$$

The advection of the horizontal wind $\boldsymbol{u}$ is still by the three-dimensional velocity $\boldsymbol{v}$.

The vertical momentum equation is the hydrostatic equation, (2.50c), and the mass conservation equation is

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho\nabla\cdot\boldsymbol{v} = 0 \qquad \text{or} \qquad \frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{v}) = 0, \tag{2.57}$$

where $\mathrm{D}/\mathrm{D}t$ is given by (2.51), and the second expression is written out in full in (2.53).

### 2.2.6 The Vector Invariant Form of the Momentum Equation

The 'vector invariant' form of the momentum equation is so-called because it appears to take the same form in all coordinate systems — there is no advective derivative of the coordinate system to worry about. With the aid of the identity $(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\boldsymbol{v}\times\boldsymbol{\omega} + \nabla(\boldsymbol{v}^2/2)$, where $\boldsymbol{\omega} \equiv \nabla\times\boldsymbol{v}$ is the relative vorticity (which we explore at greater length in Chapter 4) the three-dimensional momentum equation, (2.16), may be written:

$$\frac{\partial\boldsymbol{v}}{\partial t} + (2\boldsymbol{\Omega} + \boldsymbol{\omega})\times\boldsymbol{v} = -\frac{1}{\rho}\nabla p - \frac{1}{2}\nabla\boldsymbol{v}^2 + \boldsymbol{g}, \tag{2.58}$$

and this is the vector invariant momentum equation. In spherical coordinates the relative vorticity is given by:

$$\begin{aligned}\boldsymbol{\omega} = \nabla\times\boldsymbol{v} &= \frac{1}{r^2\cos\vartheta}\begin{vmatrix} \mathbf{i}\,r\cos\vartheta & \mathbf{j}\,r & \mathbf{k} \\ \partial/\partial\lambda & \partial/\partial\vartheta & \partial/\partial r \\ ur\cos\vartheta & rv & w \end{vmatrix} \\ &= \mathbf{i}\,\frac{1}{r}\left(\frac{\partial w}{\partial\vartheta} - \frac{\partial(rv)}{\partial r}\right) - \mathbf{j}\frac{1}{r\cos\vartheta}\left(\frac{\partial w}{\partial\lambda} - \frac{\partial}{\partial r}(ur\cos\vartheta)\right) + \mathbf{k}\frac{1}{r\cos\vartheta}\left(\frac{\partial v}{\partial\lambda} - \frac{\partial}{\partial\vartheta}(u\cos\vartheta)\right).\end{aligned} \tag{2.59}$$

We can write the horizontal momentum equations of the primitive equations in a similar way. Making the traditional and shallow fluid approximations, the horizontal components of (2.58) become

$$\frac{\partial\boldsymbol{u}}{\partial t} + (\boldsymbol{f} + \mathbf{k}\zeta)\times\boldsymbol{u} + w\frac{\partial\boldsymbol{u}}{\partial z} = -\frac{1}{\rho}\nabla_z p - \frac{1}{2}\nabla\boldsymbol{u}^2, \tag{2.60}$$

where $\boldsymbol{u} = (u, v, 0)$, $\boldsymbol{f} = \mathbf{k}\,2\Omega\sin\vartheta$ and $\nabla_z$ is the horizontal gradient operator (the gradient at a constant value of $z$). Using (2.59), $\zeta$ is given by

$$\zeta = \frac{1}{a\cos\vartheta}\frac{\partial v}{\partial\lambda} - \frac{1}{a\cos\vartheta}\frac{\partial}{\partial\vartheta}(u\cos\vartheta) = \frac{1}{a\cos\vartheta}\frac{\partial v}{\partial\lambda} - \frac{1}{a}\frac{\partial u}{\partial\vartheta} + \frac{u}{a}\tan\vartheta. \tag{2.61}$$

The separate components of the momentum equation are given by:

$$\frac{\partial u}{\partial t} - (f + \zeta)v + w\frac{\partial u}{\partial z} = -\frac{1}{a\cos\vartheta}\left(\frac{1}{\rho}\frac{\partial p}{\partial\lambda} + \frac{1}{2}\frac{\partial\boldsymbol{u}^2}{\partial\lambda}\right), \tag{2.62}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

and

$$\frac{\partial v}{\partial t} + (f + \zeta)u + w\frac{\partial v}{\partial z} = -\frac{1}{a}\left(\frac{1}{\rho}\frac{\partial p}{\partial \vartheta} + \frac{1}{2}\frac{\partial \boldsymbol{u}^2}{\partial \vartheta}\right). \tag{2.63}$$

### 2.2.7 Angular Momentum

The zonal momentum equation can be usefully expressed as a statement about axial angular momentum; that is, angular momentum about the rotation axis. The zonal angular momentum per unit mass is the component of angular momentum in the direction of the axis of rotation and it is given by, without making any shallow atmosphere approximation,

$$m = (u + \Omega r\cos\vartheta)r\cos\vartheta. \tag{2.64}$$

The evolution equation for this quantity follows from the zonal momentum equation and has the simple form

$$\frac{\mathrm{D}m}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial \lambda}, \tag{2.65}$$

where the material derivative is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \frac{u}{r\cos\vartheta}\frac{\partial}{\partial \lambda} + \frac{v}{r}\frac{\partial}{\partial \vartheta} + w\frac{\partial}{\partial r}. \tag{2.66}$$

Using the mass continuity equation, (2.65) can be written as

$$\frac{\mathrm{D}\rho m}{\mathrm{D}t} + \rho m\nabla\cdot\boldsymbol{v} = -\frac{\partial p}{\partial \lambda} \tag{2.67}$$

or

$$\frac{\partial \rho m}{\partial t} + \frac{1}{r\cos\vartheta}\frac{\partial(\rho u m)}{\partial \lambda} + \frac{1}{r\cos\vartheta}\frac{\partial}{\partial \vartheta}(\rho v m\cos\vartheta) + \frac{1}{r^2}\frac{\partial}{\partial r}(\rho m w r^2) = -\frac{\partial p}{\partial \lambda}. \tag{2.68}$$

This is an angular momentum conservation equation.

If the fluid is confined to a shallow layer near the surface of a sphere, then we may replace $r$, the radial coordinate, by $a$, the radius of the sphere, in the definition of $m$, and we define $\widetilde{m} \equiv (u + \Omega a\cos\vartheta)a\cos\vartheta$. Then (2.65) is replaced by

$$\frac{\mathrm{D}\widetilde{m}}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial \lambda}, \tag{2.69}$$

where now

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \frac{u}{a\cos\vartheta}\frac{\partial}{\partial \lambda} + \frac{v}{a}\frac{\partial}{\partial \vartheta} + w\frac{\partial}{\partial z}. \tag{2.70}$$

In the shallow fluid approximation (2.68) becomes

$$\frac{\partial \rho m}{\partial t} + \frac{1}{a\cos\vartheta}\frac{\partial(\rho u m)}{\partial \lambda} + \frac{1}{a\cos\vartheta}\frac{\partial}{\partial \vartheta}(\rho v m\cos\vartheta) + \frac{\partial}{\partial z}(\rho m w) = -\frac{\partial p}{\partial \lambda}, \tag{2.71}$$

which is an angular momentum conservation equation for a shallow atmosphere.

#### ✦ *From angular momentum to the spherical component equations*

An alternative way of deriving the three components of the momentum equation in spherical polar coordinates is to *begin* with (2.65) and the principle of conservation of energy. That is, we take the equations for conservation of angular momentum and energy as true a priori and demand that the forms of the momentum equation be constructed to satisfy these. Expanding the material

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

derivative in (2.65), noting that $\mathrm{D}r/\mathrm{D}t = w$ and $\mathrm{D}\cos\vartheta/\mathrm{D}t = -(v/r)\sin\vartheta$, immediately gives (2.47a). Multiplication by $u$ then yields

$$u\frac{\mathrm{D}u}{\mathrm{D}t} - 2\Omega uv\sin\vartheta + 2\Omega uw\cos\vartheta - \frac{u^2 v\tan\vartheta}{r} + \frac{u^2 w}{r} = -\frac{u}{\rho r\cos\vartheta}\frac{\partial p}{\partial\lambda}. \tag{2.72}$$

Now suppose that the meridional and vertical momentum equations are of the form

$$\frac{\mathrm{D}v}{\mathrm{D}t} + \text{Coriolis and metric terms} = -\frac{1}{\rho r}\frac{\partial p}{\partial\vartheta}, \tag{2.73a}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} + \text{Coriolis and metric terms} = -\frac{1}{\rho}\frac{\partial p}{\partial r}, \tag{2.73b}$$

but that we do not know what form the Coriolis and metric terms take. To determine that form, construct the kinetic energy equation by multiplying (2.73) by $v$ and $w$, respectively. Now, the metric terms must vanish when we sum the resulting equations along with (2.72), so that (2.73a) must contain the Coriolis term $2\Omega u\sin\vartheta$ as well as the metric term $u^2\tan\vartheta/r$, and (2.73b) must contain the term $-2\Omega u\cos\phi$ as well as the metric term $u^2/r$. But if (2.73b) contains the term $u^2/r$ it must also contain the term $v^2/r$ by isotropy, and therefore (2.73a) must also contain the term $vw/r$. In this way, (2.47) is precisely reproduced, although the sceptic might argue that the uniqueness of the form has not been demonstrated.

A particular advantage of this approach arises in determining the appropriate momentum equations that conserve angular momentum and energy in the shallow-fluid approximation. We begin with (2.69) and expand to obtain (2.50a). Multiplying by $u$ gives

$$u\frac{\mathrm{D}u}{\mathrm{D}t} - 2\Omega uv\sin\vartheta - \frac{u^2 v\tan\vartheta}{a} = -\frac{u}{\rho a\cos\vartheta}\frac{\partial p}{\partial\lambda}. \tag{2.74}$$

To ensure energy conservation, the meridional momentum equation must contain the Coriolis term $2\Omega u\sin\vartheta$ and the metric term $u^2\tan\vartheta/a$, but the vertical momentum equation must have neither of the metric terms appearing in (2.47c). Thus we deduce the following equations:

$$\frac{\mathrm{D}u}{\mathrm{D}t} - \left(2\Omega\sin\vartheta + \frac{u\tan\vartheta}{a}\right)v = -\frac{1}{\rho a\cos\vartheta}\frac{\partial p}{\partial\lambda}, \tag{2.75a}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t} + \left(2\Omega\sin\vartheta + \frac{u\tan\vartheta}{a}\right)u = -\frac{1}{\rho a}\frac{\partial p}{\partial\vartheta}, \tag{2.75b}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial r} - g. \tag{2.75c}$$

This equation set, when used in conjunction with the thermodynamic and mass continuity equations, conserves appropriate forms of angular momentum and energy. In the hydrostatic approximation the material derivative of $w$ in (2.75c) is *additionally* neglected. Thus, the hydrostatic approximation is mathematically and physically consistent with the shallow-fluid approximation, but it is an additional approximation with slightly different requirements that one may choose, rather than being required, to make. From an asymptotic perspective, the difference lies in the small parameter necessary for either approximation to hold, namely:

$$\text{shallow fluid and traditional approximations:} \qquad \gamma \equiv \frac{H}{a} \ll 1, \tag{2.76a}$$

$$\text{small aspect ratio for hydrostatic approximation:} \qquad \alpha \equiv \frac{H}{L} \ll 1, \tag{2.76b}$$

where $L$ is the horizontal scale of the motion and $a$ is the radius of the Earth. For hemispheric or global scale phenomena $L \sim a$ and the two approximations coincide. (Requirement (2.76b) for the hydrostatic approximation will be derived in Section 2.7.)

## 2.3 CARTESIAN APPROXIMATIONS: THE TANGENT PLANE

### 2.3.1 The $f$-plane

Although the rotation of the Earth is central for many dynamical phenomena, the sphericity of the Earth is not always so. This is especially true for phenomena on a scale somewhat smaller than global where the use of spherical coordinates becomes awkward, and it is more convenient to use a locally Cartesian representation of the equations. Referring to the red line in Fig. 2.4 we will define a plane tangent to the surface of the Earth at a latitude $\vartheta_0$, and then use a Cartesian coordinate system $(x, y, z)$ to describe motion on that plane. For small excursions on the plane, $(x, y, z) \approx (a\lambda \cos\vartheta_0, a(\vartheta - \vartheta_0), z)$. Consistently, the velocity is $\boldsymbol{v} = (u, v, w)$, so that $u$,$v$ and $w$ are the components of the velocity *in the tangent plane,* in approximately in the east–west, north–south and vertical directions, respectively.

The momentum equations for flow in this plane are then

$$\frac{\partial u}{\partial t} + (\boldsymbol{v}\cdot\nabla)u + 2(\Omega^y w - \Omega^z v) = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \tag{2.77a}$$

$$\frac{\partial v}{\partial t} + (\boldsymbol{v}\cdot\nabla)v + 2(\Omega^z u - \Omega^x w) = -\frac{1}{\rho}\frac{\partial p}{\partial y}, \tag{2.77b}$$

$$\frac{\partial w}{\partial t} + (\boldsymbol{v}\cdot\nabla)w + 2(\Omega^x v - \Omega^y u) = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g, \tag{2.77c}$$

where the rotation vector $\boldsymbol{\Omega} = \Omega^x\mathbf{i} + \Omega^y\mathbf{j} + \Omega^z\mathbf{k}$ and $\Omega^x = 0$, $\Omega^y = \Omega\cos\vartheta_0$ and $\Omega^z = \Omega\sin\vartheta_0$. If we make the traditional approximation, and so ignore the components of $\boldsymbol{\Omega}$ not in the direction of the local vertical, then the above equations become

$$\frac{\mathrm{D}u}{\mathrm{D}t} - f_0 v = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \qquad \frac{\mathrm{D}v}{\mathrm{D}t} + f_0 u = -\frac{1}{\rho}\frac{\partial p}{\partial y}, \qquad \frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g, \tag{2.78a,b,c}$$

where $f_0 = 2\Omega^z = 2\Omega\sin\vartheta_0$. Defining the horizontal velocity vector $\boldsymbol{u} = (u, v, 0)$, the first two equations may be written as

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}_0 \times \boldsymbol{u} = -\frac{1}{\rho}\nabla_z p, \tag{2.79}$$

where $\mathrm{D}\boldsymbol{u}/\mathrm{D}t = \partial\boldsymbol{u}/\partial t + \boldsymbol{v}\cdot\nabla\boldsymbol{u}$, $\boldsymbol{f}_0 = 2\Omega\sin\vartheta_0\mathbf{k} = f_0\mathbf{k}$, and $\mathbf{k}$ is the direction perpendicular to the plane. These equations are, evidently, exactly the same as the momentum equations in a system in which the rotation vector is aligned with the local vertical, as illustrated in panel (b) of Fig. 2.4. They will describe flow on the surface of a rotating sphere to a good approximation provided the flow is of limited latitudinal extent so that the effects of sphericity are unimportant; we have made what is known as the *$f$-plane*. We may in addition make the hydrostatic approximation, in which case (2.78c) becomes the familiar $\partial p/\partial z = -\rho g$.

### 2.3.2 The Beta-plane Approximation

The magnitude of the vertical component of rotation varies with latitude, and this has important dynamical consequences. We can approximate this effect by allowing the effective rotation vector to vary. Thus, noting that, for small variations in latitude,

$$f = 2\Omega\sin\vartheta \approx 2\Omega\sin\vartheta_0 + 2\Omega(\vartheta - \vartheta_0)\cos\vartheta_0, \tag{2.80}$$

then on the tangent plane we may mimic this by allowing the Coriolis parameter to vary as

$$\boxed{f = f_0 + \beta y,} \tag{2.81}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

where $f_0 = 2\Omega \sin \vartheta_0$ and $\beta = \partial f/\partial y = (2\Omega \cos \vartheta_0)/a$. This important approximation is known as the *beta-plane*, or *β-plane,* approximation; it captures the the most important *dynamical* effects of sphericity, without the complicating *geometric* effects, which are not essential to describe many phenomena. The momentum equations (2.78) are unaltered except that $f_0$ is replaced by $f_0 + \beta y$ to represent a varying Coriolis parameter. Thus, sphericity combined with rotation is dynamically equivalent to a *differentially rotating* system. For future reference, we write down the $\beta$-plane horizontal momentum equations:

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -\frac{1}{\rho}\nabla_z p, \tag{2.82}$$

where $\boldsymbol{f} = (f_0 + \beta y)\widehat{\mathbf{k}}$. In component form this equation becomes

$$\frac{\mathrm{D}u}{\mathrm{D}t} - fv = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \qquad \frac{\mathrm{D}v}{\mathrm{D}t} + fu = -\frac{1}{\rho}\frac{\partial p}{\partial y}. \tag{2.83a,b}$$

The mass conservation, thermodynamic and hydrostatic equations in the $\beta$-plane approximation are the same as the usual Cartesian, $f$-plane, forms of those equations.

## 2.4 EQUATIONS FOR A STRATIFIED OCEAN: THE BOUSSINESQ APPROXIMATION

The density variations in the ocean are quite small compared to the mean density, and we may exploit this to derive somewhat simpler but still quite accurate equations of motion. Let us first examine how much density does vary in the ocean.

### 2.4.1 Variation of Density in the Ocean

The variations of density in the ocean are due to three effects: the compression of water by pressure (which we denote as $\Delta_p \rho$), the thermal expansion of water if its temperature changes ($\Delta_T \rho$), and the haline contraction if its salinity changes ($\Delta_S \rho$). How big are these? An appropriate equation of state to approximately evaluate these effects is the linear one

$$\rho = \rho_0 \left[1 - \beta_T(T - T_0) + \beta_S(S - S_0) + \beta_p p\right], \tag{2.84}$$

where $\beta_T \approx 2 \times 10^{-4}\,\mathrm{K}^{-1}$, $\beta_S \approx 10^{-3}\,\mathrm{g/kg}^{-1}$ and $\beta_p = 1/(\rho_0 c_s^2) \approx 4.4 \times 10^{-10}\,\mathrm{Pa}^{-1}$ with $c_s \approx 1500\,\mathrm{m\,s}^{-1}$ (see Table 1.2 on page 33). The three effects may then be evaluated as follows:

*Pressure compressibility.* We have $\Delta_p \rho \approx \Delta p / c_s^2 \approx \rho_0 g H / c_s^2$ where $H$ is the depth we evaluate the pressure change (quite accurately) using the hydrostatic approximation. Thus,

$$\frac{|\Delta_p \rho|}{\rho_0} \approx \frac{gH}{c_s^2} \sim 4 \times 10^{-2}, \tag{2.85}$$

with $H = 8\,\mathrm{km}$ and $c_s^2/g \approx 200\,\mathrm{km}$. The latter quantity is the density scale height of the ocean. Thus, the pressure at the bottom of the ocean, enormous as it is, is insufficient to compress the water enough to make a significant change in its density. Changes in density due to dynamical variations of pressure are small if the Mach number is small, and this is also usually the case.

*Thermal expansion.* We have $\Delta_T \rho \approx -\beta_T \rho_0 \Delta T$ and therefore

$$\frac{|\Delta_T \rho|}{\rho_0} \approx \beta_T \Delta T \sim 4 \times 10^{-3} \tag{2.86}$$

with $\Delta T = 20\,\mathrm{K}$. Evidently we would require temperature differences of order $\beta_T^{-1}$, or 5000 K to obtain order one variations in density.

*Saline contraction.* We have $\varDelta_S \rho \approx \beta_S \rho_0 \varDelta S$ and therefore

$$\frac{|\varDelta_S \rho|}{\rho_0} \approx \beta_S \varDelta S \sim 1.5 \times 10^{-3}, \tag{2.87}$$

with $\varDelta S = 5\,g\,\mathrm{kg}^{-1}$. The fractional change in the density of seawater due to salinity variations is thus also very small.

Evidently, fractional density changes in the ocean are very small due to the above effects.

### 2.4.2 The Boussinesq Equations

The *Boussinesq equations* are a set of equations that exploit the smallness of density variations in liquids.[6] An asymptotic derivation is given in Appendix A (page 101) but in what follows we are more heuristic. To set notation we write

$$\begin{aligned} \rho &= \rho_0 + \delta\rho(x, y, z, t) && \text{(2.88a)} \\ &= \rho_0 + \widehat{\rho}(z) + \rho'(x, y, z, t) && \text{(2.88b)} \\ &= \widetilde{\rho}(z) + \rho'(x, y, z, t), && \text{(2.88c)} \end{aligned}$$

where $\rho_0$ is a constant and we assume that

$$|\widehat{\rho}|, |\rho'|, |\delta\rho| \ll \rho_0. \tag{2.89}$$

We need not assume that $|\rho'| \ll |\widehat{\rho}|$, but this is often the case in the ocean. The horizontal gradients (i.e., gradients at constant $z$, $\nabla_z$) satisfy $\nabla_z p = \nabla_z p' = \nabla_z \delta p$. To obtain the Boussinesq equations we will just use (2.88a), but (2.88c) will be useful for the anelastic equations considered later.

Associated with the reference density is a reference pressure that is defined to be in hydrostatic balance with it. That is,

$$p = p_0(z) + \delta p(x, y, z, t), \tag{2.90}$$

where $|\delta p| \ll p_0$ and

$$\frac{\mathrm{d}p_0}{\mathrm{d}z} \equiv -g\rho_0. \tag{2.91a,b}$$

*Momentum equations*

Letting $\rho = \rho_0 + \delta\rho$ the momentum equation can be written, without approximation, as

$$(\rho_0 + \delta\rho)\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\boldsymbol{\Omega} \times \boldsymbol{v}\right) = -\nabla\delta p - \frac{\partial p_0}{\partial z}\mathbf{k} - g(\rho_0 + \delta\rho)\mathbf{k}, \tag{2.92}$$

and using (2.91) this becomes, again without approximation,

$$(\rho_0 + \delta\rho)\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\boldsymbol{\Omega} \times \boldsymbol{v}\right) = -\nabla\delta p - g\delta\rho\mathbf{k}. \tag{2.93}$$

If $\delta\rho/\rho_0 \ll 1$ then we may neglect the $\delta\rho$ term on the left-hand side and the above equation becomes

$$\boxed{\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\boldsymbol{\Omega} \times \boldsymbol{v} = -\nabla\phi + b\mathbf{k},} \tag{2.94}$$

where $\phi = \delta p/\rho_0$ and $b = -g\,\delta\rho/\rho_0$ is the *buoyancy*. We should not and do not neglect the term $g\,\delta\rho$, for there is no reason to believe it to be small: $\delta\rho$ may be small, but $g$ is big! Equation (2.94) is the momentum equation in the Boussinesq approximation, and it is common to say that the

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

Boussinesq approximation ignores all variations of density of a fluid in the momentum equation, except when associated with the gravitational term.

For most large-scale motions in the ocean the *deviation* pressure and density fields are also approximately in hydrostatic balance, and in that case the vertical component of (2.94) becomes

$$\frac{\partial \phi}{\partial z} = b. \tag{2.95}$$

A condition for (2.95) to hold is that vertical accelerations are small *compared to $g\,\delta\rho/\rho_0$, and not compared to the acceleration due to gravity itself.* For more discussion of this point, see Section 2.7.

### Mass continuity

The unapproximated mass continuity equation is

$$\frac{\mathrm{D}\delta\rho}{\mathrm{D}t} + (\rho_0 + \delta\rho)\nabla\cdot\boldsymbol{v} = 0. \tag{2.96}$$

Provided that time scales advectively — that is to say that $\mathrm{D}/\mathrm{D}t$ scales in the same way as $\boldsymbol{v}\cdot\nabla$ — then we may approximate this equation by

$$\boxed{\nabla\cdot\boldsymbol{v} = 0,} \tag{2.97}$$

which is the same as that for a constant density fluid. This *absolutely does not* allow one to go back and use (2.96) to say that $D\delta\rho/Dt = 0$; the evolution of density is given by the thermodynamic equation in conjunction with an equation of state, and this should not be confused with the mass conservation equation. Note also that in eliminating the time-derivative of density we eliminate the possibility of sound waves.

### Thermodynamic equation and equation of state

The Boussinesq equations are closed by the addition of an equation of state, a thermodynamic equation and, as appropriate, a salinity equation. Neglecting salinity for the moment, a useful starting point is to write the thermodynamic equation, (1.114), as

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} - \frac{1}{c_s^2}\frac{\mathrm{D}p}{\mathrm{D}t} = \frac{\dot{Q}}{(\partial\eta/\partial\rho)_p T} \approx -\dot{Q}\left(\frac{\rho_0\beta_T}{c_p}\right) \tag{2.98}$$

using $(\partial\eta/\partial\rho)_p = (\partial\eta/\partial T)_p(\partial T/\partial\rho)_p \approx -c_p/(T\rho_0\beta_T)$. Given the expansions (2.88a) and (2.90a), (2.98) can be written to a good approximation as

$$\frac{\mathrm{D}\delta\rho}{\mathrm{D}t} - \frac{1}{c_s^2}\frac{\mathrm{D}p_0}{\mathrm{D}t} = -\dot{Q}\left(\frac{\rho_0\beta_T}{c_p}\right), \tag{2.99}$$

or, using (2.91a),

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\delta\rho + \frac{\rho_0 g}{c_s^2}z\right) = -\dot{Q}\left(\frac{\rho_0\beta_T}{c_p}\right). \tag{2.100}$$

The term in brackets on left-hand side is the potential density, as in (1.117). The severest approximation to this is to neglect the second term there, and noting that $b = -g\delta\rho/\rho_0$ we obtain

$$\boxed{\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b},} \tag{2.101}$$

where $\dot{b} = g\beta_T \dot{Q}/c_p$. The momentum equation (2.94), mass continuity equation (2.97) and thermodynamic equation (2.101) then form a closed set, called the *simple Boussinesq equations*.

In the ocean the compressibility effect can be important and it is convenient to write the thermodynamic equation as

$$\frac{\mathrm{D}b_\sigma}{\mathrm{D}t} = \dot{b}_\sigma, \tag{2.102}$$

where $b_\sigma$ is the potential buoyancy given by

$$b_\sigma \equiv -g\frac{\delta\rho_\theta}{\rho_0} = -\frac{g}{\rho_0}\left(\delta\rho + \frac{\rho_0 g z}{c_s^2}\right) = b - g\frac{z}{H_\rho}, \tag{2.103}$$

where $H_\rho = c_s^2/g$. Buoyancy itself is obtained from $b_\sigma$ by the 'equation of state', $b = b_\sigma + gz/H_\rho$.

In many applications we may need to use a still more accurate equation of state. In that case (and see Section 1.7.3) we replace (2.101) by the thermodynamic equations

$$\frac{\mathrm{D}\Theta}{\mathrm{D}t} = \dot{\Theta}, \qquad \frac{\mathrm{D}S}{\mathrm{D}t} = \dot{S}, \tag{2.104a,b}$$

where $\Theta$ is an appropriate thermodynamic state variable, such as potential enthalpy or entropy, $S$ is salinity, and an equation of state then gives the buoyancy. The equation of state has the general form $b = b(\Theta, S, p)$, but to be consistent with the level of approximation in the other Boussinesq equations we replace $p$ by the hydrostatic pressure calculated with the reference density, that is by $-\rho_0 g z$, and the equation of state then takes the general form

$$b = b(\Theta, S, z). \tag{2.105}$$

An example of (2.105) is (1.155), taken with the definition of buoyancy $b = -g\delta\rho/\rho_0$. The closed set of equations (2.94), (2.97), (2.104) and (2.105) are sometimes called the general Boussinesq equations, or, in oceanographic contexts, the seawater Boussinesq equations. Using an accurate equation of state and the Boussinesq approximation is the procedure used in many comprehensive ocean general circulation models. The Boussinesq equations, which with the hydrostatic and traditional approximations are often considered to be the oceanic primitive equations, are summarized in the shaded box on the following page.

### ✦ *Mean stratification and the buoyancy frequency*

The processes that cause density to vary in the vertical often differ from those that cause it to vary in the horizontal. For this reason it is sometimes useful to write $\rho = \rho_0 + \widehat{\rho}(z) + \rho'(x, y, z, t)$ and define $\tilde{b}(z) \equiv -g\widehat{\rho}/\rho_0$ and $b' \equiv -g\rho'/\rho_0$. Using the hydrostatic equation to evaluate pressure, the thermodynamic equation (2.98) becomes, to a good approximation,

$$\frac{\mathrm{D}b'}{\mathrm{D}t} + N^2 w = 0, \tag{2.106}$$

where $\mathrm{D}/\mathrm{D}t$ remains a three-dimensional operator and

$$N^2(z) = \left(\frac{\mathrm{d}\tilde{b}}{\mathrm{d}z} - \frac{g^2}{c_s^2}\right) = \frac{\mathrm{d}\tilde{b}_\sigma}{\mathrm{d}z}, \tag{2.107}$$

where $\tilde{b}_\sigma = \tilde{b} - gz/H_\rho$. The quantity $N^2$ is a measure of the mean stratification of the fluid, and is equal to the vertical gradient of the mean potential buoyancy. $N$ is known as the buoyancy frequency, something we return to in Section 2.10. Equations (2.106) and (2.107) also hold in the simple Boussinesq equations, but with $c_s^2 = \infty$.

### Summary of Boussinesq Equations

The simple Boussinesq equations are, for an inviscid fluid:

$$\text{momentum equations:} \quad \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{v} = -\nabla\phi + b\mathbf{k}, \tag{B.1}$$

$$\text{mass conservation:} \quad \nabla \cdot \boldsymbol{v} = 0, \tag{B.2}$$

$$\text{buoyancy equation:} \quad \frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b}. \tag{B.3}$$

A more general form replaces the buoyancy equation by:

$$\text{thermodynamic equation:} \quad \frac{\mathrm{D}\Theta}{\mathrm{D}t} = \dot{\Theta}, \tag{B.4}$$

$$\text{salinity equation:} \quad \frac{\mathrm{D}S}{\mathrm{D}t} = \dot{S}, \tag{B.5}$$

$$\text{equation of state:} \quad b = b(\Theta, S, z). \tag{B.6}$$

An equation of state of the form $b = b(\Theta, S, \phi)$ is not asymptotically correct and good conservation properties are not assured.

### 2.4.3 Energetics of the Boussinesq System

In a uniform gravitational field but with no other forcing or dissipation, we write the simple Boussinesq equations as

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\boldsymbol{\Omega} \times \boldsymbol{v} = b\mathbf{k} - \nabla\phi, \qquad \nabla \cdot \boldsymbol{v} = 0, \qquad \frac{\mathrm{D}b}{\mathrm{D}t} = 0. \tag{2.108a,b,c}$$

From (2.108a) and (2.108b) the kinetic energy density evolution (cf. Section 1.10) is given by

$$\frac{1}{2}\frac{\mathrm{D}v^2}{\mathrm{D}t} = bw - \nabla \cdot (\phi \boldsymbol{v}), \tag{2.109}$$

where the constant reference density $\rho_0$ is omitted. Let us now define the potential $\Phi \equiv -z$, so that $\nabla\Phi = -\mathbf{k}$ and

$$\frac{\mathrm{D}\Phi}{\mathrm{D}t} = \nabla \cdot (\boldsymbol{v}\Phi) = -w, \tag{2.110}$$

and using this and (2.108c) gives

$$\frac{\mathrm{D}}{\mathrm{D}t}(b\Phi) = -wb. \tag{2.111}$$

Adding (2.111) to (2.109) and expanding the material derivative gives

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\boldsymbol{v}^2 + b\Phi\right) + \nabla \cdot \left[\boldsymbol{v}\left(\frac{1}{2}\boldsymbol{v}^2 + b\Phi + \phi\right)\right] = 0. \tag{2.112}$$

This constitutes an energy equation for the Boussinesq system, and may be compared to (1.199). The energy density (divided by $\rho_0$) is just $\boldsymbol{v}^2/2 + b\Phi$. What does the term $b\Phi$ represent? Its integral, multiplied by $\rho_0$, is the potential energy of the flow minus that of the basic state, or $\int g(\rho - \rho_0) z \, \mathrm{d}z$. If there were a heating term on the right-hand side of (2.108c) this would directly provide a source of potential energy, rather than internal energy as in the compressible system. Because the fluid is incompressible, there is no conversion from kinetic and potential energy into internal energy.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

### ✦ *Energetics with a general equation of state*

Now consider the energetics of the general Boussinesq equations. Suppose first that we allow the equation of state to be a function of pressure; the equations of motion are then (2.108) except that (2.108c) is replaced by

$$\frac{\mathrm{D}\Theta}{\mathrm{D}t} = 0, \qquad \frac{\mathrm{D}S}{\mathrm{D}t} = 0, \qquad b = b(\Theta, S, \phi). \tag{2.113a,b,c}$$

where $\Theta$ is some conservative thermodynamic variable and $S$ is salinity. A little algebraic experimentation will reveal that no energy conservation law of the form (2.112) generally exists for this system! The problem arises because, by requiring the fluid to be incompressible, we eliminate the proper conversion of internal energy to kinetic energy. However, if we use the approximation $b = b(\Theta, S, z)$, the system does conserve an energy, as we now show.[7]

Define the potential, $\Pi$, as the integral of $b$ at constant potential temperature and salinity, namely

$$\Pi(\Theta, S, z) \equiv -\int_a^z b \,\mathrm{d}z', \tag{2.114}$$

where $a$ is a constant, so that $\partial\Pi/\partial z = -b$. (The quantity $\Pi$ is related to the dynamic enthalpy of Section 1.7.3.) Taking the material derivative of the left-hand side gives

$$\frac{\mathrm{D}\Pi}{\mathrm{D}t} = \left(\frac{\partial\Pi}{\partial\Theta}\right)_{S,z} \frac{\mathrm{D}\Theta}{\mathrm{D}t} + \left(\frac{\partial\Pi}{\partial S}\right)_{\Theta,z} \frac{\mathrm{D}S}{\mathrm{D}t} + \left(\frac{\partial\Pi}{\partial z}\right)_{\Theta,S} \frac{\mathrm{D}z}{\mathrm{D}t} = -bw, \tag{2.115}$$

using (2.113a,b). Combining (2.115) and (2.109) gives

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\boldsymbol{v}^2 + \Pi\right) + \nabla\cdot\left[\boldsymbol{v}\left(\frac{1}{2}\boldsymbol{v}^2 + \Pi + \phi\right)\right] = 0. \tag{2.116}$$

Thus, energetic consistency is maintained with an arbitrary equation of state, provided that the buoyancy (or density) is taken as a function of $z$ and not pressure — as Appendix A indicates is the proper thing to do.

## 2.5 EQUATIONS FOR A STRATIFIED ATMOSPHERE: THE ANELASTIC APPROXIMATION

### 2.5.1 Preliminaries

In the atmosphere the density varies significantly, especially in the vertical. However, deviations of both $\rho$ and $p$ from a statically balanced state are often quite small, and the relative vertical variation of potential temperature is also small. We can usefully exploit these observations to give a somewhat simplified set of equations, useful both for theoretical and numerical analyses because sound waves are eliminated by way of an 'anelastic' approximation.[8] To begin we set

$$\rho = \widetilde{\rho}(z) + \delta\rho(x, y, z, t), \qquad p = \widetilde{p}(z) + \delta p(x, y, z, t), \tag{2.117a,b}$$

where we assume that $|\delta\rho| \ll |\widetilde{\rho}|$ and we define $\widetilde{p}$ such that

$$\frac{\partial\widetilde{p}}{\partial z} \equiv -g\widetilde{\rho}(z). \tag{2.118}$$

The notation is similar to that for the Boussinesq case except that, importantly, the density basic state is now a (given) function of the vertical coordinate. As with the Boussinesq case, the idea is to ignore dynamic variations of density (i.e., of $\delta\rho$) except where associated with gravity. First recall a couple of ideal gas relationships involving potential temperature, $\theta$. If we define $s = \log\theta$ (so that $s$ is entropy divided by $c_p$) then

$$s = \log\theta = \log T - \frac{R}{c_p}\log p = \frac{1}{\gamma}\log p - \log\rho, \tag{2.119}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

where $\gamma = c_p/c_v$, implying

$$\delta s = \frac{\delta\theta}{\theta} = \frac{1}{\gamma}\frac{\delta p}{p} - \frac{\delta\rho}{\rho} \approx \frac{1}{\gamma}\frac{\delta p}{\widetilde{p}} - \frac{\delta\rho}{\widetilde{\rho}}. \tag{2.120}$$

Further, if $\tilde{s} \equiv \gamma^{-1}\log\widetilde{p} - \log\widetilde{\rho}$ then

$$\frac{\mathrm{d}\tilde{s}}{\mathrm{d}z} = \frac{1}{\gamma\widetilde{p}}\frac{\mathrm{d}\widetilde{p}}{\mathrm{d}z} - \frac{1}{\widetilde{\rho}}\frac{\mathrm{d}\widetilde{\rho}}{\mathrm{d}z} = -\frac{g\widetilde{\rho}}{\gamma\widetilde{p}} - \frac{1}{\widetilde{\rho}}\frac{\mathrm{d}\widetilde{\rho}}{\mathrm{d}z}. \tag{2.121}$$

In the atmosphere, the left-hand side is, typically, much smaller than either of the two terms on the right-hand side.

### 2.5.2 The Momentum Equation

The exact inviscid horizontal momentum equation is

$$(\widetilde{\rho} + \delta\rho)\left(\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u}\right) = -\nabla_z\delta p. \tag{2.122}$$

Neglecting $\delta\rho$ where it appears with $\widetilde{\rho}$ leads to

$$\boxed{\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} = -\nabla_z\phi,} \tag{2.123}$$

where $\phi = \delta p/\widetilde{\rho}$, and this is similar to the corresponding equation in the Boussinesq approximation.

The vertical component of the inviscid momentum equation is, without approximation,

$$(\widetilde{\rho} + \delta\rho)\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial\widetilde{p}}{\partial z} - \frac{\partial\delta p}{\partial z} - g\widetilde{\rho} - g\delta\rho = -\frac{\partial\delta p}{\partial z} - g\delta\rho, \tag{2.124}$$

using (2.118). Neglecting $\delta\rho$ on the left-hand side we obtain

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\widetilde{\rho}}\frac{\partial\delta p}{\partial z} - g\frac{\delta\rho}{\widetilde{\rho}} = -\frac{\partial}{\partial z}\left(\frac{\delta p}{\widetilde{\rho}}\right) - \frac{\delta p}{\widetilde{\rho}^2}\frac{\partial\widetilde{\rho}}{\partial z} - g\frac{\delta\rho}{\widetilde{\rho}}. \tag{2.125}$$

This is not a useful form for a gaseous atmosphere, since the variation of the mean density cannot be ignored. However, we may eliminate $\delta\rho$ in favour of $\delta s$ using (2.120) to give

$$\frac{\mathrm{D}w}{\mathrm{D}t} = g\,\delta s - \frac{\partial}{\partial z}\left(\frac{\delta p}{\widetilde{\rho}}\right) - \frac{g}{\gamma}\frac{\delta p}{\widetilde{p}} - \frac{\delta p}{\widetilde{\rho}^2}\frac{\partial\widetilde{\rho}}{\partial z}, \tag{2.126}$$

and using (2.121) gives

$$\frac{\mathrm{D}w}{\mathrm{D}t} = g\,\delta s - \frac{\partial}{\partial z}\left(\frac{\delta p}{\widetilde{\rho}}\right) + \frac{\mathrm{d}\tilde{s}}{\mathrm{d}z}\frac{\delta p}{\widetilde{\rho}}. \tag{2.127}$$

What have these manipulations gained us? Two things:

*(i)* The gravitational term now involves $\delta s$ rather than $\delta\rho$ which enables a more direct connection with the thermodynamic equation.

*(ii)* The potential temperature scale height (~ 100 km) in the atmosphere is much larger than the density scale height (~ 10 km), and so the last term in (2.127) is small.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

The second item thus suggests that we choose our reference state to be one of constant potential temperature. The term $\mathrm{d}\tilde{s}/\mathrm{d}z$ then vanishes and the vertical momentum equation becomes

$$\frac{\mathrm{D}w}{\mathrm{D}t} = g\,\delta s - \frac{\partial \phi}{\partial z}, \tag{2.128}$$

where $\delta s = \delta\theta/\theta_0$, where $\theta_0$ is a constant. If we define a buoyancy by $b_a \equiv g\delta s = g\delta\theta/\theta_0$, then (2.123) and (2.128) have the same form as the Boussinesq momentum equations, but with a slightly different definition of buoyancy.

### 2.5.3 Mass Conservation

Using (2.117a) the mass conservation equation may be written, without approximation, as

$$\frac{\partial \delta\rho}{\partial t} + \nabla\cdot[(\tilde{\rho} + \delta\rho)\boldsymbol{v}] = 0. \tag{2.129}$$

We neglect $\delta\rho$ where it appears with $\tilde{\rho}$ in the divergence term. Further, the local time derivative will be small if time itself is scaled advectively (i.e., $T \sim L/U$ and sound waves do not dominate), giving

$$\nabla\cdot\boldsymbol{u} + \frac{1}{\tilde{\rho}}\frac{\partial}{\partial z}(\tilde{\rho}w) = 0. \tag{2.130}$$

It is here that the eponymous anelastic approximation arises: the elastic compressibility of the fluid is neglected, and this serves to eliminate sound waves. For reference, in spherical coordinates the equation is

$$\frac{1}{a\cos\vartheta}\frac{\partial u}{\partial \lambda} + \frac{1}{a\cos\vartheta}\frac{\partial}{\partial\vartheta}(v\cos\vartheta) + \frac{1}{\tilde{\rho}}\frac{\partial(w\tilde{\rho})}{\partial z} = 0. \tag{2.131}$$

In an ideal gas, the choice of constant potential temperature determines how the reference density $\tilde{\rho}$ varies with height. In some circumstances it is convenient to let $\tilde{\rho}$ be a constant, $\rho_0$ (effectively choosing a different equation of state), in which case the anelastic equations become identical to the Boussinesq equations, albeit with the buoyancy interpreted in terms of potential temperature in the former and density in the latter.

### 2.5.4 Thermodynamic Equation

The thermodynamic equation for an ideal gas may be written

$$\frac{\mathrm{D}\ln\theta}{\mathrm{D}t} = \frac{\dot{Q}}{Tc_p}. \tag{2.132}$$

In the anelastic equations, $\theta = \theta_0 + \delta\theta$, where $\theta_0$ is constant, and the thermodynamic equation is

$$\frac{\mathrm{D}\delta s}{\mathrm{D}t} = \frac{\tilde{\theta}}{Tc_p}\dot{Q}. \tag{2.133}$$

Summarizing, the complete set of anelastic equations, with rotation but with no dissipation or diabatic terms, is

$$\begin{gathered}\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} + 2\boldsymbol{\Omega}\times\boldsymbol{v} = \mathbf{k}b_a - \nabla\phi,\\ \frac{\mathrm{D}b_a}{\mathrm{D}t} = 0,\\ \nabla\cdot(\tilde{\rho}\boldsymbol{v}) = 0,\end{gathered} \tag{2.134a,b,c}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

where $b_a = g\delta s = g\delta\theta/\theta_0$. The anelastic equations are sometimes called the 'weak Boussinesq equations', with the original incompressible set then called the 'strong Boussinesq equations'.

The main difference between the anelastic and Boussinesq sets is in the mass continuity equation, and when $\widetilde{\rho} = \rho_0 =$ constant the two equation sets are identical. However, whereas the Boussinesq approximation is a very good one for ocean dynamics, the anelastic approximation is less so for large-scale atmosphere flow: the constancy of the reference potential temperature state is not a particularly good approximation, and the deviations in density from its reference profile are not especially small, leading to inaccuracies in the momentum equation. Nevertheless, the anelastic equations have been used very productively in limited area 'large-eddy simulations' where one does not wish to make the hydrostatic approximation but where sound waves are unimportant.[9] The equations also provide a good jumping-off point for theoretical studies and for the still simpler models of Chapter 5.

### 2.5.5 ✦ Energetics of the Anelastic Equations

Conservation of energy follows in much the same way as for the Boussinesq equations, except that $\widetilde{\rho}$ enters. Take the dot product of (2.134a) with $\widetilde{\rho}\boldsymbol{v}$ to obtain

$$\widetilde{\rho}\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{1}{2}\boldsymbol{v}^2\right) = -\nabla\cdot(\phi\widetilde{\rho}\boldsymbol{v}) + b_a\widetilde{\rho}w. \tag{2.135}$$

Now, define a potential $\varPhi(z)$ such that $\nabla\varPhi = -\mathbf{k}$, and so

$$\widetilde{\rho}\frac{\mathrm{D}\varPhi}{\mathrm{D}t} = -w\widetilde{\rho}. \tag{2.136}$$

Combining this with the thermodynamic equation (2.134b) gives

$$\widetilde{\rho}\frac{\mathrm{D}(b_a\varPhi)}{\mathrm{D}t} = -wb_a\widetilde{\rho}. \tag{2.137}$$

Adding this to (2.135) gives

$$\widetilde{\rho}\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{1}{2}\boldsymbol{v}^2 + b_a\varPhi\right) = -\nabla\cdot(\phi\widetilde{\rho}\boldsymbol{v}), \tag{2.138}$$

or, expanding the material derivative,

$$\frac{\partial}{\partial t}\left[\widetilde{\rho}\left(\frac{1}{2}\boldsymbol{v}^2 + b_a\varPhi\right)\right] + \nabla\cdot\left[\widetilde{\rho}\boldsymbol{v}\left(\frac{1}{2}\boldsymbol{v}^2 + b_a\varPhi + \phi\right)\right] = 0. \tag{2.139}$$

This equation has the form

$$\frac{\partial E}{\partial t} + \nabla\cdot\left[\boldsymbol{v}(E + \widetilde{\rho}\phi)\right] = 0, \tag{2.140}$$

where $E = \widetilde{\rho}(\boldsymbol{v}^2/2 + b_a\varPhi)$ is the energy density of the flow. This is a consistent energetic equation for the system, and when integrated over a closed domain the total energy is evidently conserved. The total energy density comprises the kinetic energy and a term $\widetilde{\rho}b_a\varPhi$, which is analogous to the potential energy of a simple Boussinesq system. However, it is not exactly equal to potential energy because $b_a$ is the buoyancy based on potential temperature, not density; rather, the term combines contributions from both the internal energy and the potential energy into an enthalpy-like quantity.

## 2.6 PRESSURE AND OTHER VERTICAL COORDINATES

Although using $z$ as a vertical coordinate is a natural choice given our Cartesian worldview, it is not the only option, nor is it always the most useful one. Any variable that has a one-to-one correspondence with $z$ in the vertical, so any variable that varies monotonically with $z$, could be used; pressure and, more surprisingly, entropy, are common choices. In the atmosphere pressure almost always falls monotonically with height, and using it instead of $z$ provides a useful simplification of the mass conservation and geostrophic relations, as well as a more direct connection with observations, which are often taken at fixed values of pressure. (In the ocean pressure coordinates are essentially almost the same as height coordinates because density is almost constant.) Entropy seems an exotic vertical coordinate, but it is very useful in adiabatic flow and we consider it in Chapter 3.

### 2.6.1 General Relations

First consider a general vertical coordinate, $\xi$. Any variable $\Psi$ that is a function of the coordinates $(x, y, z, t)$ may be expressed instead in terms of $(x, y, \xi, t)$ by considering $\xi$ to be a function of the independent variables $(x, y, z, t)$. Derivatives with respect to $z$ and $\xi$ are related by

$$\frac{\partial \Psi}{\partial \xi} = \frac{\partial \Psi}{\partial z}\frac{\partial z}{\partial \xi} \qquad \text{and} \qquad \frac{\partial \Psi}{\partial z} = \frac{\partial \Psi}{\partial \xi}\frac{\partial \xi}{\partial z}. \tag{2.141a,b}$$

Horizontal derivatives in the two coordinate systems are related by the chain rule,

$$\left(\frac{\partial \Psi}{\partial x}\right)_\xi = \left(\frac{\partial \Psi}{\partial x}\right)_z + \left(\frac{\partial z}{\partial x}\right)_\xi \frac{\partial \Psi}{\partial z}, \tag{2.142}$$

and similarly for time.

The material derivative in $\xi$ coordinates may be derived by transforming the original expression in $z$ coordinates using the chain rule, but because $(x, y, \xi, t)$ are independent coordinates, and noting that the 'vertical velocity' in $\xi$ coordinates is just $\dot{\xi}$ (i.e., $\mathrm{D}\xi/\mathrm{D}t$, just as the vertical velocity in $z$ coordinates is $w = \mathrm{D}z/\mathrm{D}t$), we can write down

$$\frac{\mathrm{D}\Psi}{\mathrm{D}t} = \left(\frac{\partial \Psi}{\partial t}\right)_{x,y,\xi} + \boldsymbol{u}\cdot\nabla_\xi \Psi + \dot{\xi}\frac{\partial \Psi}{\partial \xi}, \tag{2.143}$$

where $\nabla_\xi$ is the gradient operator at constant $\xi$. The operator $\mathrm{D}/\mathrm{D}t$ is the same in $z$ or $\xi$ coordinates because it is the total derivative of some property of a fluid parcel, and this is independent of the coordinate system. However, the individual terms within it will differ between coordinate systems.

### 2.6.2 Pressure Coordinates

In pressure coordinates the analogue of the vertical velocity is $\omega \equiv Dp/Dt$, and the advective derivative itself is given by

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla_p + \omega\frac{\partial}{\partial p}. \tag{2.144}$$

Note, though, that the advective derivative is the same operator as it is in height coordinates, since it is just the total derivative of a given fluid parcel; it is just written with different coordinates.

To obtain an expression for the pressure force, now let $\xi = p$ in (2.142) and apply the relationship to $p$ itself to give

$$0 = \left(\frac{\partial p}{\partial x}\right)_z + \left(\frac{\partial z}{\partial x}\right)_p \frac{\partial p}{\partial z}, \tag{2.145}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

which, using the hydrostatic relationship, gives

$$\left(\frac{\partial p}{\partial x}\right)_z = \rho\left(\frac{\partial \Phi}{\partial x}\right)_p, \tag{2.146}$$

where $\Phi = gz$ is the *geopotential*. Thus, the horizontal pressure force in the momentum equations is

$$\frac{1}{\rho}\nabla_z p = \nabla_p \Phi, \tag{2.147}$$

where the subscripts on the gradient operator indicate that the horizontal derivatives are taken at constant $z$ or constant $p$. The horizontal momentum equation thus becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} = -\nabla_p\Phi, \tag{2.148}$$

where $\mathrm{D}/\mathrm{D}t$ is given by (2.144). The hydrostatic equation in height coordinates is $\partial p/\partial z = -\rho g$ and in pressure coordinates this becomes

$$\frac{\partial \Phi}{\partial p} = -\alpha \qquad \text{or} \qquad \frac{\partial \Phi}{\partial p} = -\frac{p}{RT}. \tag{2.149}$$

The mass continuity equation simplifies attractively in pressure coordinates, if the hydrostatic approximation is used. Recall that the mass conservation equation can be derived from the material form

$$\frac{\mathrm{D}}{\mathrm{D}t}(\rho\,\delta V) = 0, \tag{2.150}$$

where $\delta V = \delta x\,\delta y\,\delta z$ is a volume element. But by the hydrostatic relationship $\rho\delta z = -(1/g)\delta p$ and thus

$$\frac{\mathrm{D}}{\mathrm{D}t}(\delta x\,\delta y\,\delta p) = 0. \tag{2.151}$$

This is completely analogous to the expression for the material conservation of volume in an incompressible fluid, (1.15). Thus, without further ado, we write the mass conservation in pressure coordinates as

$$\nabla_p\cdot\boldsymbol{u} + \frac{\partial\omega}{\partial p} = 0, \tag{2.152}$$

where the horizontal derivative is taken at constant pressure.

The (adiabatic) thermodynamic equation is still $\mathrm{D}\theta/\mathrm{D}t = 0$, and $\theta$ may be related to pressure and temperature using its definition and the ideal gas equation to complete the equation set. However, because the hydrostatic equation is written in terms of temperature and not potential temperature it is convenient to write the thermodynamic equation accordingly. To do this we begin with the thermodynamic equation in the form of (1.99b), namely $c_p\mathrm{D}T/\mathrm{D}t - \alpha\,\mathrm{D}p/\mathrm{D}t = 0$. Since $\omega \equiv \mathrm{D}p/\mathrm{D}t$ this equation is simply

$$c_p\frac{\mathrm{D}T}{\mathrm{D}t} - \frac{RT}{p}\omega = 0, \tag{2.153}$$

which is an appropriate thermodynamic equation in pressure coordinates. It is sometimes useful to write this as

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} - \omega S_p = 0, \qquad \text{where} \qquad S_p = \frac{\kappa T}{p} - \frac{\partial T}{\partial p} = -\frac{T}{\theta}\frac{\partial\theta}{\partial p}, \tag{2.154a,b}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

## Equations of Motion in Pressure and Log-pressure Coordinates

The adiabatic, inviscid primitive equations in pressure coordinates are:

$$\frac{D\boldsymbol{u}}{Dt} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla_p \Phi, \tag{P.1}$$

$$\frac{\partial \Phi}{\partial p} = \frac{-RT}{p}, \tag{P.2}$$

$$\nabla_p \cdot \boldsymbol{u} + \frac{\partial \omega}{\partial p} = 0, \tag{P.3}$$

$$c_p \frac{DT}{Dt} - \frac{RT}{p}\omega = 0 \quad \text{or} \quad \frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} - \omega S_p = 0. \tag{P.4}$$

where $S_p = \kappa T/p - \partial T/\partial p$ and $\kappa = R/c_p$. The above equations are, respectively, the horizontal momentum equation, the hydrostatic equation, the mass continuity equation and the thermodynamic equation. Using hydrostasy and the ideal gas relation, the thermodynamic equation may also be written as

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + \omega\frac{\partial s}{\partial p} = 0, \tag{P.5}$$

where $s = T + gz/c_p$ is the dry static energy divided by $c_p$.

The corresponding equations in log-pressure coordinates are

$$\frac{D\boldsymbol{u}}{Dt} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla_Z \Phi, \tag{P.6}$$

$$\frac{\partial \Phi}{\partial Z} = \frac{RT}{H}, \tag{P.7}$$

$$\nabla_Z \cdot \boldsymbol{u} + \frac{1}{\rho_R}\frac{\partial \rho_R W}{\partial z} = 0, \tag{P.8}$$

$$c_p \frac{DT}{Dt} + W\frac{RT}{H} = 0 \quad \text{or} \quad \frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + WS_Z = 0. \tag{P.9}$$

where $\rho_R = \rho_0 \exp(-Z/H)$ and $S_Z = \kappa T/H + \partial T/\partial Z$. The thermodynamic equation may also be written as

$$\frac{\partial}{\partial t}\frac{\partial \Phi}{\partial Z} + u\frac{\partial}{\partial x}\frac{\partial \Phi}{\partial z} + v\frac{\partial}{\partial y}\frac{\partial \Phi}{\partial z} + WN_*^2 = 0, \tag{P.10}$$

where $N_*^2 = (R/H)S_Z$.

having used the ideal gas equation and the definition of potential temperature, with $\kappa = R/c_p$. Evidently, $S_p$ is an appropriate measure of static stability in pressure coordinates and it is closely related to the buoyancy frequency $N$, as we see in the next subsection.

The main practical difficulty with the pressure-coordinate equations is the lower boundary condition. Using

$$w \equiv \frac{Dz}{Dt} = \frac{\partial z}{\partial t} + \boldsymbol{u} \cdot \nabla_p z + \omega\frac{\partial z}{\partial p}, \tag{2.155}$$

and (2.149), the boundary condition of $w = 0$ at $z = z_s$ becomes

$$\frac{\partial \Phi}{\partial t} + \boldsymbol{u} \cdot \nabla_p \Phi - \alpha\omega = 0, \tag{2.156}$$

at $p(x, y, z_s, t)$. In theoretical studies, it is common to assume that the lower boundary is in fact a constant pressure surface and simply assume that $\omega = 0$, or that $\omega = -\alpha^{-1}\partial\Phi/\partial t$. For more realistic studies the fact that the level $z = 0$ is not a coordinate surface must be properly considered. For this reason, and especially if the lower boundary is uneven because of the presence of topography, so-called *sigma coordinates* are sometimes used, in which the vertical coordinate is chosen so that the lower boundary is itself a coordinate surface. Sigma coordinates may use height itself as a vertical measure (typical in oceanic applications) or use pressure (typical in atmospheric applications). In the latter case the vertical coordinate is $\sigma = p/p_s$ where $p_s(x, y, t)$ is the surface pressure. The difficulty of applying (2.156) is replaced by a prognostic equation for the surface pressure, derived from the mass conservation equation.

Interestingly, the pressure coordinate equations (collected together in the shaded box on the previous page) are isomorphic to the hydrostatic, salt-free general Boussinesq equations (see the shaded box on page 74) with $z \Leftrightarrow -p$, $w \Leftrightarrow -\omega$, $\phi \Leftrightarrow \Phi$, $b \Leftrightarrow \alpha$, $\Theta \Leftrightarrow \theta$ and an equation of state $b = b(\Theta, z) \Leftrightarrow \alpha = \alpha(\theta, p)$ (and in an ideal gas $\alpha = (\theta R/p_R)(p_R/p)^{1/\gamma}$). The dynamics of one system can often therefore be expected to have an analogue in the other.

### 2.6.3 Log-pressure Coordinates

A variant of pressure coordinates arises by using *log-pressure* coordinates, in which the vertical coordinate is $Z = -H \ln(p/p_R)$ where $p_R$ is a reference pressure (say 1000 mb) and $H$ is a constant (for example a scale height $RT_0/g$ where $T_0$ is a constant) so that $Z$ has units of length. (Uppercase letters are conventionally used for some variables in log-pressure coordinates, and these are not to be confused with scaling parameters.) The 'vertical velocity' for the system is now

$$W \equiv \frac{\mathrm{D}Z}{\mathrm{D}t}, \tag{2.157}$$

and the advective derivative is

$$\frac{\mathrm{D}}{\mathrm{D}t} \equiv \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla_p + W\frac{\partial}{\partial Z}. \tag{2.158}$$

The horizontal momentum equation is unaltered from (2.148), although we use (2.158) to evaluate the advective derivative. It is straightforward to show that the hydrostatic equation becomes

$$\frac{\partial \Phi}{\partial Z} = \frac{RT}{H}. \tag{2.159}$$

The mass continuity equation (2.152) becomes

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial W}{\partial Z} - \frac{W}{H} = 0, \tag{2.160}$$

which may be written as

$$\nabla_Z \cdot \boldsymbol{u} + \frac{1}{\rho_R}\frac{\partial(\rho_R W)}{\partial z} = 0, \tag{2.161}$$

where $\nabla_Z \cdot$ is the divergence at constant $Z$ and $\rho_R = \rho_0 \exp(-Z/H)$, so giving a form similar to the mass conservation equation in the anelastic equations. (The value of the constant $\rho_0$ may be set to one.)

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

As with pressure coordinates, it is convenient to write the thermodynamic equation in terms of temperature and not potential temperature, and in an analogous procedure to the one leading to (2.153) we obtain

$$c_p \frac{\mathrm{D}T}{\mathrm{D}t} + W \frac{RT}{H} = 0. \tag{2.162}$$

This equation may be written as

$$\frac{\partial T}{\partial t} + u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} + W S_Z = 0, \tag{2.163}$$

where

$$S_Z = \frac{\kappa T}{H} + \frac{\partial T}{\partial Z}, \tag{2.164}$$

and we may note that $S_Z = S_p p/H$. Using the hydrostatic equation we may write (2.163) as

$$\frac{\partial}{\partial t} \frac{\partial \Phi}{\partial Z} + u \frac{\partial}{\partial x} \frac{\partial \Phi}{\partial z} + v \frac{\partial}{\partial y} \frac{\partial \Phi}{\partial z} + W N_*^2 = 0, \tag{2.165}$$

where $N_*^2 = (R/H)S_Z$. The quantity $N_*$ is not exactly equal to the square of the buoyancy frequency as normally defined (for an ideal gas $N^2 = (g/\theta)\partial\theta/\partial z$), but the two can be shown to be related by $N_*/N = p/(\rho g H) = RT/gH$, and are equal for an isothermal atmosphere.[10] Integrating the hydrostatic equation between two pressure levels gives, with $\Phi = gz$,

$$z(p_2) - z(p_1) = -\frac{R}{g} \int_{p_1}^{p_2} T \,\mathrm{d} \ln p. \tag{2.166}$$

Thus, the thickness of the layer is proportional to the average temperature of the layer, and at constant temperature the geometric height increases linearly with the logarithm of pressure. At a temperature of 240 K (280 K) the scale height, $RT/g$, is about 7 km (8.2 km). A useful rule of thumb for Earth's atmosphere (and one that holds at 240 K) is that geometric height increases by about 16 km for each factor of ten decrease in pressure, and pressures of 1000 hPa, 100 hPa, 10 hPa roughly correspond to heights of 0, 16 km 32 km and so on.

## 2.7 SCALING FOR HYDROSTATIC BALANCE

We first encountered hydrostatic balance in Section 1.3.3; we now look in more detail at the conditions required for it to hold. Along with geostrophic balance, considered in the next section, it is one of the most fundamental balances in geophysical fluid dynamics. The corresponding states, hydrostasy and geostrophy, are not exactly realized, but their approximate satisfaction has profound consequences on the behaviour of the atmosphere and ocean.

### 2.7.1 Preliminaries

Consider the relative sizes of terms in (2.77c):

$$\frac{W}{T} + \frac{UW}{L} + \frac{W^2}{H} + \Omega U \sim \left| \frac{1}{\rho} \frac{\partial p}{\partial z} \right| + g. \tag{2.167}$$

For most large-scale motion in the atmosphere and ocean the terms on the right-hand side are orders of magnitude larger than those on the left, and therefore must be approximately equal. Explicitly, suppose $W \sim 1\,\mathrm{cm\,s^{-1}}$, $L \sim 10^5\,\mathrm{m}$, $H \sim 10^3\,\mathrm{m}$, $U \sim 10\,\mathrm{m\,s^{-1}}$, $T = L/U$. Then by substituting

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

into (2.167) it seems that the pressure term is the only one which could balance the gravitational term, and we are led to approximate (2.77c) by,

$$\frac{\partial p}{\partial z} = -\rho g. \tag{2.168}$$

This equation, which is a vertical momentum equation, is known as *hydrostatic balance*.

However, (2.168) is not always a useful equation! Let us suppose that the density is a constant, $\rho_0$. We can then write the pressure as

$$p(x, y, z, t) = p_0(z) + p'(x, y, z, t), \qquad \text{where} \qquad \frac{\partial p_0}{\partial z} \equiv -\rho_0 g. \tag{2.169}$$

That is, $p_0$ and $\rho_0$ are in hydrostatic balance. On the $f$-plane, the inviscid vertical momentum equation becomes, without approximation,

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho_0}\frac{\partial p'}{\partial z}. \tag{2.170}$$

Thus, *for constant density fluids the gravitational term has no dynamical effect:* there is no buoyancy force, and the pressure term in the horizontal momentum equations can be replaced by $p'$. Hydrostatic balance, and in particular (2.169), is not a useful vertical momentum equation in this case. If the fluid is stratified, we should therefore subtract off the hydrostatic pressure associated with the mean density before we can determine whether hydrostasy is a useful *dynamical* approximation, accurate enough to determine the horizontal pressure gradients. This is automatic in the Boussinesq equations, where the vertical momentum equation is

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial \phi}{\partial z} + b, \tag{2.171}$$

and the hydrostatic balance of the basic state is already subtracted out. In the more general equation,

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g, \tag{2.172}$$

we need to compare the advective term on the left-hand side with the pressure variations arising from horizontal flow in order to determine whether hydrostasy is an appropriate vertical momentum equation. Nevertheless, if we only need to determine the pressure for use in an equation of state then we simply need to compare the sizes of the dynamical terms in (2.77c) with $g$ itself, in order to determine whether a hydrostatic approximation will suffice.

### 2.7.2 Scaling and the Aspect Ratio

In a Boussinesq fluid we write the horizontal and vertical momentum equations as

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla_z \phi, \qquad \frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial \phi}{\partial z} + b. \tag{2.173a,b}$$

With $\boldsymbol{f} = 0$, (2.173a) implies the scaling

$$\phi \sim U^2. \tag{2.174}$$

If we use mass conservation, $\nabla_z \cdot \boldsymbol{u} + \partial w/\partial z = 0$, to scale vertical velocity then

$$w \sim W = \frac{H}{L}U = \alpha U, \tag{2.175}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

where $\alpha \equiv H/L$ is the aspect ratio. The advective terms in the vertical momentum equation all scale as

$$\frac{\mathrm{D}w}{\mathrm{D}t} \sim \frac{UW}{L} = \frac{U^2 H}{L^2}. \tag{2.176}$$

Using (2.174) and (2.176) the ratio of the advective term to the pressure gradient term in the vertical momentum equations then scales as

$$\frac{|\mathrm{D}w/\mathrm{D}t|}{|\partial\phi/\partial z|} \sim \frac{U^2 H/L^2}{U^2/H} \sim \left(\frac{H}{L}\right)^2. \tag{2.177}$$

Thus, the condition for hydrostasy, that $|\mathrm{D}w/\mathrm{D}t|/|\partial\phi/\partial z| \ll 1$, is:

$$\boxed{\alpha^2 \equiv \left(\frac{H}{L}\right)^2 \ll 1.} \tag{2.178}$$

The advective term in the vertical momentum may then be neglected. Thus, *hydrostatic balance arises from a small aspect ratio approximation.*

We can obtain the same result more formally by nondimensionalizing the momentum equations. Using uppercase symbols to denote scaling values we write

$$\begin{gathered}(x,y) = L(\widehat{x},\widehat{y}), \qquad z = H\widehat{z}, \qquad \boldsymbol{u} = U\widehat{\boldsymbol{u}}, \qquad w = W\widehat{w} = \frac{HU}{L}\widehat{w},\\ t = T;\hat{t} = \frac{L}{U}\hat{t}, \qquad \phi = \Phi\widehat{\phi} = U^2\widehat{\phi}, \qquad b = B\hat{b} = \frac{U^2}{H}\hat{b},\end{gathered} \tag{2.179}$$

where the hatted variables are nondimensional and the scaling for $w$ is suggested by the mass conservation equation, $\nabla_z \cdot \boldsymbol{u} + \partial w/\partial z = 0$. Substituting (2.179) into (2.173) (with $\boldsymbol{f} = 0$) gives us the nondimensional equations

$$\frac{\mathrm{D}\widehat{\boldsymbol{u}}}{\mathrm{D}\hat{t}} = -\nabla\widehat{\phi}, \qquad \alpha^2\frac{\mathrm{D}\widehat{w}}{\mathrm{D}\hat{t}} = -\frac{\partial\widehat{\phi}}{\partial\widehat{z}} + \hat{b}, \tag{2.180a,b}$$

where $\mathrm{D}/\mathrm{D}\hat{t} = \partial/\partial\hat{t} + \widehat{u}\partial/\partial\widehat{x} + \widehat{v}\partial/\partial\widehat{y} + \widehat{w}\partial/\partial\widehat{z}$ and we use the convention that when $\nabla$ operates on nondimensional quantities the operator itself is nondimensional. From (2.180b) it is clear that hydrostatic balance pertains when $\alpha^2 \ll 1$.

### 2.7.3 ✦ Effects of Stratification on Hydrostatic Balance

To include the effects of stratification we need to involve the thermodynamic equation, so let us first write down the complete set of non-rotating dimensional equations:

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\nabla_z\phi, \qquad \frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial\phi}{\partial z} + b', \tag{2.181a,b}$$

$$\frac{\mathrm{D}b'}{\mathrm{D}t} + wN^2 = 0, \qquad \nabla\cdot\boldsymbol{v} = 0. \tag{2.182a,b}$$

We have written, without approximation, $b = b'(x,y,z,t) + \tilde{b}(z)$, with $N^2 = \mathrm{d}\tilde{b}/\mathrm{d}z$; this separation is useful because the horizontal and vertical buoyancy variations may scale in different ways, and often $N^2$ may be regarded as given. (We have also redefined $\phi$ by subtracting off a static component in hydrostatic balance with $\tilde{b}$.) We nondimensionalize (2.182) by first writing

$$\begin{gathered}(x,y) = L(\widehat{x},\widehat{y}), \qquad z = H\widehat{z}, \qquad \boldsymbol{u} = U\widehat{\boldsymbol{u}}, \quad w = W\widehat{w} = \epsilon\frac{HU}{L}\widehat{w},\\ t = T\hat{t} = \frac{L}{U}\hat{t}, \qquad \phi = U^2\widehat{\phi}, \qquad b' = \Delta b\hat{b}' = \frac{U^2}{H}\hat{b}', \qquad N^2 = \overline{N}^2\widehat{N}^2,\end{gathered} \tag{2.183}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

where $\epsilon$ is, for the moment, undetermined, $\overline{N}$ is a representative constant value of the buoyancy frequency and $\Delta b$ scales only the horizontal buoyancy variations. Substituting (2.183) into (2.181) and (2.182) gives

$$\frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}\hat{t}} = -\nabla_z \widehat{\phi}, \qquad \epsilon\alpha^2 \frac{\mathrm{D}\widehat{w}}{\mathrm{D}t} = -\frac{\partial \widehat{\phi}}{\partial \widehat{z}} + \hat{b}' \tag{2.184a,b}$$

$$\frac{U^2}{\overline{N}^2 H^2} \frac{\mathrm{D}\hat{b}'}{\mathrm{D}\hat{t}} + \epsilon \widehat{w}\widehat{N}^2 = 0, \qquad \nabla \cdot \hat{\boldsymbol{u}} + \epsilon \frac{\partial \widehat{w}}{\partial \widehat{z}} = 0. \tag{2.185a,b}$$

where now $\mathrm{D}/\mathrm{D}\hat{t} = \partial/\partial\hat{t} + \hat{\boldsymbol{u}} \cdot \nabla_z + \epsilon\widehat{w}\partial/\partial\widehat{z}$. To obtain a non-trivial balance in (2.185a) we choose $\epsilon = U^2/(\overline{N}^2 H^2) \equiv Fr^2$, where $Fr$ is the *Froude number,* a measure of the stratification of the flow. A strong stratification corresponds to a small Froude number. From (2.183), the vertical velocity then scales as

$$W = \frac{Fr^2 UH}{L} \tag{2.186}$$

and if the flow is highly stratified the vertical velocity will be even smaller than a pure aspect ratio scaling might suggest. (There must, therefore, be some cancellation in horizontal divergence in the mass continuity equation; that is, $|\nabla_z \cdot \boldsymbol{u}| \ll U/L$.) With this choice of $\epsilon$ the nondimensional Boussinesq equations may be written:

$$\frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}\hat{t}} = -\nabla_z \widehat{\phi}, \qquad Fr^2\alpha^2 \frac{\mathrm{D}\widehat{w}}{\mathrm{D}\hat{t}} = -\frac{\partial \widehat{\phi}}{\partial \widehat{z}} + \hat{b}', \tag{2.187a,b}$$

$$\frac{\mathrm{D}\hat{b}'}{\mathrm{D}\hat{t}} + \widehat{w}\widehat{N}^2 = 0, \qquad \nabla \cdot \hat{\boldsymbol{u}} + Fr^2 \frac{\partial \widehat{w}}{\partial \widehat{z}} = 0. \tag{2.188a,b}$$

The nondimensional parameters in the system are the aspect ratio and the Froude number (in addition to $\widehat{N}$, but by construction this is just an order one function of $z$). From (2.187b) the condition for hydrostatic balance to hold is evidently that

$$\boxed{Fr^2\alpha^2 \ll 1,} \tag{2.189}$$

so generalizing the aspect ratio condition (2.178) to a stratified fluid. Because $Fr$ is a measure of stratification, (2.189) formalizes our intuitive expectation that the more stratified a fluid the more vertical motion is suppressed and therefore the more likely hydrostatic balance is to hold. Equation (2.189) is equivalent to

$$\frac{U^2}{\overline{N}^2 H^2} \frac{H^2}{L^2} = \frac{U^2}{L^2 \overline{N}^2} \ll 1, \tag{2.190}$$

and in a hydrostatic model the condition is always, by construction, satisfied.

Why bother with any of this scaling? Why not just say that hydrostatic balance holds when $|Dw/Dt| \ll |\partial\phi/\partial z|$? One reason is that we do not have a good idea of the value of $w$ from direct measurements, and it may change significantly in different oceanic and atmospheric parameter regimes. On the other hand the Froude number and the aspect ratio are familiar nondimensional parameters with a wide applicability in other contexts, and which we can control in a laboratory setting or estimate in the ocean or atmosphere. Still, when equations are scaled, ascertaining which parameters are to be regarded as given and which should be derived is often a choice, rather than being set a priori.

### 2.7.4 Hydrostasy in the Ocean and Atmosphere

Is the hydrostatic approximation in fact a good one in the ocean and atmosphere?

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

*In the ocean*

For the large-scale ocean circulation, let $N \sim 10^{-2}\,\mathrm{s}^{-1}$, $U \sim 0.1\,\mathrm{m\,s}^{-1}$ and $H \sim 1\,\mathrm{km}$. Then $Fr = U/(NH) \sim 10^{-2} \ll 1$. Thus, $Fr^2\alpha^2 \ll 1$ even for unit aspect-ratio motion. In fact, for larger scale flow the aspect ratio is also small; for basin-scale flow $L \sim 10^6\,\mathrm{m}$ and $Fr^2\alpha^2 \sim 0.01^2 \times 0.001^2 = 10^{-10}$ and hydrostatic balance is an extremely good approximation.

For intense convection, for example in the Labrador Sea, the hydrostatic approximation may be less appropriate, because the intense descending plumes may have an aspect ratio ($H/L$) of one or greater and the stratification is very weak. The hydrostatic condition then often becomes the requirement that the Froude number is small. Representative orders of magnitude are $U \sim W \sim 0.1\,\mathrm{m\,s}^{-1}$, $H \sim 1\,\mathrm{km}$ and $N \sim 10^{-3}\,\mathrm{s}^{-1}$ to $10^{-4}\,\mathrm{s}^{-1}$. For these values $Fr$ ranges between 0.1 and 1, and at the upper end of this range hydrostatic balance is violated.

*In the atmosphere*

Over much of the troposphere $N \sim 10^{-2}\,\mathrm{s}^{-1}$ so that with $U = 10\,\mathrm{m\,s}^{-1}$ and $H = 1\,\mathrm{km}$ we find $Fr \sim 1$. Hydrostasy is then maintained because the aspect ratio $H/L$ is much less than unity. For larger scale synoptic activity a larger vertical scale is appropriate, and with $H = 10\,\mathrm{km}$ both the Froude number and the aspect ratio are much smaller than one; indeed with $L = 1000\,\mathrm{km}$ we find $Fr^2\alpha^2 \sim 0.1^2 \times 0.01^2 = 10^{-6}$ and the flow is hydrostatic to a very good approximation indeed. However, for smaller scale atmospheric motions associated with fronts and, especially, convection, there can be little expectation that hydrostatic balance will be a good approximation.

For large-scale flows in both atmosphere and ocean, the conceptual and practical simplifications afforded by the hydrostatic approximation can hardly be overemphasized.

## 2.8 GEOSTROPHIC AND THERMAL WIND BALANCE

We now consider the dominant dynamical balance in the horizontal components of the momentum equation. In the horizontal plane (meaning along geopotential surfaces) we find that the Coriolis term is much larger than the advective terms and the dominant balance is between it and the horizontal pressure force. This balance is called *geostrophic balance,* and it occurs when the Rossby number is small, as we now investigate.

### 2.8.1 The Rossby Number

The *Rossby number* characterizes the importance of rotation in a fluid.[11] It is, essentially, the ratio of the magnitude of the relative acceleration to the Coriolis acceleration, and it is of fundamental importance in geophysical fluid dynamics. It arises from a simple scaling of the horizontal momentum equation, namely

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{u} + \boldsymbol{f}\times\boldsymbol{u} = -\frac{1}{\rho}\nabla_z p, \tag{2.191a}$$

$$U^2/L \qquad\qquad fU, \tag{2.191b}$$

where $U$ is the approximate magnitude of the horizontal velocity and $L$ is a typical length scale over which that velocity varies. (We assume that $W/H \lesssim U/L$, so that vertical advection does not dominate the advection.) The ratio of the sizes of the advective and Coriolis terms is defined to be the Rossby number,

$$\boxed{Ro \equiv \frac{U}{fL}.} \tag{2.192}$$

If the Rossby number is small then rotation effects are important and, as the values in Table 2.1 indicate, this is the case for large-scale flow in both ocean and atmosphere.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

| Variable | Scaling symbol | Meaning | Atmos. value | Ocean value |
|---|---|---|---|---|
| $(x, y)$ | $L$ | Horizontal length scale | $10^6$ m | $10^5$ m |
| $t$ | $T$ | Time scale | 1 day ($10^5$ s) | 10 days ($10^6$ s) |
| $(u, v)$ | $U$ | Horizontal velocity | $10\,\mathrm{m\,s^{-1}}$ | $0.1\,\mathrm{m\,s^{-1}}$ |
| | $Ro$ | Rossby number, $U/fL$ | 0.1 | 0.01 |

**Table 2.1** Scales of large-scale flow in atmosphere and ocean. The choices given are representative of large-scale mid-latitude eddying motion in both systems.

Another intuitive way to think about the Rossby number is in terms of time scales. The Rossby number based on a time scale is

$$Ro_T \equiv \frac{1}{fT}, \tag{2.193}$$

where $T$ is a time scale associated with the dynamics at hand. If the time scale is an advective one, meaning that $T \sim L/U$, then this definition is equivalent to (2.192). Now, $f = 2\Omega \sin\vartheta$, where $\Omega$ is the angular velocity of the rotating frame and equal to $2\pi/T_p$ where $T_p$ is the period of rotation (24 hours). Thus,

$$Ro_T = \frac{T_p}{4\pi T \sin\vartheta} = \frac{T_i}{T}, \tag{2.194}$$

where $T_i = 1/f$ is the 'inertial time scale', about three hours in mid-latitudes. Thus, for phenomena with time scales much longer than this, such as the motion of the Gulf Stream or a mid-latitude atmospheric weather system, the effects of the Earth's rotation can be expected to be important, whereas a short-lived phenomenon, such as a cumulus cloud or tornado, may be oblivious to such rotation.

### 2.8.2 Geostrophic Balance

If the Rossby number is sufficiently small in (2.191a) then the rotation term will dominate the nonlinear advection term, and if the time period of the motion scales advectively then the rotation term also dominates the local time derivative. The only term that can then balance the rotation term is the pressure term, and therefore we must have

$$\boldsymbol{f} \times \boldsymbol{u} \approx -\frac{1}{\rho}\nabla_z p, \tag{2.195}$$

or, in Cartesian component form

$$fu \approx -\frac{1}{\rho}\frac{\partial p}{\partial y}, \qquad fv \approx \frac{1}{\rho}\frac{\partial p}{\partial x}. \tag{2.196}$$

This balance is known as *geostrophic balance,* and its consequences are profound, giving geophysical fluid dynamics a special place in the broader field of fluid dynamics. We *define* the geostrophic velocity by

$$\boxed{fu_g \equiv -\frac{1}{\rho}\frac{\partial p}{\partial y}, \qquad fv_g \equiv \frac{1}{\rho}\frac{\partial p}{\partial x},} \tag{2.197}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

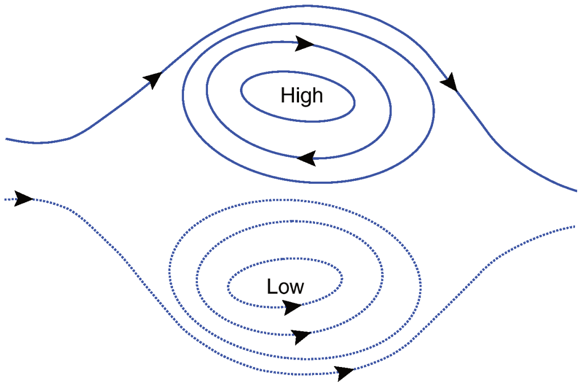


**Fig. 2.5** Geostrophic flow with a positive value of the Coriolis parameter $f$. Flow is parallel to the lines of constant pressure (isobars). Cyclonic flow is anticlockwise around a low pressure region and anticyclonic flow is clockwise around a high. If $f$ were negative, as in the Southern Hemisphere, (anti)cyclonic flow would be (anti)clockwise.

and for low Rossby number flow $u \approx u_g$ and $v \approx v_g$. In spherical coordinates the geostrophic velocity is

$$fu_g = -\frac{1}{\rho a}\frac{\partial p}{\partial \vartheta}, \qquad fv_g = \frac{1}{a\rho\cos\vartheta}\frac{\partial p}{\partial \lambda}, \tag{2.198}$$

where $f = 2\Omega\sin\vartheta$. Geostrophic balance has a number of immediate ramifications:

- Geostrophic flow is parallel to lines of constant pressure (isobars). If $f > 0$ the flow is anti-clockwise round a region of low pressure and clockwise around a region of high pressure (see Fig. 2.5).

- If the Coriolis force is constant and if the density does not vary in the horizontal the geostrophic flow is horizontally non-divergent and

$$\nabla_z \cdot \boldsymbol{u}_g = \frac{\partial u_g}{\partial x} + \frac{\partial v_g}{\partial y} = 0. \tag{2.199}$$

  We may define the *geostrophic streamfunction*, $\psi$, by

$$\psi \equiv \frac{p}{f_0\rho_0}, \qquad \text{whence} \qquad u_g = -\frac{\partial \psi}{\partial y}, \quad v_g = \frac{\partial \psi}{\partial x}. \tag{2.200}$$

  The vertical component of vorticity, $\zeta$, is then given by

$$\zeta = \mathbf{k}\cdot\nabla\times\boldsymbol{v} = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla_z^2\psi. \tag{2.201}$$

- If the Coriolis parameter is not constant, then cross-differentiating (2.197) gives, for constant density geostrophic flow,

$$v_g\frac{\partial f}{\partial y} + f\nabla_z\cdot\boldsymbol{u}_g = 0. \tag{2.202}$$

  Using the mass continuity equation, $\nabla_z \cdot \boldsymbol{u}_g = -\partial w/\partial z$, we then obtain

$$\beta v_g = f\frac{\partial w}{\partial z}, \tag{2.203}$$

  where $\beta \equiv \partial f/\partial y = 2\Omega\cos\vartheta/a$. This geostrophic vorticity balance is sometimes known as 'Sverdrup balance', although that expression is better restricted to the case when the vertical velocity comes from a wind stress, as considered in Chapter 19.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

### 2.8.3 Taylor–Proudman Effect

If $\beta = 0$, then (2.203) implies that the vertical velocity is not a function of height. In fact, in that case none of the components of velocity vary with height if density is also constant. To show this, in the limit of zero Rossby number we first write the three-dimensional momentum equation as

$$\boldsymbol{f}_0 \times \boldsymbol{v} = -\nabla\phi - \nabla\chi, \tag{2.204}$$

where $\boldsymbol{f}_0 = 2\boldsymbol{\Omega} = 2\Omega\mathbf{k}$, $\phi = p/\rho_0$, and $\nabla\chi$ represents other potential forces. If $\chi = gz$ then the vertical component of this equation represents hydrostatic balance, and the horizontal components represent geostrophic balance. On taking the curl of this equation, the terms on the right-hand side vanish and the left-hand side becomes

$$(\boldsymbol{f}_0 \cdot \nabla)\boldsymbol{v} - \boldsymbol{f}_0\nabla \cdot \boldsymbol{v} - (\boldsymbol{v} \cdot \nabla)\boldsymbol{f}_0 + \boldsymbol{v}\nabla \cdot \boldsymbol{f}_0 = 0. \tag{2.205}$$

But $\nabla \cdot \boldsymbol{v} = 0$ by mass conservation, and because $\boldsymbol{f}_0$ is constant both $\nabla \cdot \boldsymbol{f}_0$ and $(\boldsymbol{v} \cdot \nabla)\boldsymbol{f}_0$ vanish. Equation (2.205) thus reduces to

$$(\boldsymbol{f}_0 \cdot \nabla)\boldsymbol{v} = 0, \tag{2.206}$$

which, since $\boldsymbol{f}_0 = f_0\mathbf{k}$, implies $f_0\partial\boldsymbol{v}/\partial z = 0$, and in particular we have

$$\frac{\partial u}{\partial z} = 0, \quad \frac{\partial v}{\partial z} = 0, \quad \frac{\partial w}{\partial z} = 0. \tag{2.207}$$

All three components of velocity are uniform along the axis of rotation.

A different presentation of this argument proceeds as follows. If the flow is exactly in geostrophic and hydrostatic balance then

$$v = \frac{1}{f_0}\frac{\partial\phi}{\partial x}, \quad u = -\frac{1}{f_0}\frac{\partial\phi}{\partial y}, \qquad \frac{\partial\phi}{\partial z} = -g. \tag{2.208a,b,c}$$

Differentiating (2.208a,b) with respect to $z$, and using (2.208c) yields

$$\frac{\partial v}{\partial z} = \frac{-1}{f_0}\frac{\partial g}{\partial x} = 0, \qquad \frac{\partial u}{\partial z} = \frac{1}{f_0}\frac{\partial g}{\partial y} = 0. \tag{2.209}$$

Noting that the geostrophic velocities are horizontally non-divergent ($\nabla_z \cdot \boldsymbol{u} = 0$), and using mass continuity then gives $\partial w/\partial z = 0$, as before.

If there is a solid horizontal boundary anywhere in the fluid, for example at the surface, then $w = 0$ at that surface and thus $w = 0$ everywhere. Hence the motion occurs in planes that lie perpendicular to the axis of rotation, and the flow is effectively two dimensional. This result is known as the *Taylor–Proudman effect,* namely that for constant density flow in geostrophic and hydrostatic balance the vertical derivatives of the horizontal and the vertical velocities are zero.[12] At zero Rossby number, if the vertical velocity is zero somewhere in the flow then it is zero everywhere in that vertical column; furthermore, the horizontal flow has no vertical shear, and the fluid moves like a slab. The effects of rotation have provided a *stiffening* of the fluid in the vertical.

In neither the atmosphere nor the ocean do we observe precisely such vertically coherent flow, mainly because of the effects of stratification. However, it is typical of geophysical fluid dynamics that the assumptions underlying a derivation are not fully satisfied, yet there are manifestations of it in real flow. For example, one might have naïvely expected, because $\partial w/\partial z = -\nabla_z \cdot \boldsymbol{u}$, that the scales of the various variables would be related by $W/H \sim U/L$. However, if the flow is rapidly rotating we expect that the horizontal flow will be in near geostrophic balance and therefore nearly divergence free; thus $\nabla_z \cdot \boldsymbol{u} \ll U/L$, and $W \ll HU/L$.

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

### 2.8.4 Thermal Wind Balance

Thermal wind balance arises by combining the geostrophic and hydrostatic approximations, and this is most easily done in the context of the anelastic (or Boussinesq) equations, or in pressure coordinates. For the anelastic equations, geostrophic balance may be written

$$-fv_g = -\frac{\partial \phi}{\partial x} = -\frac{1}{a\cos\vartheta}\frac{\partial \phi}{\partial \lambda}, \qquad fu_g = -\frac{\partial \phi}{\partial y} = -\frac{1}{a}\frac{\partial \phi}{\partial \vartheta}. \tag{2.210a,b}$$

Combining these relations with hydrostatic balance, $\partial\phi/\partial z = b$, gives

$$\boxed{\begin{aligned} -f\frac{\partial v_g}{\partial z} &= -\frac{\partial b}{\partial x} = -\frac{1}{a\cos\lambda}\frac{\partial b}{\partial \lambda}, \\ f\frac{\partial u_g}{\partial z} &= -\frac{\partial b}{\partial y} = -\frac{1}{a}\frac{\partial b}{\partial \vartheta}. \end{aligned}} \tag{2.211a,b}$$

These equations represent *thermal wind balance*, and the vertical derivative of the geostrophic wind is the 'thermal wind'. Equation (2.211) may be written in terms of the zonal angular momentum as

$$\frac{\partial m_g}{\partial z} = -\frac{a}{2\Omega\tan\vartheta}\frac{\partial b}{\partial y}, \tag{2.212}$$

where $m_g = (u_g + \Omega a\cos\vartheta)a\cos\vartheta$. Potentially more accurate than geostrophic balance is the so-called gradient wind balance, discussed more in Section 2.9, which retains a centrifugal term in the momentum equation. The meridional momentum equation (2.50b) becomes

$$2u\Omega\sin\vartheta + \frac{u^2}{a}\tan\vartheta \approx -\frac{\partial \phi}{\partial y} = -\frac{1}{a}\frac{\partial \phi}{\partial \vartheta}. \tag{2.213}$$

For large-scale flow this only differs significantly from geostrophic balance very close to the equator. Taking the vertical derivative of (2.213) and using the hydrostatic relation $\partial\phi/\partial z = b$ gives a modified thermal wind relation, and this may be put in the simple form

$$\frac{\partial m^2}{\partial z} = -\frac{a^3\cos^3\vartheta}{\sin\vartheta}\frac{\partial b}{\partial y}, \tag{2.214}$$

where $m = (u + \Omega a\cos\vartheta)a\cos\vartheta$ is the annular angular momentum.

If the density or buoyancy is constant then there is no shear and (2.211) or (2.214) give the Taylor–Proudman result. But suppose that the temperature falls in the poleward direction. Then thermal wind balance implies that the (eastward) wind will increase with height — just as is observed in the atmosphere! In general, a vertical shear of the horizontal wind is associated with a horizontal temperature gradient, and this is one of the most simple and far-reaching effects in geophysical fluid dynamics. The underlying physical mechanism is illustrated in Fig. 2.6.

#### *Geostrophic and thermal wind balance in pressure coordinates*

In pressure coordinates geostrophic balance is just

$$\boldsymbol{f} \times \boldsymbol{u}_g = -\nabla_p \Phi, \tag{2.215}$$

where $\Phi$ is the geopotential and $\nabla_p$ is the gradient operator taken at constant pressure. If $f$ is constant, it follows from (2.215) that the geostrophic wind is non-divergent on pressure surfaces.

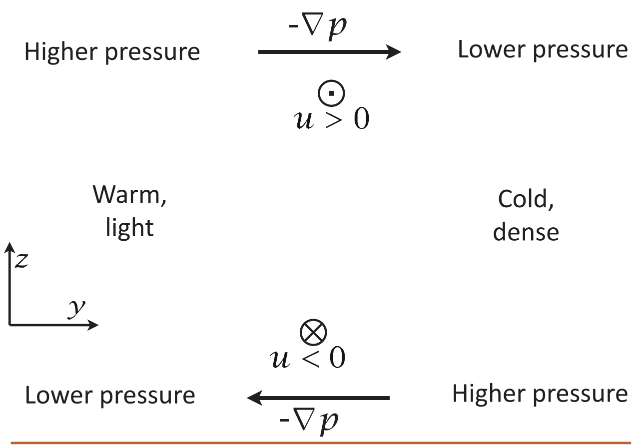


**Fig. 2.6** The mechanism of thermal wind. A cold fluid is denser than a warm fluid, so by hydrostasy the vertical pressure gradient is greater where the fluid is cold. Thus, pressure gradients form as shown, where 'higher' and 'lower' mean relative to the average at that height. The horizontal pressure gradients are balanced by the Coriolis force, producing (for $f > 0$) the horizontal winds shown. Only the wind *shear* is given by the thermal wind.

Taking the vertical derivative of (2.215) (that is, its derivative with respect to $p$) and using the hydrostatic equation, $\partial\Phi/\partial p = -\alpha$, gives the thermal wind equation

$$\boldsymbol{f} \times \frac{\partial \boldsymbol{u}_g}{\partial p} = \nabla_p \alpha = \frac{R}{p}\nabla_p T, \tag{2.216}$$

where the last equality follows using the ideal gas equation and because the horizontal derivative is at constant pressure. In component form this is

$$-f\frac{\partial v_g}{\partial p} = \frac{R}{p}\frac{\partial T}{\partial x}, \qquad f\frac{\partial u_g}{\partial p} = \frac{R}{p}\frac{\partial T}{\partial y}. \tag{2.217}$$

In log-pressure coordinates, with $Z = -H\ln(p/p_R)$, thermal wind is

$$\boldsymbol{f} \times \frac{\partial \boldsymbol{u}_g}{\partial Z} = -\frac{R}{H}\nabla_Z T. \tag{2.218}$$

The effect in all these cases is the same: a horizontal temperature gradient, or a temperature gradient along an isobaric surface, is accompanied by a vertical shear of the horizontal wind.

### 2.8.5 ✦ Vertical Velocity and Hydrostatic Balance

*Scaling for vertical velocity*

If the Coriolis parameter is constant then flows that are in geostrophic balance have zero horizontal divergence ($\nabla_x \cdot \boldsymbol{u} = 0$) and zero vertical velocity. We can therefore expect that any flow with small Rossby number will have a correspondingly small vertical velocity. Let us make this statement more precise using the rotating Boussinesq equations, (2.173) with constant Coriolis parameter. Let $\boldsymbol{u} = \boldsymbol{u}_g + \boldsymbol{u}_a$ where the geostrophic flow satisfies $\boldsymbol{f}_0 \times \boldsymbol{u}_g = -\nabla\phi$. The horizontal momentum equation, with corresponding scales for each term, then becomes

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + w\frac{\partial \boldsymbol{u}}{\partial z} + \boldsymbol{f}_0 \times \boldsymbol{u}_a = 0, \tag{2.219}$$

$$\frac{U^2}{L} \qquad \frac{U^2}{L} \qquad \frac{WU}{H} \qquad f_0 U_a. \tag{2.220}$$

This equation suggests a scaling for the ageostrophic flow of

$$U_a = \frac{U}{f_0 L}U = Ro\, U. \tag{2.221}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

That is, the ageostrophic flow is Rossby number smaller (at least) than the geostrophic flow. To obtain a scaling for the vertical velocity we look to the mass continuity equation written in the form

$$\frac{\partial w}{\partial z} = -\nabla \cdot \boldsymbol{u}_a, \tag{2.222}$$

since only the ageostrophic flow has a divergence. Equations (2.221) and (2.222) suggest the scaling

$$W = Ro\,\frac{HU}{L}. \tag{2.223}$$

That is, the vertical velocity is order Rossby number smaller than an estimate based purely on the mass continuity equation would suggest.

If the Coriolis parameter is not constant then the geostrophic flow itself is divergent and this induces a vertical velocity, as in (2.203). The scaling for vertical velocity is now

$$W = \frac{\beta}{f} HU = Ro_\beta\,\frac{HU}{L}, \tag{2.224}$$

where $Ro_\beta = \beta L/f$ is the *beta Rossby number*. It is less than one for all flows except those with a truly global scale.

### *Scaling for hydrostatic balance*

Let us nondimensionalize the rotating Boussinesq equations, (2.173), by writing

$$\begin{gathered}(x, y) = L(\hat{x}, \hat{y}), \qquad z = H\hat{z}, \qquad \boldsymbol{u} = U\hat{\boldsymbol{u}}, \qquad t = T\hat{t} = \frac{L}{U}\hat{t}, \qquad \boldsymbol{f} = f_0\widehat{\boldsymbol{f}}, \\ w = \frac{\epsilon HU}{L}\widehat{w}, \qquad \phi = \Phi\widehat{\phi} = f_0 UL\widehat{\phi}, \qquad b = B\hat{b} = \frac{f_0 UL}{H}\hat{b}.\end{gathered} \tag{2.225}$$

These relations are almost the same as (2.179), except for the factor of $\epsilon$ in the scaling of $w$. If the Coriolis parameter is constant or nearly so then, from (2.223), $\epsilon = Ro$, whereas if the Coriolis parameter varies then $\epsilon = Ro_\beta$, as in (2.223). The scaling for $\phi$ and $b'$ are suggested by geostrophic and thermal wind balance with $f_0$ a representative value of $f$. Substituting these values into (2.173) we obtain the scaled momentum equations:

$$Ro\,\frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}\hat{t}} + \widehat{\boldsymbol{f}} \times \hat{\boldsymbol{u}} = -\nabla\widehat{\phi}, \qquad Ro\,\epsilon\alpha^2\,\frac{\mathrm{D}\widehat{w}}{\mathrm{D}\hat{t}} = -\frac{\partial\widehat{\phi}}{\partial\hat{z}} - \hat{b}, \tag{2.226a,b}$$

where $\mathrm{D}/\mathrm{D}\hat{t} = \partial/\partial\hat{t} + \hat{\boldsymbol{u}}\cdot\nabla_z + \epsilon\widehat{w}\partial/\partial\hat{z}$. There are two notable aspects to these equations. First and most obviously, when $Ro \ll 1$, (2.226a) reduces to geostrophic balance, $\boldsymbol{f} \times \boldsymbol{u} \approx -\nabla\widehat{\phi}$. Second, the material derivative in (2.226b) is multiplied by three nondimensional parameters, and we can understand the appearance of each as follows:

*(i)* The aspect ratio dependence ($\alpha^2$) arises in the same way as for non-rotating flows — that is, because of the presence of $w$ and $z$ in the vertical momentum equation as opposed to $(u, v)$ and $(x, y)$ in the horizontal equations.

*(ii)* The Rossby number dependence ($Ro$) arises because in rotating flow the pressure gradient is balanced by the Coriolis force, and the advective terms are Rossby-number smaller.

*(iii)* The factor $\epsilon$ arises because in rotating flow $w$ is smaller than $u$ by $\epsilon$ times the aspect ratio. The factor may be the Rossby number itself, or the beta Rossby number.

The factor $Ro\,\epsilon\alpha^2$ is very small for large-scale flow; the reader is invited to calculate representative values. Evidently, a rapidly rotating fluid is more likely to be in hydrostatic balance than a non-rotating fluid, other conditions being equal. The combined effects of rotation and stratification are, not surprisingly, quite subtle and we leave that topic for Chapter 5.

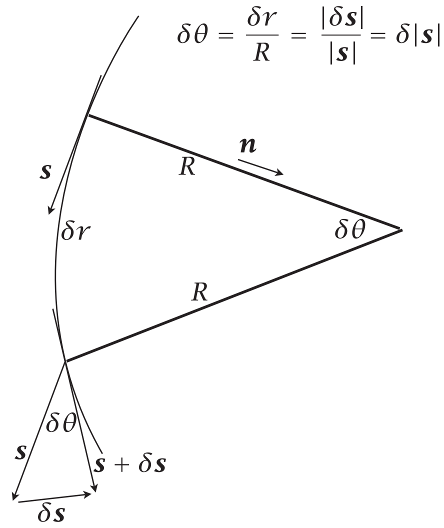


**Fig. 2.7** A parcel tracing a curved path with radius of curvature $R$, moving a small distance $\delta r$ and through a small angle $\delta\theta$. The parcel may experience centrifugal forces along $\boldsymbol{n}$ as well as Coriolis forces due to Earth's rotation.

## 2.9 ✦ GRADIENT WIND BALANCE

If a flow follows a curved path then our intuition suggests that it will experience a centrifugal force of some kind in addition to the Coriolis force. We can easily imagine that a parcel then experiences a three-way balance, between Coriolis, centrifugal and pressure forces, and this balance is called *gradient wind balance.* It is a more general, and more accurate, balance than geostrophic balance but it is not always as useful. To illustrate it we will keep matters simple and consider purely horizontal flow of constant density. We first introduce the notion of *natural coordinates* and show how gradient wind balance emerges straightforwardly.[13] We then discuss how gradient wind balance arises in the Eulerian equations of motion in a fixed coordinate system.

### 2.9.1 Natural Coordinates

For most purposes a coordinate system that is fixed in space, or rotating coincidentally with the Earth, is the most practically useful. However, it is sometimes useful to use a coordinate system that is moving with the local flow. To that end, and restricting our attention to horizontal flow, consider a parcel of fluid moving with velocity $\boldsymbol{u}$. We define a *natural coordinate system* by the set of unit vectors $\boldsymbol{s}$, $\boldsymbol{n}$ and $\mathbf{k}$, where $\boldsymbol{s}$ is the unit vector tangential to the flow, $\boldsymbol{n}$ is the horizontal unit vector normal to the flow and $\mathbf{k}$ is the unit vector in the vertical. Apart from $\mathbf{k}$, all these vectors evolve with the flow. Our goal is to split the horizontal momentum equation into components parallel to and normal to the direction of the local flow and thereby to discern the force balances in either direction. (Readers who trust their intuition may skip ahead to (2.231), which they may find obvious.)

If $U$ is the speed of the parcel then $\boldsymbol{u} = U\boldsymbol{s}$ and so

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = \boldsymbol{s}\frac{\mathrm{d}U}{\mathrm{d}t} + U\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t}. \tag{2.227}$$

Furthermore, if the flow of a parcel follows the curve $r(t)$ — that is, $r$ gives the distance moved by the parcel — then $U = \mathrm{d}r/\mathrm{d}t$. To obtain a useful expression for $\mathrm{d}\boldsymbol{s}/\mathrm{d}t$ we note that, as in Fig. 2.7,

$$\theta \equiv \frac{\delta r}{R} = \frac{|\delta\boldsymbol{s}|}{|\boldsymbol{s}|} = |\delta\boldsymbol{s}|, \tag{2.228}$$

where $R$ is the *radius of curvature* and $|\boldsymbol{s}| = 1$. [The radius of curvature may be evaluated geometrically as follows. Draw a line tangent to the path at some point, and draw a line perpendicular to the

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

tangent through that point. An infinitesimal distance along the curve, construct another perpendicular line in the same manner. The two perpendicular lines meet at the center of curvature, and the distance along one of the perpendicular lines from the center of curvature to the curve itself is the radius of curvature. By convention, the radius of curvature is positive (negative) if the parcel is curving to the left (right).]

The change of $\boldsymbol{s}$ is directed along $\boldsymbol{n}$, and so from (2.228) we infer that

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}r} = \frac{\boldsymbol{n}}{R}. \tag{2.229}$$

Thus, the rate of change of $\boldsymbol{s}$ is given by

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}r}\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\boldsymbol{n}}{R}U, \tag{2.230}$$

and using this expression in (2.227), the acceleration of the fluid parcel is given by

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = \boldsymbol{s}\frac{\mathrm{d}U}{\mathrm{d}t} + \boldsymbol{n}\frac{U^2}{R}. \tag{2.231}$$

The two terms on the right-hand side may be interpreted as being, respectively, the change in speed of the parcel along its path and the centripetal acceleration owing to the curvature of the path (if the path is a straight line then $R$ is infinite and the centripetal acceleration is zero).

### 2.9.2 Application to Fluids

Now consider the application of the above to a fluid in a rotating frame of reference. The total time derivatives of (2.231) may be replaced by material derivatives, and additional pressure and Coriolis forces act on the flow. The pressure-gradient force, $-\nabla\phi$, has components along and perpendicular to the flow so that

$$-\nabla\phi = -\left(\frac{\partial\phi}{\partial r}\boldsymbol{s} + \frac{\partial\phi}{\partial n}\boldsymbol{n}\right). \tag{2.232}$$

Given our convention, $\partial\phi/\partial n$ is positive if pressure increases to the left of the trajectory of the flow. The Coriolis force, $-\boldsymbol{f}\times\boldsymbol{u}$, has only a component perpendicular to the flow so that

$$-\boldsymbol{f}\times\boldsymbol{u} = -fU\boldsymbol{n}. \tag{2.233}$$

Because $\boldsymbol{n}$ is directed to the left of the flow, for positive $f$ the Coriolis force is directed to the right of the flow. Using (2.231), (2.232) and (2.233), and neglecting friction, we can write down the components of the momentum equation parallel and perpendicular to the flow, namely

$$\frac{\mathrm{D}U}{\mathrm{D}t} = -\frac{\partial\phi}{\partial r}, \qquad \frac{U^2}{R} + fU = -\frac{\partial\phi}{\partial n}. \tag{2.234a,b}$$

These equations tell us, at least in principle, how a fluid field will evolve from some initial conditions. The radius of curvature can only be determined from the instantaneous velocity field for steady flow: if the flow is unsteady then the radius of curvature for a parcel depends on the pressure field and differs from the radius of curvature of a streamline.

#### *Gradient wind balance, cyclostrophic balance and inertial flow*

Gradient flow (or, as it is commonly called, the gradient wind) is the flow that satisfies (2.234b). We may always define a gradient wind at a given point by using this equation, and the only forces that are neglected are frictional ones. Given the pressure field $\phi$ we may use (2.234b) to calculate the gradient wind using the quadratic formula. Gradient wind balance is thus a better approximation to the real flow than is geostrophic balance (which omits the centrifugal term). The gradient wind will not in general be a steady solution to the equations of motion — the flow will accelerate and the pressure field will change.

If the flow is in a straight line then the radius of curvature is infinite and (2.234b) reduces to $fU = -\partial\phi/\partial n$, which is geostrophic balance. For this to hold exactly the flow need not be steady, just in a straight line. It will be a good approximation to the real flow when $fU \gg U^2/R$ or $U/fR \ll 1$, taking all quantities as positive, which is similar to a condition of low Rossby number. In contrast, cyclostrophic flow occurs when the quantity $U/fR$ is large and the Coriolis force may be neglected, giving $U^2/R = -\partial\phi/\partial n$.

Inertial flow arises when the pressure gradient vanishes and (2.234b) reduces to $U/R + f = 0$. Now, if the pressure gradient vanishes then, from (2.234a) the speed of fluid parcels is constant. Thus, if $f$ is constant the fluid parcels execute circles, known as inertia circles, of radius $U/f$ and period $2\pi/f = 1\,\mathrm{day}/(2\sin\vartheta)$ where $\vartheta$ is latitude. Such a motion is evidently not truly inertial, for inertial motion is in a straight line and at a constant speed; other forces — gravitational and centrifugal — must act to keep the flow in the horizontal plane. Inertial motion is rarely a good approximation to flow in the atmosphere or ocean, although occasionally parcels are observed to trace approximations to inertia circles, especially in the ocean.

### 2.9.3 Gradient Wind Balance in the Two-dimensional Eulerian Equations

Let us now consider gradient wind balance in the Eulerian equations of motion. The basic ideas can be exposed by considering unforced constant-density two dimensional flow in a rotating reference frame, for which the equations of motion are

$$\frac{\mathrm{D}u}{\mathrm{D}t} - fv = -\frac{\partial\phi}{\partial x}, \qquad \frac{\mathrm{D}v}{\mathrm{D}t} + fu = -\frac{\partial\phi}{\partial y}, \qquad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0. \tag{2.235a,b,c}$$

We may introduce a stream function $\psi$ such that $u = -\partial\psi/\partial y$ and $v = \partial\psi/\partial x$. If the flow is geostrophic then $f \times (u, v) = (-\partial\phi/\partial y, \partial\phi/\partial x)$; however, unless the Coriolis parameter $f$ is constant the streamfunction is not proportional to the pressure field. The vorticity (considered further in Chapter 4), which is defined to be the curl of the velocity, is given by

$$\boldsymbol{\omega} \equiv \nabla \times \boldsymbol{u} = \mathbf{k}\zeta, \qquad \text{where} \qquad \zeta = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} \equiv \nabla_z^2\psi. \tag{2.236}$$

Let us form the evolution equations for vorticity and divergence by taking the curl and divergence of (2.235a,b). After a little algebra the vorticity equation is found to be

$$\frac{\partial\zeta}{\partial t} + \boldsymbol{u}\cdot\nabla(\zeta + f) = 0, \qquad \text{or equivalently} \qquad \frac{\partial\zeta}{\partial t} + J(\psi, \zeta + f) = 0, \tag{2.237}$$

where $J(a, b) \equiv (\partial a/\partial x)(\partial b/\partial y) - (\partial a/\partial y)(\partial b/\partial x)$ is the *Jacobian*. If $f = f(y)$ then $J(\psi, f) = \boldsymbol{u}\cdot\nabla f = \beta v$ where $\beta = \partial f/\partial y$. Because the velocity is divergence-free the vorticity equation above is closed; that is, it may be integrated without the need to solve any other evolution equations, and in particular without the need to solve for the pressure field. (In general the vorticity equation alone is not closed.)

If we take the divergence of (2.235a,b) then, using (2.235c) the time derivative disappears and again after a little algebra we obtain the divergence equation,

$$2J(u, v) + \nabla\cdot(f\nabla\psi) = \nabla^2\phi. \tag{2.238}$$

In this context, (2.238) may be regarded as an equation for pressure, $\phi$, given the velocity or the streamfunction. Equation (2.238) is commonly referred to as the *gradient wind balance relation*,

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

and it is analogous to (2.234b) in that it generalizes and is more accurate than geostrophic balance. Equation (2.238) is *exact* for two-dimensional, incompressible, inviscid and unforced flow, even when time-dependent. If the Rossby number is small then the second term on the left-hand side of (2.238) dominates the first, and $\nabla \cdot (f \nabla \psi) \approx \nabla^2 \phi$. This is equivalent to geostrophic balance, and if $f$ is constant the streamfunction and pressure field are proportional to each other.

In the more general case the two-dimensional divergence is not zero; that is $\partial_t(\partial_x u + \partial_y v) \neq 0$. However, if the Rossby number is small and if $f$ is nearly constant, then the geostrophic relation implies that the two-dimensional divergence will be small compared to the two-dimensional vorticity. In this case, gradient wind balance will be a good *approximation* to the flow, and a somewhat better approximation, if not more useful, than geostrophic balance alone.

## 2.10 STATIC INSTABILITY AND THE PARCEL METHOD

In this and the next couple of sections we consider how a fluid might oscillate if it were perturbed away from a resting state. Our focus is on vertical displacements, and the restoring force is gravity. Given that, the simplest way to approach the problem is to consider from first principles the pressure and gravitational forces on a displaced parcel, as in Fig. 2.8. To this end, consider a fluid initially at rest in a constant gravitational field, and therefore in hydrostatic balance. Suppose that a small parcel of the fluid is adiabatically displaced upwards by the small distance $\delta z$, without altering the overall pressure field; that is, the fluid parcel instantly assumes the pressure of its environment. If after the displacement the parcel is lighter than its environment, it will accelerate upwards, because the upward pressure gradient force is now greater than the downward gravity force on the parcel; that is, the parcel is *buoyant* (a manifestation of Archimedes' principle) and the fluid is *statically unstable*. If on the other hand the fluid parcel finds itself heavier than its surroundings, the downward gravitational force will be greater than the upward pressure force and the fluid will sink back towards its original position and an oscillatory motion will develop. Such an equilibrium is *statically stable*. Using such simple parcel arguments we will now develop criteria for the stability of the environmental profile.

### 2.10.1 Stability and the Profile of Potential Density

Consider the case of a stationary fluid whose density varies with altitude. We denote this background state with a tilde, as in $\widetilde{\rho}(z)$. We then displace a fluid parcel adiabatically a small distance from $z$ to $\delta z$, as in Fig. 2.8. In such a displacement it is the *potential density* $\rho_\theta$ (not the actual density) that is materially conserved, because potential density takes into account the effects of pressure compressibility. Let us also use the pressure at level $z + \delta z$ as the reference level, where potential density equals in situ density.

The parcel at $z$ takes on the potential density of its environment so that $\rho_\theta(z) = \widetilde{\rho}_\theta(z)$ and it preserves this as it rises, so that $\rho_\theta(z+\delta z) = \rho_\theta(z)$. But since $z+\delta z$ is the reference level, the *in situ* density of the displaced parcel, $\rho(z+\delta z)$, is equal to its potential density $\rho_\theta(z+\delta z)$, which is equal to $\widetilde{\rho}_\theta(z)$. Thus, at $z+\delta z$ the environment has in situ density equal to $\widetilde{\rho}_\theta(z+\delta z)$ and the parcel has in situ density equal to $\widetilde{\rho}_\theta(z)$. Putting all this together in a single equation, the difference between the parcel density and the environmental density, $\delta\rho$, is given by

$$\begin{aligned} \delta\rho &= \rho(z+\delta z) - \widetilde{\rho}(z+\delta z) = \rho_\theta(z+\delta z) - \widetilde{\rho}_\theta(z+\delta z) \\ &= \rho_\theta(z) - \widetilde{\rho}_\theta(z+\delta z) = \widetilde{\rho}_\theta(z) - \widetilde{\rho}_\theta(z+\delta z). \end{aligned} \tag{2.239}$$

Thus, for small $\delta z$,

$$\delta\rho = -\frac{\partial \widetilde{\rho}_\theta}{\partial z}\delta z, \tag{2.240}$$

where the derivative on the right-hand side is the environmental gradient of potential density. If the right-hand side is positive, the parcel is heavier than its surroundings and the displacement is

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

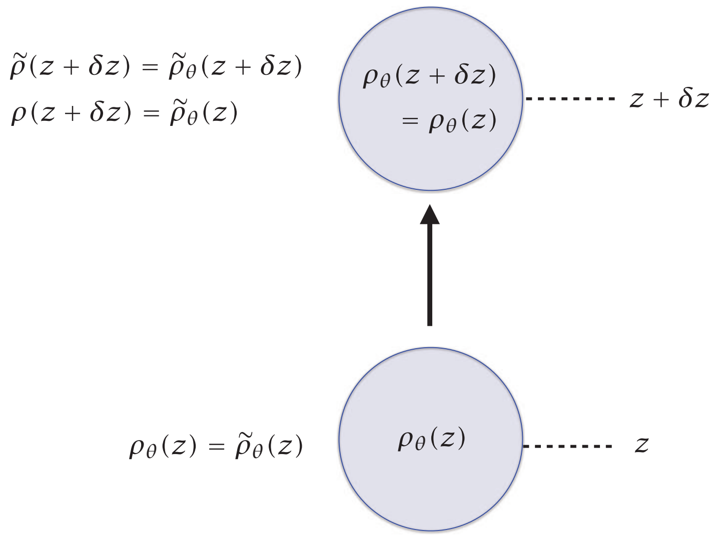


**Fig. 2.8** A parcel is adiabatically displaced upward from level $z$ to $z + \delta z$. A tilde denotes the value in the environment, and variables without tildes are those in the parcel.

The parcel preserves its potential density, $\rho_\theta$, which it takes from the environment at level $z$. If $z+\delta z$ is the reference level, the potential density there is equal to the actual density. The parcel's stability is determined by the difference between its density and the environmental density, as in (2.239). If the difference is positive the displacement is stable, and if negative the displacement is unstable.

stable. That is, *the stability of a parcel of fluid is determined by the gradient of the locally-referenced potential density.*

The conditions for stability are

$$\text{Stability}: \quad \frac{\partial \tilde{\rho}_\theta}{\partial z} < 0, \qquad \text{Instability}: \quad \frac{\partial \tilde{\rho}_\theta}{\partial z} > 0. \tag{2.241a,b}$$

The equation of motion of the fluid parcel is then given by a direct application of Newton's second law, that the mass times the acceleration is given by the force acting on the parcel. The force is the above-derived buoyancy force so we have

$$\frac{\partial^2 \delta z}{\partial t^2} = \frac{g}{\rho}\left(\frac{\partial \tilde{\rho}_\theta}{\partial z}\right)\delta z = -N^2 \delta z, \tag{2.242}$$

where, noting that $\rho(z) = \tilde{\rho}_\theta(z)$ to within $O(\delta z)$,

$$N^2 = -\frac{g}{\tilde{\rho}_\theta}\left(\frac{\partial \tilde{\rho}_\theta}{\partial z}\right). \tag{2.243}$$

A parcel that is displaced in a stably stratified fluid will thus oscillate at the *buoyancy frequency* $N$, proportional to the vertical gradient of potential density. (The buoyancy frequency is also known as the Brunt–Väisälä frequency, after its discovers.) The above expression for the buoyancy frequency is a general one, true in both liquids and gases in a constant gravitational field. The quantity $\tilde{\rho}_\theta$ is the *locally-referenced* potential density of the environment. The reference level turns out not to be important for the atmosphere, but it is for the ocean: parcels at the same level with the same in situ density may have different potential densities if their salinity differs. In contrast, for fresh water in a laboratory setting potential density is virtually equal to in situ density.

### 2.10.2 A Dry Ideal-gas Atmosphere

*Buoyancy frequency*

In the atmosphere potential density is related to potential temperature by $\rho_\theta = p_R/(\theta R)$, where $p_R$ is the reference level for potential temperature. Using this expression in (2.243) gives

$$\boxed{N^2 = \frac{g}{\widetilde{\theta}}\left(\frac{\partial \widetilde{\theta}}{\partial z}\right),} \tag{2.244}$$

where $\widetilde{\theta}$ is the environmental potential temperature. The reference value $p_R$ does not appear, and we are free to choose this value arbitrarily — the surface pressure is a common choice. The conditions for stability, (2.241), then correspond to $N^2 > 0$ for stability and $N^2 < 0$ for instability. On average the atmosphere is stable and in the troposphere (the lowest several kilometres of the atmosphere) the average $N$ is about $0.01\,\mathrm{s}^{-1}$, with a corresponding period, $(2\pi/N)$, of about 10 minutes. In the stratosphere (which lies above the troposphere) $N$ is a few times higher than this.

*Dry adiabatic lapse rate*

The negative of the rate of change of the (real) temperature in the vertical is known as the *temperature lapse rate*, or often just the lapse rate, and denoted $\Gamma$. The lapse rate corresponding to $\partial\theta/\partial z = 0$ is called the *dry adiabatic lapse rate* and denoted $\Gamma_d$. Using $\theta = T(p_0/p)^{R/c_p}$ and $\partial p/\partial z = -\rho g$ we find that the lapse rate and the potential temperature lapse rate are related by

$$\frac{T}{\theta}\frac{\partial \theta}{\partial z} = \frac{\partial T}{\partial z} + \frac{g}{c_p}, \tag{2.245}$$

so that the dry adiabatic lapse rate is given by

$$\Gamma_d = \frac{g}{c_p}, \tag{2.246}$$

as we derived in (1.131). The conditions for static stability corresponding to (2.241) are thus:

$$\boxed{\begin{aligned} &\text{Stability}: && \frac{\partial \widetilde{\theta}}{\partial z} > 0, \quad \text{or} \quad -\frac{\partial \widetilde{T}}{\partial z} < \Gamma_d. \\ &\text{Instability}: && \frac{\partial \widetilde{\theta}}{\partial z} < 0, \quad \text{or} \quad -\frac{\partial \widetilde{T}}{\partial z} > \Gamma_d. \end{aligned}} \tag{2.247a,b}$$

The observed lapse rate (look ahead to Fig. 15.25) is often less than $7\,\mathrm{K\,km}^{-1}$ (corresponding to a buoyancy frequency of about $10^{-2}\,\mathrm{s}^{-1}$) whereas a dry adiabatic lapse rate is about $10\,\mathrm{K\,km}^{-1}$. Why the discrepancy? Why is the atmosphere so apparently stable? One reason is that in mid-latitudes heat is transferred upwards by in large-scale weather systems that keep the atmosphere stable even in the absence of convection. A second reason is that the atmosphere contains water vapour and a column of air that contains water may be unstable even if its lapse rate is less than dry adiabatic. If a moist parcel rises then, as it enters a cooler environment, water vapour may condense releasing more heat and leading to more ascent; that is, a moist atmosphere may be unstable when a dry atmosphere is stable. We defer more discussion to Chapters 15 and 18.

### 2.10.3 A Liquid Ocean

No simple, accurate, analytic expression is available for computing static stability in the ocean. If the ocean had no salt, then the potential density referenced to the surface would generally be a

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

measure of the sign of stability of a fluid column, if not of the buoyancy frequency. However, in the presence of salinity, the surface-referenced potential density is not necessarily even a measure of the sign of stability, because the coefficients of compressibility $\beta_T$ and $\beta_S$ vary in different ways with pressure. To see this, suppose two neighbouring fluid elements at the surface have the same potential density, but different salinities and temperatures, and displace them both adiabatically to the deep ocean. Although their potential densities referenced to the surface are still equal, we can say little about their actual densities, and hence their stability relative to each other, without doing a detailed calculation because they will each have been compressed by different amounts. It is the profile of the *locally-referenced* potential density that determines the stability.

A useful expression for stability arises by noting that in an adiabatic displacement

$$\delta\rho_\theta = \delta\rho - \frac{1}{c_s^2}\delta p = 0. \tag{2.248}$$

If the fluid is hydrostatic $\delta p = -\rho g \delta z$, so that if a parcel is displaced adiabatically its density changes according to

$$\left(\frac{\partial \rho}{\partial z}\right)_{\rho_\theta} = -\frac{\rho g}{c_s^2}. \tag{2.249}$$

If a parcel is displaced a distance $\delta z$ upwards then the density difference between it and its new surroundings is

$$\delta\rho = -\left[\left(\frac{\partial \rho}{\partial z}\right)_{\rho_\theta} - \left(\frac{\partial \widetilde{\rho}}{\partial z}\right)\right]\delta z = \left[\frac{\rho g}{c_s^2} + \left(\frac{\partial \widetilde{\rho}}{\partial z}\right)\right]\delta z, \tag{2.250}$$

where the tilde again denotes the environmental field. It follows that the stratification is given by

$$\boxed{N^2 = -g\left[\frac{g}{c_s^2} + \frac{1}{\widetilde{\rho}}\left(\frac{\partial \widetilde{\rho}}{\partial z}\right)\right].} \tag{2.251}$$

This expression holds for both liquids and gases, and it is proportional to the vertical gradient of potential density. For ideal gases it is the same as (2.244), as a little algebra will show, using $c_s^2 = \gamma p/\rho$. In seawater the expression may be compared to the gradient of (1.144). The factor of $g/c_s^2$ is small but not negligible; it is a slightly destabilising factor in the sense that a density profile with an in situ density that increases with depth is not necessarily stable. In liquids, a good approximation is to use a reference value $\rho_0$ for the undifferentiated density in the denominator, whence (2.251) becomes equal to the Boussinesq expression (2.107). On average the ocean is statically stable, with typical values of $N$ in the upper ocean being about 0.01 s$^{-1}$, falling to 0.001 s$^{-1}$ in the more homogeneous abyssal ocean. These frequencies correspond to periods of about 10 and 100 minutes, respectively.

### 2.10.4 Gravity Waves and Convection Using the Equations of Motion

The parcel approach to oscillations and stability, while simple and direct, seems divorced from the fluid-dynamical equations of motion. To remedy this, we now use the equations of motion for a stratified Boussinesq fluid to analyze the motion resulting from a small disturbance. Our treatment here is brief and introductory, with a fuller treatment given in Chapter 7.

Consider a Boussinesq fluid, initially at rest, in which the buoyancy varies linearly with height. Linearizing the equations of motion about this basic state gives the linear momentum equations,

$$\frac{\partial u'}{\partial t} = -\frac{\partial \phi'}{\partial x}, \qquad \frac{\partial w'}{\partial t} = -\frac{\partial \phi'}{\partial z} + b', \tag{2.252a,b}$$

https://doi.org/10.1017/9781107588417.003 Published online by Cambridge University Press

the mass continuity and thermodynamic equations,

$$\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0, \qquad \frac{\partial b'}{\partial t} + w' N^2 = 0, \tag{2.253a,b}$$

where $N^2 = \mathrm{d}\tilde{b}/\mathrm{d}z$ is the basic state buoyancy profile, and we assume that the flow is a function only of $x$ and $z$. A little algebra reduces the above equations to a single one for $w'$,

$$\left[\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)\frac{\partial^2}{\partial t^2} + N^2 \frac{\partial^2}{\partial x^2}\right] w' = 0. \tag{2.254}$$

Seeking solutions of the form $w' = \mathrm{Re}\, W \exp[\mathrm{i}(kx + mz - \omega t)]$ yields the dispersion relationship for gravity waves:

$$\boxed{\omega^2 = \frac{k^2 N^2}{k^2 + m^2}.} \tag{2.255}$$

The frequency (look ahead to Fig. 7.2) is always less than $N$, approaching $N$ for small horizontal scales, $k \gg m$.

Consider two special cases. First, if we neglect pressure perturbations, as in the parcel argument, then the two equations,

$$\frac{\partial w'}{\partial t} = b', \qquad \frac{\partial b'}{\partial t} + w' N^2 = 0, \tag{2.256}$$

form a closed set and give $\omega^2 = N^2$, as in the parcel argument. Second, if we make the hydrostatic approximation and omit $\partial w'/\partial t$ in (2.252b) then the dispersion relation becomes $\omega^2 = k^2 N^2/m^2$. The frequency then grows, artifactually, without bound as the horizontal scale becomes smaller.

If the basic state density increases with height then $N^2 < 0$ and we expect this state to be unstable. Indeed, the disturbance grows exponentially according to $\exp(\sigma t)$ where $\sigma = \mathrm{i}\omega = \pm k\widetilde{N}/(k^2 + m^2)^{1/2}$, and where $\widetilde{N}^2 = -N^2$. We have reproduced the result previously obtained by parcel theory, namely that if the basic state density (or more generally potential density) increases with height the flow is unstable. Most convective activity in the ocean and atmosphere is, in the end, related to an instability of this form.

## APPENDIX A: ASYMPTOTIC DERIVATION OF THE BOUSSINESQ EQUATIONS

The Boussinesq equations are those equations that are appropriate when the density variations are very small but gravitational effects are large, and here we provide an asymptotic derivation. Two key results are that the velocity field is divergence-free, and the buoyancy should be taken as a function of $z$ and not $p$ in the equation of state. The first result follows from fact that density variations are presumptively small, and the second follows because the lowest order balance in the vertical momentum equation is $\partial p_0/\partial z = -\rho_0 g$, whence $p_0 = -\rho_0 g z$, and $p_0$ and not $p$ should be used in the equation of state at lowest order. The following derivation, which assumes familiarity with elementary asymptotics (or which can be taken as a gentle introduction to asymptotics) mainly just formalizes these results.

Let us suppose that the density varies like $\rho(x, y, z, t) = \rho_0 + \delta\rho(x, y, z, t)$, where $\rho_0$ is a constant and $|\delta\rho| \ll \rho_0$. Specifically, let $\epsilon\rho_0$ be a typical magnitude for $\delta\rho$ where $\epsilon \ll 1$ so that

$$\delta\rho = (\epsilon\rho_0)\delta\widehat{\rho} \qquad \text{and} \qquad \rho = \rho_0(1 + \epsilon\delta\widehat{\rho}) \tag{2.257}$$

where a hat denotes a nondimensional quantity and $\delta\widehat{\rho}$ is an $\mathcal{O}(1)$ quantity.

The dimensional vertical momentum equation, omitting rotation and viscosity for simplicity, is

$$(\rho_0 + \delta\rho)\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial p}{\partial z} - (\rho_0 + \delta\rho)g. \tag{2.258}$$

Now, $g$ is 'big' and variations in density are 'small' and so $g\delta\rho$ is taken to be the same approximate size as the advection term $\rho_0 \mathrm{D}w/\mathrm{D}t$ on the left-hand side. The term $\rho_0 g$ must then be balanced by the pressure gradient. Also, there is no necessary difference between vertical and horizontal scales and velocities. With these points in mind we nondimensionalize with the following scales:

$$(u, v, w) = U(\widehat{u}, \widehat{v}, \widehat{w}), \quad (x, y, z) = L(\widehat{x}, \widehat{y}, \widehat{z}), \quad t = \frac{L}{U}\widehat{t}, \quad p = \rho_0\frac{U^2}{\epsilon}\widehat{p}, \quad g = \frac{U^2}{\epsilon L}\widehat{g}, \tag{2.259}$$

where the hatted quantities are nondimensional and are presumptively $\mathcal{O}(1)$. Equation (2.258) becomes

$$(1 + \epsilon\delta\widehat{\rho})\frac{\mathrm{D}\widehat{w}}{\mathrm{D}\widehat{t}} = -\frac{1}{\epsilon}\frac{\partial \widehat{p}}{\partial \widehat{z}} - \frac{1}{\epsilon}(1 + \epsilon\delta\widehat{\rho})\widehat{g}. \tag{2.260}$$

We now take $\epsilon$ as an asymptotic ordering parameter and expand the nondimensional fields as series in $\epsilon$. Then, with subscripts denoting the asymptotic order (for nondimensional quantities only), we have

$$\delta\widehat{\rho} = \delta\widehat{\rho}_0 + \epsilon\delta\widehat{\rho}_1 + \epsilon^2\delta\widehat{\rho}_2 \ldots, \qquad \widehat{p} = \widehat{p}_0 + \epsilon\widehat{p}_1 + \epsilon^2\widehat{p}_2 \ldots, \qquad \widehat{w} = \widehat{w}_0 + \epsilon\widehat{w}_1 + \epsilon^2\widehat{w}_2 \ldots, \tag{2.261}$$

and similarly for $\widehat{u}$ and $\widehat{v}$. If we substitute the above series into (2.260) and equate terms with the same power of $\epsilon$, the first two orders are

$$\frac{\partial \widehat{p}_0}{\partial \widehat{z}} = -\widehat{g}, \qquad \frac{\mathrm{D}\widehat{w}_0}{\mathrm{D}\widehat{t}} = -\frac{\partial \widehat{p}_1}{\partial \widehat{z}} - \delta\widehat{\rho}_0\, g. \tag{2.262a,b}$$

Evidently the leading order pressure, $\widehat{p}_0$, is hydrostatic. We may now revert to dimensional variables and (2.262a) gives $p_0(z) = -\rho_0 g z$, and (2.262b) becomes

$$\frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{1}{\rho_0}\frac{\partial p'}{\partial z} - \frac{\delta\rho}{\rho_0}g \qquad \text{or} \qquad \frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial \phi}{\partial z} + b, \tag{2.263}$$

where $p'$ is the deviation from the hydrostatic pressure, $\phi = p'/\rho_0$ and $b = -g\delta\rho/\rho_0$.

In the horizontal momentum equations only the perturbation pressure $p'$ appears since $p_0$ is a function of $z$ only. Furthermore, the density must be taken to be the constant $\rho_0$ (since this is $1/\epsilon$ larger than $\delta\rho$). Then, if $\boldsymbol{u}$ is the horizontal velocity, $(u, v)$, and $\boldsymbol{v}$ is the three-dimensional one, $(u, v, w)$, we obtain

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{v}\cdot\nabla\boldsymbol{u} = -\frac{1}{\rho_0}\nabla_z p' \qquad \text{or} \qquad \frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\nabla_z\phi, \tag{2.264}$$

where the gradients (i.e., $\nabla_z$) on the right-hand side are horizontal, at constant $z$. Equations (2.263) and (2.264) combine to give

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla\phi + b\mathbf{k}, \tag{2.265}$$

where $\mathbf{k}$ is the vertical unit vector and frictional and rotational terms may be added as needed.

The mass continuity equation is

$$\frac{\mathrm{D}}{\mathrm{D}t}(\rho_0 + \delta\rho) + (\rho_0 + \delta\rho)\nabla\cdot\boldsymbol{v} = 0. \tag{2.266}$$

Since $\mathrm{D}\rho_0/\mathrm{D}t = 0$ and $\rho_0$ is $1/\epsilon$ larger than $\delta\rho$, we see without further ado that the lowest order mass conservation equation is that of an incompressible, divergence-free fluid, namely $\nabla\cdot\boldsymbol{v} = 0$.

### *Boussinesq thermodynamics*

The Boussinesq equations are completed with a thermal equation of state, an equation of evolution for the composition (e.g., $S$, the salinity), and an equation that provides a thermodynamic state variable $\Theta$ (e.g., entropy or internal energy). The thermal equation of state gives the density in terms of pressure, composition and the thermodynamic variable and so is of the form $b = b(p, \Theta, S)$, meaning $b$ is a function of $(p, \Theta, S)$. However, to be consistent with the derivation of (2.263) we should use the lowest order variables on the right-hand side of (2.267), and specifically *we should use $p_0$ and not $p$ itself.* The equation of state becomes

$$b = b(p_0, \Theta, S) \qquad \text{or} \qquad b = b(z, \Theta, S), \tag{2.267}$$

and an example is (1.155); that equation gives the inverse density, $\alpha$, in terms of $z$, potential temperature and salinity, and from which an expression for $b$ immediately follows.

Evolution of the thermodynamic variable is obtained from the first law, and the discussions of Sections 1.6 and 1.7.3 generally apply. The internal energy equation is, with no external source or diffusion,

$$\frac{\mathrm{D}I}{\mathrm{D}t} + p\alpha \nabla \cdot \boldsymbol{v} = 0. \tag{2.268}$$

The lowest order velocities are divergence-free and at lowest order we thus have $\mathrm{D}I/\mathrm{D}t = 0$. (To determine the actual magnitudes of the terms, we can obtain the internal energy from the Gibbs function using $I = g + \eta T - p\alpha = g - T\partial g/\partial T - p\partial g/\partial p$, and with the seawater Gibbs function, (1.146), we obtain $I = c_{p0}T$ + smaller terms. Referring to the values in table 1.2, $c_{p0}\Delta T$ is roughly comparable to $p\alpha$, so that $p\alpha\nabla \cdot \boldsymbol{v}$ is indeed much smaller than $\mathrm{D}I/\mathrm{D}t$.) If we evolved $I$, we would then need an equation of state of the form $b = b(z, I, S)$ to complete the system.

However, as for the full (non-Boussinesq) system it is often advantageous to evolve potential enthalpy, or potential temperature or entropy, rather than internal energy. We then obtain buoyancy using an accurate equation of state but with the hydrostatic pressure instead of the full pressure. For example, and for idealized or laboratory work with fresh water, a thermodynamic equation and an approximate equation of state might use (1.128) and a simplified form of (1.155) giving

$$\frac{\mathrm{D}\theta}{\mathrm{D}t} = 0 \qquad \text{and} \qquad b = g\frac{\delta\alpha}{\alpha_0} = g\left[\frac{gz}{c_s^2} + \beta_T(\theta - \theta_0)\right], \tag{2.269}$$

where $\theta$ is potential temperature, $\theta_0$ is a constant reference value and $c_s$ is the speed of sound. The potential temperature is related to the actual temperature by $\theta = T + \beta_T g\theta_0 z/c_p$, but $T$ is not needed to evolve the system. The term in $gz/c_s^2$ is often small enough to be neglected, in which case the thermodynamic equation becomes an evolution of buoyancy itself, $\mathrm{D}b/\mathrm{D}t = 0$.

## *Notes*

1 The geocentric view was slowly and contentiously being replaced by the Copernican or heliocentric view during Shakespeare's lifetime, with the upheaval of matters thought settled and stable. Galileo, whose telescopes helped confirm the heliocentric view, was born in the same year as Shakespeare, 1564. In the geocentric view Earth's surface is an inertial frame and there is no Coriolis force.

2 The distinction between Coriolis force and acceleration is not always made in the literature, even after noting that the force is considered as a force per unit mass. For a fluid in geostrophic balance, one might either say that there is a balance between the pressure force and the Coriolis force, with no net acceleration, or that the pressure force produces a Coriolis acceleration. The descriptions are equivalent, because of Newton's second law, but should not be conflated.

The Coriolis effect is named after Gaspard Gustave de Coriolis (1792–1843), who discussed the eponymous force in the context of rotating mechanical systems (Coriolis 1832, 1835), but Euler was aware of the effect almost a century before. Persson (1998) provides a historical account.

3 Phillips (1973). A related discussion can be found in Stommel & Moore (1989).

4 Phillips (1966) and White (2002, 2003) form a pleasing set of review articles that synthesize the various forms and approximations of the equations of motion. In the early days of numerical modelling the primitive equations were indeed the most primitive — i.e., the least filtered — equations that could practically be integrated numerically. Associated with increasing computer power there is a tendency for comprehensive numerical models to use non-hydrostatic equations of motion that do not make the shallow-fluid or traditional approximations, and it is conceivable that the meaning of the word 'primitive' may evolve to accommodate them.

5 It is nevertheless possible to derive dynamically consistent equations for a shallow atmosphere that do not make the traditional approximation (Tort & Dubos 2014, Dellar 2011). See White *et al.* (2005) for a related discussion.

6 The Boussinesq approximation is named for Boussinesq (1903), although similar approximations were used earlier by Oberbeck (1879, 1888). Spiegel & Veronis (1960) give a physically based derivation for an ideal gas, and Mihaljan (1962) and Gray & Giorgini (1976) provide more systematic derivations that include the effects of viscosity and diffusion and discussions of the energetics.

7 Young (2010).

8 Various versions of anelastic and pseudo-incompressible equations exist — see Batchelor (1953a), Ogura & Phillips (1962), Gough (1969), Gilman & Glatzmaier (1981), Lipps & Hemler (1982), and Durran (1989), although not all have potential vorticity and energy conservation laws (Bannon 1995, 1996; Scinocca & Shepherd 1992). The system we derive is most similar to that of Ogura & Phillips (1962) and unpublished notes by J. S. A. Green. The connection between the Boussinesq and anelastic equations is discussed by Lilly (1996) and Ingersoll (2005), and the extension to a complex equation of state and inclusion of moisture is discussed by Pauluis (2008).

9 A numerical model that explicitly includes sound waves must take very small timesteps in order to maintain numerical stability, in particular to satisfy the Courant–Friedrichs–Lewy (CFL) criterion. An alternative is to use implicit timestepping that effectively lets the numerics filter the sound waves. If we make the hydrostatic approximation then all sound waves except those that propagate horizontally are eliminated, and there is little numerical need to make the anelastic approximation.

10 Gill (1982) provides a longer discussion. Not all authors differentiate between $N$ and $N_*$.

11 The Rossby number is named for C.-G. Rossby (see endnote 2 on page 211), but it was also used by Kibel (1940) and is sometimes called the Kibel or Rossby–Kibel number. The notion of geostrophic balance and so, implicitly, that of a small Rossby number, predates both Rossby and Kibel.

12 After Taylor (1921b) and Proudman (1916). The Taylor–Proudman effect is sometimes called the Taylor–Proudman 'theorem', but it is more usefully thought of as a physical effect, with manifestations even when the conditions for its satisfaction are not precisely met. In fact, Hough (1897) seems to have been aware of the effect well before Taylor and Proudman.

13 This discussion owes much to that in Holton (1992). Inertial motion is discussed by Durran (1993).

Jim Holton (1938–2004) made many contributions to atmospheric dynamics over the course of a distinguished career spent almost entirely at the University of Washington in Seattle. In his early career he elucidated, with Richard Lindzen of MIT, the essential mechanism of the quasi-biennial oscillation, or QBO, and he continued to make important contributions to wave–mean-flow interaction, stratosphere-troposphere interaction and stratospheric dynamics more generally throughout his career. He is also known, both to scientists and students, for his popular textbook *An Introduction to Dynamical Meteorology*.

*Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration … he should seek an economical description of natural phenomena.*

George E. Box, *Science and Statistics*, 1976.

*The sciences do not try to explain … they mainly make models … a mathematical construct the justification [of which] is that it is expected to work.*

John von Neumann, *Methods in the Physical Sciences*, 1955.

# CHAPTER 3

# Shallow Water Systems

Conventionally, 'the' shallow water equations describe a thin layer of constant density fluid in hydrostatic balance, rotating or not, bounded from below by a rigid surface and from above by a free surface, above which we suppose is another fluid of negligible inertia. Such a configuration can be generalized to multiple layers of immiscible fluids of different densities lying one on top of another, forming a stably-stratified 'stacked shallow water' system, which in many ways behaves like a continuously stratified fluid. These types of systems are the main subject of this chapter. We also introduce the notion of available potential energy, which involves thinking about a continuously stratified system as if it were a stacked shallow water system.

The single-layer model is one of the simplest useful models in geophysical fluid dynamics because it allows for a consideration of the effects of rotation in a simple framework without the complicating effects of stratification. A model with just two layers is not only a simple model of a stratified fluid, it is a surprisingly good model of many phenomena in the ocean and atmosphere. Such models are more than just pedagogical tools — we will find that there is a close physical and mathematical analogy between the shallow water equations and a description of the continuously stratified ocean or atmosphere written in isopycnal or isentropic coordinates, with a meaning beyond a coincidental similarity in the equations. Let us begin with the single-layer case.

## 3.1 DYNAMICS OF A SINGLE SHALLOW LAYER OF FLUID

Shallow water dynamics apply, by definition, to a fluid layer of constant density in which the horizontal scale of the flow is much greater than the layer depth. The fluid motion is fully determined by the momentum and mass continuity equations, and because of the assumed small aspect ratio the hydrostatic approximation is well satisfied, and we invoke this from the outset. Consider, then, fluid in a container above which is another fluid of negligible density (and therefore negligible inertia) relative to the fluid of interest, as illustrated in Fig. 3.1. Our notation is that $\boldsymbol{v} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ is the three-dimensional velocity and $\boldsymbol{u} = u\mathbf{i} + v\mathbf{j}$ is the horizontal velocity. $h(x, y)$ is the thickness of the liquid column, $H$ is its mean height, and $\eta$ is the height of the free surface. In a flat-bottomed container $\eta = h$, whereas in general $h = \eta - \eta_b$, where $\eta_b$ is the height of the floor of the container.



https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

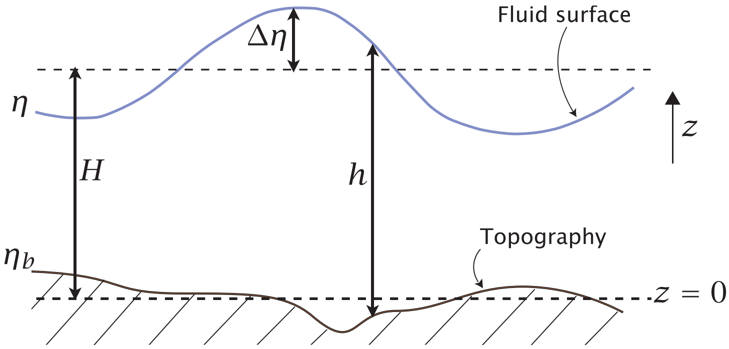


**Fig. 3.1** A shallow water system. $h$ is the thickness of a water column, $H$ its mean thickness, $\eta$ the height of the free surface and $\eta_b$ is the height of the lower, rigid, surface above some arbitrary origin, typically chosen such that the average of $\eta_b$ is zero. $\Delta\eta$ is the deviation free surface height, so we have $\eta = \eta_b + h = H + \Delta\eta$.

### 3.1.1 Momentum Equations

The vertical momentum equation is just the hydrostatic equation,

$$\frac{\partial p}{\partial z} = -\rho_0 g, \tag{3.1}$$

and, because density is assumed constant, we may integrate this to

$$p(x, y, z, t) = -\rho_0 g z + p_o. \tag{3.2}$$

At the top of the fluid, $z = \eta$, the pressure is determined by the weight of the overlying fluid and this is assumed to be negligible. Thus, $p = 0$ at $z = \eta$, giving

$$p(x, y, z, t) = \rho_0 g(\eta(x, y, t) - z). \tag{3.3}$$

The consequence of this is that the horizontal gradient of pressure is independent of height. That is

$$\nabla_z p = \rho_0 g \nabla_z \eta, \tag{3.4}$$

where

$$\nabla_z = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} \tag{3.5}$$

is the gradient operator at constant $z$. (In the rest of this chapter we will drop the subscript $z$ unless that causes ambiguity. The three-dimensional gradient operator will be denoted by $\nabla_3$. We will also mostly use Cartesian coordinates, but the shallow water equations may certainly be applied over a spherical planet — 'Laplace's tidal equations' are essentially the shallow water equations on a sphere.) The horizontal momentum equations therefore become

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho_0}\nabla p = -g\nabla\eta. \tag{3.6}$$

The right-hand side of this equation is independent of the vertical coordinate $z$. Thus, if the flow is initially independent of $z$, it must stay so. (This $z$-independence is unrelated to that arising from the rapid rotation necessary for the Taylor–Proudman effect.) The velocities $u$ and $v$ are functions of $x$, $y$ and $t$ only, and the horizontal momentum equation is therefore

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \frac{\partial \boldsymbol{u}}{\partial t} + u\frac{\partial \boldsymbol{u}}{\partial x} + v\frac{\partial \boldsymbol{u}}{\partial y} = -g\nabla\eta. \tag{3.7}$$

That the horizontal velocity is independent of $z$ is a consequence of the hydrostatic equation, which ensures that the horizontal pressure gradient is independent of height. (Another starting point

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

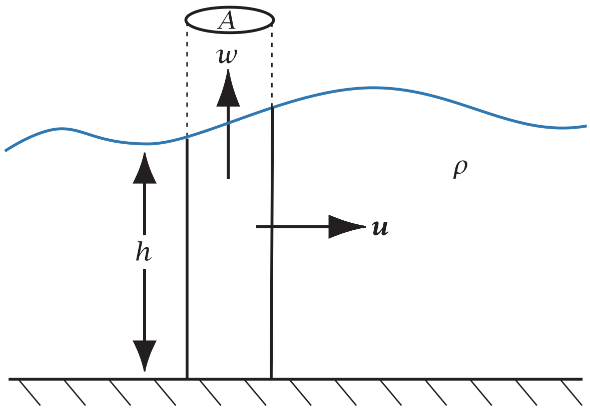


**Fig. 3.2** The mass budget for a column of area $A$ in a shallow water system. The fluid leaving the column is $\oint \rho_0 h\boldsymbol{u} \cdot \boldsymbol{n}\, \mathrm{d}l$ where $\boldsymbol{n}$ is the unit vector normal to the boundary of the fluid column. There is a non-zero vertical velocity at the top of the column if the mass convergence into the column is non-zero.

would be to take this independence of the horizontal motion with height as the *definition* of shallow water flow. In real physical situations such independence does not hold exactly — for example, friction at the bottom may induce a vertical dependence of the flow in a boundary layer.) In the presence of rotation, (3.7) easily generalizes to

$$\boxed{\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta,} \tag{3.8}$$

where $\boldsymbol{f} = f\mathbf{k}$. Just as with the primitive equations, $f$ may be constant or may vary with latitude, so that on a spherical planet $f = 2\Omega \sin\vartheta$ and on the $\beta$-plane $f = f_0 + \beta y$.

### 3.1.2 Mass Continuity Equation

*From first principles*

The mass contained in a fluid column of height $h$ and cross-sectional area $A$ is given by $\int_A \rho_0 h\, \mathrm{d}\boldsymbol{A}$ (see Fig. 3.2). If there is a net flux of fluid across the column boundary (by advection) then this must be balanced by a net increase in the mass in $A$, and therefore a net increase in the height of the water column. The mass convergence into the column is given by

$$F_m = \text{mass flux in} = -\int_S \rho_0 \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{S}, \tag{3.9}$$

where $S$ is the area of the vertical boundary of the column. The surface area of the column is composed of elements of area $h\boldsymbol{n}\,\delta l$, where $\delta l$ is a line element circumscribing the column and $\boldsymbol{n}$ is a unit vector perpendicular to the boundary, pointing outwards. Thus (3.9) becomes

$$F_m = -\oint \rho_0 h\boldsymbol{u} \cdot \boldsymbol{n}\, \mathrm{d}l. \tag{3.10}$$

Using the divergence theorem in two dimensions, (3.10) simplifies to

$$F_m = -\int_A \nabla \cdot (\rho_0 \boldsymbol{u} h)\, \mathrm{d}A, \tag{3.11}$$

where the integral is over the cross-sectional area of the fluid column (looking down from above). This is balanced by the local increase in height of the water column, given by

$$F_m = \frac{\mathrm{d}}{\mathrm{d}t}\int \rho_0\, \mathrm{d}V = \frac{\mathrm{d}}{\mathrm{d}t}\int_A \rho_0 h\, \mathrm{d}A = \int_A \rho_0 \frac{\partial h}{\partial t}\, \mathrm{d}A. \tag{3.12}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

Because $\rho_0$ is constant, the balance between (3.11) and (3.12) leads to

$$\int_A \left[ \frac{\partial h}{\partial t} + \nabla \cdot (\boldsymbol{u} h) \right] \mathrm{d}A = 0, \tag{3.13}$$

and because the area is arbitrary the integrand itself must vanish, whence,

$$\frac{\partial h}{\partial t} + \nabla \cdot (\boldsymbol{u} h) = 0 \qquad \text{or} \qquad \frac{\mathrm{D}h}{\mathrm{D}t} + h \nabla \cdot \boldsymbol{u} = 0. \tag{3.14a,b}$$

This derivation holds whether or not the lower surface is flat. If it is, then $h = \eta$, and if not $h = \eta - \eta_b$. Equations (3.8) and (3.14) form a complete set, summarized in the shaded box on the facing page.

*From the 3D mass conservation equation*

Since the fluid is incompressible, the three-dimensional mass continuity equation is just $\nabla \cdot \boldsymbol{v} = 0$. Writing this out in component form

$$\frac{\partial w}{\partial z} = -\left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = -\nabla \cdot \boldsymbol{u}. \tag{3.15}$$

Integrate this from the bottom of the fluid ($z = \eta_b$) to the top ($z = \eta$), noting that the right-hand side is independent of $z$, to give

$$w(\eta) - w(\eta_b) = -h \nabla \cdot \boldsymbol{u}. \tag{3.16}$$

At the top the vertical velocity is the material derivative of the position of a particular fluid element. But the position of the fluid at the top is just $\eta$, and therefore (see Fig. 3.2)

$$w(\eta) = \frac{\mathrm{D}\eta}{\mathrm{D}t}. \tag{3.17a}$$

At the bottom of the fluid we have similarly

$$w(\eta_b) = \frac{\mathrm{D}\eta_b}{\mathrm{D}t}, \tag{3.17b}$$

where, apart from earthquakes and the like, $\partial \eta_b / \partial t = 0$. Using (3.17a,b), (3.16) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}(\eta - \eta_b) + h \nabla \cdot \boldsymbol{u} = 0 \tag{3.18}$$

or, as in (3.14b),

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h \nabla \cdot \boldsymbol{u} = 0. \tag{3.19}$$

### 3.1.3 A Rigid Lid

The case where the *upper* surface is held flat by the imposition of a rigid lid is sometimes of interest. The ocean suggests one such example, since the bathymetry at the bottom of the ocean provides much larger variations in fluid thickness than do the small variations in the height of the ocean surface. If we suppose that the upper surface is at a constant height $H$, then from (3.14a) with $\partial h / \partial t = 0$ the mass conservation equation is

$$\nabla_h \cdot (\boldsymbol{u} h_b) = 0, \tag{3.20}$$

### The Shallow Water Equations

For a single-layer fluid, and including the Coriolis term, the inviscid shallow water equations are

$$\text{momentum:} \qquad \frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta. \tag{SW.1}$$

$$\text{mass continuity:} \qquad \frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla\cdot\boldsymbol{u} = 0 \qquad \text{or} \qquad \frac{\partial h}{\partial t} + \nabla\cdot(h\boldsymbol{u}) = 0, \tag{SW.2}$$

where $\boldsymbol{u}$ is the horizontal velocity, $h$ is the total fluid thickness, $\eta$ is the height of the upper free surface and $\eta_b$ is the height of the lower surface (the bottom topography). Thus,

$$h(x, y, t) = \eta(x, y, t) - \eta_b(x, y) \tag{SW.3}$$

The material derivative is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}, \tag{SW.4}$$

with the rightmost expression holding in Cartesian coordinates.

where $h_b = H - \eta_b$. Note that (3.20) allows us to define an incompressible *mass-transport velocity*, $\boldsymbol{U} \equiv h_b\boldsymbol{u}$.

Although the upper surface is flat, the pressure there is no longer constant because a force must be provided by the rigid lid to keep the surface flat. The horizontal momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho_0}\nabla p_{lid}, \tag{3.21}$$

where $p_{lid}$ is the pressure at the lid, and the complete equations of motion are then (3.20) and (3.21).[1] If the lower surface is flat, the two-dimensional flow itself is divergence-free, and the equations reduce to the two-dimensional incompressible Euler equations.

### 3.1.4 Stretching and the Vertical Velocity

Because the horizontal velocity is depth independent, the vertical velocity plays no role in advection. However, $w$ is certainly not zero for then the free surface would be unable to move up or down, but because of the vertical independence of the horizontal flow $w$ does have a simple vertical structure; to determine this we write the mass conservation equation as

$$\frac{\partial w}{\partial z} = -\nabla\cdot\boldsymbol{u}, \tag{3.22}$$

and integrate upwards from the bottom to give

$$w = w_b - (\nabla\cdot\boldsymbol{u})(z - \eta_b). \tag{3.23}$$

Thus, the vertical velocity is a linear function of height. Equation (3.23) can be written as

$$\frac{\mathrm{D}z}{\mathrm{D}t} = \frac{\mathrm{D}\eta_b}{\mathrm{D}t} - (\nabla\cdot\boldsymbol{u})(z - \eta_b), \tag{3.24}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

and at the upper surface $w = \mathrm{D}\eta/\mathrm{D}t$ so that here we have

$$\frac{\mathrm{D}\eta}{\mathrm{D}t} = \frac{\mathrm{D}\eta_b}{\mathrm{D}t} - (\nabla \cdot \boldsymbol{u})(\eta - \eta_b). \tag{3.25}$$

Eliminating the divergence term from the last two equations gives

$$\frac{\mathrm{D}}{\mathrm{D}t}(z - \eta_b) = \frac{z - \eta_b}{\eta - \eta_b}\frac{\mathrm{D}}{\mathrm{D}t}(\eta - \eta_b), \tag{3.26}$$

which in turn gives

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{z - \eta_b}{\eta - \eta_b}\right) = \frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{z - \eta_b}{h}\right) = 0. \tag{3.27}$$

This means that the ratio of the height of a fluid parcel above the floor to the total depth of the column is fixed; that is, the fluid stretches uniformly in a column, and this is a kinematic property of the shallow water system.

### 3.1.5 Analogy with Compressible Flow

The shallow water equations (3.8) and (3.14) are analogous to the compressible gas dynamic equations in two dimensions, namely

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho}\nabla p \tag{3.28}$$

and

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\boldsymbol{u}\rho) = 0, \tag{3.29}$$

along with an equation of state which we take to be $p = f(\rho)$. The mass conservation equations (3.14) and (3.29) are identical, with the replacement $\rho \leftrightarrow h$. If $p = C\rho^\gamma$, then (3.28) becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}\rho}\nabla\rho = -C\gamma\rho^{\gamma-2}\nabla\rho. \tag{3.30}$$

If $\gamma = 2$ then the momentum equations (3.8) and (3.30) become equivalent, with $\rho \leftrightarrow h$ and $C\gamma \leftrightarrow g$. In an ideal gas $\gamma = c_p/c_v$ and values typically are in fact less than 2 (in air $\gamma \approx 7/5$); however, if the equations are linearized, then the analogy is exact for all values of $\gamma$, for then (3.30) becomes $\partial \boldsymbol{v}'/\partial t = -\rho_0^{-1}c_s^2\nabla\rho'$ where $c_s^2 = \mathrm{d}p/\mathrm{d}\rho$, and the linearized shallow water momentum equation is $\partial \boldsymbol{u}'/\partial t = -H^{-1}(gH)\nabla h'$, so that $\rho_0 \leftrightarrow H$ and $c_s^2 \leftrightarrow gH$. The sound waves of a compressible fluid are then analogous to shallow water waves, which are considered in Section 3.8.

## 3.2 REDUCED GRAVITY EQUATIONS

Consider now a single shallow moving layer of fluid on top of a deep, quiescent fluid layer (Fig. 3.3), and beneath a fluid of negligible inertia. This configuration is often used as a model of the upper ocean: the upper layer represents flow in perhaps the upper few hundred metres of the ocean, the lower layer being the near-stagnant abyss. If we turn the model upside-down we have a perhaps slightly less realistic model of the atmosphere: the lower layer represents motion in the troposphere above which lies an inactive stratosphere. The equations of motion are virtually the same in both cases.

### 3.2.1 Pressure Gradient in the Active Layer

We will derive the equations for the oceanic case (active layer on top) in two cases, which differ slightly in the assumption made about the upper surface.

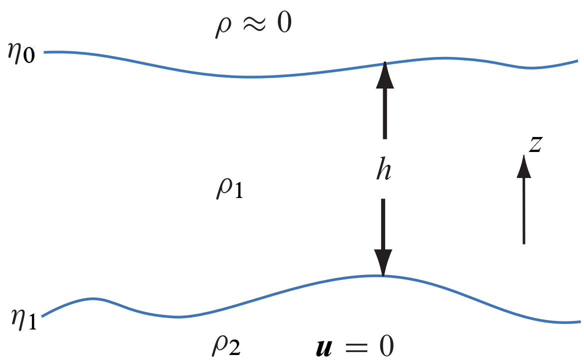


**Fig. 3.3** The reduced gravity shallow water system. An active layer lies over a deep, denser, quiescent layer. In a common variation the upper surface is held flat by a rigid lid, and $\eta_0 = 0$.

### *I Free upper surface*

The pressure in the upper layer is given by integrating the hydrostatic equation down from the upper surface. Thus, at a height $z$ in the upper layer

$$p_1(z) = g\rho_1(\eta_0 - z), \tag{3.31}$$

where $\eta_0$ is the height of the upper surface. Hence, everywhere in the upper layer,

$$\frac{1}{\rho_1}\nabla p_1 = g\nabla\eta_0, \tag{3.32}$$

and the momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta_0. \tag{3.33}$$

In the lower layer the pressure is also given by the weight of the fluid above it. Thus, at some level $z$ in the lower layer,

$$p_2(z) = \rho_1 g(\eta_0 - \eta_1) + \rho_2 g(\eta_1 - z). \tag{3.34}$$

But if this layer is motionless the horizontal pressure gradient in it is zero and therefore

$$\rho_1 g\eta_0 = -\rho_1 g'\eta_1 + \text{constant}, \tag{3.35}$$

where $g' = g(\rho_2 - \rho_1)/\rho_1$ is the *reduced gravity*, and normally $(\rho_2 - \rho_1)/\rho \ll 1$ and $g' \ll g$. The momentum equation becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = g'\nabla\eta_1. \tag{3.36}$$

The equations are completed by the usual mass conservation equation,

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla\cdot\boldsymbol{u} = 0, \tag{3.37}$$

where $h = \eta_0 - \eta_1$. Because $g \gg g'$, (3.35) shows that surface displacements are *much smaller* than the displacements at the interior interface. We see this in the real ocean where the mean interior isopycnal displacements may be several tens of metres but variations in the mean height of ocean surface are of the order of centimetres.

### *II The rigid lid approximation*

The smallness of the upper surface displacement suggests that we will make little error if we impose a *rigid lid* at the top of the fluid. Displacements are no longer allowed, but the lid will in general

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

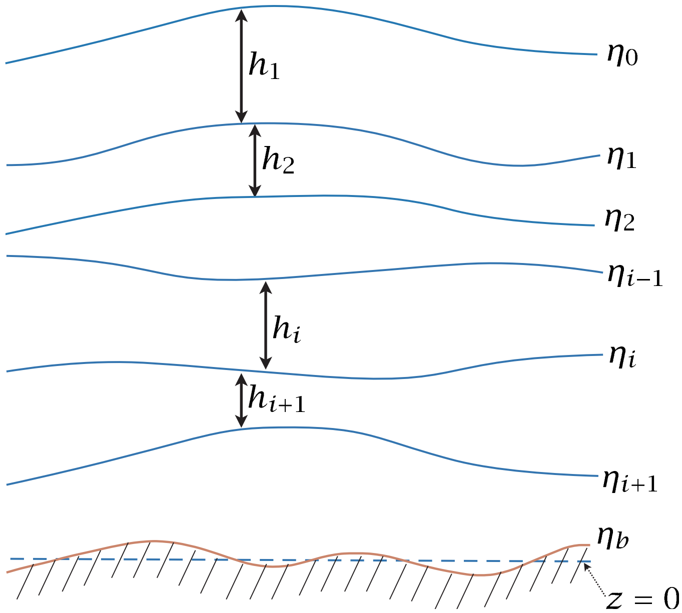


**Fig. 3.4** The multi-layer shallow water system. The layers are numbered from the top down. The coordinates of the interfaces are denoted by $\eta$, and the layer thicknesses by $h$, so that $h_i = \eta_{i-1} - \eta_i$.

impart a pressure force to the fluid. Suppose that this is $P(x, y, t)$, then the horizontal pressure gradient in the upper layer is simply

$$\nabla p_1 = \nabla P. \tag{3.38}$$

The pressure in the lower layer is again given by hydrostasy, and is

$$p_2 = -\rho_1 g\eta_1 + \rho_2 g(\eta_1 - z) + P = \rho_1 gh - \rho_2 g(h + z) + P, \tag{3.39}$$

so that

$$\nabla p_2 = -g(\rho_2 - \rho_1)\nabla h + \nabla P. \tag{3.40}$$

Then if $\nabla p_2 = 0$ (because the lower layer is stationary) we have $g(\rho_2 - \rho_1)\nabla h = \nabla P$, and the momentum equation for the upper layer is just

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g'\nabla h, \tag{3.41}$$

where $g' = g(\rho_2 - \rho_1)/\rho_1$. These equations differ from the usual shallow water equations only in the use of a reduced gravity $g'$ in place of $g$ itself. It is the density *difference* between the two layers that is important. Similarly, if we take a shallow water system, with the moving layer on the bottom, and we suppose that overlying it is a stationary fluid of finite density, then we would easily find that the fluid equations for the moving layer are the same as if the fluid on top had zero inertia, except that $g$ would be replaced by an appropriate reduced gravity.

## 3.3 MULTI-LAYER SHALLOW WATER EQUATIONS

We now consider the dynamics of multiple layers of fluid stacked on top of each other. This is a crude representation of continuous stratification, but it turns out to be a powerful model of many geophysically interesting phenomena as well as being physically realizable in the laboratory. The pressure is continuous across the interface, but the density jumps discontinuously and this allows the horizontal velocity to have a corresponding discontinuity. The set up is illustrated in Fig. 3.4.

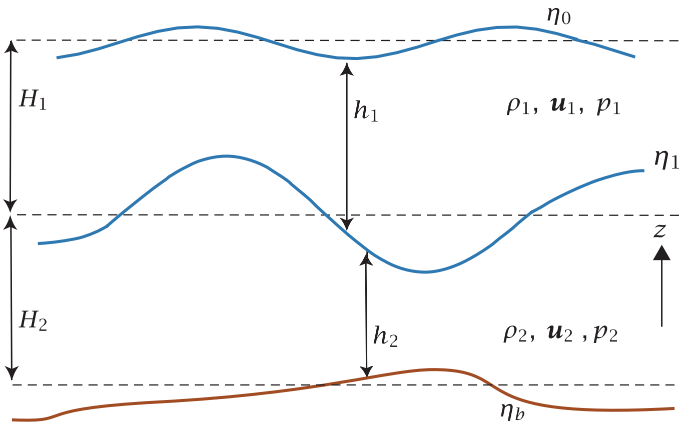


**Fig. 3.5** The two-layer shallow water system. A fluid of density $\rho_1$ lies over a denser fluid of density $\rho_2$. In the reduced gravity case the lower layer is arbitrarily thick and is assumed stationary and so has no horizontal pressure gradient. In the 'rigid-lid' approximation the top surface displacement is neglected, but there is then a non-zero pressure gradient induced by the lid.

In each layer pressure is given by the hydrostatic approximation, and so anywhere in the interior we can find the pressure by integrating down from the top. Thus, at a height $z$ in the first layer we have

$$p_1 = \rho_1 g(\eta_0 - z), \tag{3.42}$$

and in the second layer,

$$p_2 = \rho_1 g(\eta_0 - \eta_1) + \rho_2 g(\eta_1 - z) = \rho_1 g\eta_0 + \rho_1 g_1'\eta_1 - \rho_2 gz, \tag{3.43}$$

where $g_1' = g(\rho_2 - \rho_1)/\rho_1$, and so on. The term involving $z$ is irrelevant for the dynamics, because only the horizontal derivative enters the equation of motion. Omitting this term, for the $n$th layer the dynamical pressure is given by the sum from the top down:

$$p_n = \rho_1 \sum_{i=0}^{n-1} g_i'\eta_i, \tag{3.44}$$

where $g_i' = g(\rho_{i+1} - \rho_i)/\rho_1$ (and $g_0 = g$). The interface displacements may be expressed in terms of the layer thicknesses by summing from the bottom up:

$$\eta_n = \eta_b + \sum_{i=n+1}^{i=N} h_i. \tag{3.45}$$

The momentum equation for each layer may then be written, in general,

$$\frac{\mathrm{D}\boldsymbol{u}_n}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u}_n = -\frac{1}{\rho_n}\nabla p_n, \tag{3.46}$$

where the pressure is given by (3.44) and in terms of the layer depths using (3.46). If we make the Boussinesq approximation then $\rho_n$ on the right-hand side of (3.46) is replaced by $\rho_1$.

Finally, the mass conservation equation for each layer has the same form as the single-layer case, and is

$$\frac{\mathrm{D}h_n}{\mathrm{D}t} + h_n\nabla\cdot\boldsymbol{u}_n = 0. \tag{3.47}$$

*The two- and three-layer cases*

The two-layer model (Fig. 3.5) is the simplest model to capture the effects of stratification. Evaluating the pressures using (3.44) and (3.45) we find:

$$p_1 = \rho_1 g\eta_0 = \rho_1 g(h_1 + h_2 + \eta_b), \tag{3.48a}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

$$p_2 = \rho_1[g\eta_0 + g_1'\eta_1] = \rho_1\left[g(h_1 + h_2 + \eta_b) + g_1'(h_2 + \eta_b)\right]. \tag{3.48b}$$

The momentum equations for the two layers are then

$$\frac{\mathrm{D}\boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_1 = -g\nabla\eta_0 = -g\nabla(h_1 + h_2 + \eta_b), \tag{3.49a}$$

and in the bottom layer

$$\begin{aligned}\frac{\mathrm{D}\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_2 &= -\frac{\rho_1}{\rho_2}\left(g\nabla\eta_0 + g_1'\nabla\eta_1\right)\\ &= -\frac{\rho_1}{\rho_2}\left[g\nabla(\eta_b + h_1 + h_2) + g_1'\nabla(h_2 + \eta_b)\right].\end{aligned} \tag{3.49b}$$

In the Boussinesq approximation $\rho_1/\rho_2$ is replaced by unity.

In a three-layer model the dynamical pressures are found to be

$$p_1 = \rho_1 g h, \tag{3.50a}$$
$$p_2 = \rho_1\left[gh + g_1'(h_2 + h_3 + \eta_b)\right], \tag{3.50b}$$
$$p_3 = \rho_1\left[gh + g_1'(h_2 + h_3 + \eta_b) + g_2'(h_3 + \eta_b)\right], \tag{3.50c}$$

where $h = \eta_0 = \eta_b + h_1 + h_2 + h_3$ and $g_2' = g(\rho_3 - \rho_2)/\rho_1$. More layers can obviously be added in a systematic fashion.

### 3.3.1 Reduced-gravity Multi-layer Equation

As with a single active layer, we may envision multiple layers of fluid overlying a deeper stationary layer. This is a useful model of the stratified upper ocean overlying a nearly stationary and nearly unstratified abyss. Indeed we use such a model to study the 'ventilated thermocline' in Chapter 20 and a detailed treatment may be found there. If we suppose there is a lid at the top, then the model is almost the same as that of the previous section. However, now the horizontal pressure gradient in the lowest model layer is zero, and so we may obtain the pressures in all the active layers by integrating the hydrostatic equation upwards from this layer. Suppose we have $N$ moving layers, then the reader may verify that the dynamic pressure in the $n$th layer is given by

$$p_n = -\sum_{i=n}^{i=N} \rho_1 g_i' \eta_i, \tag{3.51}$$

where as before $g_i' = g(\rho_{i+1} - \rho_i)/\rho_1$. If we have a lid at the top, and take $\eta_0 = 0$, then the interface displacements are related to the layer thicknesses by

$$\eta_n = -\sum_{i=1}^{i=n} h_i. \tag{3.52}$$

From these expressions the momentum equation in each layer is easily constructed.

## 3.4 ✦ FROM CONTINUOUS STRATIFICATION TO SHALLOW WATER

In this section we show that the *continuously stratified* equations have a close correspondence to the shallow water equations, without breaking the fluid into discrete layers of differing densities. In particular, if the continuous equations are linearized and the flow is stably stratified, then each vertical mode of the continuous equations has the same form as the shallow water equations, with the modes being distinguished by the phase speed of the associated gravity waves.[2]

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

### 3.4.1 Vertical Normal Modes of the Linear Equations

We begin with a hydrostatic Boussinesq system, linearized about a state of rest and with fixed stratification, $N(z)$, noting that a similar derivation can be applied to an ideal gas using pressure coordinates. The equations are

$$\frac{\partial u}{\partial t} - fv = -\frac{\partial \phi}{\partial x}, \qquad \frac{\partial v}{\partial t} + fu = -\frac{\partial \phi}{\partial y}, \qquad 0 = -\frac{\partial \phi}{\partial z} + b, \tag{3.53a,b,c}$$

$$\nabla \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0, \qquad \frac{\partial b}{\partial t} + wN^2 = 0. \tag{3.53c,d}$$

The first line above contains the $u$ and $v$ momentum equations and the hydrostatic equation, and the second line contains the mass continuity equation and the buoyancy or thermodynamic equation, with the $\nabla$ operator being purely horizontal (or at constant pressure), and we will take $N^2 > 0$. We assume a lid at the bottom and top of the domain. Including a free surface at the top, as appropriate for an ocean, is a slight extension. Including a 'leaky' tropopause with a stratosphere above is a more major extension.

The difficulty with these equations is that there are five independent variables in three spatial coordinates so that even the linear problems are algebraically complex, especially when $f$ is variable. The equations are more general than is needed, because it is often observed that the vertical structure of solutions is relatively simple, especially in linear problems. A solution is to project the vertical structure onto appropriate eigenfunctions, and then to retain a very small number — often only one — of these eigenfunctions.

To determine what those eigenfunctions should be, we first combine the hydrostatic and buoyancy equations to give

$$\frac{\partial}{\partial t}\left(\frac{\phi_z}{N^2}\right) + w = 0. \tag{3.54}$$

Differentiating with respect to $z$ and using the mass continuity equation gives

$$\frac{\partial}{\partial t}\left(\frac{\phi_z}{N^2}\right)_z - \nabla \cdot \boldsymbol{u} = 0. \tag{3.55}$$

It is this equation that motivates our choice of basis functions: we choose to expand the pressure and horizontal components of velocity in terms of an eigenfunction that satisfies the following Sturm–Liouville problem:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{N^2}\frac{\mathrm{d}C_m}{\mathrm{d}z}\right) + \frac{1}{c_m^2}C_m = 0, \qquad \frac{\mathrm{d}}{\mathrm{d}z}C_m(0) = \frac{\mathrm{d}}{\mathrm{d}z}C_m(-H) = 0. \tag{3.56}$$

The eigenfunctions $C_m$ are orthogonal in the sense that

$$\int_{-H}^{0} C_m C_n \,\mathrm{d}z = \frac{c_m^2}{g}\delta_{mn}\,, \tag{3.57}$$

where $\delta_{mn} = 0$ unless $m = n$, in which case it equals one. The normalization is by convention and the factor of $g$ makes the functions $C_m$ nondimensional. There are an infinite number of eigenvalues, $c_m$, namely $c_0, c_1, c_2 \ldots$, normally arranged in descending order of size, and for each there is a corresponding eigenfunction $C_m$. The pressure and horizontal velocity components are then expressed as

$$[u, v, \phi] = \sum_{0}^{\infty} \left[u_m(x, y, t), v_m(x, y, t), \phi_m(x, y, t)\right] C_m(z). \tag{3.58}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

The benefit of this procedure is that the $z$-derivatives in the equations of motion are replaced by multiplications, and in particular (3.55) becomes

$$\frac{\partial \phi_m}{\partial t} + c_m^2 \nabla \cdot \boldsymbol{u}_m = 0 \quad \text{or} \quad \frac{\partial \eta_m^*}{\partial t} + H_m \nabla \cdot \boldsymbol{u}_m = 0, \tag{3.59a,b}$$

where $\eta^* \equiv \phi/g$. The quantity $H_m = c_m^2/g$ is the *equivalent depth* associated with the eigenmode. Equations (3.59) are evidently of the same form as the familiar linear mass continuity equation in the shallow water equations, namely

$$\frac{\partial \hat{\eta}}{\partial t} + c^2 \nabla \cdot \boldsymbol{u} = 0 \qquad \text{or} \qquad \frac{\partial \eta}{\partial t} + H \nabla \cdot \boldsymbol{u} = 0, \tag{3.60a,b}$$

where $c = \sqrt{gH}$ and $\hat{\eta} = g\eta$.

The horizontal momentum equations are simply,

$$\frac{\partial u_m}{\partial t} - f v_m = -\frac{\partial \phi_m}{\partial x}, \qquad \frac{\partial v_m}{\partial t} + f u_m = -\frac{\partial \phi_m}{\partial y}. \tag{3.61a,b}$$

Equations (3.59) and (3.61) are a closed set, once we have calculated the equivalent depth $H_m$ for each mode. If there is a forcing in the momentum equation then the transformed forcing appears on the right-hand sides of (3.61). If there is a source in the buoyancy equation then a corresponding term appears on the right-hand side of (3.59), analogous to a mass source term in the shallow water equations. Note that a thermodynamic source affects $\partial\phi/\partial z$ and not $\phi$ itself.

### *Eigenfunctions for buoyancy and vertical velocity*

The vertical velocity and the buoyancy do not satisfy the same boundary conditions and so should not be expanded in the same way. Rather, we let

$$\left[w, \frac{b}{N^2}\right] = \sum_0^\infty \left[w_m(x,y,t), \hat{b}_m(x,y,t)\right] S_m(z), \tag{3.62}$$

where the eigenfunctions satisfy

$$\frac{1}{N^2}\frac{\mathrm{d}^2 S_m}{\mathrm{d}z^2} + \frac{1}{c_m^2} S_m = 0, \qquad S_m(0) = S_m(-H) = 0, \tag{3.63a,b}$$

where $S_m = 0$ if $N = 0$, and we may use the orthonormalization,

$$\int_{-H}^{0} N^2 S_m S_n \,\mathrm{d}z = g\delta_{mn}. \tag{3.64}$$

The functions $S_m$ and $C_m$ are related by

$$C_m = \frac{c_m^2}{g}\frac{\mathrm{d}S_m}{\mathrm{d}z}, \qquad N^2 S_m = -g\frac{\mathrm{d}C_m}{\mathrm{d}z}, \tag{3.65}$$

and it is these relationships that motivate the form of (3.63). The vertical velocity may be evaluated from the mass continuity equation, $\partial w/\partial z = -\nabla \cdot \boldsymbol{u}$, which becomes

$$w_m \frac{\mathrm{d}S_m}{\mathrm{d}z} = -C_m \nabla \cdot \boldsymbol{u}_m \qquad \Longrightarrow \qquad w_m = -\frac{c_m^2}{g}\nabla \cdot \boldsymbol{u}_m. \tag{3.66a,b}$$

Buoyancy is obtained from (3.53c) which, using (3.65), gives $\hat{b}_m = -\phi_m/g$.

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

### 3.4.2 Examples and Approximations

The values of $c_m$ can be computed by solving the eigenvalue problem for the given stratification, although in general this must be carried out numerically. Consider, though, the simplest case in which $N$ is constant, which is a reasonable approximation for the troposphere, less so for the ocean. The normal modes are sines and cosines, and for $m = 1, 2 \ldots$ we have

$$C_m(z) = A_m \cos\frac{m\pi z}{H}, \qquad S_m(z) = B_m \sin\frac{m\pi z}{H}, \qquad c_m = \frac{NH}{m\pi}, \tag{3.67}$$

where, for $m > 0$, $A_m = c_m/\sqrt{gH/2}$ and $B_m = \sqrt{2g/HN^2}$. The equivalent depth is given by

$$H_m = \frac{N^2H^2}{gm^2\pi^2} = \frac{g'H}{gm^2\pi^2}, \tag{3.68}$$

where $g' \equiv HN^2$ and for a Boussinesq fluid $g' = (gH/\rho_0)\partial\rho/\partial z$. Using (3.67) we see that $c_m = \sqrt{gH_m} = \sqrt{g'H}/m\pi$, and note the factors of $\pi$ are significant in these expressions. The mode with $m = 0$ is a special one and is called the *barotropic mode* with

$$C_0 = A_0/2, \qquad c_0^2 = gH. \tag{3.69}$$

The above expressions allow us to estimate equivalent depths and phase speeds for the atmosphere and ocean, with some caveats. For the atmosphere we should properly take into account its compressibility and a leaky tropopause, but proceeding nevertheless let us take $H = 10\,\mathrm{km}$ and $N = 10^{-2}\,\mathrm{s}^{-1}$ (a typical tropospheric value), whence

$$c_0 \approx 300\,\mathrm{m\,s^{-1}},\ c_1 \approx 30\,\mathrm{m\,s^{-1}},\ c_2 \approx 15\,\mathrm{m\,s^{-1}} \quad \text{and} \quad H_1 \approx 100\,\mathrm{m},\ H_2 \approx 25\,\mathrm{m}. \tag{3.70}$$

These equivalent depths are much smaller than the actual depth of the atmosphere, a fact that transcends our approximations and that greatly affects the properties of atmospheric gravity waves, as we discover in later chapters. The best fits to observations of internal gravity waves in the atmosphere are often in fact made with an equivalent depth of 50 m or less and a speed of about $20\,\mathrm{m\,s^{-1}}$.

The oceanic stratification is in fact not constant, but decreases significantly below the thermocline, which is about 1 km thick. We might proceed by simply using values appropriate for the thermocline in the above, and if we take $N = 10^{-2}\,\mathrm{s}^{-1}$ and $H = 1\,\mathrm{km}$ we find, using (3.67) and (3.68),

$$c_0 \approx 200\,\mathrm{m\,s^{-1}},\ c_1 \approx 3\,\mathrm{m\,s^{-1}},\ c_2 \approx 1.5\,\mathrm{m\,s^{-1}} \quad \text{and} \quad H_1 \approx 1\,\mathrm{m},\ H_2 \approx 0.25\,\mathrm{m}. \tag{3.71}$$

The speed $c_0$ is (as for the atmosphere) vastly larger than any parcel speed in the ocean. In contrast, the equivalent depths are very small, but this just reflects the smallness of the density variations in the ocean and the fact that $H_m$ is proportional to $g'/g$.

If the oceanic stratification varies reasonably slowly we can use WKB methods (page 247) to good effect to better evaluate the eigenvalues and eigenfunctions.[3] Roughly speaking $NH$ is replaced by $\int N\,\mathrm{d}z$ in (3.67), and the WKB solution, for $m \geq 1$, is

$$S_m \sim S_0 \sin\left(\frac{1}{c_m}\int_{-H}^{z} N(z)\,\mathrm{d}z\right), \qquad C_m \sim \left(\frac{c_m N S_0}{g}\right)\cos\left(\frac{1}{c_m}\int_{-H}^{z} N(z)\,\mathrm{d}z\right),$$
$$c_m \approx \frac{1}{m\pi}\int_{-H}^{0} N\,\mathrm{d}z, \tag{3.72a,b,c}$$

where $S_0 = (c_m/N)^{1/2}$. Using (3.72c) still gives values of $c_1$ of around 2–3 $\mathrm{m\,s^{-1}}$ over the ocean gyres, less in equatorial regions, providing some post facto justification for using $H = 1\,\mathrm{km}$ previously. The eigenfunctions, (3.72a,b), are 'stretched' sines and cosines, with local wavenumbers

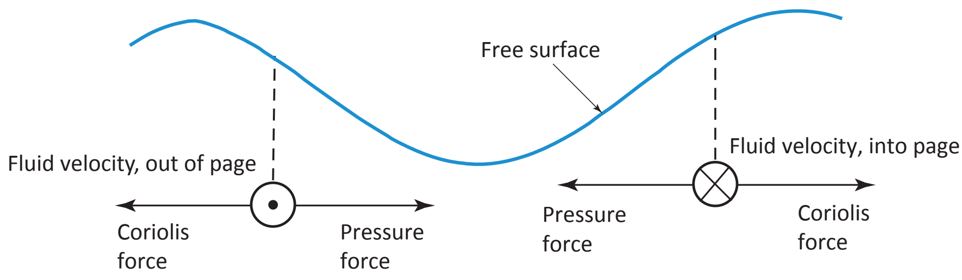


**Fig. 3.6** Geostrophic flow in a shallow water system, with a positive value of the Coriolis parameter $f$, as in the Northern Hemisphere. The pressure force is directed down the gradient of the height field, and this can be balanced by the Coriolis force if the fluid velocity is at right angles to it. If $f$ were negative, the geostrophic flow would be reversed.

proportional to $N(z)$ and so varying more rapidly in the upper ocean than at depth (look ahead to Fig. 12.12). The vertical velocity eigenfunctions, $S_m$, have a smaller amplitude in the upper ocean but the pressure and horizontal velocity amplitudes are larger.

For the remainder of this chapter we will use the shallow water equations in their conventional form, for if there *is* a region where density changes rapidly in the vertical then the layered equations are quite natural, and allow for the incorporation of nonlinearities more easily.

## 3.5 GEOSTROPHIC BALANCE AND THERMAL WIND

We now turn our attention to the *dynamics* of shallow water systems, beginning with the effects of rotation. Geostrophic balance occurs in the shallow water equations, just as in the continuously stratified equations, when the Rossby number $U/fL$ is small and the Coriolis term dominates the advective terms in the momentum equation. In the single-layer shallow water equations the geostrophic flow is:

$$\boldsymbol{f} \times \boldsymbol{u}_g = -g\nabla\eta. \tag{3.73}$$

Thus, the geostrophic velocity is proportional to the slope of the surface, as sketched in Fig. 3.6. (For the rest of this section we drop the subscript $g$, and take all velocities to be geostrophic.)

In both the single-layer and multi-layer cases, the slope of an interfacial surface is directly related to the difference in pressure gradient on either side and so, by geostrophic balance, to the shear of the flow. This is the shallow water analogue of the thermal wind relation. To obtain an expression for this, consider the interface, $\eta$, between two layers labelled 1 and 2. The pressure in two layers is given by the hydrostatic relation and so,

$$p_1 = A(x, y) - \rho_1 g z \qquad \text{(at some } z \text{ in layer 1)}, \tag{3.74a}$$

$$\begin{aligned} p_2 &= A(x, y) - \rho_1 g\eta + \rho_2 g(\eta - z) \\ &= A(x, y) + \rho_1 g_1'\eta - \rho_2 g z \qquad \text{(at some } z \text{ in layer 2)}, \end{aligned} \tag{3.74b}$$

where $A(x, y)$ is a function of integration. Thus we find

$$\frac{1}{\rho_1}\nabla(p_1 - p_2) = -g_1'\nabla\eta. \tag{3.75}$$

If the flow is geostrophically balanced and Boussinesq then, in each layer, the velocity obeys

$$f\boldsymbol{u}_i = \frac{1}{\rho_1}\mathbf{k} \times \nabla p_i\,. \tag{3.76}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

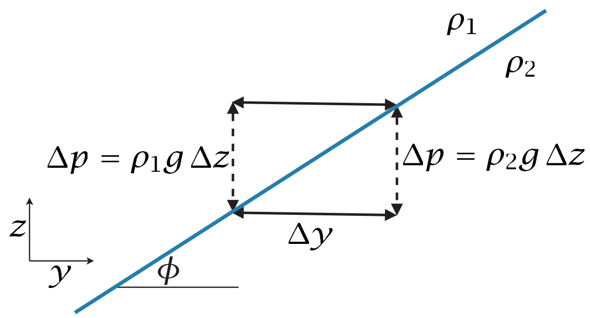


**Fig. 3.7** Margules' relation: using hydrostasy, the difference in the horizontal pressure gradient between the upper and the lower layer is given by $-g' \rho_1 s$, where $s = \tan\phi = \Delta z / \Delta y$ is the interface slope and $g' = g(\rho_2 - \rho_1)/\rho_1$. Geostrophic balance then gives $f(u_1 - u_2) = g's$, which is a special case of (3.78).

Using (3.75) then gives

$$f(\boldsymbol{u}_1 - \boldsymbol{u}_2) = -\mathbf{k} \times g_1' \nabla \eta, \tag{3.77}$$

or in general

$$f(\boldsymbol{u}_n - \boldsymbol{u}_{n+1}) = -\mathbf{k} \times g_n' \nabla \eta. \tag{3.78}$$

This is the thermal wind equation for the shallow water system. It applies at any interface, and it implies *the shear is proportional to the interface slope,* a result known as the Margules relation[4] (Fig. 3.7).

Suppose that we represent the atmosphere by two layers of fluid; a meridionally decreasing temperature may then be represented by an interface that slopes upwards toward the pole. Then, in either hemisphere, we have

$$u_1 - u_2 = \frac{g_1'}{f} \frac{\partial \eta}{\partial y} > 0, \tag{3.79}$$

and the temperature gradient is associated with a positive shear.

## 3.6 FORM STRESS

When the interface between two layers varies with position — that is, when it is wavy — the layers exert a pressure force on each other. Similarly, if the bottom of the fluid is not flat then the topography and the bottom layer will in general exert forces on each other. This kind of force (normally arising as a force per unit area) is known as *form stress,* and it is an important means whereby momentum can be added to or extracted from a flow.[5] Consider a layer confined between two interfaces, $\eta_1(x, y)$ and $\eta_2(x, y)$. Then over some zonal interval $L$ the average zonal pressure force on that fluid layer is given by

$$F_p = -\frac{1}{L} \int_{x_1}^{x_2} \int_{\eta_2}^{\eta_1} \frac{\partial p}{\partial x} \, \mathrm{d}x \, \mathrm{d}z. \tag{3.80}$$

Integrating by parts first in $z$ and then in $x$, and noting that by hydrostasy $\partial p / \partial z$ does not depend on horizontal position within the layer, we obtain

$$F_p = -\frac{1}{L} \int_{x_1}^{x_2} \left[ \frac{\partial p}{\partial x} z \right]_{\eta_2}^{\eta_1} \mathrm{d}x = -\overline{\eta_1 \frac{\partial p_1}{\partial x}} + \overline{\eta_2 \frac{\partial p_2}{\partial x}} = +\overline{p_1 \frac{\partial \eta_1}{\partial x}} - \overline{p_2 \frac{\partial \eta_2}{\partial x}}, \tag{3.81}$$

where $p_1$ is the pressure at $\eta_1$, and similarly for $p_2$, and to obtain the second line we suppose that the integral is around a closed path, such as a circle of latitude, and the average is denoted with an overbar. These terms represent the transfer of momentum from one layer to the next, and at a particular interface, $i$, we may define the form stress, $\tau_i$, by

$$\tau_i \equiv \overline{p_i \frac{\partial \eta_i}{\partial x}} = -\overline{\eta_i \frac{\partial p_i}{\partial x}}. \tag{3.82}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

The form stress is a force per unit area and its vertical derivative, $\partial\tau/\partial z$, is the force (per unit volume) on the fluid. Form stress is a particularly important means for the vertical transfer of momentum and its ultimate removal in an eddying fluid, and is one of the main mechanisms whereby the wind stress at the top of the ocean is communicated to the ocean bottom. At the fluid bottom the form stress is $\overline{p\partial_x\eta_b}$, where $\eta_b$ is the bottom topography, and this is proportional to the momentum exchange with the solid Earth. This is a significant mechanism for the ultimate removal of momentum in the ocean, especially in the Antarctic Circumpolar Current where it is likely to be much larger than bottom (or Ekman) drag arising from small-scale turbulence and friction. In the two-layer, flat-bottomed case the only form stress occurring is that at the interface, and the momentum transfer between the layers is just $\overline{p_1\partial\eta_1/\partial x}$ or $-\overline{\eta_1\partial p_1/\partial x}$; then, the force on each layer due to the other is equal and opposite, as we would expect from momentum conservation. (Form stress is discussed more in an oceanographic context in Sections 19.6.3 and 21.7.2.)

For flows in geostrophic balance, the form stress is related to the meridional heat flux. The pressure gradient and velocity are related by $\rho f v' = \partial p'/\partial x$ and the interfacial displacement is proportional to the temperature perturbation, $b'$ — in fact one may show that $\eta' \approx -b'/(\partial\overline{b}/\partial z)$. Thus $-\overline{\eta'\partial p'/\partial x} \propto \overline{v'b'}$, a correspondence that will recur when we consider the *Eliassen–Palm flux* in Chapter 10.

## 3.7 CONSERVATION PROPERTIES OF SHALLOW WATER SYSTEMS

There are two common types of conservation property in fluids: (i) material invariants; and (ii) integral invariants. Material invariance occurs when a property ($\varphi$ say) is conserved on each fluid element, and so obeys the equation $\mathrm{D}\varphi/\mathrm{D}t = 0$. An integral invariant is one that is conserved after an integration over some, usually closed, volume; energy is an example.

### 3.7.1 Potential Vorticity: a Material Invariant

The vorticity of a fluid (considered at greater length in chapter 4), denoted $\boldsymbol{\omega}$, is defined to be the curl of the velocity field. Let us also define the shallow water vorticity, $\boldsymbol{\omega}^*$, as the curl of the horizontal velocity. We therefore have:

$$\boldsymbol{\omega} \equiv \nabla\times\boldsymbol{v}, \qquad \boldsymbol{\omega}^* \equiv \nabla\times\boldsymbol{u}. \tag{3.83}$$

Because $\partial u/\partial z = \partial v/\partial z = 0$, only the vertical component of $\boldsymbol{\omega}^*$ is non-zero and

$$\boldsymbol{\omega}^* = \mathbf{k}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = \mathbf{k}\,\zeta. \tag{3.84}$$

Considering first the non-rotating case, we use the vector identity

$$(\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \frac{1}{2}\nabla(\boldsymbol{u}\cdot\boldsymbol{u}) - \boldsymbol{u}\times(\nabla\times\boldsymbol{u}), \tag{3.85}$$

to write the momentum equation, (3.8) with $f = 0$, as

$$\frac{\partial\boldsymbol{u}}{\partial t} + \boldsymbol{\omega}^*\times\boldsymbol{u} = -\nabla\left(g\eta + \frac{1}{2}\boldsymbol{u}^2\right). \tag{3.86}$$

To obtain an evolution equation for the vorticity we take the curl of (3.86), and make use of the vector identity

$$\begin{aligned}\nabla\times(\boldsymbol{\omega}^*\times\boldsymbol{u}) &= (\boldsymbol{u}\cdot\nabla)\boldsymbol{\omega}^* - (\boldsymbol{\omega}^*\cdot\nabla)\boldsymbol{u} + \boldsymbol{\omega}^*\nabla\cdot\boldsymbol{u} - \boldsymbol{u}\nabla\cdot\boldsymbol{\omega}^*\\ &= (\boldsymbol{u}\cdot\nabla)\boldsymbol{\omega}^* + \boldsymbol{\omega}^*\nabla\cdot\boldsymbol{u},\end{aligned} \tag{3.87}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

using the fact that $\nabla \cdot \boldsymbol{\omega}^*$ is the divergence of a curl and therefore zero, and $(\boldsymbol{\omega}^* \cdot \nabla)\boldsymbol{u} = 0$ because $\boldsymbol{\omega}^*$ is perpendicular to the surface in which $\boldsymbol{u}$ varies. Taking the curl of (3.86) gives

$$\frac{\partial \zeta}{\partial t} + (\boldsymbol{u} \cdot \nabla)\zeta = -\zeta \nabla \cdot \boldsymbol{u}, \tag{3.88}$$

where $\zeta = \mathbf{k} \cdot \boldsymbol{\omega}^*$. Now, the mass conservation equation may be written as

$$-\zeta \nabla \cdot \boldsymbol{u} = \frac{\zeta}{h}\frac{\mathrm{D}h}{\mathrm{D}t}, \tag{3.89}$$

and using this (3.88) becomes

$$\frac{\mathrm{D}\zeta}{\mathrm{D}t} = \frac{\zeta}{h}\frac{\mathrm{D}h}{\mathrm{D}t}, \tag{3.90}$$

which simplifies to

$$\boxed{\frac{\mathrm{D}Q}{\mathrm{D}t} = 0 \qquad \text{where} \qquad Q = \left(\frac{\zeta}{h}\right).} \tag{3.91}$$

The important quantity $Q$ is known as the *potential vorticity*, and (3.91) is the potential vorticity equation. We re-derive this conservation law in a different way in Section 4.6.

Because $Q$ is conserved on parcels, then so is any function of $Q$; that is, $F(Q)$ is a material invariant, where $F$ is any function. To see this algebraically, multiply (3.91) by $F'(Q)$, the derivative of $F$ with respect to $Q$, giving

$$F'(Q)\frac{\mathrm{D}Q}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t}F(Q) = 0. \tag{3.92}$$

Since $F$ is arbitrary there are an infinite number of material invariants corresponding to different choices of $F$.

### *Effects of rotation*

In a rotating frame of reference, the shallow water momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta, \tag{3.93}$$

where (as before) $\boldsymbol{f} = f\mathbf{k}$. This may be written in vector invariant form as

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{\omega}^* + \boldsymbol{f}) \times \boldsymbol{u} = -\nabla\left(g\eta + \frac{1}{2}\boldsymbol{u}^2\right), \tag{3.94}$$

and taking the curl of this gives the vorticity equation

$$\frac{\partial \zeta}{\partial t} + (\boldsymbol{u} \cdot \nabla)(\zeta + f) = -(f + \zeta)\nabla \cdot \boldsymbol{u}. \tag{3.95}$$

This is the same as the shallow water vorticity equation in a non-rotating frame, save that $\zeta$ is replaced by $\zeta + f$, the reason for this being that $f$ is the vorticity that the fluid has by virtue of the background rotation. Thus, (3.95) is simply the equation of motion for the total or absolute vorticity, $\boldsymbol{\omega}_a = \boldsymbol{\omega}^* + \boldsymbol{f} = (\zeta + f)\mathbf{k}$.

The potential vorticity equation in the rotating case follows, much as in the non-rotating case, by combining (3.95) with the mass conservation equation, giving

$$\boxed{\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\zeta + f}{h}\right) = 0.} \tag{3.96}$$

That is, the potential vorticity in a rotating shallow system is given by $Q = (\zeta + f)/h$ and is a material invariant. (The same symbol, $Q$, is commonly used for many of the manifestations of potential vorticity.)

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

*Vorticity and circulation*

Although vorticity itself is not a material invariant, its integral over a horizontal material area is invariant. To demonstrate this in the non-rotating case, consider the integral

$$C = \int_A \zeta \, \mathrm{d}A = \int_A Qh \, \mathrm{d}A, \tag{3.97}$$

over a surface $A$, the cross-sectional area of a column of height $h$ (as in Fig. 3.2). Taking the material derivative of this gives

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \int_A \frac{\mathrm{D}Q}{\mathrm{D}t} h \, \mathrm{d}A + \int_A Q \frac{\mathrm{D}}{\mathrm{D}t}(h \, \mathrm{d}A). \tag{3.98}$$

On the right-hand side the first term is zero, by (3.91), and the second term is just the derivative of the volume of a column of fluid of constant density and so it too is zero. Thus,

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t} \int_A \zeta \, \mathrm{d}A = 0. \tag{3.99}$$

Thus, the integral of the vorticity over some cross-sectional area of the fluid is unchanging, although both the vorticity and area of the fluid may individually change. Using Stokes' theorem, it may be written as

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t} \oint \boldsymbol{u} \cdot \, \mathrm{d}\boldsymbol{l}, \tag{3.100}$$

where the line integral is around the boundary of $A$. This is an example of Kelvin's circulation theorem, which we shall meet again in a more general form in Chapter 4, where we also consider the rotating case.

A slight generalization of (3.99) is possible. Consider the integral $I = \int F(Q) h \, \mathrm{d}A$ where again $F$ is any differentiable function of its argument. It is clear that

$$\frac{\mathrm{D}}{\mathrm{D}t} \int_A F(Q) h \, \mathrm{d}A = 0. \tag{3.101}$$

If the area of integration in (3.86) or (3.101) is the whole domain (enclosed by frictionless walls, for example) then it is clear that the integral of $hF(Q)$ is a constant, including as a special case the integral of $\zeta$.

### 3.7.2 Energy Conservation: an Integral Invariant

Since we have made various simplifications in deriving the shallow water system, it is not self-evident that energy should be conserved, or indeed what form the energy takes. The kinetic energy density (KE), meaning the kinetic energy per unit area, is $\rho_0 h \boldsymbol{u}^2/2$. The potential energy density of the fluid is

$$\mathrm{PE} = \int_0^h \rho_0 g z \, \mathrm{d}z = \frac{1}{2} \rho_0 g h^2. \tag{3.102}$$

The factor $\rho_0$ appears in both kinetic and potential energies and, because it is a constant, we will omit it. For algebraic simplicity we also assume the bottom is flat, at $z = 0$.

Using the mass conservation equation (3.14b) we obtain an equation for the evolution of potential energy density, namely

$$\frac{\mathrm{D}}{\mathrm{D}t} \frac{gh^2}{2} + gh^2 \nabla \cdot \boldsymbol{u} = 0 \tag{3.103a}$$

or

$$\frac{\partial}{\partial t} \frac{gh^2}{2} + \nabla \cdot \left( \boldsymbol{u} \frac{gh^2}{2} \right) + \frac{gh^2}{2} \nabla \cdot \boldsymbol{u} = 0. \tag{3.103b}$$

From the momentum and mass continuity equations we obtain an equation for the evolution of kinetic energy density, namely

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{h\boldsymbol{u}^2}{2} + \frac{\boldsymbol{u}^2 h}{2}\nabla\cdot\boldsymbol{u} = -g\boldsymbol{u}\cdot\nabla\frac{h^2}{2} \tag{3.104a}$$

or

$$\frac{\partial}{\partial t}\frac{h\boldsymbol{u}^2}{2} + \nabla\cdot\left(\boldsymbol{u}\frac{h\boldsymbol{u}^2}{2}\right) + g\boldsymbol{u}\cdot\nabla\frac{h^2}{2} = 0. \tag{3.104b}$$

Adding (3.103b) and (3.104b) we obtain

$$\frac{\partial}{\partial t}\frac{1}{2}\left(h\boldsymbol{u}^2 + gh^2\right) + \nabla\cdot\left[\frac{1}{2}\boldsymbol{u}\left(gh^2 + h\boldsymbol{u}^2 + gh^2\right)\right] = 0, \tag{3.105}$$

or

$$\frac{\partial E}{\partial t} + \nabla\cdot\boldsymbol{F} = 0, \tag{3.106}$$

where $E = \mathrm{KE} + \mathrm{PE} = (h\boldsymbol{u}^2 + gh^2)/2$ is the density of the total energy and $\boldsymbol{F} = \boldsymbol{u}(h\boldsymbol{u}^2/2 + gh^2)$ is the energy flux. If the fluid is confined to a domain bounded by rigid walls, on which the normal component of velocity vanishes, then on integrating (3.105) over that area and using Gauss's theorem, the total energy is seen to be conserved; that is

$$\frac{\mathrm{d}\widehat{E}}{\mathrm{d}t} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_A \left(h\boldsymbol{u}^2 + gh^2\right)\mathrm{d}A = 0. \tag{3.107}$$

Such an energy principle also holds in the case with bottom topography. Just as we found in the case for a compressible fluid in Chapter 2, the energy flux in (3.106) is not just the energy density multiplied by the velocity; it contains an additional term $guh^2/2$, and this represents the energy transfer occurring when the fluid does work against the pressure force.

## 3.8 SHALLOW WATER WAVES

Let us now look at the gravity waves that occur in shallow water. To isolate the essence we will consider waves in a single fluid layer, with a flat bottom and a free upper surface, in which gravity provides the sole restoring force.

### 3.8.1 Non-rotating Shallow Water Waves

Given a flat bottom the fluid thickness is equal to the free surface displacement (Fig. 3.1), and taking the basic state of the fluid to be at rest we let

$$h(x, y, t) = H + h'(x, y, t) = H + \eta'(x, y, t), \tag{3.108a}$$
$$\boldsymbol{u}(x, y, t) = \boldsymbol{u}'(x, y, t). \tag{3.108b}$$

The mass conservation equation, (3.14b), then becomes

$$\frac{\partial \eta'}{\partial t} + (H + \eta')\nabla\cdot\boldsymbol{u}' + \boldsymbol{u}'\cdot\nabla\eta' = 0, \tag{3.109}$$

and neglecting squares of small quantities this yields the linear equation

$$\frac{\partial \eta'}{\partial t} + H\nabla\cdot\boldsymbol{u}' = 0. \tag{3.110}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

Similarly, linearizing the momentum equation, (3.8) with $\boldsymbol{f} = 0$, yields

$$\frac{\partial \boldsymbol{u}'}{\partial t} = -g\nabla\eta'. \tag{3.111}$$

Eliminating velocity by differentiating (3.110) with respect to time and taking the divergence of (3.111) leads to

$$\frac{\partial^2\eta'}{\partial t^2} - gH\nabla^2\eta' = 0, \tag{3.112}$$

which may be recognized as a wave equation. We can find the dispersion relationship for this by substituting the trial solution

$$\eta' = \mathrm{Re}\,\tilde{\eta}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}, \tag{3.113}$$

where $\tilde{\eta}$ is a complex constant, $\boldsymbol{k} = \mathbf{i}k + \mathbf{j}l$ is the horizontal wavenumber and Re indicates that the real part of the solution should be taken. If, for simplicity, we restrict attention to the one-dimensional problem, with no variation in the $y$-direction, then substituting into (3.112) leads to the dispersion relationship

$$\omega = \pm ck, \tag{3.114}$$

where $c = \sqrt{gH}$; that is, the wave speed is proportional to the square root of the mean fluid depth and is independent of the wavenumber — the waves are dispersionless. The general solution is a superposition of all such waves, with the amplitudes of each wave (or Fourier component) being determined by the Fourier decomposition of the initial conditions.

Because the waves are dispersionless, the general solution can be written as

$$\eta'(x,t) = \frac{1}{2}\left[F(x-ct) + F(x+ct)\right], \tag{3.115}$$

where $F(x)$ is the height field at $t = 0$. From this, it is easy to see that the shape of an initial disturbance is preserved as it propagates both to the right and to the left at speed $c$.

### 3.8.2 Rotating Shallow Water (Poincaré) Waves

We now consider the effects of rotation on shallow water waves. Linearizing the rotating, flat-bottomed $f$-plane shallow water equations, (SW.1) and (SW.2) on page 109, about a state of rest we obtain

$$\frac{\partial u'}{\partial t} - f_0 v' = -g\frac{\partial \eta'}{\partial x}, \qquad \frac{\partial v'}{\partial t} + f_0 u' = -g\frac{\partial \eta'}{\partial y}, \qquad \frac{\partial \eta'}{\partial t} + H\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = 0. \tag{3.116a,b,c}$$

To obtain a dispersion relationship we let

$$(u, v, \eta) = (\tilde{u}, \tilde{v}, \tilde{\eta})\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}, \tag{3.117}$$

and substitute into (3.116), giving

$$\begin{pmatrix} -\mathrm{i}\,\omega & -f_0 & \mathrm{i}\,gk \\ f_0 & -\mathrm{i}\,\omega & \mathrm{i}\,gl \\ \mathrm{i}\,Hk & \mathrm{i}\,Hl & -\mathrm{i}\,\omega \end{pmatrix} \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{\eta} \end{pmatrix} = 0. \tag{3.118}$$

This homogeneous equation has non-trivial solutions only if the determinant of the matrix vanishes, and that condition gives

$$\omega(\omega^2 - f_0^2 - c^2K^2) = 0, \tag{3.119}$$

where $K^2 = k^2 + l^2$ and $c^2 = gH$. There are two classes of solution to (3.119). The first is simply $\omega = 0$, i.e., time-independent flow corresponding to geostrophic balance in (3.116). Because

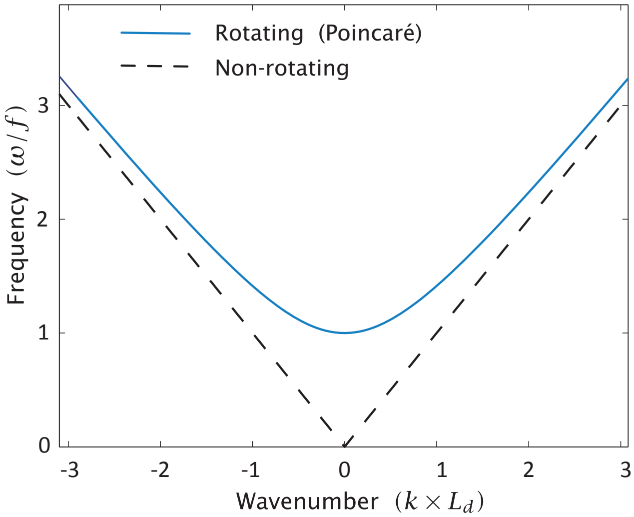


**Fig. 3.8** Dispersion relation for Poincaré waves and non-rotating shallow water waves. Frequency is scaled by the Coriolis frequency $f$, and wavenumber by the inverse deformation radius $\sqrt{gH}/f$. For small wavenumbers the frequency of the Poincaré waves is approximately $f$, and for high wavenumbers it asymptotes to that of non-rotating waves.

geostrophic balance gives a divergence-free velocity field for a constant Coriolis parameter the equations are satisfied by a time-independent solution. (If the Coriolis parameter varies in space then the $\omega = 0$ solution morphs into a non-trivial dispersion relation for *Rossby waves,* considered in Chapter 5.) The second set of solutions gives the dispersion relation

$$\omega^2 = f_0^2 + c^2(k^2 + l^2), \tag{3.120}$$

or

$$\boxed{\omega^2 = f_0^2 + gH(k^2 + l^2).} \tag{3.121}$$

The corresponding waves are known as *Poincaré* waves,[6] and the dispersion relationship is illustrated in Fig. 3.8. Note that the frequency is always greater than the Coriolis frequency $f_0$. There are two interesting limits:

(i) *The short wave limit.* If

$$K^2 \gg \frac{f_0^2}{gH}, \tag{3.122}$$

where $K^2 = k^2 + l^2$, then the dispersion relationship reduces to that of the non-rotating case (3.114). This condition is equivalent to requiring that the wavelength be much shorter than the *deformation radius,* $L_d \equiv \sqrt{gH}/f$. Specifically, if $l = 0$ and $\lambda = 2\pi/k$ is the wavelength, the condition is

$$\lambda^2 \ll L_d^2(2\pi)^2. \tag{3.123}$$

The numerical factor of $(2\pi)^2$ is more than an order of magnitude, so care must be taken when deciding if the condition is satisfied in particular cases. Furthermore, the wavelength must still be longer than the depth of the fluid, otherwise the shallow water condition is not met.

(ii) *The long wave limit.* If

$$K^2 \ll \frac{f_0^2}{gH}, \tag{3.124}$$

that is if the wavelength is much longer than the deformation radius $L_d$, then the dispersion relationship is

$$\omega = f_0. \tag{3.125}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

These are known as *inertial oscillations*. The equations of motion giving rise to them are

$$\frac{\partial u'}{\partial t} - f_0 v' = 0, \qquad \frac{\partial v'}{\partial t} + f_0 u' = 0, \tag{3.126}$$

which are equivalent to material equations for free particles in a rotating frame, unconstrained by pressure forces, namely

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} - f_0 v = 0, \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + f_0 u = 0. \tag{3.127}$$

### 3.8.3 Kelvin Waves

The Kelvin wave is a particular type of gravity wave that exists in the presence of both rotation and a lateral boundary. Suppose there is a solid boundary at $y = 0$; clearly harmonic solutions in the $y$-direction are not allowable, as these would not satisfy the condition of no normal flow at the boundary. Do any wave-like solutions exist? The affirmative answer to this question was provided by W. Thomson and the associated waves are now eponymously known as *Kelvin waves*.[7] We begin with the linearized shallow water equations, namely

$$\frac{\partial u'}{\partial t} - f_0 v' = -g\frac{\partial \eta'}{\partial x}, \qquad \frac{\partial v'}{\partial t} + f_0 u' = -g\frac{\partial \eta'}{\partial y}, \qquad \frac{\partial \eta'}{\partial t} + H\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = 0. \tag{3.128a,b,c}$$

The fact that $v' = 0$ at $y = 0$ suggests that we look for a solution with $v' = 0$ everywhere, whence these equations become

$$\frac{\partial u'}{\partial t} = -g\frac{\partial \eta'}{\partial x}, \qquad f_0 u' = -g\frac{\partial \eta'}{\partial y}, \qquad \frac{\partial \eta'}{\partial t} + H\frac{\partial u'}{\partial x} = 0. \tag{3.129a,b,c}$$

Equations (3.129a) and (3.129c) lead to the standard wave equation

$$\frac{\partial^2 u'}{\partial t^2} = c^2\frac{\partial^2 u'}{\partial x^2}, \tag{3.130}$$

where $c = \sqrt{gH}$, the usual wave speed of shallow water waves. The solution of (3.130) is

$$u' = F_1(x + ct, y) + F_2(x - ct, y), \tag{3.131}$$

with corresponding surface displacement

$$\eta' = \sqrt{H/g}\left[-F_1(x + ct, y) + F_2(x - ct, y)\right]. \tag{3.132}$$

The solution represents the superposition of two waves, one ($F_1$) travelling in the negative $x$-direction, and the other in the positive $x$-direction. To obtain the $y$ dependence of these functions we use (3.129b) which gives

$$\frac{\partial F_1}{\partial y} = \frac{f_0}{\sqrt{gH}}F_1, \qquad \frac{\partial F_2}{\partial y} = -\frac{f_0}{\sqrt{gH}}F_2, \tag{3.133}$$

with solutions

$$F_1 = F(x + ct)\mathrm{e}^{y/L_d}, \qquad F_2 = G(x - ct)\mathrm{e}^{-y/L_d}, \tag{3.134}$$

where $L_d = \sqrt{gH}/f_0$ is the radius of deformation. If we consider flow in the half-plane in which $y > 0$, then for positive $f_0$ the solution $F_1$ grows exponentially away from the wall, and so fails

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

to satisfy the condition of boundedness at infinity. It thus must be eliminated, leaving the general solution

$$u' = \mathrm{e}^{-y/L_d} G(x - ct), \qquad v' = 0, \qquad \eta' = \sqrt{H/g}\,\mathrm{e}^{-y/L_d} G(x - ct). \tag{3.135a,b,c}$$

These are Kelvin waves, and they decay exponentially away from the boundary. In general, for $f_0$ positive the boundary is to the right of an observer moving with the wave. Given a constant Coriolis parameter, we could equally well have obtained a solution on a meridional wall, in which case we would find that the wave again moves such that the wall is to the right of the wave direction. (This is obvious once it is realized that $f$-plane dynamics are isotropic in $x$ and $y$.) Thus, in the Northern Hemisphere the wave moves anticlockwise round a basin, and conversely in the Southern Hemisphere, and in both hemispheres the direction is cyclonic.

## 3.9 GEOSTROPHIC ADJUSTMENT

We noted in Chapter 2 that the large-scale, extratropical circulation of the atmosphere is in near-geostrophic balance. Why is this? Why should the Rossby number be small? Arguably, the magnitude of the velocity in the atmosphere and ocean is ultimately given by the strength of the forcing, and so ultimately by the differential heating between pole and equator (although even this argument is not satisfactory, since the forcing mainly determines the energy throughput, not directly the energy itself, and the forcing is itself dependent on the atmosphere's response). But even supposing that the velocity magnitudes are given, there is no a-priori guarantee that the forcing or the dynamics will produce length scales that are such that the Rossby number is small. However, there is in fact a powerful and ubiquitous process whereby a fluid in an initially unbalanced state naturally evolves toward a state of geostrophic balance, namely *geostrophic adjustment*. This process occurs quite generally in rotating fluids, whether stratified or not. To pose the problem in a simple form we consider the free evolution of a single shallow layer of fluid whose initial state is manifestly unbalanced, and we suppose that surface displacements are small so that the evolution of the system is described by the linearized shallow equations of motion. These are

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta, \qquad \frac{\partial \eta}{\partial t} + H\nabla\cdot\boldsymbol{u} = 0, \tag{3.136a,b}$$

where $\eta$ is the free surface displacement and $H$ is the mean fluid depth, and we omit the primes on the linearized variables.

### 3.9.1 Non-rotating Flow

We consider first the non-rotating problem set, with little loss of generality, in one dimension. We suppose that initially the fluid is at rest but with a simple discontinuity in the height field so that

$$\eta(x, t = 0) = \begin{cases} +\eta_0 & x < 0 \\ -\eta_0 & x > 0, \end{cases} \tag{3.137}$$

and $u(x, t = 0) = 0$ everywhere. We can realize these initial conditions physically by separating two fluid masses of different depths by a thin dividing wall, and then quickly removing the wall. What is the subsequent evolution of the fluid? The general solution to the linear problem is given by (3.115) where the functional form is determined by the initial conditions so that here

$$F(x) = \eta(x, t = 0) = -\eta_0\,\mathrm{sgn}(x). \tag{3.138}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

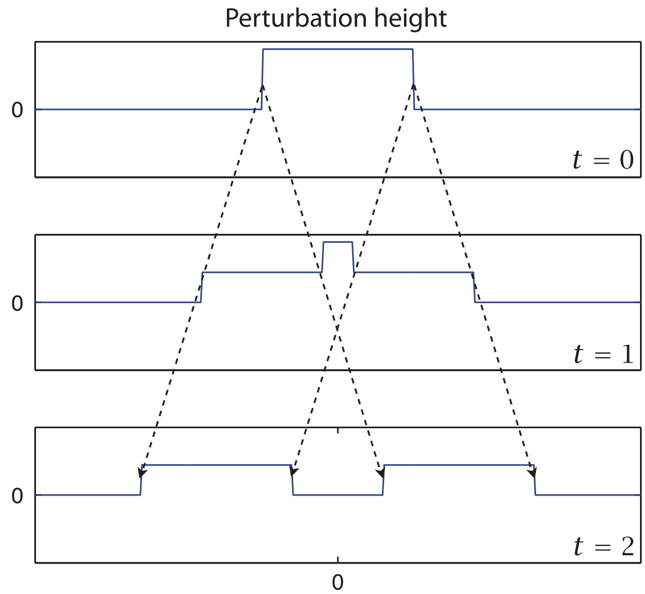


**Fig. 3.9** The time development of an initial 'top hat' height disturbance, with zero initial velocity, in non-rotating flow. Fronts propagate in both directions, and the velocity is non-zero between fronts, but ultimately the disturbances are radiated away to infinity, and the fluid is left at rest with zero perturbation height.

Equation (3.115) states that this initial pattern is propagated to the right and to the left. That is, two discontinuities in fluid height move to the right and left at a speed $c = \sqrt{gH}$. Specifically, the solution is

$$\eta(x,t) = -\frac{1}{2}\eta_0[\mathrm{sgn}(x+ct) + \mathrm{sgn}(x-ct)]. \tag{3.139}$$

The initial conditions may be much more complex than a simple front, but, because the waves are dispersionless, the solution is still simply a sum of the translation of those initial conditions to the right and to the left at speed $c$. The velocity field in this class of problem is obtained from

$$\frac{\partial u}{\partial t} = -g\frac{\partial \eta}{\partial x}, \tag{3.140}$$

which gives, using (3.115),

$$u = -\frac{g}{2c}[F(x+ct) - F(x-ct)]. \tag{3.141}$$

Consider the case with initial conditions given by (3.137). At a given location, away from the initial disturbance, the fluid remains at rest and undisturbed until the front arrives. After the front has passed, the fluid surface is again undisturbed and the velocity is uniform and non-zero. Specifically:

$$\eta = \begin{cases} -\eta_0 \mathrm{sgn}(x) \\ 0 \end{cases} \qquad u = \begin{cases} 0 & |x| > ct \\ (\eta_0 g/c) & |x| < ct. \end{cases} \tag{3.142}$$

The solution with 'top-hat' initial conditions in the height field, and zero initial velocity, is a superposition of two discontinuities similar to (3.142) and is illustrated in Fig. 3.9. Two fronts propagate in either direction from each discontinuity and, in this case, the final velocity, as well as the fluid displacement, is zero after all the fronts have passed. That is, the disturbance is radiated completely away.

### 3.9.2 Rotating Flow

Rotation makes a profound difference to the adjustment problem of the shallow water system, because a steady, adjusted, solution can exist with non-zero gradients in the height field — the

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

associated pressure gradients being balanced by the Coriolis force — and potential vorticity conservation provides a powerful constraint on the fluid evolution.[8] In a rotating shallow fluid that conservation is represented by

$$\frac{\partial Q}{\partial t} + \boldsymbol{u} \cdot \nabla Q = 0, \tag{3.143}$$

where $Q = (\zeta + f)/h$. In the linear case with constant Coriolis parameter, (3.143) becomes

$$\frac{\partial q}{\partial t} = 0, \qquad q = \left(\zeta - f_0 \frac{\eta}{H}\right). \tag{3.144}$$

This equation may be obtained either from the linearized velocity and mass conservation equations, (3.136), or from (3.143) directly. In the latter case, we write

$$Q = \frac{\zeta + f_0}{H + \eta} \approx \frac{1}{H}(\zeta + f_0)\left(1 - \frac{\eta}{H}\right) \approx \frac{1}{H}\left(f_0 + \zeta - f_0 \frac{\eta}{H}\right) = \frac{f_0}{H} + \frac{q}{H}, \tag{3.145}$$

having used $f_0 \gg |\zeta|$ and $H \gg |\eta|$. The term $f_0/H$ is a constant and so dynamically unimportant, as is the $H^{-1}$ factor multiplying $q$. Further, the advective term $\boldsymbol{u} \cdot \nabla Q$ becomes $\boldsymbol{u} \cdot \nabla q$, and this is second order in perturbed quantities and so is neglected. Thus, making these approximations, (3.143) reduces to (3.144). The potential vorticity field is therefore fixed in space! Of course, this was also true in the non-rotating case where the fluid is initially at rest. Then $q = \zeta = 0$ and the fluid remains irrotational throughout the subsequent evolution of the flow. However, this is rather a weak constraint on the subsequent evolution of the fluid; it does nothing, for example, to prevent the conversion of all the potential energy to kinetic energy. In the rotating case the potential vorticity is non-zero, and potential vorticity conservation and geostrophic balance are all we need to infer the final steady state, assuming it exists, without solving for the details of the flow evolution, as we now see.

With an initial condition for the height field given by (3.137), the initial potential vorticity is given by

$$q(x, y) = \begin{cases} -f_0 \eta_0 / H & x < 0 \\ \phantom{-}f_0 \eta_0 / H & x > 0, \end{cases} \tag{3.146}$$

and this remains unchanged throughout the adjustment process. The final steady state is then the solution of the equations

$$\zeta - f_0 \frac{\eta}{H} = q(x, y), \qquad f_0 u = -g \frac{\partial \eta}{\partial y}, \qquad f_0 v = g \frac{\partial \eta}{\partial x}, \tag{3.147a,b,c}$$

where $\zeta = \partial v/\partial x - \partial u/\partial y$. Because the Coriolis parameter is constant, the velocity field is horizontally non-divergent and we may define a streamfunction $\psi = g\eta/f_0$. Equations (3.147) then reduce to

$$\left(\nabla^2 - \frac{1}{L_d^2}\right)\psi = q(x, y), \tag{3.148}$$

where $L_d = \sqrt{gH}/f_0$ is known as the *Rossby radius of deformation* or often just the 'deformation radius' or the 'Rossby radius'. It is a naturally occurring length scale in problems involving both rotation and gravity, and arises in a slightly different form in stratified fluids.

The initial conditions (3.146) admit of a nice analytic solution, for the flow will remain uniform in $y$, and (3.148) reduces to

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{L_d^2}\psi = \frac{f_0 \eta_0}{H} \operatorname{sgn}(x). \tag{3.149}$$

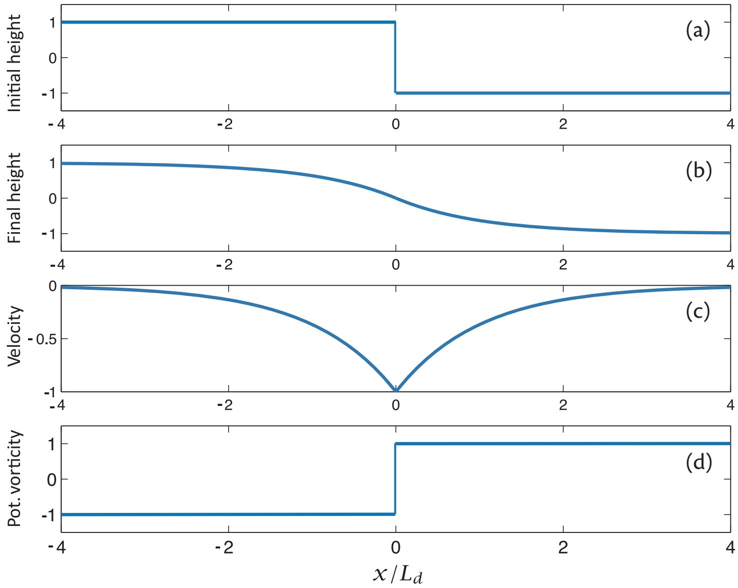


**Fig. 3.10** Solutions of a linear geostrophic adjustment problem. (a) Initial height field, given by (3.137) with $\eta_0 = 1$. (b) Equilibrium (final) height field, $\eta$ given by (3.150) and $\eta = f_0\psi/g$. (c) Equilibrium geostrophic velocity, normal to the gradient of height field, given by (3.151). (d) Potential vorticity, given by (3.146), and this does not evolve.

The distance, $x$ is nondimensionalized by the deformation radius $L_d$ and the velocity by $\eta_0(g/f_0L_d)$. Changes to the initial state occur within $\mathcal{O}(L_d)$ of the initial discontinuity.

We solve this separately for $x > 0$ and $x < 0$ and then match the solutions and their first derivatives at $x = 0$, also imposing the condition that the velocity decays to zero as $x \to \pm\infty$. The solution is

$$\psi = \begin{cases} -(g\eta_0/f_0)(1 - e^{-x/L_d}) & x > 0 \\ +(g\eta_0/f_0)(1 - e^{x/L_d}) & x < 0. \end{cases} \tag{3.150}$$

The velocity field associated with this is obtained from (3.147b,c), and is

$$u = 0, \qquad v = -\frac{g\eta_0}{f_0 L_d} e^{-|x|/L_d}. \tag{3.151}$$

The velocity is perpendicular to the slope of the free surface, and a jet forms along the initial discontinuity, as illustrated in Fig. 3.10.

The important point of this problem is that the variations in the height and field are not radiated away to infinity, as in the non-rotating problem. Rather, potential vorticity conservation constrains the influence of the adjustment to within a deformation radius (we see now why this name is appropriate) of the initial disturbance. This property is a general one in geostrophic adjustment — it also arises if the initial condition consists of a velocity jump.

A snapshot of the time evolution of flow, obtained by a numerical integration of the shallow water equations for both rotating and non-rotating flow, is illustrated in Fig. 3.11. The initial conditions are a jump in the height field, as in Fig. 3.10. Fronts propagate away at a speed $\sqrt{gH} = 1$ in both cases, but in the rotating flow they leave behind a geostrophically balanced state with a non-zero meridional velocity.

### 3.9.3 ✦ Energetics of Adjustment

How much of the initial potential energy of the flow is lost to infinity by gravity wave radiation, and how much is converted to kinetic energy? The linear equations (3.136) lead to

$$\frac{1}{2}\frac{\partial}{\partial t}(H\boldsymbol{u}^2 + g\eta^2) + gH\nabla \cdot (\boldsymbol{u}\eta) = 0, \tag{3.152}$$

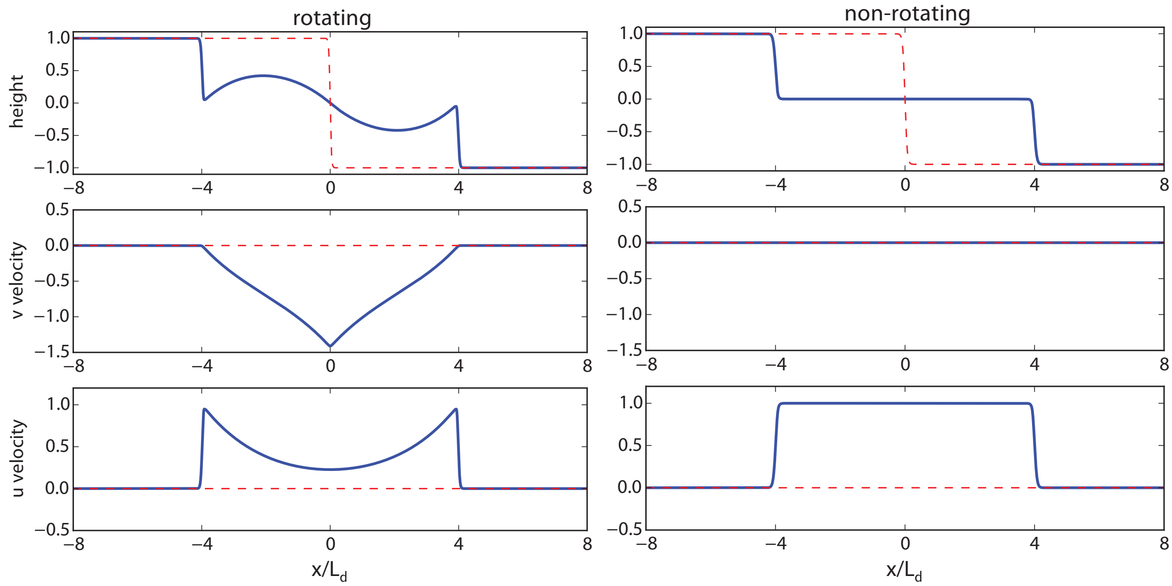


**Fig. 3.11** The solutions of the shallow water equations obtained by numerically integrating the equations of motion with and without rotation. The panels show snapshots of the state of the fluid (solid lines) soon after being released from a stationary initial state (red dashed lines) with a height discontinuity. The rotating flow is evolving toward an end state similar to Fig. 3.10 whereas the non-rotating flow will eventually become stationary. In the non-rotating case $L_d$ is defined using the rotating parameters.[9]

so that energy conservation holds in the form

$$E = \frac{1}{2}\int (H\boldsymbol{u}^2 + g\eta^2)\,\mathrm{d}\boldsymbol{x}, \qquad \frac{\mathrm{d}E}{\mathrm{d}t} = 0, \tag{3.153}$$

provided the integral of the divergence term vanishes, as it normally will in a closed domain. The fluid has a non-zero potential energy, $(1/2)\int_{-\infty}^{\infty} g\eta^2\,\mathrm{d}x$, if there are variations in fluid height, and with the initial conditions (3.137) the initial potential energy is

$$PE_I = \int_0^{\infty} g\eta_0^2\,\mathrm{d}x. \tag{3.154}$$

This is nominally infinite if the fluid has no boundaries, and the initial potential energy density is $g\eta_0^2/2$ everywhere.

In the non-rotating case, and with initial conditions (3.137), after the front has passed, the potential energy density is zero and the kinetic energy density is $Hu^2/2 = g\eta_0^2/2$, using (3.142) and $c^2 = gH$. Thus, all the potential energy is locally converted to kinetic energy as the front passes, and eventually the kinetic energy is distributed uniformly along the line. In the case illustrated in Fig. 3.9, the potential energy and kinetic energy are both radiated away from the initial disturbance. (Note that although we can superpose the solutions from different initial conditions, we cannot superpose their potential and kinetic energies.) The general point is that the evolution of the disturbance is not confined to its initial location.

In contrast, in the rotating case the conversion from potential to kinetic energy *is largely confined to within a deformation radius of the initial disturbance,* and at locations far from the initial disturbance the initial state is essentially unaltered. The conservation of potential vorticity has prevented the complete conversion of potential energy to kinetic energy, a result that is not sensitive to the precise form of the initial conditions.

In fact, in the rotating case, some of the initial potential energy is converted to kinetic energy, some remains as potential energy and some is lost to infinity; let us calculate these amounts. The final potential energy, after adjustment, is, using (3.150),

$$PE_F = \frac{1}{2} g\eta_0^2 \left[ \int_0^\infty \left(1 - \mathrm{e}^{-x/L_d}\right)^2 \mathrm{d}x + \int_{-\infty}^0 \left(1 - \mathrm{e}^{x/L_d}\right)^2 \mathrm{d}x \right]. \tag{3.155}$$

This is nominally infinite, but the change in potential energy is finite and is given by

$$PE_I - PE_F = g\eta_0^2 \int_0^\infty (2\mathrm{e}^{-x/L_d} - \mathrm{e}^{-2x/L_d})\,\mathrm{d}x = \frac{3}{2} g\eta_0^2 L_d. \tag{3.156}$$

The initial kinetic energy is zero, because the fluid is at rest, and its final value is, using (3.151),

$$KE_F = \frac{1}{2} H \int \boldsymbol{u}^2 \,\mathrm{d}x = H \left( \frac{g\eta_0}{fL_d} \right)^2 \int_0^\infty \mathrm{e}^{-2x/L_d}\,\mathrm{d}x = \frac{g\eta_0^2 L_d}{2}. \tag{3.157}$$

Thus one-third of the difference between the initial and final potential energies is converted to kinetic energy, and this is trapped within a distance of the order of a deformation radius of the disturbance; the remainder, an amount $gL_d\eta_0^2$ is radiated away and lost to infinity. In any finite region surrounding the initial discontinuity the final energy is less than the initial energy.

### 3.9.4 ✦ General Initial Conditions

Because of the linearity of the (linear) adjustment problem a spectral viewpoint is useful, in which the fields are represented as the sum or integral of *non-interacting* Fourier modes. For example, suppose that the height field of the initial disturbance is a two-dimensional field given by

$$\eta(0) = \iint \widetilde{\eta}_{k,l}(0)\mathrm{e}^{\mathrm{i}(kx+ly)}\,\mathrm{d}k\,\mathrm{d}l, \tag{3.158}$$

where the Fourier coefficients $\widetilde{\eta}_{k,l}(0)$ are given, and the initial velocity field is zero. Then the initial (and final) potential vorticity field is given by

$$q = -\frac{f_0}{H} \iint \widetilde{\eta}_{k,l}(0)\mathrm{e}^{\mathrm{i}(kx+ly)}\,\mathrm{d}k\,\mathrm{d}l. \tag{3.159}$$

To obtain an expression for the final height and velocity fields, we express the potential vorticity field as

$$q = \iint \widetilde{q}_{k,l}\,\mathrm{d}k\,\mathrm{d}l. \tag{3.160}$$

The potential vorticity field does not evolve, and it is related to the initial height field by

$$\widetilde{q}_{k,l} = -\frac{f_0}{H}\eta_{k,l}(0). \tag{3.161}$$

In the final, geostrophically balanced state, the potential vorticity is related to the height field by

$$q = \frac{g}{f_0}\nabla^2\eta - \frac{f_0}{H}\eta \qquad \text{and} \qquad \widetilde{q}_{k,l} = \left( -\frac{g}{f_0}K^2 - \frac{f_0}{H} \right)\widetilde{\eta}_{k,l}, \tag{3.162a,b}$$

where $K^2 = k^2 + l^2$. Using (3.161) and (3.162), the Fourier components of the final height field satisfy

$$\left( -\frac{g}{f_0}K^2 - \frac{f_0}{H} \right)\widetilde{\eta}_{k,l} = -\frac{f_0}{H}\widetilde{\eta}_{k,l}(0) \tag{3.163}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

or

$$\widetilde{\eta}_{k,l} = \frac{\widetilde{\eta}_{k,l}(0)}{K^2 L_d^2 + 1}. \tag{3.164}$$

In physical space the final height field is just the spectral integral of this, namely

$$\eta = \iint \widetilde{\eta}_{k,l} \, \mathrm{e}^{\mathrm{i}(kx+ly)} \, \mathrm{d}k \, \mathrm{d}l = \iint \frac{\widetilde{\eta}_{k,l}(0) \mathrm{e}^{\mathrm{i}(kx+ly)}}{K^2 L_d^2 + 1} \, \mathrm{d}k \, \mathrm{d}l. \tag{3.165}$$

We see that at large scales ($K^2 L_d^2 \ll 1$) $\eta_{k,l}$ is almost unchanged from its initial state; the velocity field, which is then determined by geostrophic balance, thus adjusts to the pre-existing height field. At large scales most of the energy in geostrophically balanced flow is potential energy; thus, it is energetically easier for the velocity to change to come into balance with the height field than vice versa. At small scales, however, the final height field has much less variability than it did initially.

Conversely, at small scales the height field adjusts to the velocity field. To see this, let us suppose that the initial conditions contain vorticity but have zero height displacement. Specifically, if the initial vorticity is $\nabla^2 \psi(0)$, where $\psi(0)$ is the initial streamfunction, then it is straightforward to show that the final streamfunction is given by

$$\psi = \iint \widetilde{\psi}_{k,l} \, \mathrm{e}^{\mathrm{i}(kx+ly)} \, \mathrm{d}k \, \mathrm{d}l = \iint \frac{K^2 L_d^2 \widetilde{\psi}_{k,l}(0) \mathrm{e}^{\mathrm{i}(kx+ly)}}{K^2 L_d^2 + 1} \, \mathrm{d}k \, \mathrm{d}l. \tag{3.166}$$

The final height field is then obtained from this, via geostrophic balance, by $\eta = (f_0/g)\psi$. Evidently, for small scales ($K^2 L_d^2 \gg 1$) the streamfunction, and hence the vortical component of the velocity field, are almost unaltered from their initial values. On the other hand, at large scales the final streamfunction has much less variability than it does initially, and so the height field is largely governed by whatever variation it (and not the velocity field) had initially. In general, the final state is a superposition of the states given by (3.165) and (3.166). The divergent component of the initial velocity field does not affect the final state because it has no potential vorticity, and so all of the associated energy is eventually lost to infinity.

Finally, we remark that just as in the problem with a discontinuous initial height profile, the change in total energy during adjustment is negative — this can be seen from the form of the integrals above, although we leave the specifics as a problem to the reader. That is, some of the initial potential and kinetic energy is lost to infinity, but some is trapped by the potential vorticity constraint.

### 3.9.5 A Variational Perspective

In the non-rotating problem, all of the initial potential energy is eventually radiated away to infinity. In the rotating problem, the final state contains both potential and kinetic energy. Why is the energy not all radiated away to infinity? It is because potential vorticity conservation on parcels prevents all of the energy being dispersed. This suggests that it may be informative to think of the geostrophic adjustment problem as a *variational problem:* we seek to minimize the energy consistent with the conservation of potential vorticity. We stay in the linear approximation in which, because the advection of potential vorticity is neglected, potential vorticity remains constant at each point.

The energy of the flow is given by the sum of potential and kinetic energies, namely

$$\text{energy} = \int (H\boldsymbol{u}^2 + g\eta^2) \, \mathrm{d}A, \tag{3.167}$$

(where $\mathrm{d}A \equiv \mathrm{d}x \, \mathrm{d}y$) and the potential vorticity field is

$$q = \zeta - f_0 \frac{\eta}{H} = (v_x - u_y) - f_0 \frac{\eta}{H}, \tag{3.168}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

where the subscripts $x, y$ denote derivatives. The problem is then to extremize the energy subject to potential vorticity conservation. This is a constrained problem in the calculus of variations, sometimes called an *isoperimetric* problem because of its origins in maximizing the area of a surface for a given perimeter.[10] The mathematical problem is to extremize the integral

$$I = \int \left\{ H(u^2 + v^2) + g\eta^2 + \lambda(x, y)\left[(v_x - u_y) - f_0\eta/H\right] \right\} \mathrm{d}A, \tag{3.169}$$

where $\lambda(x, y)$ is a Lagrange multiplier, undetermined at this stage. It is a function of space: if it were a constant, the integral would merely extremize energy subject to a given integral of potential vorticity, and rearrangements of potential vorticity (which here we wish to disallow) would leave the integral unaltered.

As there are three independent variables there are three Euler–Lagrange equations that must be solved in order to minimize $I$. These are

$$\begin{gathered} \frac{\partial L}{\partial \eta} - \frac{\partial}{\partial x}\frac{\partial L}{\partial \eta_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial \eta_y} = 0, \\ \frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\frac{\partial L}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial u_y} = 0, \qquad \frac{\partial L}{\partial v} - \frac{\partial}{\partial x}\frac{\partial L}{\partial v_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial v_y} = 0, \end{gathered} \tag{3.170}$$

where $L$ is the integrand on the right-hand side of (3.169). Substituting the expression for $L$ into (3.170) gives, after a little algebra,

$$2g\eta - \frac{\lambda f_0}{H} = 0, \qquad 2Hu + \frac{\partial \lambda}{\partial y} = 0, \qquad 2Hv - \frac{\partial \lambda}{\partial x} = 0, \tag{3.171}$$

and then eliminating $\lambda$ gives the simple relationships

$$u = -\frac{g}{f_0}\frac{\partial \eta}{\partial y}, \qquad v = \frac{g}{f_0}\frac{\partial \eta}{\partial x}, \tag{3.172}$$

which are the equations of geostrophic balance. Thus, in the linear approximation, *geostrophic balance is the minimum energy state for a given field of potential vorticity.*

## 3.10 ISENTROPIC COORDINATES

We now return to the continuously stratified primitive equations, and consider the use of potential density as a vertical coordinate. In practice this means using potential temperature in the atmosphere and (for simple equations of state) buoyancy in the ocean; such coordinate systems are generically called *isentropic coordinates,* and sometimes *isopycnal coordinates* if density is used. This may seem an odd thing to do but for adiabatic flow the resulting equations of motion have an attractive form that aids the interpretation of large-scale flow. The thermodynamic equation becomes a statement for the conservation of the mass of fluid with a given value of potential density and, because the flow of both the atmosphere and the ocean is largely along isentropic surfaces, the momentum and vorticity equations have a quasi-two-dimensional form.

The particular choice of vertical coordinate is determined by the form of the thermodynamic equation in the equation-set at hand; thus, if the thermodynamic equation is $\mathrm{D}\theta/\mathrm{D}t = \dot{\theta}$, we transform the equations from $(x, y, z)$ coordinates to $(x, y, \theta)$ coordinates. The material derivative in this coordinate system is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + u\left(\frac{\partial}{\partial x}\right)_\theta + v\left(\frac{\partial}{\partial y}\right)_\theta + \frac{\mathrm{D}\theta}{\mathrm{D}t}\frac{\partial}{\partial \theta} = \frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla_\theta + \dot{\theta}\frac{\partial}{\partial \theta}, \tag{3.173}$$

where the last term on the right-hand side is zero for adiabatic flow.

### 3.10.1 A Hydrostatic Boussinesq Fluid

In the simple Boussinesq equations (see the table on page 74) the buoyancy is the relevant thermodynamic variable. With hydrostatic balance the horizontal and vertical momentum equations are, in height coordinates,

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla\phi, \qquad b = \frac{\partial\phi}{\partial z}, \tag{3.174}$$

where $b$ is the buoyancy, the variable analogous to the potential temperature $\theta$ of an ideal gas. The thermodynamic equation is

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b}, \tag{3.175}$$

and because $b = -g\delta\rho/\rho_0$, isentropic coordinates are analogous to isopycnal coordinates.

Using (2.142) the horizontal pressure gradient may be transformed to isentropic coordinates:

$$\left(\frac{\partial\phi}{\partial x}\right)_z = \left(\frac{\partial\phi}{\partial x}\right)_b - \left(\frac{\partial z}{\partial x}\right)_b \frac{\partial\phi}{\partial z} = \left(\frac{\partial\phi}{\partial x}\right)_b - b\left(\frac{\partial z}{\partial x}\right)_b = \left(\frac{\partial M}{\partial x}\right)_b, \tag{3.176}$$

where

$$M \equiv \phi - zb. \tag{3.177}$$

Thus, the horizontal momentum equation becomes

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla_b M. \tag{3.178}$$

where the material derivative is given by (3.173), with $b$ replacing $\theta$. Using (3.177) the hydrostatic equation becomes

$$\frac{\partial M}{\partial b} = -z. \tag{3.179}$$

The mass continuity equation may be derived by noting that for a Boussinesq fluid the mass element may be written as

$$\delta m = \rho_0 \frac{\partial z}{\partial b} \delta b\, \delta x\, \delta y. \tag{3.180}$$

The mass continuity equation, $\mathrm{D}\delta m/\mathrm{D}t = 0$, becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{\partial z}{\partial b} + \frac{\partial z}{\partial b}\nabla_3 \cdot \boldsymbol{v} = 0, \tag{3.181}$$

where $\nabla_3 \cdot \boldsymbol{v} = \nabla_b \cdot \boldsymbol{u} + \partial\dot{b}/\partial b$ is the three-dimensional derivative of the velocity in isentropic coordinates. Equation (3.181) may thus be written

$$\frac{\mathrm{D}\sigma}{\mathrm{D}t} + \sigma\nabla_b \cdot \boldsymbol{u} = -\sigma\frac{\partial\dot{b}}{\partial b}, \tag{3.182}$$

where $\sigma \equiv \partial z/\partial b$ is a measure of the thickness between two isentropic surfaces and the material derivative is given by (3.173) with $\theta$ replaced by $b$. Equations (3.178), (3.179) and (3.182) comprise a closed set, with dependent variables $\boldsymbol{u}$, $M$ and $z$ in the space of independent variables $x$, $y$ and $b$.

### 3.10.2 A Hydrostatic Ideal Gas

Deriving the equations of motion for this system requires a little more work than in the Boussinesq case but the idea is the same. For an ideal gas in hydrostatic balance we have, using (1.110),

$$\frac{\delta\theta}{\theta} = \frac{\delta T}{T} - \kappa\frac{\delta p}{p} = \frac{\delta T}{T} + \frac{\delta\Phi}{c_p T} = \frac{1}{c_p T}\delta M, \tag{3.183}$$

where $\delta\Phi = g\delta z$ and $M \equiv c_p T + \Phi$ is the 'Montgomery potential', equal to the dry static energy. (We use some of the same symbols as in the Boussinesq case to facilitate comparison, but their meanings are slightly different.) From this

$$\frac{\partial M}{\partial\theta} = \Pi, \tag{3.184}$$

where $\Pi \equiv c_p T/\theta = c_p(p/p_R)^{R/c_p}$ is the 'Exner function'. Equation (3.184) represents the hydrostatic relation in isentropic coordinates. Note also that $M = \theta\Pi + \Phi$.

To obtain an appropriate form for the horizontal pressure gradient force first note that, in the usual height coordinates, it is given by

$$\frac{1}{\rho}\nabla_z p = \theta\nabla_z\Pi, \tag{3.185}$$

where $\Pi = c_p T/\theta$. Using (2.142) gives

$$\theta\nabla_z\Pi = \theta\nabla_\theta\Pi - \frac{\theta}{g}\frac{\partial\Pi}{\partial z}\nabla_\theta\Phi. \tag{3.186}$$

Then, using the definition of $\Pi$ and the hydrostatic approximation to help evaluate the vertical derivative, we obtain

$$\frac{1}{\rho}\nabla_z p = c_p\nabla_\theta T + \nabla_\theta\Phi = \nabla_\theta M. \tag{3.187}$$

Thus, the horizontal momentum equation is

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} = -\nabla_\theta M. \tag{3.188}$$

Much as in the Boussinesq case, the mass continuity equation may be derived by noting that the mass element may be written as

$$\delta m = -\frac{1}{g}\frac{\partial p}{\partial\theta}\delta\theta\,\delta x\,\delta y. \tag{3.189}$$

The mass continuity equation, $\mathrm{D}\delta m/\mathrm{D}t = 0$, becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{\partial p}{\partial\theta} + \frac{\partial p}{\partial\theta}\nabla_3\cdot\boldsymbol{v} = 0 \quad \text{or} \quad \frac{\mathrm{D}\sigma}{\mathrm{D}t} + \sigma\nabla_\theta\cdot\boldsymbol{u} = -\sigma\frac{\partial\dot\theta}{\partial\theta}, \tag{3.190a,b}$$

where now $\sigma \equiv \partial p/\partial\theta$ is a measure of the (pressure) thickness between two isentropic surfaces. Equations (3.184), (3.188) and (3.190b) form a closed set, analogous to (3.179), (3.178) and (3.182).

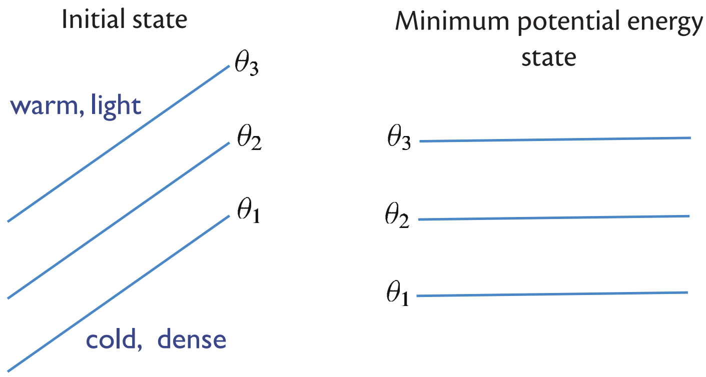


**Fig. 3.12** If a stably stratified initial state with sloping isentropes (left) is adiabatically rearranged then the state of minimum potential energy has flat isentropes, as on the right, but the amount of fluid contained between each isentropic surface is unchanged. The difference between the potential energies of the two states is the *available potential energy*.

### 3.10.3 ✦ Analogy to Shallow Water Equations

The equations of motion in isentropic coordinates have an analogy with the shallow water equations, and we may think of the shallow water equations as a finite-difference representation of the primitive equations written in isentropic coordinates, or think of the latter as the continuous limit of the shallow water equations as the number of layers increases. For example, consider a two-isentropic-level representation of (3.184), (3.188) and (3.190), in which the lower boundary is an isentrope. A natural finite differencing gives

$$-M_1 = \Pi_0 \Delta\theta_0, \qquad M_1 - M_2 = \Pi_1 \Delta\theta_1, \tag{3.191a,b}$$

where the $\Delta\theta$s are constants, and the momentum equations for each layer become

$$\frac{\mathrm{D}\boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_1 = -\Delta\theta_0 \nabla \Pi_0, \qquad \frac{\mathrm{D}\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_2 = -\Delta\theta_0 \nabla \Pi_0 - \Delta\theta_1 \nabla \Pi_1. \tag{3.192}$$

Together with the mass continuity equation for each level these are similar to the two-layer shallow water equations (3.49).

## 3.11 AVAILABLE POTENTIAL ENERGY

We now revisit the issue of the internal and potential energy in stratified flow, motivated by the following remarks. In adiabatic, inviscid flow the total amount of energy is conserved, and there are conversions between internal energy, potential energy and kinetic energy. In an ideal gas the potential energy and the internal energy of a column extending throughout the atmosphere are in a constant ratio to each other — their sum is called the total potential energy. In a simple Boussinesq fluid, energetic conversions involve only the potential and kinetic energy, and not the internal energy. Yet, plainly, in neither a Boussinesq fluid nor an ideal gas can *all* the total potential energy in a fluid be converted to kinetic energy, for then all of the fluid would be adjacent to the ground and the fluid would have no thickness. Given a state of the atmosphere or ocean, how much of its total potential energy is available for conversion to kinetic energy? In particular, because energy is conserved only in adiabatic flow, we may usefully ask: how much potential energy is available for conversion to kinetic energy under an adiabatic rearrangement of fluid parcels?

Suppose that at any given time the flow is stably stratified, but that the isentropes (or more generally the surfaces of constant potential density) are sloping, as in Fig. 3.12. The potential energy of the system would be reduced if the isentropes were flattened, for then heavier fluid would be moved to lower altitudes, with lighter fluid replacing it at higher altitudes. In an adiabatic rearrangement the amount of fluid between the isentropes would remain constant, and a state with flat isentropes (meaning parallel to the geopotential surfaces) evidently constitutes a state of minimum total potential energy. The difference between the total potential energy of the fluid and the

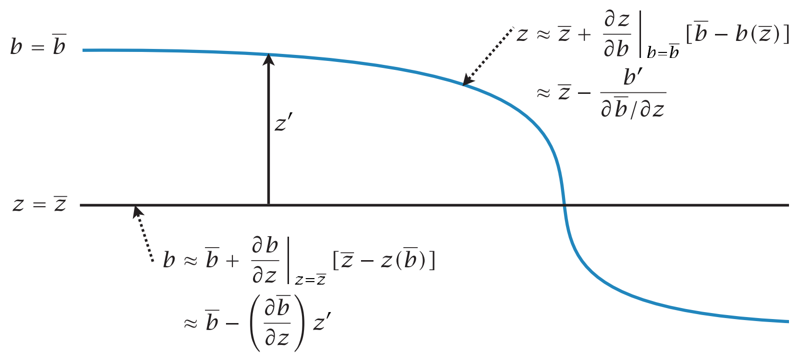


**Fig. 3.13** An isopycnal surface, $b = \overline{b}$, and the constant height surface, $z = \overline{z}$, where $\overline{z}$ is the height of the isopycnal surface after a rearrangement to a minimum potential energy state, equal to the average height of the isopycnal surface. The values of $z$ on the isopycnal surface, and of $b$ on the constant height surface, can be obtained by the Taylor expansions shown. For an ideal gas in pressure coordinates, replace $z$ by $p$ and $b$ by $\theta$.

total potential energy after an adiabatic rearrangement to a state in which the isentropic surfaces are flat is called the *available potential energy*, or APE.[11]

### 3.11.1 A Boussinesq Fluid

The potential energy of a column of a Boussinesq fluid of unit area is given by

$$P = -\int_0^H bz \, dz = -\int_0^H \frac{b}{2} \, dz^2. \tag{3.193}$$

and the potential energy of the entire fluid is given by the horizontal integral of this. The minimum potential energy of the fluid arises after an adiabatic rearrangement in which the isopycnals are flattened, and the resulting buoyancy is only a function of $z$. The available potential energy is then the difference between the energy of the initial state and of this minimum state, and to obtain an approximate expression for this we first integrate (3.193) by parts to give

$$P = \frac{1}{2}\int_0^{b_m} z^2 \, db - \left[\frac{bz^2}{2}\right]_0^H = \frac{1}{2}\int_0^{b_m} z^2 \, db - \frac{b_m H^2}{2}, \tag{3.194}$$

where $b_m$ is the maximum value of $b$ in the domain, and we may formally take the upper boundary to have this value of $b$ without affecting the final result. The minimum potential energy state arises when $z$ is a function only of $b$, $z = Z(b)$ say. Because mass is conserved in the rearrangement, $Z$ is equal to the horizontally averaged value of $z$ on a given isopycnal surface, $\overline{z}$, and the surfaces $\overline{z}$ and $\overline{b}$ thus define each other completely. The average available potential energy, per unit area, is then given by

$$\text{APE} = \frac{1}{2}\int_0^{b_m} (\overline{z^2} - \overline{z}^2) \, db = \frac{1}{2}\int_0^{b_m} \overline{z'^2} \, db, \tag{3.195}$$

where $z = \overline{z} + z'$; that is, $z'$ is the height variation of an isopycnal surface, and the last term on the right-hand side of (3.194) has cancelled with an identical term in the expression for the potential energy of the re-arranged state. The available potential energy is thus proportional to the integral of the variance of the altitude of such a surface, and it is a positive-definite quantity. To obtain an expression in $z$-coordinates, we express the height variations on an isopycnal surface in terms of buoyancy variations on a surface of constant height by Taylor-expanding the height about its value on the isopycnal surface. Referring to Fig. 3.13 this gives

$$z(\overline{b}) = \overline{z} + \left.\frac{\partial z}{\partial b}\right|_{b=\overline{b}} [\overline{b} - b(\overline{z})] = \overline{z} - \left.\frac{\partial z}{\partial b}\right|_{b=\overline{b}} b', \tag{3.196}$$

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

where $b' = b(\overline{z}) - \overline{b}$ is corresponding buoyancy perturbation on the $\overline{z}$ surface and $\overline{b}$ is the average value of $b$ on the $\overline{z}$ surface. Furthermore, $\partial z/\partial b|_{z=\overline{b}} \approx \partial\overline{z}/\partial b \approx (\partial\overline{b}/\partial z)^{-1}$, and (3.196) thus becomes

$$z' = z(\overline{b}) - \overline{z} \approx -b'\left(\frac{\partial\overline{z}}{\partial b}\right) \approx -\frac{b'}{(\partial\overline{b}/\partial z)}, \tag{3.197}$$

where $z' = z(b) - \overline{z}$ is the height perturbation of the isopycnal surface, from its average value. Using (3.197) in (3.195) we obtain an expression for the APE per unit area, to wit

$$\text{APE} \approx \frac{1}{2}\int_0^H \frac{\overline{b'^2}}{\partial\overline{b}/\partial z}\,\mathrm{d}z. \tag{3.198}$$

The total APE of the fluid is the horizontal integral of the above, and so is proportional to the variance of the buoyancy on a height surface. We emphasize that APE is not defined for a single column of fluid, for it depends on the variations of buoyancy over a horizontal surface. Note too that the derivation neglects the effects of topography; this, and the use of a basic-state stratification, effectively restrict the use of (3.198) to a single ocean basin, and even for that the approximations used limit the accuracy of the expressions.

### 3.11.2 An Ideal Gas

The expression for the APE for an ideal gas is obtained, *mutatis mutandis,* in the same way as for a Boussinesq fluid and the trusting reader may skip directly to (3.206). The internal energy of an ideal gas column of unit area is given by

$$I = \int_0^\infty c_v T\rho\,\mathrm{d}z = \int_0^{p_s}\frac{c_v}{g}T\,\mathrm{d}p, \tag{3.199}$$

where $p_s$ is the surface pressure, and the corresponding potential energy is given by

$$P = \int_0^\infty \rho g z\,\mathrm{d}z = \int_0^{p_s} z\,\mathrm{d}p = \int_0^\infty p\,\mathrm{d}z = \int_0^{p_s}\frac{R}{g}T\,\mathrm{d}p. \tag{3.200}$$

In (3.199) we use hydrostasy, and in (3.200) the equalities make successive use of hydrostasy, an integration by parts, hydrostasy and the ideal gas relation. Thus, the total potential energy (TPE) is given by

$$\text{TPE} \equiv I + P = \frac{c_p}{g}\int_0^{p_s} T\,\mathrm{d}p. \tag{3.201}$$

Using the ideal gas equation of state we can write this as

$$\text{TPE} = \frac{c_p}{g}\int_0^{p_s}\left(\frac{p}{p_s}\right)^\kappa \theta\,\mathrm{d}p = \frac{c_p p_s}{g(1+\kappa)}\int_0^\infty \left(\frac{p}{p_s}\right)^{\kappa+1}\mathrm{d}\theta, \tag{3.202}$$

after an integration by parts. (We omit a term proportional to $p_s\theta_s$ that arises in the integration by parts, because it cancels in a similar fashion to the boundary term in the Boussinesq derivation; or take $\theta_s = 0$.) The total potential energy of the entire fluid is equal to a horizontal integral of (3.202). The minimum total potential energy arises when the pressure in (3.202) is a function only of $\theta$, $p = P(\theta)$, where by conservation of mass $P$ is the average value of the original pressure on the isentropic surface, $P = \overline{p}$. The average available potential energy per unit area is then given by the difference between the initial state and this minimum, namely

$$\text{APE} = \frac{c_p p_s}{g(1+\kappa)}\int_0^\infty \left[\overline{\left(\frac{p}{p_s}\right)^{\kappa+1}} - \left(\frac{\overline{p}}{p_s}\right)^{\kappa+1}\right]\mathrm{d}\theta, \tag{3.203}$$

which is a positive-definite quantity. A useful approximation to this is obtained by expressing the right-hand side in terms of the variance of the potential temperature on a pressure surface. We first use the binomial expansion to expand $p^{\kappa+1} = (\overline{p} + p')^{\kappa+1}$. Neglecting third- and higher-order terms (3.203) becomes

$$\mathrm{APE} = \frac{R\overline{p}_s}{2g} \int_0^\infty \left(\frac{\overline{p}}{\overline{p}_s}\right)^{\kappa+1} \overline{\left(\frac{p'}{\overline{p}}\right)^2} \, \mathrm{d}\theta. \tag{3.204}$$

The variable $p' = p(\theta) - \overline{p}$ is a pressure perturbation on an isentropic surface, and is related to the potential temperature perturbation on an isobaric surface by [cf. (3.197)]

$$p' \approx -\theta' \frac{\partial \overline{p}}{\partial \theta} \approx -\frac{\theta'}{\partial \overline{\theta}/\partial p}, \tag{3.205}$$

where $\theta' = \theta(p) - \theta(\overline{p})$ is the potential temperature perturbation on the $\overline{p}$ surface. Using (3.205) in (3.204) we finally obtain

$$\boxed{\mathrm{APE} = \frac{R\overline{p}_s^{-\kappa}}{2} \int_0^{p_s} p^{\kappa-1} \left(-g\frac{\partial \overline{\theta}}{\partial p}\right)^{-1} \overline{\theta'^2} \, \mathrm{d}p.} \tag{3.206}$$

The APE is thus proportional to the variance of the potential temperature on the pressure surface or, from (3.204), proportional to the variance of the pressure on an isentropic surface.

### 3.11.3 Use and Interpretation

The potential energy of a fluid is reduced when the dynamics acts to flatten the isentropes. Consider, for example, Earth's atmosphere, with isentropes sloping upwards toward the pole (as in the left panel of Fig. 3.12 with the pole on the right). Flattening these isentropes amounts to a sinking of dense air and a rising of light air, and this reduction of potential energy leads to a corresponding production of kinetic energy. Thus, if the dynamics is such as to reduce the temperature gradient between equator and pole by flattening the isentropes then APE is converted to KE by that process. A statistically steady state is achieved because the heating from the Sun continually acts to restore the horizontal temperature gradient between equator and pole, thus replenishing the pool of APE, and to this extent the large-scale atmospheric circulation acts like a heat engine.

It is a useful exercise to calculate the total potential energy, the available potential energy and the kinetic energy of atmosphere and the ocean. One finds

$$\mathrm{TPE} \gg \mathrm{APE} > \mathrm{KE} \tag{3.207}$$

with, very approximately, TPE $\sim$ 100 APE and APE $\sim$ 10 KE. The first inequality should not surprise us (as it was this that led us to define APE in the first instance), but the second inequality is not obvious (and in fact the ratio is larger in the ocean). It is related to the fact that the instabilities of the atmosphere and ocean occur at a scale smaller than the size of the domain, and are unable to release all the potential energy that might be available. Understanding this more fully is the topic of Chapters 9 and 12.

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

## Notes

1 The algorithm to solve these equations numerically differs from that of the free-surface shallow water equations because the mass conservation equation can no longer be stepped forward in time. Rather, an elliptic equation for $p_{lid}$ must be derived by eliminating time derivatives between (3.21) using (3.20), and this is then solved at each timestep.

2 This correspondence was known to Matsuno (1966). Gill & Clarke (1974), McCreary (1985) and others also provide derivations of various kinds.

3 Chelton *et al.* (1998), who also provide maps of the first deformation radius and related quantities for the world's oceans.

4 After Margules (1903). Margules sought to relate the energy of fronts to their slope. In this same paper the notion of available potential energy arose.

5 'Form stress' is an expression derived from 'form drag', an expression commonly used in aerodynamics. In aerodynamics, form drag is the force due to the pressure difference between the front and rear of an object, or any other 'form', moving through a fluid. Aerodynamic form drag may, albeit uncommonly, also include frictional effects between the wind and the surface itself.

6 (Jules) Henri Poincaré (1854–1912) was a prodigious French mathematician, physicist and philosopher, certainly one of the greatest mathematicians living at the turn of the twentieth century. He is remembered for his original work in (among other things) algebra, topology, dynamical systems and celestial mechanics, obtaining many results in what would be called nonlinear dynamics and chaos when these fields re-emerged some 60 years later — the notion of 'sensitive dependence on initial conditions', for example, is present in his work. He obtained a number of the results of special relativity independently of Einstein, and worked on the theory of rotating fluids — hence the Poincaré waves of this chapter. He also wrote extensively and successfully for the general public on the meaning, importance and philosophy of science. Among other things he discussed whether scientific knowledge was an arbitrary convention, a notion that remains discussed and controversial to this day. (His answer: 'convention', in part, yes; 'arbitrary', no.) He was a proponent of the role of intuition in mathematical and scientific progress, and did not believe that mathematics could ever be wholly reduced to pure logic.

7 Thomson (1869). William Thompson later became Lord Kelvin.

8 As was considered by Rossby (1938).

9 The code (available from the author's web site) will also reproduce Fig. 3.9.

10 An introduction to variational problems may be found in Weinstock (1952) and a number of other textbooks. Applications to many traditional problems in mechanics are discussed by Lanczos (1970).

11 Margules (1903) introduced the concept of potential energy that is available for conversion to kinetic energy, Lorenz (1955) clarified its meaning and derived useful, approximate formulae for its computation, and there has since been a host of papers on the subject. Thus, for example, Shepherd (1993) showed that the APE is just the non-kinetic part of the pseudoenergy, Huang (1998) looked at some of the limitations of the approximate expressions in an oceanic context, and on the atmospheric side Pauluis (2007) looked at the effects of moisture.

In addition to his formulation of available potential energy, Edward Lorenz (1917–2008) made enormous contributions to the atmospheric sciences over the course of a long career spent almost entirely at MIT. He is perhaps most famous for being one of the modern founders of chaos theory as it emerged in the 1960s, and his paper *Deterministic non-periodic flow*, published in the *Journal of the Atmospheric Sciences* in 1963, was not only a watershed in meteorology but it changed the way we think about irregular systems — see also the endnotes on page 443. Lorenz also introduced the idea of empirical orthogonal functions to meteorology and wrote with clarity and insight about the atmospheric general circulation in a monograph in 1967.

In common with a number of meteorologists of his generation (Eric Eady was another), Lorenz was first educated in mathematics and then, because of World War II, was trained as a weather forecaster. After the war he moved to MIT in 1946 for a PhD, and in 1953 was hired on the MIT faculty,

https://doi.org/10.1017/9781107588417.004 Published online by Cambridge University Press

*Strange as it may sound, the power of mathematics rests on its evasion of all unnecessary thought and on its wonderful saving of mental operations.*
Ernst Mach (1838–1916), quoted in Bell (1937).

# CHAPTER 4

# Vorticity and Potential Vorticity

VORTICITY AND POTENTIAL VORTICITY both play a central role in geophysical fluid dynamics, especially in the dynamics of the large scale circulation. In this chapter we define and discuss these quantities and deduce some of their dynamical properties and effects. Along the way we will come across *Kelvin's circulation theorem,* one of the most fundamental conservation laws in all of fluid mechanics, to which the conservation of potential vorticity is intimately tied.

## 4.1 VORTICITY AND CIRCULATION

### 4.1.1 Preliminaries

*Vorticity,* $\boldsymbol{\omega}$, is defined to be the curl of velocity and so is given by

$$\boldsymbol{\omega} \equiv \nabla \times \boldsymbol{v}. \tag{4.1}$$

*Circulation,* $C$, is defined to be the integral of velocity around a closed fluid loop and so is given by

$$C \equiv \oint \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r} = \int_S \boldsymbol{\omega} \cdot \mathrm{d}\boldsymbol{S}, \tag{4.2}$$

where the second expression uses Stokes' theorem and $S$ is any surface bounded by the loop. The circulation around the path is equal to the integral of the normal component of vorticity over *any* surface bounded by that path. The circulation is not a field like vorticity and velocity; rather, we think of the circulation around a particular material line of finite length, and so its value generally depends on the path chosen. If $\delta S$ is an infinitesimal surface element whose normal points in the direction of the unit vector $\hat{\boldsymbol{n}}$, then

$$\hat{\boldsymbol{n}} \cdot (\nabla \times \boldsymbol{v}) = \frac{1}{\delta S} \oint_{\delta r} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r}, \tag{4.3}$$

where the line integral is around the infinitesimal area. Thus at a point the component of vorticity in the direction of $\boldsymbol{n}$ is proportional to the circulation around the surrounding infinitesimal fluid element, divided by the elemental area bounded by the path of the integral. A heuristic test for the presence of vorticity is to imagine a small paddle wheel in the flow; the paddle wheel acts as a 'circulation-meter', and rotates if the vorticity is non-zero. Vorticity might seem to be similar



https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

to angular momentum, in that it is a measure of spin. However, unlike angular momentum, *the value of vorticity at a point does not depend on the particular choice of an axis of rotation;* indeed, the definition of vorticity makes no reference at all to an axis of rotation or to a coordinate system. Rather, vorticity is a measure of the *local* spin of a fluid element.

### 4.1.2 Simple Axisymmetric Examples

Consider axisymmetric motion in two dimensions, so that the flow is confined to a plane. We use cylindrical coordinates $(r, \phi, z)$, where $z$ is the direction perpendicular to the plane, with velocity components $(u^r, u^\phi, u^z)$. For axisymmetric flow $u^z = u^r = 0$ but $u^\phi \neq 0$. The following two examples are quite instructive. (A third example that the reader may wish to consider is solid body rotation on a sphere, which has a vorticity gradient in latitude.)

#### *Rigid body motion*

For a body in rigid body rotation, the velocity distribution is given by

$$u^\phi = \Omega r, \tag{4.4}$$

where $\Omega$ is the angular velocity of the fluid and $r$ is the distance from the axis of rotation. Associated with this rotation is a vorticity given by

$$\boldsymbol{\omega} = \nabla \times \boldsymbol{v} = \omega^z \mathbf{k}, \tag{4.5}$$

where

$$\omega^z = \frac{1}{r}\frac{\partial}{\partial r}(r u^\phi) = \frac{1}{r}\frac{\partial}{\partial r}(r^2 \Omega) = 2\Omega. \tag{4.6}$$

The vorticity of a fluid in solid body rotation is thus twice the angular velocity of the fluid about the axis of rotation, and is pointed in a direction orthogonal to the plane of rotation.

#### *The ‘$v r$’ vortex*

This vortex is so-called because the tangential velocity (historically denoted by ‘$v$’ in this context) is such that the product $vr$ is constant. In our notation we would have

$$u^\phi = \frac{K}{r}, \tag{4.7}$$

where $K$ is a constant determining the vortex strength. Evaluating the $z$-component of vorticity gives

$$\omega^z = \frac{1}{r}\frac{\partial}{\partial r}(r u^\phi) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{K}{r}\right) = 0, \tag{4.8}$$

except where $r = 0$, at which the expression is singular and the vorticity is infinite. Our paddle wheel rotates when placed at the vortex center, but, less obviously, does not if placed elsewhere.

The circulation around a circle that encloses the origin is given by

$$C = \oint \frac{K}{r} r \, \mathrm{d}\phi = 2\pi K. \tag{4.9}$$

This does not depend on the radius, and so it is true even as the radius tends to zero. Since the vorticity is the circulation divided by the area, the vorticity at the origin must be infinite. Consider now an integration path that does *not* enclose the origin, for example the contour $A$–$B$–$C$–$D$–$A$ in Fig. 4.1. Over the segments $A$–$B$ and $C$–$D$ the velocity is orthogonal to the contour, and so the contribution is zero. Over $B$–$C$ and $D$–$A$ we have

$$C_{BC} = \frac{K}{r_2}\phi r_2 = K\phi, \qquad C_{DA} = -\frac{K}{r_1}\phi r_1 = -K\phi. \tag{4.10}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

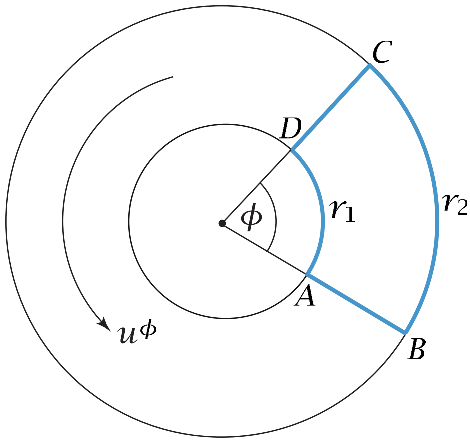


**Fig. 4.1** Evaluation of circulation in the axisymmetric $vr$ vortex. The circulation around the path $A$–$B$–$C$–$D$ is zero. This result does not depend on the radii $r_1$ or $r_2$ or the angle $\phi$, and the circulation around any infinitesimal path not enclosing the origin is zero. Thus the vorticity is zero everywhere except at the origin.

Adding these two expressions we see that the net circulation around the contour $C_{ABCDA}$ is zero. If we shrink the integration path to an infinitesimal size then, within the path, by Stokes' theorem, the vorticity is zero. We can of course place the path anywhere we wish, except surrounding the origin, and obtain this result. Thus the vorticity is everywhere zero, except at the origin.

## 4.2 THE VORTICITY EQUATION

Using the vector identity $\boldsymbol{v} \times (\nabla \times \boldsymbol{v}) = \nabla(\boldsymbol{v} \cdot \boldsymbol{v})/2 - (\boldsymbol{v} \cdot \nabla)\boldsymbol{v}$, we write the momentum equation as

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{\omega} \times \boldsymbol{v} = -\frac{1}{\rho}\nabla p - \frac{1}{2}\nabla \boldsymbol{v}^2 + \boldsymbol{F}, \tag{4.11}$$

where $\boldsymbol{F}$ represents viscous and body forces. Taking the curl of (4.11) gives the vorticity equation

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + \nabla \times (\boldsymbol{\omega} \times \boldsymbol{v}) = \frac{1}{\rho^2}(\nabla \rho \times \nabla p) + \nabla \times \boldsymbol{F}. \tag{4.12}$$

Now, the vector identity

$$\nabla \times (\boldsymbol{a} \times \boldsymbol{b}) = (\boldsymbol{b} \cdot \nabla)\boldsymbol{a} - (\boldsymbol{a} \cdot \nabla)\boldsymbol{b} + \boldsymbol{a}\nabla \cdot \boldsymbol{b} - \boldsymbol{b}\nabla \cdot \boldsymbol{a}, \tag{4.13}$$

implies that the second term on the left-hand side of (4.12) may be written as

$$\nabla \times (\boldsymbol{\omega} \times \boldsymbol{v}) = (\boldsymbol{v} \cdot \nabla)\boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \nabla)\boldsymbol{v} + \boldsymbol{\omega}\nabla \cdot \boldsymbol{v} - \boldsymbol{v}\nabla \cdot \boldsymbol{\omega}. \tag{4.14}$$

Because vorticity is the curl of velocity its divergence vanishes, and so (4.12) becomes

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{\omega} = (\boldsymbol{\omega} \cdot \nabla)\boldsymbol{v} - \boldsymbol{\omega}\nabla \cdot \boldsymbol{v} + \frac{1}{\rho^2}(\nabla \rho \times \nabla p) + \nabla \times \boldsymbol{F}. \tag{4.15}$$

The divergence term may be eliminated with the aid of the mass-conservation equation to give

$$\boxed{\frac{\mathrm{D}\widetilde{\boldsymbol{\omega}}}{\mathrm{D}t} = (\widetilde{\boldsymbol{\omega}} \cdot \nabla)\boldsymbol{v} + \frac{1}{\rho^3}(\nabla \rho \times \nabla p) + \frac{1}{\rho}\nabla \times \boldsymbol{F},} \tag{4.16}$$

where $\widetilde{\boldsymbol{\omega}} \equiv \boldsymbol{\omega}/\rho$. We will set $\boldsymbol{F} = 0$ in most of what follows.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

The third term on the right-hand side of (4.15), as well as the second term on the right-hand side of (4.16), is variously called the *baroclinic* term, the *non-homentropic* term, or the *solenoidal* term. (A solenoidal vector has no divergence, hence the name.) The solenoidal vector, $\boldsymbol{S}_o$, is defined by

$$\boldsymbol{S}_o \equiv \frac{1}{\rho^2}\nabla\rho \times \nabla p = -\nabla\alpha \times \nabla p. \tag{4.17}$$

A solenoid is a tube directed perpendicular to both $\nabla\alpha$ and $\nabla p$, with elements of length proportional to $\nabla p \times \nabla\alpha$. If the isolines of $p$ and $\alpha$ are parallel to each other, then solenoids do not exist. This occurs when the density is a function only of pressure, for then

$$\nabla\rho \times \nabla p = \nabla\rho \times \nabla\rho \,\frac{\mathrm{d}p}{\mathrm{d}\rho} = 0. \tag{4.18}$$

The solenoidal vector may also be written

$$\boldsymbol{S}_o = -\nabla\eta \times \nabla T. \tag{4.19}$$

This follows most easily by first writing the momentum equation in the form $\partial\boldsymbol{v}/\partial t + \boldsymbol{\omega} \times \boldsymbol{v} = T\nabla\eta - \nabla B$, and taking its curl. Evidently the solenoidal term vanishes if: (i) isolines of pressure and density are parallel; (ii) isolines of temperature and entropy are parallel; (iii) density, entropy, temperature or pressure are constant. A *barotropic* fluid has by definition $\rho = \rho(p)$ and therefore no solenoids. A *baroclinic* fluid is one for which $\nabla p$ is not parallel to $\nabla\rho$. From (4.16) we see that the baroclinic term must be balanced by terms involving velocity or its tendency and therefore, in general, *a baroclinic fluid is a moving fluid,* even in the presence of viscosity.

For a barotropic fluid the vorticity equation takes the simple form,

$$\frac{\mathrm{D}\tilde{\boldsymbol{\omega}}}{\mathrm{D}t} = (\tilde{\boldsymbol{\omega}} \cdot \nabla)\boldsymbol{v}. \tag{4.20}$$

If the fluid is also incompressible, meaning that $\nabla \cdot \boldsymbol{v} = 0$, then we have the even simpler form,

$$\frac{\mathrm{D}\boldsymbol{\omega}}{\mathrm{D}t} = (\boldsymbol{\omega} \cdot \nabla)\boldsymbol{v}. \tag{4.21}$$

When expanded into components, the terms on the right-hand side of (4.20) or (4.21) can be divided into 'stretching' and 'tipping' (or 'tilting') terms, and we return to that in Section 4.3.1.

### *An integral conservation property*

Consider a single Cartesian component in (4.15). Then, using superscripts to denote components,

$$\begin{aligned}\frac{\partial\omega^x}{\partial t} &= -\boldsymbol{v} \cdot \nabla\omega^x - \omega^x\nabla \cdot \boldsymbol{v} + (\boldsymbol{\omega} \cdot \nabla)v^x + S_o^x \\ &= -\nabla \cdot (\boldsymbol{v}\omega^x) + \nabla \cdot (\boldsymbol{\omega}v^x) + S_o^x,\end{aligned} \tag{4.22}$$

where $S_o^x$ is the ($x$-component of the) solenoidal term. Equation (4.22) may be written as

$$\frac{\partial\omega^x}{\partial t} + \nabla \cdot (\boldsymbol{v}\omega^x - \boldsymbol{\omega}v^x) = S_o^x, \tag{4.23}$$

and this implies the Cartesian tensor form of the vorticity equation, namely

$$\frac{\partial\omega_i}{\partial t} + \frac{\partial}{\partial x_j}(v_j\omega_i - v_i\omega_j) = S_{oi}, \tag{4.24}$$

with summation over repeated indices. The tendency of the components of vorticity is thus given by the solenoidal term plus the divergence of a vector field, and if the solenoidal term vanishes the volume integrated vorticity can only be altered by boundary effects. However, in both the atmosphere and the ocean the solenoidal term *is* important, although we will see in Section 4.5 that a useful conservation law for a scalar quantity can still be obtained.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

### 4.2.1 Two-dimensional Flow

In two-dimensional flow the fluid is confined to a surface, and independent of the dimension normal to that surface. In the simplest case in Cartesian geometry the flow is on a flat plane, and the velocity normal to the plane and the rate of change of any quantity normal to that plane are zero. Let the normal direction be the $z$-direction; the fluid velocity in the plane, $\boldsymbol{u}$, is $\boldsymbol{u} = u\mathbf{i} + v\mathbf{j}$, and the velocity normal to the plane, $w$, is zero. Only one component of vorticity is non-zero and this is

$$\boldsymbol{\omega} = \mathbf{k}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right). \tag{4.25}$$

That is, in two-dimensional flow the vorticity is perpendicular to the velocity. We let $\zeta \equiv \omega^z = \boldsymbol{\omega}\cdot\mathbf{k}$. Both the stretching and tilting terms vanish in two-dimensional flow, and the two-dimensional vorticity equation becomes, for incompressible flow,

$$\frac{\mathrm{D}\zeta}{\mathrm{D}t} = 0, \tag{4.26}$$

where $\mathrm{D}\zeta/\mathrm{D}t = \partial\zeta/\partial t + \boldsymbol{u}\cdot\nabla\zeta$. That is, in two-dimensional flow vorticity is conserved following the fluid elements; each material parcel of fluid keeps its value of vorticity even as it is being advected around. Furthermore, specification of the vorticity completely determines the flow field. To see this, we use the incompressibility condition to define a streamfunction $\psi$ such that

$$u = -\frac{\partial \psi}{\partial y}, \qquad v = \frac{\partial \psi}{\partial x}, \qquad \zeta = \nabla^2\psi. \tag{4.27a,b,c}$$

Given the vorticity, the Poisson equation (4.27c) can be solved for the streamfunction and the velocity fields obtained through (4.27a,b), and this process is called 'inverting the vorticity'.

Numerical integration of (4.26) is then a process of timestepping plus inversion. The vorticity equation may then be written as an advection equation for vorticity,

$$\frac{\partial \zeta}{\partial t} + \boldsymbol{u}\cdot\nabla\zeta = 0, \tag{4.28}$$

in conjunction with (4.27). The vorticity is stepped forward one timestep using a finite-difference representation of (4.28), and the vorticity inverted to obtain a velocity using (4.27).

Two-dimensional flow is not restricted to a Cartesian plane — it exists on the surface of a sphere for example. In that case the velocity normal to the spherical surface (the 'vertical velocity') vanishes, and the equations are naturally expressed in spherical coordinates. Nevertheless, vorticity (absolute vorticity if the sphere is rotating) is still conserved on parcels as they move over the spherical surface.

## 4.3 VORTICITY AND CIRCULATION THEOREMS

### 4.3.1 The 'Frozen-in' Property of Vorticity

Let us first consider some simple topological properties of the vorticity field and its evolution. We define a *vortex line* to be a line drawn through the fluid which is everywhere in the direction of the local vorticity. This definition is analogous to that of a streamline, which is everywhere in the direction of the local velocity. A *vortex tube* is formed by the collection of vortex lines passing through a closed curve (Fig. 4.2). A *material line* is just a line that connects material fluid elements. Suppose we draw a vortex line through the fluid; such a line obviously connects fluid elements and therefore defines a coincident material line. As the fluid moves the material line deforms, and the vortex line also evolves in a manner determined by the equations of motion. A remarkable

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

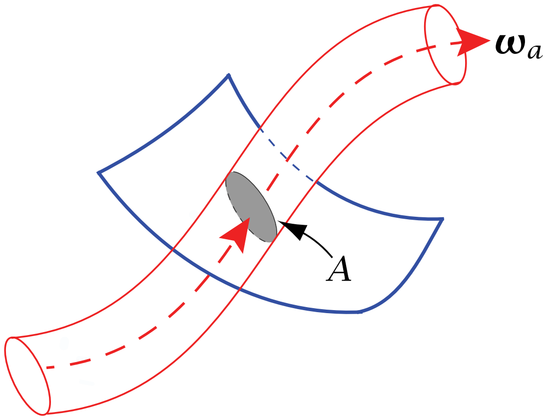


**Fig. 4.2** A vortex tube passing through a material sheet. The circulation is the integral of the velocity around the boundary of $A$, and is equal to the integral of the normal component of vorticity over $A$.

property of vorticity is that, for an unforced and inviscid barotropic fluid, the flow evolution is such that a vortex line remains coincident with the material line that it was initially associated with. Put another way, a vortex line always contains the same material elements — the vorticity is 'frozen' or 'glued' to the material fluid.[1]

To prove this we consider how an infinitesimal material line element $\delta\boldsymbol{l}$ evolves, $\delta\boldsymbol{l}$ being the infinitesimal material element connecting $\boldsymbol{l}$ with $\boldsymbol{l}+\delta\boldsymbol{l}$. The rate of change of $\delta\boldsymbol{l}$ following the flow is given by

$$\frac{\mathrm{D}\delta\boldsymbol{l}}{\mathrm{D}t} = \frac{1}{\delta t}\left[\delta\boldsymbol{l}(t+\delta t) - \delta\boldsymbol{l}(t)\right], \tag{4.29}$$

which follows from the definition of the material derivative in the limit $\delta t \to 0$. From the Taylor expansion of $\delta\boldsymbol{l}(t)$ and the definition of velocity it is also apparent that

$$\delta\boldsymbol{l}(t+\delta t) = \boldsymbol{l}(t) + \delta\boldsymbol{l}(t) + (\boldsymbol{v}+\delta\boldsymbol{v})\delta t - (\boldsymbol{l}(t) + \boldsymbol{v}\delta t) = \delta\boldsymbol{l} + \delta\boldsymbol{v}\,\delta t, \tag{4.30}$$

as illustrated in Fig. 4.3. Substituting (4.30) into (4.29) gives $\mathrm{D}\delta\boldsymbol{l}/\mathrm{D}t = \delta\boldsymbol{v}$, as expected, and because $\delta\boldsymbol{v} = (\delta\boldsymbol{l}\cdot\nabla)\boldsymbol{v}$ we obtain

$$\frac{\mathrm{D}\delta\boldsymbol{l}}{\mathrm{D}t} = (\delta\boldsymbol{l}\cdot\nabla)\boldsymbol{v}. \tag{4.31}$$

Comparing this with (4.16), we see that vorticity evolves in the same way as a line element in an unforced barotropic fluid. To see what this means, at some initial time we can define an infinitesimal material line element parallel to the vorticity at that location, that is,

$$\delta\boldsymbol{l}(\boldsymbol{x}, t=0) = A\boldsymbol{\omega}(\boldsymbol{x}, t=0), \tag{4.32}$$

where $A$ is a constant. Then, for all subsequent times the magnitude of the vorticity of that fluid element, even as it moves to a new location $\boldsymbol{x}'$, remains proportional to the length of the fluid element at that point and is oriented in the same way; that is $\boldsymbol{\omega}(\boldsymbol{x}',t) = A^{-1}\delta\boldsymbol{l}(\boldsymbol{x}',t)$.

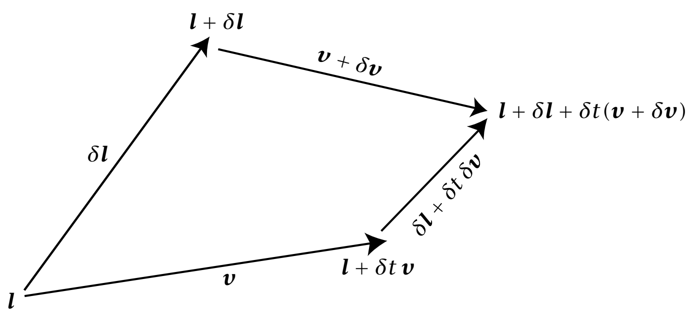


**Fig. 4.3** Evolution of an infinitesimal material line $\delta\boldsymbol{l}$ from time $t$ to time $t+\delta t$. It can be seen from the diagram that $\mathrm{D}\delta\boldsymbol{l}/\mathrm{D}t = \delta\boldsymbol{v}$.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

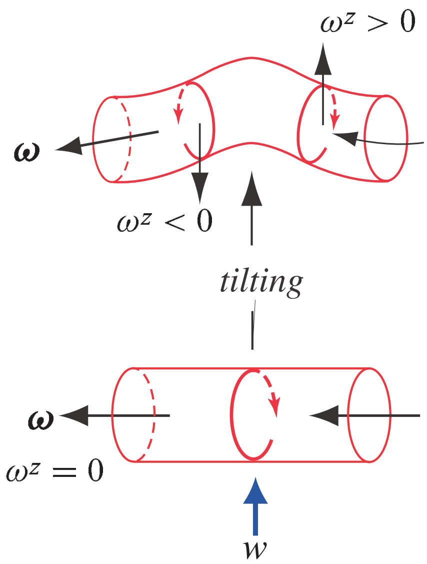


**Fig. 4.4** The tilting of vorticity. Suppose that the vorticity, $\boldsymbol{\omega}$ is initially directed horizontally, as in the lower figure, so that $\omega^z$, its vertical component, is zero. The material lines, and therefore also the vortex lines, are tilted by the positive vertical velocity $w$ thus creating a non-zero vertically oriented vorticity. This mechanism is important in creating vertical vorticity in the atmospheric boundary layer, and is connected to the $\beta$-effect in large-scale flow.

To verify this result in a different way note that a vortex line element is determined by the condition $\delta\boldsymbol{l} = A\boldsymbol{\omega}$ which implies $\boldsymbol{\omega} \times \delta\boldsymbol{l} = 0$. Now, for any line element we have that

$$\frac{\mathrm{D}}{\mathrm{D}t}(\boldsymbol{\omega} \times \delta\boldsymbol{l}) = \frac{\mathrm{D}\boldsymbol{\omega}}{\mathrm{D}t} \times \delta\boldsymbol{l} - \frac{\mathrm{D}\delta\boldsymbol{l}}{\mathrm{D}t} \times \boldsymbol{\omega}. \tag{4.33}$$

We also have that

$$\frac{\mathrm{D}\delta\boldsymbol{l}}{\mathrm{D}t} = \delta\boldsymbol{v} = (\delta\boldsymbol{l} \cdot \nabla)\boldsymbol{v} \quad \text{and} \quad \frac{\mathrm{D}\boldsymbol{\omega}}{\mathrm{D}t} = (\boldsymbol{\omega} \cdot \nabla)\boldsymbol{v}. \tag{4.34}$$

If the line element is initially a vortex line element then, at $t = 0$, $\delta\boldsymbol{l} = A\boldsymbol{\omega}$ and, using (4.34), the right-hand side of (4.33) vanishes. Thus, the *tendency* of $\boldsymbol{\omega}\times\delta\boldsymbol{l}$ is zero, and the vortex line continues to be a material line.

### *Stretching and tilting*

The terms on the right-hand side of (4.20) or (4.21) may be interpreted in terms of 'stretching' and 'tipping' (or 'tilting'). Consider a single Cartesian component of (4.21),

$$\frac{\mathrm{D}\omega^x}{\mathrm{D}t} = \omega^x \frac{\partial u}{\partial x} + \omega^y \frac{\partial u}{\partial y} + \omega^z \frac{\partial u}{\partial z}. \tag{4.35}$$

The second and third terms on the right-hand side are the tilting or tipping terms because they involve changes in the orientation of the vorticity vector. They tell us that vorticity in the $x$-direction may be generated from vorticity in the $y$- and $z$-directions if the advection acts to tilt the material lines. Because vorticity is tied to these lines, vorticity oriented in one direction becomes oriented in another, as in Fig. 4.4.

The first term on the right-hand side of (4.35) is the stretching term, and it acts to intensify the $x$-component of vorticity if the velocity is increasing in the $x$-direction — that is, if the material lines are being stretched (Fig. 4.5). The effect arises because a vortex line is tied to a material line, and therefore vorticity is amplified in proportion to the stretching of the material line aligned with it. This effect is important in tornadoes, to give one example. If the fluid is incompressible, stretching of a fluid mass in one direction must be accompanied by convergence in another, and this leads to the conservation of circulation, as we now discuss.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

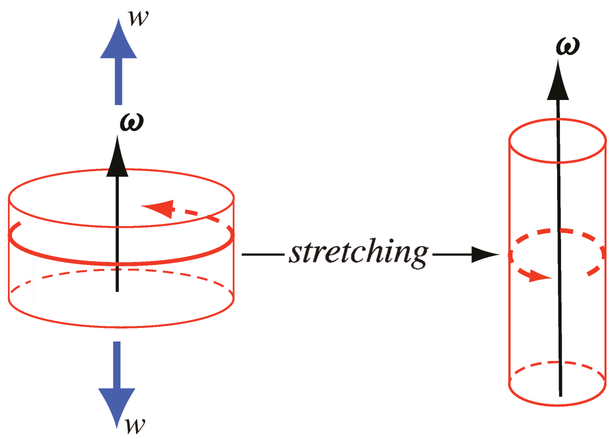


**Fig. 4.5** A vertical velocity, $w$, stretches the cylinder. Vorticity is tied to material lines and so is amplified in the direction of the stretching. However, because the volume of fluid is conserved, the end surfaces shrink, the material lines through the cylinder ends converge and the integral of vorticity over a material surface (the circulation) remains constant.

### 4.3.2 Kelvin's Circulation Theorem

Kelvin's circulation theorem states that under certain circumstances the circulation around a material fluid parcel is conserved; that is, the circulation is conserved 'moving with the flow'.[2] The primary restrictions are that body forces are conservative (i.e., they are representable as potential forces, and therefore that the flow be inviscid), and that the fluid is barotropic with $\rho = \rho(p)$. Of these, the latter is more restrictive for geophysical fluids. The circulation in the theorem is defined with respect to an inertial frame of reference; specifically, the velocity in (4.39) is the velocity relative to an inertial frame. To prove the theorem, we begin with the inviscid momentum equation,

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\frac{1}{\rho}\nabla p - \nabla\Phi, \tag{4.36}$$

where $\nabla\Phi$ represents the conservative body forces on the system. Applying the material derivative to the circulation, (4.2), gives

$$\begin{aligned}\frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\mathrm{D}}{\mathrm{D}t}\oint \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} &= \oint\left(\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t}\cdot\mathrm{d}\boldsymbol{r} + \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{v}\right)\\ &= \oint\left[\left(-\frac{1}{\rho}\nabla p - \nabla\Phi\right)\cdot\mathrm{d}\boldsymbol{r} + \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{v}\right]\\ &= \oint -\frac{1}{\rho}\nabla p\cdot\mathrm{d}\boldsymbol{r},\end{aligned} \tag{4.37}$$

using (4.36) and $D(\mathrm{d}\boldsymbol{r})/Dt = \mathrm{d}\boldsymbol{v}$, where $\mathrm{d}\boldsymbol{r}$ is the line element and with the line integration being over a closed, material, circuit. The second and third terms on the second line vanish separately, because they are exact differentials integrated around a closed loop. The term on the last line vanishes if the density is constant or, more generally, if the density is a function of pressure alone, in which case $\nabla p$ is parallel to $\nabla\rho$. To see this, note that

$$\oint \frac{1}{\rho}\nabla p\cdot\mathrm{d}\boldsymbol{r} = \int_S \nabla\times\left(\frac{\nabla p}{\rho}\right)\cdot\mathrm{d}\boldsymbol{S} = \int_S \frac{-\nabla\rho\times\nabla p}{\rho^2}\cdot\mathrm{d}\boldsymbol{S}, \tag{4.38}$$

using Stokes' theorem where $S$ is any surface bounded by the path of the line integral. The integral evidently vanishes identically if $p$ is a function of $\rho$ alone. The right-most expression above is the integral of the solenoidal vector, and if it is zero (4.37) becomes

$$\boxed{\frac{\mathrm{D}}{\mathrm{D}t}\oint \boldsymbol{v}\cdot\mathrm{d}\boldsymbol{r} = 0.} \tag{4.39}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

This is Kelvin's circulation theorem. In words, *the circulation around a material loop is invariant for a barotropic fluid that is subject only to conservative forces.* Using Stokes' theorem, the circulation theorem may also be written as

$$\boxed{\frac{\mathrm{D}}{\mathrm{D}t}\int_S \boldsymbol{\omega}\cdot \mathrm{d}\boldsymbol{S} = 0.} \tag{4.40}$$

That is, the area integral of the normal component of vorticity across any material surface is constant, under the same conditions. This form is both natural and useful, and it arises because of the way vorticity is tied to material fluid elements. Kelvin's circulation theorem is the one conservation law that is unique to fluids. Unlike, say, the conservation of energy, it has no analogue in solid body mechanics. Potential vorticity conservation, which we come to later on, is an extension of circulation conservation.

### *Stretching and circulation*

Let us informally consider how vortex stretching and mass conservation work together to give the circulation theorem. Let the fluid be incompressible so that the volume of a fluid mass is constant, and consider a surface normal to a vortex tube, as in Fig. 4.5. Let the volume of a small material box around the surface be $\delta V$, the length of the material lines be $\delta l$ and the surface area be $\delta A$. Then

$$\delta V = \delta l\,\delta A. \tag{4.41}$$

Because of the frozen-in property, the vorticity passing through the surface is proportional to the length of the material lines. That is, $\omega \propto \delta l$, and

$$\delta V \propto \omega\,\delta A. \tag{4.42}$$

The right-hand side is just the circulation around the surface. Now, if the corresponding material tube is stretched $\delta l$ increases, but the volume, $\delta V$, remains constant by mass conservation. Thus, the circulation given by the right-hand side of (4.42) also remains constant. In other words, because of the frozen-in property vorticity is amplified by the stretching, but the vortex lines get closer together in such a way that the product $\omega\,\delta A$ remains constant and circulation is conserved.

## 4.3.3 Baroclinic Flow and the Solenoidal Term

In baroclinic flow, the circulation is not generally conserved, and from (4.37) we have

$$\frac{\mathrm{D}C}{\mathrm{D}t} = -\oint \frac{\nabla p}{\rho}\cdot \mathrm{d}\boldsymbol{r} = -\oint \frac{\mathrm{d}p}{\rho}, \tag{4.43}$$

and this is called the baroclinic circulation theorem.[3] Recalling the fundamental thermodynamic relation $T\,\mathrm{d}\eta = \mathrm{d}I + p\,\mathrm{d}\alpha$, where $\alpha = \rho^{-1}$, we have

$$\alpha\,\mathrm{d}p = \mathrm{d}(p\alpha) - T\,\mathrm{d}\eta + \mathrm{d}I, \tag{4.44}$$

and the first and last terms on the right-hand side will vanish upon integration around a circuit. The solenoidal term on the right-hand side of (4.43) may therefore be written as

$$S_o \equiv -\oint \alpha\,\mathrm{d}p = \oint T\,\mathrm{d}\eta = -\oint \eta\,\mathrm{d}T = -R\oint T\,\mathrm{d}\log p, \tag{4.45}$$

where the last equality holds only for an ideal gas. Using Stokes' theorem, $S_o$ can also be written as

$$S_o = -\int_S \nabla\alpha\times\nabla p\cdot \mathrm{d}\boldsymbol{S} = -\int_S \left(\frac{\partial \alpha}{\partial T}\right)_p \nabla T\times\nabla p\cdot \mathrm{d}\boldsymbol{S} = \int_S \nabla T\times\nabla\eta\cdot \mathrm{d}\boldsymbol{S}. \tag{4.46}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

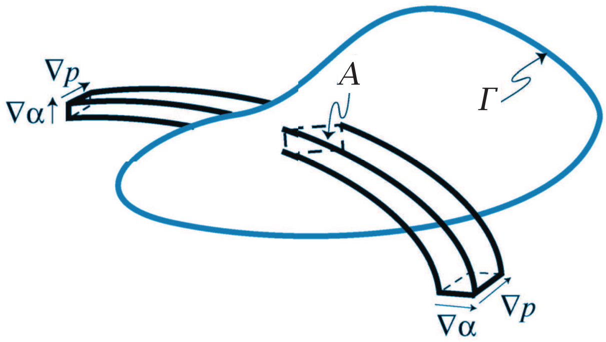


**Fig. 4.6** Solenoids and the circulation theorem. Solenoids are tubes perpendicular to both $\nabla\alpha$ and $\nabla p$, and they have a non-zero cross-sectional area if isolines of $\alpha$ and $p$ do not coincide. The rate of change of circulation over a material surface is given by the sum of all the solenoidal areas crossing the surface. If $\nabla\alpha \times \nabla p = 0$ there are no solenoids.

The rate of change of the circulation across a surface depends on the existence of this solenoidal term (Fig. 4.6). However, even if the solenoidal vector is in non-zero, circulation *is* conserved if the material path is in a surface of constant entropy, $\eta$, *and if* $\mathrm{D}\eta/\mathrm{D}t = 0$. The solenoidal term then vanishes and, because $\mathrm{D}\eta/\mathrm{D}t = 0$, entropy remains constant on that same material loop as it evolves. This result gives rise to the conservation of potential vorticity, discussed in Section 4.5.

### 4.3.4 Circulation in a Rotating Frame

The absolute and relative velocities are related by $\boldsymbol{v}_a = \boldsymbol{v}_r + \boldsymbol{\Omega} \times \boldsymbol{r}$, so that in a rotating frame the rate of change of circulation is given by

$$\frac{\mathrm{D}}{\mathrm{D}t}\oint(\boldsymbol{v}_r + \boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{r} = \oint\left[\left(\frac{\mathrm{D}\boldsymbol{v}_r}{\mathrm{D}t} + \boldsymbol{\Omega}\times\boldsymbol{v}_r\right)\cdot\mathrm{d}\boldsymbol{r} + (\boldsymbol{v}_r + \boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{v}_r\right]. \tag{4.47}$$

But $\oint \boldsymbol{v}_r \cdot \mathrm{d}\boldsymbol{v}_r = 0$ and, integrating by parts,

$$\begin{aligned}\oint(\boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{v}_r &= \oint\left\{\mathrm{d}[(\boldsymbol{\Omega}\times\boldsymbol{r})\cdot\boldsymbol{v}_r] - (\boldsymbol{\Omega}\times\mathrm{d}\boldsymbol{r})\cdot\boldsymbol{v}_r\right\}\\ &= \oint\left\{\mathrm{d}[(\boldsymbol{\Omega}\times\boldsymbol{r})\cdot\boldsymbol{v}_r] + (\boldsymbol{\Omega}\times\boldsymbol{v}_r)\cdot\mathrm{d}\boldsymbol{r}\right\}.\end{aligned} \tag{4.48}$$

The first term on the right-hand side is zero and so (4.47) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\oint(\boldsymbol{v}_r + \boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{r} = \oint\left(\frac{\mathrm{D}\boldsymbol{v}_r}{\mathrm{D}t} + 2\boldsymbol{\Omega}\times\boldsymbol{v}_r\right)\cdot\mathrm{d}\boldsymbol{r} = -\oint\frac{\mathrm{d}p}{\rho}, \tag{4.49}$$

where the second equality uses the momentum equation. The last term vanishes if the fluid is barotropic, and if so the circulation theorem is, unsurprisingly,

$$\frac{\mathrm{D}}{\mathrm{D}t}\oint(\boldsymbol{v}_r + \boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{r} = 0, \qquad \text{or} \qquad \frac{\mathrm{D}}{\mathrm{D}t}\int_S(\boldsymbol{\omega}_r + 2\boldsymbol{\Omega})\cdot\,\mathrm{d}\boldsymbol{S} = 0, \tag{4.50a,b}$$

where the second equation uses Stokes' theorem and we have used $\nabla\times(\boldsymbol{\Omega}\times\boldsymbol{r}) = 2\boldsymbol{\Omega}$, and where $\boldsymbol{\omega}_r = \nabla\times\boldsymbol{v}_r$ is the *relative vorticity*.[4]

### 4.3.5 The Circulation Theorem for Hydrostatic Flow

Kelvin's circulation theorem holds for hydrostatic flow, with a slightly different form. For simplicity we restrict attention to the $f$-plane, and start with the hydrostatic momentum equations,

$$\frac{\mathrm{D}\boldsymbol{u}_r}{\mathrm{D}t} + 2\boldsymbol{\Omega}\times\boldsymbol{u}_r = -\frac{1}{\rho}\nabla_z p, \qquad 0 = -\frac{1}{\rho}\frac{\partial p}{\partial z} - \nabla\Phi, \tag{4.51a,b}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

where $\Phi = gz$ is the gravitational potential and $\boldsymbol{\Omega} = \Omega\mathbf{k}$. The advecting field is three-dimensional, and in particular we still have $\mathrm{D}\delta\boldsymbol{r}/\mathrm{D}t = \delta\boldsymbol{v} = (\delta\boldsymbol{r}\cdot\nabla)\boldsymbol{v}$. Thus, using (4.51) we have

$$\begin{aligned}\frac{\mathrm{D}}{\mathrm{D}t}\oint(\boldsymbol{u}_r + \boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{r} &= \oint\left[\left(\frac{\mathrm{D}\boldsymbol{u}_r}{\mathrm{D}t} + \boldsymbol{\Omega}\times\boldsymbol{v}_r\right)\cdot\mathrm{d}\boldsymbol{r} + (\boldsymbol{u}_r + \boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{v}_r\right] \\ &= \oint\left(\frac{\mathrm{D}\boldsymbol{u}_r}{\mathrm{D}t} + 2\boldsymbol{\Omega}\times\boldsymbol{u}_r\right)\cdot\mathrm{d}\boldsymbol{r} \\ &= \oint\left(-\frac{1}{\rho}\nabla p - \nabla\Phi\right)\cdot\mathrm{d}\boldsymbol{r}, \end{aligned}\tag{4.52}$$

as with (4.49), having used $\boldsymbol{\Omega}\times\boldsymbol{v}_r = \boldsymbol{\Omega}\times\boldsymbol{u}_r$, and where the gradient operator $\nabla$ is three-dimensional. The last term on the right-hand side vanishes because it is the integral of the gradient of a potential around a closed path. The first term vanishes if the fluid is barotropic, so that the circulation theorem is

$$\frac{\mathrm{D}}{\mathrm{D}t}\oint(\boldsymbol{u}_r + \boldsymbol{\Omega}\times\boldsymbol{r})\cdot\mathrm{d}\boldsymbol{r} = 0. \tag{4.53}$$

Using Stokes' theorem we have the equivalent form

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_S(\boldsymbol{\omega}_{hy} + 2\boldsymbol{\Omega})\cdot\,\mathrm{d}\boldsymbol{S} = 0, \tag{4.54}$$

where the subscript '$hy$' denotes hydrostatic and, in Cartesian coordinates,

$$\boldsymbol{\omega}_{hy} = \nabla\times\boldsymbol{u}_r = -\mathbf{i}\frac{\partial v_r}{\partial z} + \mathbf{j}\frac{\partial u_r}{\partial z} + \mathbf{k}\left(\frac{\partial v_r}{\partial x} - \frac{\partial u_r}{\partial y}\right). \tag{4.55}$$

## 4.4 VORTICITY EQUATION IN A ROTATING FRAME

Perhaps the easiest way to derive the vorticity equation appropriate for a rotating reference frame is to begin with the momentum equation in the form

$$\frac{\partial\boldsymbol{v}_r}{\partial t} + (2\boldsymbol{\Omega} + \boldsymbol{\omega}_r)\times\boldsymbol{v}_r = -\frac{1}{\rho}\nabla p - \nabla\left(\Phi + \frac{1}{2}\boldsymbol{v}_r^2\right), \tag{4.56}$$

where the potential $\Phi$ contains the gravitational and centrifugal forces. Take the curl of this and use the identity (4.13), which here implies

$$\nabla\times[(2\boldsymbol{\Omega}+\boldsymbol{\omega}_r)\times\boldsymbol{v}_r] = (2\boldsymbol{\Omega}+\boldsymbol{\omega}_r)\nabla\cdot\boldsymbol{v}_r + (\boldsymbol{v}_r\cdot\nabla)(2\boldsymbol{\Omega}+\boldsymbol{\omega}_r) - [(2\boldsymbol{\Omega}+\boldsymbol{\omega}_r)\cdot\nabla]\boldsymbol{v}_r, \tag{4.57}$$

(noting that $\nabla\cdot(2\boldsymbol{\Omega}+\boldsymbol{\omega}) = 0$), to give the vorticity equation

$$\frac{\mathrm{D}\boldsymbol{\omega}_r}{\mathrm{D}t} = [(2\boldsymbol{\Omega}+\boldsymbol{\omega}_r)\cdot\nabla]\,\boldsymbol{v} - (2\boldsymbol{\Omega}+\boldsymbol{\omega}_r)\nabla\cdot\boldsymbol{v}_r + \frac{1}{\rho^2}(\nabla\rho\times\nabla p). \tag{4.58}$$

If the rotation rate, $\boldsymbol{\Omega}$, is a constant then $\mathrm{D}\boldsymbol{\omega}_r/\mathrm{D}t = \mathrm{D}\boldsymbol{\omega}_a/\mathrm{D}t$ where $\boldsymbol{\omega}_a = 2\boldsymbol{\Omega} + \boldsymbol{\omega}_r$ is the absolute vorticity. The only difference between the vorticity equation in the rotating and inertial frames of reference is in the presence of the solid-body vorticity $2\boldsymbol{\Omega}$ on the right-hand side. The second term on the right-hand side may be folded into the material derivative using mass continuity, and after a little manipulation (4.58) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\boldsymbol{\omega}_a}{\rho}\right) = \frac{1}{\rho}(2\boldsymbol{\Omega}+\boldsymbol{\omega}_r)\cdot\nabla\boldsymbol{v}_r + \frac{1}{\rho^3}(\nabla\rho\times\nabla p). \tag{4.59}$$

However, note that it is the absolute vorticity, $\boldsymbol{\omega}_a$, that now appears on the left-hand side. If $\rho$ is constant, $\boldsymbol{\omega}_a$ may be replaced by $\boldsymbol{\omega}_r$.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

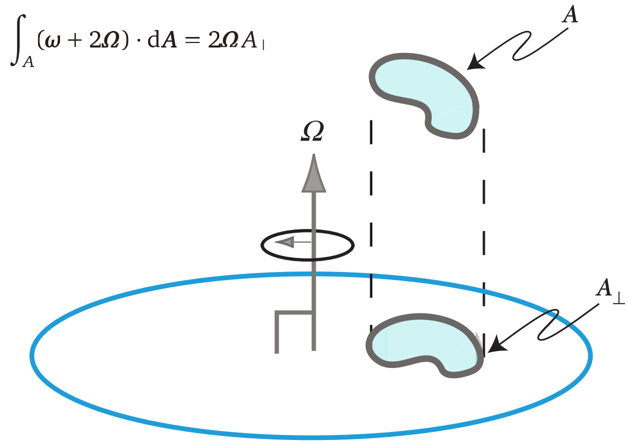


**Fig. 4.7** The projection of a material circuit on to the equatorial plane. If a fluid element moves poleward, keeping its orientation to the local vertical fixed (i.e., it stays horizontal), then the area of its projection on to the equatorial plane increases. If its total (absolute) circulation is to be maintained, then the vertical component of its relative vorticity must diminish; that is, $\int_A (\boldsymbol{\omega} + 2\boldsymbol{\Omega}) \cdot \mathrm{d}\boldsymbol{A} = \int_A (\zeta + f)\,\mathrm{d}A = \text{constant}$. Thus, the $\beta$ term in $\mathrm{D}(\zeta + f)/\mathrm{D}t = \mathrm{D}\zeta/\mathrm{D}t + \beta v = 0$ arises from the *tilting* of a parcel relative to the axis of rotation as it moves meridionally.

### 4.4.1 The Circulation Theorem and the Beta Effect

What are the implications of the circulation theorem on a rotating, spherical planet? Let us define relative circulation over some material loop as

$$C_r \equiv \oint \boldsymbol{v}_r \cdot \mathrm{d}\boldsymbol{r}. \tag{4.60}$$

Because $\boldsymbol{v}_r = \boldsymbol{v}_a - \boldsymbol{\Omega} \times \boldsymbol{r}$ (where $\boldsymbol{r}$ is the distance from the axis of rotation), we use Stokes' theorem to give

$$C_r = C_a - \int 2\boldsymbol{\Omega} \cdot \mathrm{d}\boldsymbol{S} = C_a - 2\Omega A_\perp, \tag{4.61}$$

where $C_a$ is the total or absolute circulation and $A_\perp$ is the area enclosed by the projection of the material circuit on to the plane normal to the rotation vector; that is, on to the equatorial plane (Fig. 4.7). If the solenoidal term is zero, then the circulation theorem, (4.50), may be written as

$$\frac{\mathrm{D}}{\mathrm{D}t}(C_r + 2\Omega A_\perp) = 0. \tag{4.62}$$

Thus, the relative circulation around a circuit changes if the orientation of the plane changes; that is, if the area of its projection on to the equatorial plane changes. In large scale dynamics the most common cause of this is when a fluid parcel changes its latitude. For example, consider the flow of a two-dimensional, infinitesimal, horizontal (i.e., tangent to the the radial vector), constant-density fluid parcel at a latitude $\vartheta$ with area $A$, so that the projection of its area on to the equatorial plane is $A_\perp = A \sin\vartheta$ and $C_r = \zeta_r A$. If the fluid surface moves, but remains horizontal, its area is preserved (because it is incompressible) and directly from (4.62) its relative vorticity changes as

$$\frac{\mathrm{D}\zeta_r}{\mathrm{D}t} = -\frac{2\Omega}{A}\frac{\mathrm{D}A_\perp}{\mathrm{D}t} = -2\Omega\frac{\mathrm{D}}{\mathrm{D}t}\sin\vartheta = -v_r\frac{2\Omega\cos\vartheta}{a} = -\beta v_r, \tag{4.63}$$

where

$$\beta \equiv \frac{\mathrm{d}f}{\mathrm{d}y} = \frac{2\Omega}{a}\cos\vartheta. \tag{4.64}$$

The means by which the vertical component of the relative vorticity of a parcel changes by virtue of its latitudinal displacement is known as the *beta effect*, or the *β-effect*. It is a manifestation of the tilting term in the vorticity equation, and it is often the most important means by which relative vorticity does change in large-scale flow. The $\beta$-effect arises in the full vorticity equation, as we now see.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

### 4.4.2 The Vertical Component of the Vorticity Equation

In large-scale dynamics, the most important, although not the largest, component of the vorticity is often the vertical one, because this contains much of the information about the horizontal flow. We can obtain an explicit expression for its evolution by taking the vertical component of (4.58), although care must be taken because the unit vectors ($\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$) are functions of position (see Section 2.2).

An alternative derivation begins with the horizontal momentum equations,

$$\frac{\partial u}{\partial t} - v(\zeta + f) + w\frac{\partial u}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial x} - \frac{1}{2}\frac{\partial}{\partial x}(u^2 + v^2) + F^x \tag{4.65a}$$

$$\frac{\partial v}{\partial t} + u(\zeta + f) + w\frac{\partial v}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - \frac{1}{2}\frac{\partial}{\partial y}(u^2 + v^2) + F^y, \tag{4.65b}$$

where in this section we again drop the subscript $r$ on variables measured in the rotating frame. Cross-differentiating gives, after a little algebra,

$$\begin{aligned}\frac{\mathrm{D}}{\mathrm{D}t}(\zeta + f) = &-(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \left(\frac{\partial u}{\partial z}\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\frac{\partial w}{\partial x}\right)\\ &+ \frac{1}{\rho^2}\left(\frac{\partial \rho}{\partial x}\frac{\partial p}{\partial y} - \frac{\partial \rho}{\partial y}\frac{\partial p}{\partial x}\right) + \left(\frac{\partial F^y}{\partial x} - \frac{\partial F^x}{\partial y}\right).\end{aligned} \tag{4.66}$$

We interpret the various terms as follows:

$\mathrm{D}\zeta/\mathrm{D}t = \partial\zeta/\partial t + \boldsymbol{v}\cdot\nabla\zeta$. The material derivative of the vertical component of the vorticity.

$\mathrm{D}f/\mathrm{D}t = v\partial f/\partial y = v\beta$. The $\beta$-effect. The vorticity is affected by the meridional motion of the fluid, so that, apart from the terms on the right-hand side, $(\zeta + f)$ is conserved on parcels. Because the Coriolis parameter changes with latitude this is like saying that the system has differential rotation. This effect is precisely that due to the change in orientation of fluid surfaces with latitude, as discussed in Section 4.4.1 and illustrated Fig. 4.7.

$-(\zeta + f)(\partial u/\partial x + \partial v/\partial y)$. The divergence term, which gives rise to vortex stretching. In an incompressible fluid this may be written $(\zeta + f)\partial w/\partial z$, so that vorticity is amplified if the vertical velocity increases with height, so stretching the material lines and the vorticity.

$(\partial u/\partial z)(\partial w/\partial y) - (\partial v/\partial z)(\partial w/\partial x)$. The tilting term, whereby a vertical component of vorticity may be generated by a vertical velocity acting on a horizontal vorticity. See Fig. 4.4.

$\rho^{-2}\left[(\partial\rho/\partial x)(\partial p/\partial y) - (\partial\rho/\partial y)(\partial p/\partial x)\right] = \rho^{-2}J(\rho, p)$. The solenoidal term, also called the non-homentropic or baroclinic term, arising when isosurfaces of pressure and density are not parallel.

$(\partial F^y/\partial x - \partial F^x/\partial y)$. The forcing and friction term. If the only contribution is from molecular viscosity then this term is $\nu\nabla^2\zeta$.

#### *Two-dimensional and shallow water vorticity equations*

In an inviscid two-dimensional incompressible flow, all of the terms on the right-hand side of (4.66) vanish and we have the simple equation

$$\frac{\mathrm{D}(\zeta + f)}{\mathrm{D}t} = 0, \tag{4.67}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

implying that the absolute vorticity, $\zeta_a \equiv \zeta + f$, is materially conserved. If $f$ is a constant, then (4.67) reduces to (4.28), and background rotation plays no role. If $f$ varies linearly with $y$, so that $f = f_0 + \beta y$, then (4.67) becomes

$$\frac{\partial \zeta}{\partial t} + \boldsymbol{u} \cdot \nabla \zeta + \beta v = 0, \tag{4.68}$$

which is known as the two-dimensional $\beta$-plane vorticity equation.

For inviscid shallow water flow, we can show that (see Chapter 3)

$$\frac{\mathrm{D}(\zeta + f)}{\mathrm{D}t} = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right). \tag{4.69}$$

In this equation the vanishing of the tilting term is perhaps the only aspect which is not immediately apparent, but this succumbs to a little thought.

## 4.5 POTENTIAL VORTICITY CONSERVATION

*Too much of a good thing is wonderful.*
Mae West (1892–1990).

Although Kelvin's circulation theorem is a general statement about vorticity conservation, in its original form it is not always a practically useful statement for two reasons. First, it is not a statement about a *field*, such as vorticity itself. Second, it is not satisfied for baroclinic flow, such as is found in the atmosphere and ocean. (Non-conservative forces such as viscosity also lead to circulation non-conservation, but this applies to virtually all conservation laws and does not diminish them.) It turns out that it is possible to derive a beautiful conservation law that overcomes both of these failings and one that, furthermore, is extraordinarily useful in geophysical fluid dynamics. This is the conservation of *potential vorticity* (PV) introduced first by Rossby and then in a more general form by Ertel.[5] The idea is that we can use a scalar field that is being advected by the flow to keep track of, or to take care of, the evolution of fluid elements. For a baroclinic fluid this scalar field must be chosen in a special way (it must be a function of the density and pressure alone), but there is no restriction to a barotropic fluid. Then using the scalar evolution equation in conjunction with the vorticity equation gives us a scalar conservation equation. In the next few subsections we derive the equation for potential vorticity conservation in a number of superficially different ways — different explications but the same explanation.[6]

### 4.5.1 PV Conservation from the Circulation Theorem

*Barotropic fluids*

Let us begin with the simple case of a barotropic fluid. For an infinitesimal volume we write Kelvin's theorem as

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[(\boldsymbol{\omega}_a \cdot \boldsymbol{n})\delta A\right] = 0, \tag{4.70}$$

where $\boldsymbol{n}$ is a unit vector normal to an infinitesimal surface $\delta A$. Now consider a volume bounded by two isosurfaces of values $\chi$ and $\chi + \delta\chi$, where $\chi$ is any materially conserved tracer, thus satisfying $\mathrm{D}\chi/\mathrm{D}t = 0$, so that $\delta A$ initially lies in an isosurface of $\chi$ (see Fig. 4.8). Since $\boldsymbol{n} = \nabla\chi/|\nabla\chi|$ and the infinitesimal volume $\delta V = \delta h\, \delta A$, where $\delta h$ is the separation between the two surfaces, we have

$$\boldsymbol{\omega}_a \cdot \boldsymbol{n}\,\delta A = \boldsymbol{\omega}_a \cdot \frac{\nabla\chi}{|\nabla\chi|}\frac{\delta V}{\delta h}. \tag{4.71}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

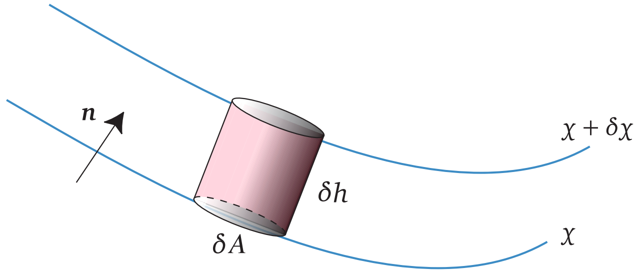


**Fig. 4.8** An infinitesimal fluid element, bounded by two isosurfaces of the conserved tracer $\chi$. As $\mathrm{D}\chi/\mathrm{D}t = 0$, then $\mathrm{D}\delta\chi/\mathrm{D}t = 0$.

Now, the value of $\delta h$ may be obtained from

$$\delta\chi = \delta\boldsymbol{x} \cdot \nabla\chi = \delta h |\nabla\chi|, \tag{4.72}$$

and using this in (4.70) we obtain

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\frac{(\boldsymbol{\omega}_a \cdot \nabla\chi)\delta V}{\delta\chi}\right] = 0. \tag{4.73}$$

Since $\chi$ is conserved on material elements then so is $\delta\chi$ and it may be taken out of the differentiation. The mass of the volume element $\rho\,\delta V$ is also conserved, so that (4.73) becomes

$$\frac{\rho\delta V}{\delta\chi}\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\boldsymbol{\omega}_a}{\rho} \cdot \nabla\chi\right) = 0 \tag{4.74}$$

or

$$\frac{\mathrm{D}}{\mathrm{D}t}(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla\chi) = 0, \tag{4.75}$$

where $\widetilde{\boldsymbol{\omega}}_a = \boldsymbol{\omega}_a/\rho$. Equation (4.75) is a statement of potential vorticity conservation for a barotropic fluid. The field $\chi$ may be chosen arbitrarily, provided that it is materially conserved.

### *The general case*

For a baroclinic fluid the above derivation fails simply because the statement of the conservation of circulation, (4.70) is not, in general, true: there are solenoidal terms on the right-hand side and from (4.43) and (4.45) we have

$$\frac{\mathrm{D}}{\mathrm{D}t}[(\boldsymbol{\omega}_a \cdot \boldsymbol{n})\delta A] = \boldsymbol{S}_o \cdot \boldsymbol{n}\,\delta A, \qquad \boldsymbol{S}_o = -\nabla\alpha \times \nabla p = -\nabla\eta \times \nabla T. \tag{4.76a,b}$$

However, the right-hand side of (4.76a) may be annihilated by choosing the circuit around which we evaluate the circulation to be such that the solenoidal term is identically zero. Given the form of $\boldsymbol{S}_o$, this occurs if the values of any of $p, \rho, \eta, T$ are constant on that circuit; that is, if $\chi = p, \rho, \eta$ or $T$. But the derivation also demands that $\chi$ be a materially conserved quantity, which usually restricts the choice of $\chi$ to be $\eta$ (or potential temperature), or to be $\rho$ itself if the thermodynamic equation is $\mathrm{D}\rho/\mathrm{D}t = 0$. Thus, the conservation of potential vorticity for inviscid, adiabatic flow is

$$\frac{\mathrm{D}}{\mathrm{D}t}(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla\theta) = 0, \tag{4.77}$$

where $\mathrm{D}\theta/\mathrm{D}t = 0$. For diabatic flow source terms appear on the right-hand side, and we derive these later on. A summary of this derivation is provided by Fig. 4.9.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

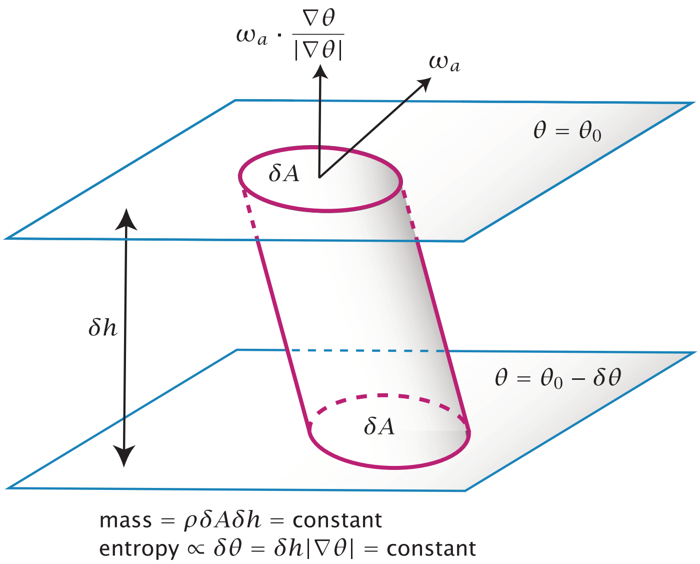


**Fig. 4.9** Geometry of potential vorticity conservation. The circulation equation is $\mathrm{D}[(\boldsymbol{\omega}_a \cdot \boldsymbol{n})\delta A]/\mathrm{D}t = \boldsymbol{S}_o \cdot \boldsymbol{n}\delta A$, where $\boldsymbol{S}_o \propto \nabla\theta \times \nabla T$. We choose $\boldsymbol{n} = \nabla\theta/|\nabla\theta|$, where $\theta$ is materially conserved, to annihilate the solenoidal term on the right-hand side, and we note that $\delta A = \delta V/\delta h$, where $\delta V$ is the volume of the cylinder, and the height of the column is $\delta h = \delta\theta/|\nabla\theta|$. The circulation is $C \equiv \boldsymbol{\omega}_a \cdot \boldsymbol{n}\delta A = \boldsymbol{\omega}_a \cdot (\nabla\theta/|\nabla\theta|)(\delta V/\delta h) = [\rho^{-1}\boldsymbol{\omega}_a \cdot \nabla\theta](\delta M/\delta\theta)$, where $\delta M = \rho\delta V$ is the mass of the cylinder. As $\delta M$ and $\delta\theta$ are materially conserved, so is the potential vorticity $\rho^{-1}\boldsymbol{\omega}_a \cdot \nabla\theta$.

### 4.5.2 PV Conservation from the Frozen-in Property

In this section we show that potential vorticity conservation is a consequence of the frozen-in property of vorticity. This is not surprising, because the circulation theorem itself has a similar origin. Thus, this derivation is not independent of the derivation in the previous section, just a re-expression of it. We first consider the case in which the solenoidal term vanishes from the outset.

#### *Barotropic fluids*

If $\chi$ is a materially conserved tracer then the difference in $\chi$ between two infinitesimally close fluid elements is also conserved and

$$\frac{\mathrm{D}}{\mathrm{D}t}(\chi_1 - \chi_2) = \frac{\mathrm{D}\delta\chi}{\mathrm{D}t} = 0. \tag{4.78}$$

But $\delta\chi = \nabla\chi \cdot \delta\boldsymbol{l}$, where $\delta\boldsymbol{l}$ is the infinitesimal vector connecting the two fluid elements. Thus

$$\frac{\mathrm{D}}{\mathrm{D}t}(\nabla\chi \cdot \delta\boldsymbol{l}) = 0. \tag{4.79}$$

However, as the line element and the vorticity (divided by density) obey the same equation, we can replace the line element by vorticity (divided by density) in (4.79) to obtain again

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\nabla\chi \cdot \boldsymbol{\omega}_a}{\rho}\right) = 0. \tag{4.80}$$

That is, the potential vorticity, $Q = (\widetilde{\boldsymbol{\omega}}_a \cdot \nabla\chi)$ is a material invariant, where $\chi$ is any scalar quantity that satisfies $\mathrm{D}\chi/\mathrm{D}t = 0$.

#### *Baroclinic fluids*

In baroclinic fluids we cannot casually substitute the vorticity for that of a line element in (4.79) because of the presence of the solenoidal term, and in any case a little more detail would not be amiss. From (4.79) we obtain

$$\delta\boldsymbol{l} \cdot \frac{\mathrm{D}\nabla\chi}{\mathrm{D}t} + \nabla\chi \cdot \frac{\mathrm{D}\delta\boldsymbol{l}}{\mathrm{D}t} = 0, \tag{4.81}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

or, using (4.31),

$$\delta \boldsymbol{l} \cdot \frac{\mathrm{D}\nabla\chi}{\mathrm{D}t} + \nabla\chi \cdot [(\delta\boldsymbol{l} \cdot \nabla)\boldsymbol{v}] = 0. \tag{4.82}$$

Now, let us choose $\delta\boldsymbol{l}$ to correspond to a vortex line, so that at the initial time $\delta\boldsymbol{l} = \epsilon\widetilde{\boldsymbol{\omega}}_a$. (Note that in this case the association of $\delta\boldsymbol{l}$ with a vortex line can only be made instantaneously, and we cannot set $\mathrm{D}\delta\boldsymbol{l}/\mathrm{D}t \propto \mathrm{D}\boldsymbol{\omega}_a/\mathrm{D}t$.) Then,

$$\widetilde{\boldsymbol{\omega}}_a \cdot \frac{\mathrm{D}\nabla\chi}{\mathrm{D}t} + \nabla\chi \cdot [(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla)\boldsymbol{v}] = 0, \tag{4.83}$$

or, using the vorticity equation (4.16),

$$\widetilde{\boldsymbol{\omega}}_a \cdot \frac{\mathrm{D}\nabla\chi}{\mathrm{D}t} + \nabla\chi \cdot \left( \frac{\mathrm{D}\widetilde{\boldsymbol{\omega}}_a}{\mathrm{D}t} - \frac{1}{\rho^3}\nabla\rho \times \nabla p \right) = 0. \tag{4.84}$$

This may be written as

$$\frac{\mathrm{D}}{\mathrm{D}t}\widetilde{\boldsymbol{\omega}}_a \cdot \nabla\chi = \frac{1}{\rho^3}\nabla\chi \cdot (\nabla\rho \times \nabla p). \tag{4.85}$$

The term on the right-hand side is, in general, non-zero for an arbitrary choice of scalar, but it will evidently vanish if $\nabla p$, $\nabla\rho$ and $\nabla\chi$ are coplanar. If $\chi$ is any function of $p$ and $\rho$ this will be satisfied, but $\chi$ must also be a materially conserved scalar. If, as for an ideal gas, $\rho = \rho(\eta, p)$ (or $\eta = \eta(p, \rho)$) where $\eta$ is the entropy (which is materially conserved), and if $\chi$ is a function of entropy $\eta$ alone, then $\chi$ satisfies both conditions. Explicitly, the solenoidal term vanishes because

$$\nabla\chi \cdot (\nabla\rho \times \nabla p) = \frac{\mathrm{d}\chi}{\mathrm{d}\eta}\nabla\eta \cdot \left[ \left( \frac{\partial\rho}{\partial p}\nabla p + \frac{\partial\rho}{\partial\eta}\nabla\eta \right) \times \nabla p \right] = 0. \tag{4.86}$$

Thus, provided $\chi$ satisfies the two conditions

$$\frac{\mathrm{D}\chi}{\mathrm{D}t} = 0 \quad \text{and} \quad \chi = \chi(p, \rho), \tag{4.87}$$

then (4.85) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\left( \frac{\boldsymbol{\omega}_a \cdot \nabla\chi}{\rho} \right) = 0. \tag{4.88}$$

The natural choice for $\chi$ is potential temperature, whence

$$\boxed{\frac{\mathrm{D}}{\mathrm{D}t}\left( \frac{\boldsymbol{\omega}_a \cdot \nabla\theta}{\rho} \right) = 0.} \tag{4.89}$$

The presence of a density term in the denominator is not necessary for incompressible flows (i.e., if $\nabla \cdot \boldsymbol{v} = 0$).

### 4.5.3 PV Conservation: an Algebraic Derivation

Finally, we give an algebraic derivation of potential vorticity conservation. We will take the opportunity to include frictional and, diabatic processes, although these may also be included in the derivations above.[7] We begin with the frictional vorticity equation in the form

$$\frac{\mathrm{D}\widetilde{\boldsymbol{\omega}}_a}{\mathrm{D}t} = (\widetilde{\boldsymbol{\omega}}_a \cdot \nabla)\boldsymbol{v} + \frac{1}{\rho^3}(\nabla\rho \times \nabla p) + \frac{1}{\rho}(\nabla \times \boldsymbol{F}), \tag{4.90}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

where $\boldsymbol{F}$ represents any non-conservative force term on the right-hand side of the momentum equation (i.e., $\mathrm{D}\boldsymbol{v}/\mathrm{D}t = -\rho^{-1}\nabla p + \boldsymbol{F}$). We have also the equation for our materially conserved scalar $\chi$,

$$\frac{\mathrm{D}\chi}{\mathrm{D}t} = \dot{\chi}, \tag{4.91}$$

where $\dot{\chi}$ represents any sources and sinks of $\chi$. Now

$$(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla)\frac{\mathrm{D}\chi}{\mathrm{D}t} = \widetilde{\boldsymbol{\omega}}_a \cdot \frac{\mathrm{D}\nabla\chi}{\mathrm{D}t} + [(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla)\boldsymbol{v}] \cdot \nabla\chi, \tag{4.92}$$

which may be obtained just by expanding the left-hand side. Thus, using (4.91),

$$\widetilde{\boldsymbol{\omega}}_a \cdot \frac{\mathrm{D}\nabla\chi}{\mathrm{D}t} = (\widetilde{\boldsymbol{\omega}}_a \cdot \nabla)\,\dot{\chi} - [(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla)\boldsymbol{v}] \cdot \nabla\chi. \tag{4.93}$$

Now take the dot product of (4.90) with $\nabla\chi$:

$$\nabla\chi \cdot \frac{\mathrm{D}\widetilde{\boldsymbol{\omega}}_a}{\mathrm{D}t} = \nabla\chi \cdot [(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla)\boldsymbol{v}] + \nabla\chi \cdot \left[\frac{1}{\rho^3}(\nabla\rho \times \nabla p)\right] + \nabla\chi \cdot \left[\frac{1}{\rho}(\nabla \times \boldsymbol{F})\right]. \tag{4.94}$$

The sum of the last two equations yields

$$\frac{\mathrm{D}}{\mathrm{D}t}(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla\chi) = \widetilde{\boldsymbol{\omega}}_a \cdot \nabla\dot{\chi} + \nabla\chi \cdot \left[\frac{1}{\rho^3}(\nabla\rho \times \nabla p)\right] + \frac{\nabla\chi}{\rho} \cdot (\nabla \times \boldsymbol{F}). \tag{4.95}$$

This equation reprises (4.85), but with the addition of frictional and diabatic terms. As before, the solenoidal term is annihilated if we choose $\chi = \theta(p, \rho)$, so giving the evolution equation for potential vorticity in the presence of forcing and diabatic terms, namely

$$\boxed{\frac{\mathrm{D}}{\mathrm{D}t}(\widetilde{\boldsymbol{\omega}}_a \cdot \nabla\theta) = \widetilde{\boldsymbol{\omega}}_a \cdot \nabla\dot{\theta} + \frac{\nabla\theta}{\rho} \cdot (\nabla \times \boldsymbol{F}).} \tag{4.96}$$

### 4.5.4 Effects of Salinity and Moisture

For seawater the equation of state may be written as

$$\theta = \theta(\rho, p, S), \tag{4.97}$$

where $\theta$ is the potential temperature and $S$ is the salinity. In the absence of diabatic terms and saline diffusion the potential temperature is a materially conserved quantity. However, because of the presence of salinity, potential temperature cannot be used to annihilate the solenoidal term; that is

$$\nabla\theta \cdot (\nabla\rho \times \nabla p) = \left(\frac{\partial\theta}{\partial S}\right)_{p,\rho} \nabla S \cdot (\nabla\rho \times \nabla p) \neq 0. \tag{4.98}$$

Strictly speaking then, *there is no potential vorticity conservation principle for seawater.* However, such a blunt statement overemphasizes the non-conservation of potential vorticity because the saline effect is small. In fact, we can derive an approximate potential vorticity conservation law, as follows.[8]

Suppose that we use potential density to try to annihilate the solenoidal term. Potential density is adiabatically conserved but, like $\theta$, it is a function of salinity so that

$$\nabla\rho_\theta \cdot (\nabla\rho \times \nabla p) = \left(\frac{\partial\rho_\theta}{\partial S}\right)_{p,\rho} \nabla S \cdot (\nabla\rho \times \nabla p) \neq 0. \tag{4.99}$$

Now, potential density may be written as function of salinity and potential temperature (or entropy) with no pressure dependence and therefore we can rewrite the above expression as

$$(\nabla\rho \times \nabla p)\cdot\nabla\rho_\theta = (\nabla\rho_\theta \times \nabla\rho)\cdot\nabla p \tag{4.100a}$$

$$= \left[\left(\frac{\partial\rho_\theta}{\partial S}\nabla S + \frac{\partial\rho_\theta}{\partial\theta}\nabla\theta\right)\times\left(\frac{\partial\rho}{\partial S}\nabla S + \frac{\partial\rho}{\partial\theta}\nabla\theta + \frac{\partial\rho}{\partial p}\nabla p\right)\right]\cdot\nabla p \tag{4.100b}$$

$$= \left[\frac{\partial\rho_\theta}{\partial S}\nabla S \times \frac{\partial\rho}{\partial\theta}\nabla\theta + \frac{\partial\rho_\theta}{\partial\theta}\nabla\theta \times \frac{\partial\rho}{\partial S}\nabla S\right]\cdot\nabla p \tag{4.100c}$$

$$= \left[\frac{\partial\rho_\theta}{\partial S}\frac{\partial\rho}{\partial\theta} - \frac{\partial\rho_\theta}{\partial\theta}\frac{\partial\rho}{\partial S}\right](\nabla S \times \nabla\theta)\cdot\nabla p, \tag{4.100d}$$

If the term in square brackets in (4.100d) is zero then potential vorticity is conserved, and this is the case if the density and potential density are related by

$$\rho(S,\theta,p) = \rho_\theta(S,\theta) + F(p), \tag{4.101}$$

where $F$ is some function of $p$; the result also follows directly using (4.101) and (4.100a). Equation (4.101) does not exactly hold because the compressibility of seawater is not in fact just a function of pressure (this is the thermobaric effect). However, as can be seen from (1.156b), the equation holds to a good approximation, a result related to the fact that the speed of sound in seawater is nearly constant. Thus, to this approximation, potential vorticity is adiabatically conserved in seawater if potential density is used as the scalar variable. The derivation does not care whether density itself is a function of salinity; rather, it asks that the difference between density and potential density is a function only of pressure.

Similarly, in a moist atmosphere there is, strictly, no conservation of a conventional potential vorticity because potential temperature is a function of density, pressure and water vapour, although the moisture dependence is usually weak. Condensational heating also provides a diabatic source term that provides a source of potential vorticity. These effects may accounted for in part by using a virtual potential temperature in the definition of potential vorticity, and including the contribution of liquid water to the entropy.[9] In any case, and as with seawater, the compositional effects are fairly small, especially in mid-latitudes, and the dynamics of potential vorticity conservation play a central role in the large-scale dynamics of both atmosphere and ocean.

### 4.5.5 Effects of Rotation, and Summary Remarks

In a rotating frame the potential vorticity conservation equation is obtained simply by replacing $\boldsymbol{\omega}_a$ by $\boldsymbol{\omega} + 2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the rotation rate of the rotating frame. The operator $\mathrm{D}/\mathrm{D}t$ is reference-frame invariant, and so may be evaluated using the usual formulae with velocities measured in the rotating frame.

We have generally referred to the quantity $\boldsymbol{\omega}_a \cdot \nabla\theta/\rho$ as the potential vorticity; however, this form (often referred to as the Ertel or Rossby–Ertel potential vorticity) is not unique. If $\theta$ is a materially conserved variable, then so is $g(\theta)$ where $g$ is any function, so that $\boldsymbol{\omega}_a \cdot \nabla g(\theta)/\rho$ is also a potential vorticity. In the atmosphere $\theta$ itself is in fact commonly used, whereas in the ocean potential density is the more appropriate scalar, with $f\partial\rho_\theta/\partial z$ being a common approximation for low Rossby number flows.

The conservation of potential vorticity has profound consequences in fluid dynamics, especially in a rotating, stratified fluid. The non-conservative terms are often small, and large-scale flow in both the ocean and the atmosphere is well characterized by conservation of potential vorticity. Such conservation is a very powerful constraint on the flow, and indeed it turns out that potential vorticity is usually a more useful quantity for baroclinic, or non-homentropic, fluids than for barotropic fluids, because the required use of a special conserved scalar imparts additional information; in barotropic fluids potential vorticity has little more power than vorticity itself.

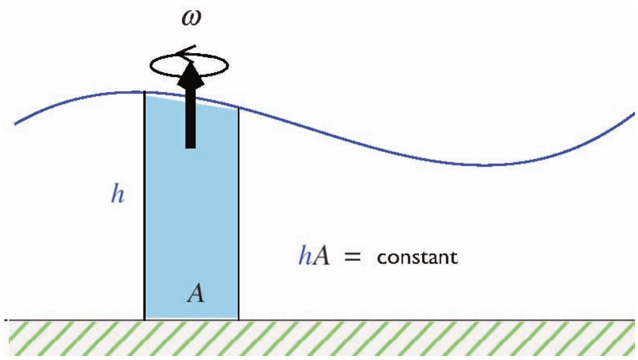


**Fig. 4.10** The volume of a column of fluid, $hA$, is conserved. Furthermore, the vorticity is tied to material lines so that $\zeta A$ is also a material invariant, where $\zeta = \boldsymbol{\omega} \cdot \mathbf{k}$ is the vertical component of the vorticity. From this, $\zeta/h$ must be materially conserved, or $\mathrm{D}(\zeta/h)/\mathrm{D}t = 0$, which is the conservation of potential vorticity in a shallow water system. With rotation this generalizes to $\mathrm{D}[(\zeta + f)/h]/\mathrm{D}t = 0$.

## 4.6 ✦ POTENTIAL VORTICITY IN THE SHALLOW WATER SYSTEM

In Chapter 3 we derived potential vorticity conservation by direct manipulation of the shallow water equations. In this short section we show that shallow water potential vorticity is also derivable from the conservation of circulation. Specifically, we will begin with the three-dimensional form of Kelvin's theorem, and then make the small aspect ratio assumption (which is the key assumption underlying shallow water dynamics), and thereby recover shallow water potential vorticity conservation (see also Fig. 4.10).

We begin with

$$\frac{\mathrm{D}}{\mathrm{D}t}(\boldsymbol{\omega}_3 \cdot \delta \boldsymbol{S}) = 0, \tag{4.102}$$

where $\boldsymbol{\omega}_3$ is the curl of the three-dimensional velocity and $\delta \boldsymbol{S} = \boldsymbol{n}\,\delta S$ is an arbitrary infinitesimal vector surface element, with $\boldsymbol{n}$ being a unit vector pointing in the direction normal to the surface. If we separate the vorticity and the surface element into vertical and horizontal components we can write (4.102) as

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[(\zeta + f)\delta A + \boldsymbol{\omega}_h \cdot \delta \boldsymbol{S}_h\right] = 0, \tag{4.103}$$

where $\boldsymbol{\omega}_h$ and $\delta \boldsymbol{S}_h$ are the horizontally directed components of the vorticity and the surface element, and $\delta A = \mathbf{k}\,\delta \boldsymbol{S}$ is the area of a horizontal cross-section of a fluid column. In Cartesian form the horizontal component of the vorticity is

$$\boldsymbol{\omega}_h = \mathbf{i}\left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right) - \mathbf{j}\left(\frac{\partial w}{\partial x} - \frac{\partial u}{\partial z}\right) = \mathbf{i}\frac{\partial w}{\partial y} - \mathbf{j}\frac{\partial w}{\partial x}, \tag{4.104}$$

where vertical derivatives of the horizontal velocity are zero by virtue of the nature of the shallow water system. Now, the vertical velocity in the shallow water system is smaller than the horizontal velocity by the order of the aspect ratio — the ratio of the fluid depth to the horizontal scale of the motion. Furthermore, the size of the horizontally directed surface element is also smaller than the vertically-directed component by the aspect ratio; that is,

$$|\boldsymbol{\omega}_h| \sim \alpha|\zeta| \qquad \text{and} \qquad |\delta \boldsymbol{S}_h| \sim \alpha|\delta A|, \tag{4.105}$$

where $\alpha = H/L$ is the aspect ratio. Thus $\boldsymbol{\omega}_h \cdot \delta \boldsymbol{S}_h$ is smaller than the term $\zeta \delta A$ by the aspect number squared, and in the small aspect ratio approximation should be neglected. Kelvin's circulation theorem, (4.103), becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}[(\zeta + f)\delta A] = 0 \qquad \text{or} \qquad \frac{\mathrm{D}}{\mathrm{D}t}\left[\frac{(\zeta + f)}{h} h\,\delta A\right] = 0, \tag{4.106a,b}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

where $h$ is the depth of the fluid column. But $h\,\delta A$ is the volume of the fluid column, and this is constant. Thus, (4.106b) gives, as in (3.96),

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\zeta + f}{h}\right) = 0, \tag{4.107}$$

where, because horizontal velocities are independent of the vertical coordinate, the advection is purely horizontal.

## 4.7 POTENTIAL VORTICITY IN APPROXIMATE, STRATIFIED MODELS

If approximate models of stratified flow (Boussinesq, hydrostatic and so on) are to be useful then they should conserve an appropriate form of potential vorticity, and we consider a few such cases.

### 4.7.1 The Boussinesq Equations

A Boussinesq fluid is incompressible; that is, the volume of a fluid element is conserved and the flow is divergence-free, with $\nabla \cdot \boldsymbol{v} = 0$. The equation for vorticity itself is then isomorphic to that for a line element. However, the Boussinesq equations are not barotropic — $\nabla\rho$ is not parallel to $\nabla p$ — and although the pressure gradient term $\nabla\phi$ disappears on taking its curl (or equivalently disappears on integration around a closed path) the buoyancy term $\mathbf{k}b$ does not, and it is this term that prevents Kelvin's circulation theorem from holding. Specifically, the evolution of circulation in the Boussinesq equations obeys

$$\frac{\mathrm{D}}{\mathrm{D}t}[(\boldsymbol{\omega}_a \cdot \boldsymbol{n})\delta A] = (\nabla \times b\mathbf{k}) \cdot \boldsymbol{n}\,\delta A, \tag{4.108}$$

where here, as in (4.70), $\boldsymbol{n}$ is a unit vector orthogonal to an infinitesimal surface element of area $\delta A$. The right-hand side is annihilated if we choose $\boldsymbol{n}$ to be parallel to $\nabla b$, because $\nabla b \cdot \nabla \times (b\mathbf{k}) = 0$. In the simple Boussinesq equations the thermodynamic equation is

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0, \tag{4.109}$$

and potential vorticity conservation is therefore (with $\boldsymbol{\omega}_a = \boldsymbol{\omega} + 2\boldsymbol{\Omega}$)

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = 0, \qquad Q = (\boldsymbol{\omega} + 2\boldsymbol{\Omega}) \cdot \nabla b. \tag{4.110a,b}$$

Expanding (4.110b) in Cartesian coordinates with $2\boldsymbol{\Omega} = f\mathbf{k}$ we obtain:

$$Q = (v_x - u_y)b_z + (w_y - v_z)b_x + (u_z - w_x)b_y + f b_z. \tag{4.111}$$

In the general Boussinesq equations $b$ itself is not materially conserved. We cannot expect to obtain a conservation law if salinity is present, but if the equation of state and the thermodynamic equation are:

$$b = b(\theta, z), \qquad \frac{\mathrm{D}\theta}{\mathrm{D}t} = 0, \tag{4.112}$$

then potential vorticity conservation follows, because taking $\boldsymbol{n}$ to be parallel to $\nabla\theta$ will cause the right-hand side of (4.108) to vanish; that is,

$$\nabla\theta \cdot \nabla \times (b\mathbf{k}) = \left(\frac{\partial\theta}{\partial z}\nabla z + \frac{\partial\theta}{\partial b}\nabla b\right) \cdot \nabla \times (b\mathbf{k}) = 0. \tag{4.113}$$

The materially conserved potential vorticity in the Boussinesq approximation, $Q_B$, is thus

$$\boxed{Q_B = \boldsymbol{\omega}_a \cdot \nabla\theta.} \tag{4.114}$$

Note that if the equation of state is $b = b(\theta, \phi)$, where $\phi$ is the pressure, then potential vorticity is not conserved because then, in general, $\nabla\phi \cdot \nabla \times (b\mathbf{k}) \neq 0$.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

### 4.7.2 The Hydrostatic Equations

Making the hydrostatic approximation has no effect on whether or not the circulation theorem is satisfied. Thus, in a baroclinic hydrostatic fluid we have

$$\frac{\mathrm{D}}{\mathrm{D}t}\int(\boldsymbol{\omega}_{hy}+2\boldsymbol{\Omega})\cdot\,\mathrm{d}\boldsymbol{S}=-\int\nabla\alpha\times\nabla p\cdot\,\mathrm{d}\boldsymbol{S}, \tag{4.115}$$

where, from (4.55) $\boldsymbol{\omega}_{hy} = \nabla\times\boldsymbol{u} = -\mathbf{i}v_z + \mathbf{j}u_z + \mathbf{k}(v_x - u_y)$, but the gradient operator and material derivative are fully three-dimensional. Derivation of potential vorticity conservation then proceeds, as in Section 4.5.1, by choosing the circuit over which the circulation is calculated to be such that the right-hand side vanishes; that is, to be such that the solenoidal term is annihilated. Precisely as before, this occurs if the circuit is barotropic, and without further ado we write

$$\frac{\mathrm{D}Q_{hy}}{\mathrm{D}t}=\frac{\mathrm{D}}{\mathrm{D}t}\left[\frac{(\boldsymbol{\omega}_{hy}+2\boldsymbol{\Omega})\cdot\nabla\theta}{\rho}\right]=0. \tag{4.116}$$

Expanding the expression for $Q_{hy}$ in Cartesian coordinates gives

$$Q_{hy}=\frac{1}{\rho}\left[(v_x-u_y)\theta_z-v_z\theta_x+u_z\theta_y+2\Omega\theta_z\right]. \tag{4.117}$$

In spherical coordinates the hydrostatic approximation is usually accompanied by the traditional approximation and the expanded expression for a conserved potential vorticity is more complicated. It can still be derived from Kelvin's theorem, but this is left as an exercise for the reader.

### 4.7.3 Potential Vorticity on Isentropic Surfaces

If we begin with the primitive equations in isentropic coordinates then potential vorticity conservation follows quite simply. Cross-differentiating the horizontal momentum equations (3.178) gives the vorticity equation

$$\frac{\mathrm{D}}{\mathrm{D}t}(\zeta+f)+(\zeta+f)\nabla_\theta\cdot\boldsymbol{u}=0, \tag{4.118}$$

where $\mathrm{D}/\mathrm{D}t = \partial/\partial t + \boldsymbol{u}\cdot\nabla_\theta$. The thermodynamic equation is

$$\frac{\mathrm{D}\sigma}{\mathrm{D}t}+\sigma\nabla\cdot\boldsymbol{u}=0, \tag{4.119}$$

where $\sigma = \partial z/\partial b$ (Boussinesq) or $\partial p/\partial\theta$ (ideal gas) is the thickness of an isopycnal layer. Eliminating the divergence between (4.118) and (4.119) gives

$$\frac{\mathrm{D}Q_{IS}}{\mathrm{D}t}=0,\quad\text{where}\quad Q_{IS}=\left(\frac{\zeta+f}{\sigma}\right). \tag{4.120}$$

The derivation, and the result, are precisely the same as with the shallow water equations (Sections 3.7.1 and 4.6).

*A connection between isentropic and height coordinates*

The hydrostatic potential vorticity written in height coordinates may be transformed into a form that reveals its intimate connection with isentropic surfaces. Let us make the Boussinesq approximation for which the hydrostatic potential vorticity is, with no rotation,

$$Q_{hy}=(v_x-u_y)b_z-v_zb_x+u_zb_y, \tag{4.121}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

where $b$ is the buoyancy. We can write this as

$$Q_{hy} = b_z \left[ \left( v_x - v_z \frac{b_x}{b_z} \right) - \left( u_y - u_z \frac{b_y}{b_z} \right) \right]. \tag{4.122}$$

But the terms in the inner brackets are just the horizontal velocity derivatives at constant $b$. To see this, note that

$$\left( \frac{\partial v}{\partial x} \right)_b = \left( \frac{\partial v}{\partial x} \right)_z + \frac{\partial v}{\partial z} \left( \frac{\partial z}{\partial x} \right)_b = \left( \frac{\partial v}{\partial x} \right)_z - \frac{\partial v}{\partial z} \left( \frac{\partial b}{\partial x} \right)_z \bigg/ \frac{\partial b}{\partial z}, \tag{4.123}$$

with a similar expression for $(\partial u/\partial y)_b$. (These relationships follow from standard rules of partial differentiation. Derivatives with respect to $z$ are taken at constant $x$ and $y$.) Thus, we obtain

$$Q_{hy} = \frac{\partial b}{\partial z} \left[ \left( \frac{\partial v}{\partial x} \right)_b - \left( \frac{\partial u}{\partial y} \right)_b \right] = \frac{\partial b}{\partial z} \zeta_b. \tag{4.124}$$

Thus, potential vorticity is simply the horizontal vorticity evaluated on a surface of constant buoyancy, multiplied by the vertical derivative of buoyancy. An analogous derivation, with a similar result, proceeds for the ideal gas equations, with potential temperature replacing buoyancy.

## 4.8 ✦ THE IMPERMEABILITY OF ISENTROPES TO POTENTIAL VORTICITY

An interesting property of isentropic surfaces is that they are 'impermeable' to potential vorticity, meaning that the mass integral of potential vorticity ($\int Q\rho \, \mathrm{d}V$) over a volume bounded by an isentropic surface remains constant, even in the presence of diabatic sources, provided the surfaces do not intersect a non-isentropic surface such as the ground.[10] This may seem surprising, especially because unlike most conservation laws the result does not require adiabatic flow, and for that reason it leads to interesting interpretations of a number of phenomena. However, impermeability is a consequence of the definition of potential vorticity rather than the equations of motion, and in that sense it is a kinematic and not dynamical property.

To derive the result we define $s \equiv \rho Q = \nabla \cdot (\theta \boldsymbol{\omega}_a)$, where $\boldsymbol{\omega}_a$ is the absolute vorticity, and integrate over some volume $V$ to give

$$I = \int_V s \, \mathrm{d}V = \int_V \nabla \cdot (\theta \boldsymbol{\omega}_a) \, \mathrm{d}V = \int_S \theta \boldsymbol{\omega}_a \cdot \mathrm{d}\boldsymbol{S}, \tag{4.125}$$

using the divergence theorem, where $S$ is the surface surrounding the volume $V$. If this is an isentropic surface then we have

$$I = \theta \int_S \boldsymbol{\omega}_a \cdot \mathrm{d}\boldsymbol{S} = \theta \int_V \nabla \cdot \boldsymbol{\omega}_a \, \mathrm{d}V = 0, \tag{4.126}$$

again using the divergence theorem. That is, over a volume wholly enclosed by a single isentropic surface the integral of $s$ vanishes. If the volume is bounded by more than one isentropic surface none of which intersect the surface, for example by concentric spheres of different radii as in Fig. 4.11(a), the result still holds. The quantity $s$ is called 'potential vorticity concentration', or 'PV concentration'. The integral of $s$ over a volume is akin to the total amount of a conserved material property, such as salt content, and so may be called 'PV substance'. That is, the PV concentration is the amount of potential vorticity substance per unit volume and

$$\text{PV substance} = \int s \, \mathrm{d}V = \int \rho Q \, \mathrm{d}V. \tag{4.127}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

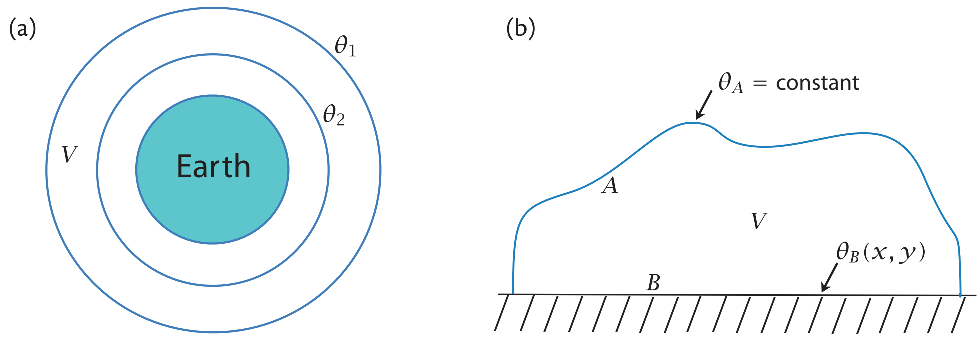


**Fig. 4.11** (a) Two isentropic surfaces that do not intersect the ground. The integral of PV concentration over the volume between them, $V$, is zero, even if there is heating and the contours move. (b) An isentropic surface, $A$, intersects the ground, $B$, thus enclosing a volume $V$. The rate of change of PV concentration over the volume is given by an integral over $B$.

Suppose now that the fluid volume is enclosed by an isentrope that intersects the ground, as in Fig. 4.11(b). Let $A$ denote the isentropic surface, $B$ denote the ground, $\theta_A$ the constant value of $\theta$ on the isentrope, and $\theta_B(x, y, t)$ the non-constant value of $\theta$ on the ground. The integral of $s$ over the volume is then

$$\begin{aligned} I = \int_V \nabla \cdot (\theta \boldsymbol{\omega}_a)\,\mathrm{d}V &= \theta_A \int_A \boldsymbol{\omega}_a \cdot \mathrm{d}\boldsymbol{S} + \int_B \theta_B \boldsymbol{\omega}_a \cdot \mathrm{d}\boldsymbol{S} \\ &= \theta_A \int_{A+B} \boldsymbol{\omega}_a \cdot \mathrm{d}\boldsymbol{S} + \int_B (\theta_B - \theta_A)\boldsymbol{\omega}_a \cdot \mathrm{d}\boldsymbol{S} \\ &= \int_B (\theta_B - \theta_A)\boldsymbol{\omega}_a \cdot \mathrm{d}\boldsymbol{S}. \end{aligned} \tag{4.128}$$

The first term on the second line vanishes after using the divergence theorem. Thus, the value of $I$, and hence its rate of change, is a function *only of an integral over the surface B*, and the PV flux there must be calculated using the full equations of motion. However, we do not need to be concerned with any flux of PV concentration through the isentropic surface; put another way, the PV substance in a volume can change only when isentropes enclosing the volume intersect a boundary such as the Earth's surface.

### 4.8.1 Interpretation and Application

*Motion of the isentropic surface*

How can the above results hold in the presence of heating? The isentropic surfaces must move in such a way that the total amount of PV concentration contained between them nevertheless stays fixed, and we now demonstrate this explicitly. The potential vorticity equation may be written

$$\frac{\partial Q}{\partial t} + \boldsymbol{v} \cdot \nabla Q = S_Q, \tag{4.129}$$

where, from (4.96), $S_Q = (\boldsymbol{\omega}_a/\rho) \cdot \nabla\dot{\theta} + \nabla\theta \cdot (\nabla \times \boldsymbol{F})/\rho$. Using mass continuity this may be written as

$$\frac{\partial s}{\partial t} + \nabla \cdot \boldsymbol{J} = 0, \tag{4.130}$$

where $\boldsymbol{J} \equiv \rho \boldsymbol{v} Q + \boldsymbol{N}$ and $\nabla \cdot \boldsymbol{N} = -\rho S_Q$. Written in this way, the quantity $\boldsymbol{J}/s$ is a notional velocity, $\boldsymbol{v}_Q$ say, and $s$ satisfies

$$\frac{\partial s}{\partial t} + \nabla \cdot (\boldsymbol{v}_Q s) = 0. \tag{4.131}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

That is, $s$ evolves as if it were being fluxed by the velocity $\boldsymbol{v}_Q$. The concentration of a chemical tracer $\chi$ (i.e., $\chi$ is the amount of tracer per unit volume) obeys a similar equation, to wit

$$\frac{\partial \chi}{\partial t} + \nabla \cdot (\boldsymbol{v}\chi) = 0. \tag{4.132}$$

However, whereas (4.132) implies that $\mathrm{D}(\chi/\rho)/\mathrm{D}t = 0$, (4.131) does not imply that $\partial Q/\partial t + \boldsymbol{v}_Q \cdot \nabla Q = 0$ because $\partial \rho/\partial t + \nabla \cdot (\rho \boldsymbol{v}_Q) \neq 0$.

Now, the impermeability result tells us that there can be no notional velocity across an isentropic surface. How can this be satisfied by the equations of motion? We write the right-hand side of (4.129) as

$$\rho S_Q = \nabla \cdot (\dot{\theta}\boldsymbol{\omega}_a + \theta \nabla \times \boldsymbol{F}) = \nabla \cdot (\dot{\theta}\boldsymbol{\omega}_a + \boldsymbol{F} \times \nabla\theta). \tag{4.133}$$

Thus, $\boldsymbol{N} = -\dot{\theta}\boldsymbol{\omega}_a - \boldsymbol{F} \times \nabla\theta$ and we may write the $\boldsymbol{J}$ vector as

$$\boldsymbol{J} = \rho \boldsymbol{v} Q - \dot{\theta}\boldsymbol{\omega}_a - \boldsymbol{F} \times \nabla\theta = \rho Q(\boldsymbol{v}_\perp + \boldsymbol{v}_\parallel) - \dot{\theta}\boldsymbol{\omega}_\parallel - \boldsymbol{F} \times \nabla\theta, \tag{4.134}$$

where, making use of the thermodynamic equation,

$$\boldsymbol{v}_\parallel = \boldsymbol{v} - \frac{\boldsymbol{v} \cdot \nabla\theta}{|\nabla\theta|^2}\nabla\theta, \qquad \boldsymbol{v}_\perp = -\frac{\partial\theta/\partial t}{|\nabla\theta|^2}\nabla\theta, \tag{4.135a}$$

$$\boldsymbol{\omega}_\parallel = \boldsymbol{\omega}_a - \frac{\boldsymbol{\omega}_a \cdot \nabla\theta}{|\nabla\theta|^2}\nabla\theta = \boldsymbol{\omega}_a - \frac{Q\rho}{|\nabla\theta|^2}\nabla\theta. \tag{4.135b}$$

The subscripts '$\perp$' and '$\|$' denote components perpendicular and parallel to the local isentropic surface, and $\boldsymbol{v}_\perp$ is the velocity of the isentropic surface normal to itself. Equation (4.134) may be verified by using (4.135) and $\mathrm{D}\theta/\mathrm{D}t = \dot{\theta}$.

The 'parallel' terms in (4.135) are all vectors parallel to the local isentropic surface, and therefore do not lead to any flux of PV concentration across that surface. Furthermore, the term $\rho Q \boldsymbol{v}_\perp$ is $\rho Q$ multiplied by the normal velocity of the surface. That is to say, the notional velocity associated with the flux normal to the isentropic surface is equal to the normal velocity of the isentropic surface itself, and so it too provides no flux of PV concentration across that surface (even though there may well be a mass flux across the surface). Put simply, the isentropic surface always moves in such a way as to ensure that there is no flux of PV concentration across it. In our proof of the impermeability result in the previous subsection we used the fact that the potential vorticity multiplied by the density is the divergence of a vector. In the demonstration above we used the fact that the terms *forcing* potential vorticity are the divergence of a vector.

### † *Dynamical choices of PV flux and a connection to Bernoulli's theorem*

If we add a non-divergent vector to the flux, $\boldsymbol{J}$, then it has no effect on the evolution of $s$. This gauge invariance means that the notional velocity, $v_Q = \boldsymbol{J}/(\rho Q)$ is similarly non-unique, although it does not mean that there are not dynamical choices for it that are more appropriate in given circumstances. To explore this, let us obtain a general expression for $\boldsymbol{J}$ by starting with the definition of $s$, so that

$$\begin{aligned}\frac{\partial s}{\partial t} &= \nabla\theta \cdot \frac{\partial \boldsymbol{\omega}_a}{\partial t} + \boldsymbol{\omega}_a \cdot \nabla \frac{\partial\theta}{\partial t} \\ &= \nabla\theta \cdot \nabla \times \frac{\partial \boldsymbol{v}}{\partial t} + \nabla \cdot \left(\boldsymbol{\omega}_a \frac{\partial\theta}{\partial t}\right) = -\nabla \cdot \boldsymbol{J}',\end{aligned} \tag{4.136}$$

where

$$\boldsymbol{J}' = \nabla\theta \times \frac{\partial \boldsymbol{v}}{\partial t} - \frac{\partial\theta}{\partial t}\boldsymbol{\omega}_a + \nabla\phi \times \nabla\chi. \tag{4.137}$$

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

The last term in this expression is an arbitrary divergence-free vector. If we choose $\phi = \theta$ and $\chi = B$, where $B$ is the Bernoulli function given by $B = I + \boldsymbol{v}^2/2 + p/\rho$ where $I$ is the internal energy per unit mass, then

$$\boldsymbol{J}' = \nabla\theta \times \left(\nabla B + \frac{\partial \boldsymbol{v}}{\partial t}\right) - \boldsymbol{\omega}_a(\dot{\theta} - \boldsymbol{v}\cdot\nabla\theta), \tag{4.138}$$

having used the thermodynamic equation $\mathrm{D}\theta/\mathrm{D}t = \dot{\theta}$. Now, the momentum equation may be written, without approximation, in the form

$$\frac{\partial \boldsymbol{v}}{\partial t} = -\boldsymbol{\omega}_a \times \boldsymbol{v} + T\nabla\eta + \boldsymbol{F} - \nabla B, \tag{4.139}$$

where $\eta$ is the specific entropy ($\mathrm{d}\eta = c_p\,\mathrm{d}\ln\theta$). Using (4.138) and (4.139) gives

$$\boldsymbol{J}' = \rho Q\boldsymbol{v} - \dot{\theta}\boldsymbol{\omega}_a + \nabla\theta \times \boldsymbol{F}, \tag{4.140}$$

which is the same as (4.134). Furthermore, using (4.137) for steady flow,

$$\boldsymbol{J} = \nabla\theta \times \nabla B. \tag{4.141}$$

That is, the flux of potential vorticity (in this gauge) is aligned with the intersection of $\theta$- and $B$-surfaces. For steady *inviscid and adiabatic* flow the Bernoulli function is constant along streamlines; that is, surfaces of constant Bernoulli function are aligned with streamlines, and, because $\theta$ is materially conserved, streamlines are formed at intersecting $\theta$- and $B$-surfaces, as in (1.204). In the presence of forcing, this property is replaced by (4.141), and the flux of PV concentration is along such intersections.

This choice of gauge leading to (4.140) is physical in that it reduces to the true advective flux $\boldsymbol{v}\rho Q$ for unforced, adiabatic flow, but it is not a unique choice, nor is it mandated by the dynamics. Choosing $\chi = 0$ leads to the flux

$$\boldsymbol{J}_1 = \rho Q\boldsymbol{v} - \dot{\theta}\boldsymbol{\omega}_a + \nabla\theta \times (\boldsymbol{F} - \nabla B), \tag{4.142}$$

and using (4.137) this vanishes for steady flow, which is a potentially useful property.

### *Summary remarks*

The impermeability result is kinematic, but can provide an interesting point of view and useful diagnostic tool.[11] We make the following summary remarks:

- There can be no net transport of potential vorticity across an isentropic surface, and the total amount of potential vorticity in a volume wholly enclosed by isentropic surfaces is zero. Thus, and with hindsight trivially, the amount of potential vorticity contained between two isentropes isolated from the Earth's surface in the Northern Hemisphere is the negative of the corresponding amount in the Southern Hemisphere.
- Potential vorticity flux lines (i.e., lines everywhere parallel to $\boldsymbol{J}$) can either close in on themselves or begin and end at boundaries (e.g., the ground or the ocean surface). However, $\boldsymbol{J}$ may change its character. Thus, for example, at the base of the oceanic mixed layer $\boldsymbol{J}$ may change from being a diabatic flux above to an adiabatic advective flux below. There may be a similar change in character at the atmospheric tropopause.
- The flux vector $\boldsymbol{J}$ is defined only to within the curl of a vector. Thus the vector $\boldsymbol{J}' = \boldsymbol{J} + \nabla\times\boldsymbol{A}$, where $\boldsymbol{A}$ is arbitrary, is as valid as is $\boldsymbol{J}$ in the above derivations.

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

## Notes

1 The frozen-in property — that vortex lines are material lines — was derived by Helmholtz (1858) and is sometimes called Helmholtz's theorem.

2 The theorem originates with William Thomson (1824–1907), who became Lord Kelvin in 1892. The circulation theorem was published in Thomson (1869) and is a conservation law that is unique to a fluid: unlike, for example, the conservation of energy, it has no analogue in solid-body mechanics. Thomson was born in Belfast but spent most of his life in Scotland, becoming a professor at the University of Glasgow in 1846 (at the age of 22!) and staying there for 53 years. A prolific and creative scientist, he made a lasting impact on both fluid dynamics and thermodynamics — among other achievements he proposed an absolute temperature scale and a formulation of the second law of thermodynamics. Later in life he turned to engineering and was one of the proponents of a telegraph cable under the Atlantic. To his credit he also had some grand failures — his estimates of the age of Earth and how long oxygen would last in the atmosphere were both wrong by orders of magnitude.

3 Silberstein (1896) proved that 'the necessary and sufficient condition for the generation of vortical flow…influenced only by conservative forces …is that the surface of constant pressure and surface of constant density…intersect', as we derived in Section 4.2, and this leads to (4.43). Bjerknes (1898a,b) explicitly put this into the form of a circulation theorem and applied it to problems of meteorological and oceanographic importance (see Thorpe *et al.* 2003), and the theorem is sometimes called the Bjerknes theorem or the Bjerknes–Silberstein theorem.

Vilhelm Bjerknes (1862–1951) was a physicist and hydrodynamicist who in 1917 moved to the University of Bergen as founding head of the Bergen Geophysical Institute. Here he did what was probably his most influential work in meteorology, setting up and contributing to the 'Bergen School of Meteorology'. Among other things he and his colleagues were among the first to consider, as a practical proposition, the use of numerical methods — initial data in conjunction with the fluid equations of motion — to forecast the state of the atmosphere, based on earlier work describing how that task might be done (Abbe 1901, Bjerknes 1904). Inaccurate initial velocity fields compounded with the shear complexity of the effort ultimately defeated them, but the effort was continued (also unsuccessfully) by L. F. Richardson (Richardson 1922), before J. Charney, R. Fjørtoft and J. Von Neumann eventually made what may be regarded as the first successful numerical forecast (Charney *et al.* 1950). Their success can be attributed to the used of a simplified, filtered, set of equations and the use of an electronic computer.

4 The result (4.50) was given by Poincaré (1893) although it is sometimes attributed to Bjerknes (1902).

5 The first derivation of the PV conservation law was given for the hydrostatic shallow water equations by Rossby (1936), with a generalization to the stratified case, via the use of isentropic coordinates, in Rossby (1938) and Rossby (1940). In the 1936 paper Rossby noted — his Eq. (75) — that a fluid column satisfies $f + \zeta = cD$, where $c$ is a constant and $D$ is the thickness of a fluid column; equivalently, $(f + \zeta)/D$ is a material invariant. The expression 'potential vorticity' was introduced in Rossby (1940), as follows: *'This quantity, which may be called the potential vorticity, represents the vorticity the air column would have it it were brought, isopycnally or isentropically, to a standard latitude ($f_0$) and stretched or shrunk vertically to a standard depth $D_0$ or weight $\Delta_0$.'* (Rossby's italics.) That is,

$$\text{potential vorticity} = \zeta_0 = \left(\frac{\zeta + f}{D}\right) D_0 - f_0, \tag{4.143}$$

which follows from his Eq. (11), and this is the sense he uses it in that paper. However, potential vorticity has come to mean the quantity $(\zeta + f)/D$, which of course does not have the dimensions of vorticity. We use it in this latter, now conventional, sense throughout this book. Ironically, quasi-geostrophic potential vorticity as usually defined does have the dimensions of vorticity.

The expression for potential vorticity in a non-hydrostatic, continuously stratified fluid was given by Ertel (1942a), and its relationship to circulation was given by Ertel (1942b). It is now commonly known as the *Ertel potential vorticity*, or the *Rossby–Ertel* potential vorticity. Interestingly, in Rossby

https://doi.org/10.1017/9781107588417.005 Published online by Cambridge University Press

(1940) we find the Fermat-like comment 'It is possible to derive corresponding results for an atmosphere in which the potential temperature varies continuously with elevation.... The generalized treatment will be presented in another place.' Given his prior (1938) derivation of the stratified quantity in isentropic coordinates, he must not have regarded his own derivations as very general. Opinions differ as to whether Rossby's and Ertel's derivations were independent, and Cressman (1996) remarks that the origin of the concept of potential vorticity is a 'delicate one that has aroused some passion in private correspondences'. In fact, Ertel visited MIT in autumn 1937 and presumably talked to Rossby and became aware of his work. It seems almost certain that Ertel knew of Rossby's shallow water and isentropic theorems, but it is also clear that Ertel subsequently provided a significant generalization, most likely independently. Rossby and Ertel apparently remained on good terms, but further collaboration was stymied by World War II. They later published a pair of short joint papers, one in German and the other in English, describing related conservation theorems (Ertel & Rossby 1949a,b). English translations of a number of Ertel's papers are to be found in Schubert *et al.* (2004). I thank Roger Samelson for enlightening me about the history of Rossby and Ertel.

6 Native French speakers may be confused by the difference. In English, an explication is a particular way of performing an analysis or presenting an explanation, so there can be different explications of the same mechanism.

7 Truesdell (1951, 1954) and Obukhov (1962) were early explorers of the consequences of heating and friction on potential vorticity. The work of F. P. Bretherton, R. E. Dickinson and J. S. A. Green (e.g., Bretherton 1966a, Dickinson 1969, Green 1970) helped bring potential vorticity ideas further into the mainstream of GFD.

8 Many thanks to Stephen Griffies for pointing out this argument, and for many other comments and discussions on matters related to this book. A study of saline effects on potential vorticity is to be found in Straub (1999).

9 Schubert *et al.* (2001) provide more discussion of these matters. They derive a 'moist PV' that is an extension of the dry Ertel PV to moist atmospheres and that has invertibility and impermeability properties.

10 Haynes & McIntyre (1987, 1990). See also Danielsen (1990), Schär (1993), who obtained the result (4.141), Bretherton & Schär (1993) and Davies-Jones (2003).

11 See, for example, McIntyre & Norton (1990) and Marshall & Nurser (1992). The latter use $\boldsymbol{J}$ vectors to study the creation and transport of potential vorticity in the oceanic thermocline.

*A little inaccuracy sometimes saves a ton of explanation.*
H. M. Munro (Saki), *The Square Egg*, 1924.

*Every decoding is another encoding.*
David Lodge, in the voice of Morris Zapp, *Small World*, 1984.

# CHAPTER 5

# Geostrophic Theory

LARGE-SCALE FLOW IN THE OCEAN AND THE ATMOSPHERE is characterized by an approximate balance in the vertical direction between the pressure gradient and gravity (hydrostatic balance), and in the horizontal direction between the pressure gradient and the Coriolis force (geostrophic balance). In this chapter we exploit these balances to simplify the Navier–Stokes equations and thereby obtain various sets of simplified 'geostrophic equations'. Depending on the precise nature of the assumptions we make, we are led to the *quasi-geostrophic* (QG) system for horizontal scales similar to that on which most synoptic activity takes place and, for very large-scale motion, to the *planetary-geostrophic* (PG) set of equations. By eliminating unwanted or unimportant modes of motion, in particular sound waves and gravity waves, and by building in the important balances between flow fields, these filtered equation sets allow the investigator to better focus on a particular class of phenomena and to potentially achieve a deeper understanding than might otherwise be possible.[1]

Simplifying the equations in this way relies first on scaling the equations. The idea is that we *choose* the scales we wish to describe, typically either on some a-priori basis or by using observations as a guide. We then attempt to derive a set of equations that is simpler than the original set but that consistently describes motion of the chosen scale. An asymptotic method is one way to achieve this, for it systematically tells us which terms we can drop and which we should keep. The combined approach — scaling plus asymptotics — has proven enormously useful, but we should always remember two things: (i) that scaling is a choice; (ii) that the approach does not explain the existence of particular scales of motion, it just describes the motion that might occur on such scales. We have already employed this general approach in deriving the hydrostatic primitive equations, but now we go further.

## 5.1 GEOSTROPHIC SCALING

### 5.1.1 Scaling in the Shallow Water Equations

Postponing the complications that come with stratification, we begin with the shallow water equations. With the odd exception, we will denote the scales of variables by capital letters; thus, if $L$ is a typical length scale of the motion we wish to describe, and $U$ is a typical velocity scale, and



https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

assuming the scales are horizontally isotropic, we write

$$
\begin{aligned}
(x, y) &\sim L \qquad \text{or} \qquad (x, y) = \mathcal{O}(L) \\
(u, v) &\sim U \qquad \text{or} \qquad (u, v) = \mathcal{O}(U),
\end{aligned} \tag{5.1}
$$

and similarly for other variables. We may then nondimensionalize the variables by writing

$$
(x, y) = L(\widehat{x}, \widehat{y}), \qquad (u, v) = U(\widehat{u}, \widehat{v}), \tag{5.2}
$$

where the hatted variables are nondimensional and, by supposition, are $\mathcal{O}(1)$. The various terms in the momentum equation then scale as:

$$
\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} + \boldsymbol{f} \times \boldsymbol{u} = -g \nabla \eta, \tag{5.3a}
$$

$$
\frac{U}{T} \qquad \frac{U^2}{L} \qquad fU \;\sim\; g\frac{\mathcal{H}}{L}, \tag{5.3b}
$$

where the $\nabla$ operator acts in the $x$–$y$ plane and $\mathcal{H}$ is the amplitude of the variations in the surface displacement. (We use $\eta$ to denote the height of the free surface above some arbitrary reference level, as in Fig. 3.1. Thus, $\eta = H + \Delta\eta$, where $\Delta\eta$ denotes the variation of $\eta$ about its mean position.)

The ratio of the advective term to the rotational term in the momentum equation (5.3) is $(U^2/L)/(fU) = U/fL$; this is the Rossby number,[2] first encountered in Chapter 2. Using values typical of the large-scale circulation (e.g., from Table 2.1) we find that $Ro \approx 0.1$ for the atmosphere and $Ro \approx 0.01$ for the ocean: small in both cases. If we are interested in motion that has the advective time scale $T = L/U$ then we scale time by $L/U$ so that

$$
t = \frac{L}{U}\hat{t}, \tag{5.4}
$$

and the local time derivative and the advective term then both scale as $U^2/L$, and both are smaller than the rotation term by a factor of the order of the Rossby number. Then, either the Coriolis term is the dominant term in the equation, in which case we have a state of no motion with $-fv = 0$, or else the Coriolis force is balanced by the pressure force, and the dominant balance is

$$
-fv = -g\frac{\partial \eta}{\partial x}, \tag{5.5}
$$

namely *geostrophic balance*, as encountered in Chapter 2. If we make this non-trivial choice, then the equation informs us that variations in $\eta$ (i.e., $\Delta\eta$) scale according to

$$
\Delta\eta \sim \mathcal{H} = \frac{fUL}{g}. \tag{5.6}
$$

We can also write $\mathcal{H}$ as

$$
\mathcal{H} = Ro\,\frac{f^2 L^2}{g} = Ro\; H\frac{L^2}{L_d^2}, \tag{5.7}
$$

where $L_d = \sqrt{gH}/f$ is the deformation radius and $H$ is the mean depth of the fluid. The variations in fluid height thus scale as

$$
\frac{\Delta\eta}{H} \sim Ro\,\frac{L^2}{L_d^2}, \tag{5.8}
$$

and the height of the fluid may be written as

$$
\eta = H\left(1 + Ro\,\frac{L^2}{L_d^2}\widehat{\eta}\right) \qquad \text{and} \qquad \Delta\eta = Ro\,\frac{L^2}{L_d^2}H\widehat{\eta}, \tag{5.9}
$$

where $\widehat{\eta}$ is the $\mathcal{O}(1)$ nondimensional value of the surface height deviation.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

### Nondimensional momentum equation

If we use (5.9) to scale height variations, (5.2) to scale lengths and velocities, and (5.4) to scale time, then the momentum equation (5.3) becomes

$$Ro\left[\frac{\partial \hat{\boldsymbol{u}}}{\partial \hat{t}} + (\hat{\boldsymbol{u}}\cdot\nabla)\hat{\boldsymbol{u}}\right] + \widehat{\boldsymbol{f}}\times\hat{\boldsymbol{u}} = -\nabla\hat{\eta}, \tag{5.10}$$

where $\widehat{\boldsymbol{f}} = \mathbf{k}\widehat{f} = \mathbf{k}f/f_0$, where $f_0$ is a representative value of the Coriolis parameter. (If $f$ is a constant, then $\widehat{f} = 1$, but it is informative to explicitly write $\widehat{f}$ in the equations. Also, where the operator $\nabla$ operates on a nondimensional variable then the differentials are taken with respect to the nondimensional variables $\hat{x}$, $\hat{y}$.) All the variables in (5.10) will be assumed to be of order unity, and the Rossby number multiplying the local time derivative and the advective terms indicates the smallness of those terms. By construction, the dominant balance in this equation is the geostrophic balance between the last two terms.

### Nondimensional mass continuity (height) equation

The (dimensional) mass continuity equation can be written as

$$\frac{1}{H}\frac{\mathrm{D}\eta}{\mathrm{D}t} + \left(1 + \frac{\Delta\eta}{H}\right)\nabla\cdot\boldsymbol{u} = 0. \tag{5.11}$$

Using (5.2), (5.4) and (5.9) this equation may be written

$$Ro\left(\frac{L}{L_d}\right)^2\frac{\mathrm{D}\hat{\eta}}{\mathrm{D}\hat{t}} + \left[1 + Ro\left(\frac{L}{L_d}\right)^2\hat{\eta}\right]\nabla\cdot\hat{\boldsymbol{u}} = 0. \tag{5.12}$$

Equations (5.10) and (5.12) are the nondimensional versions of the full shallow water equations of motion. Evidently, some terms in the equations of motion are small and may be eliminated with little loss of accuracy, and the way this is done will depend on the size of the second nondimensional parameter, $(L/L_d)^2$, which we come to shortly.

### Froude and Burger numbers

The Froude number may be generally defined as the ratio of a fluid particle speed to a wave speed. In a shallow water system this gives

$$Fr \equiv \frac{U}{\sqrt{gH}} = \frac{U}{f_0 L_d} = Ro\,\frac{L}{L_d}\,. \tag{5.13}$$

The Burger number[3] is a useful measure of the scale of motion of the fluid, relative to the deformation radius, and may be defined by

$$Bu \equiv \left(\frac{L_d}{L}\right)^2 = \frac{gH}{f_0^2L^2} = \left(\frac{Ro}{Fr}\right)^2. \tag{5.14}$$

It is also useful to define the parameter $F \equiv Bu^{-1}$, which is like the square of a Froude number but uses the rotational speed $fL$ instead of $U$ in the numerator.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

### 5.1.2 Geostrophic Scaling in the Stratified Equations

We now apply the same scaling ideas, *mutatis mutandis,* to the stratified primitive equations. We use the hydrostatic anelastic equations, which we write as

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u} = -\nabla_z \phi, \tag{5.15a}$$

$$\frac{\partial \phi}{\partial z} = b, \tag{5.15b}$$

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0, \tag{5.15c}$$

$$\nabla \cdot (\widetilde{\rho} \boldsymbol{v}) = 0, \tag{5.15d}$$

where $b$ is the buoyancy and $\widetilde{\rho}$ is a reference density profile. Anticipating that the average stratification may not scale in the same way as the deviation from it, let us separate out the contribution of the advection of a reference stratification in (5.15c) by writing

$$b = \tilde{b}(z) + b'(x, y, z, t). \tag{5.16}$$

The thermodynamic equation then becomes

$$\frac{\mathrm{D}b'}{\mathrm{D}t} + N^2 w = 0, \tag{5.17}$$

where $N^2 \equiv \partial\tilde{b}/\partial z$ (and the advective derivative is still three-dimensional). We then let $\phi = \widetilde{\phi}(z) + \phi'$, where $\widetilde{\phi}$ is hydrostatically balanced by $\tilde{b}$, and the hydrostatic equation becomes

$$\frac{\partial \phi'}{\partial z} = b'. \tag{5.18}$$

Equations (5.17) and (5.18) replace (5.15c) and (5.15b), and $\phi'$ is used in (5.15a).

*Nondimensional equations*

We scale the basic variables by supposing that

$$(x, y) \sim L, \quad (u, v) \sim U, \quad t \sim \frac{L}{U}, \quad z \sim H, \quad f \sim f_0, \quad N \sim N_0 \tag{5.19}$$

where the scaling variables (capitalized, except for $f_0$) are chosen to be such that the nondimensional variables have magnitudes of the order of unity, and the parameters $N_0$ and $f_0$ are representative values of $N$ and $f$. The scales chosen are such that the Rossby number is small; that is $Ro = U/(f_0 L) \ll 1$. In the momentum equation the pressure term then balances the Coriolis force,

$$|\boldsymbol{f} \times \boldsymbol{u}| \sim |\nabla \phi'|, \tag{5.20}$$

and so the pressure scales as

$$\phi' \sim \Phi = f_o U L. \tag{5.21}$$

Using the hydrostatic relation, (5.21) implies that the buoyancy scales as

$$b' \sim B = \frac{f_0 U L}{H}, \tag{5.22}$$

and from this we obtain

$$\frac{(\partial b'/\partial z)}{N^2} \sim Ro \frac{L^2}{L_d^2}, \tag{5.23}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

where $L_d = N_0 H/f_0$ is the deformation radius in the continuously stratified fluid, analogous to the quantity $\sqrt{gH}/f_0$ in the shallow water system, and we use the same symbol, $L_d$, for both. In the continuously stratified system, *if the scale of motion is the same as or smaller than the deformation radius, and the Rossby number is small, then the variations in stratification are small.* The choice of scale is the key difference between the planetary-geostrophic and quasi-geostrophic equations.

Finally, we will nondimensionalize the vertical velocity by using the mass conservation equation,

$$\frac{1}{\tilde{\rho}}\frac{\partial \tilde{\rho} w}{\partial z} = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \tag{5.24}$$

and we suppose that this implies

$$w \sim W = \frac{UH}{L}\,. \tag{5.25}$$

This is a naïve scaling for rotating flow: if the Coriolis parameter is nearly constant the geostrophic velocity is nearly horizontally non-divergent and the right-hand side of (5.24) is small, and $W \ll UH/L$. We might then estimate $w$ by cross-differentiating geostrophic balance (with $\tilde{\rho}$ constant for simplicity) to obtain the linear geostrophic vorticity equation and corresponding scaling:

$$\beta v \approx f\frac{\partial w}{\partial z}, \qquad w \sim W = \frac{\beta U H}{f_0}. \tag{5.26a,b}$$

However, rather than using (5.26b) from the outset, we will use (5.25) and let the asymptotics guide us to a proper scaling in the fullness of time. Note that if variations in the Coriolis parameter are large and $\beta \sim f_0/L$, then (5.26b) is the same as (5.25).

Given the scalings above (using (5.25) for $w$) we nondimensionalize by setting

$$\begin{gathered}(\hat{x},\hat{y}) = L^{-1}(x,y), \qquad \hat{z} = H^{-1}z, \quad (\hat{u},\hat{v}) = U^{-1}(u,v), \quad \hat{t} = \frac{U}{L}t, \\ \widehat{w} = \frac{L}{UH}w, \quad \widehat{f} = \frac{f}{f_0}, \quad \widehat{N} = \frac{N}{N_0}, \quad \widehat{\phi} = \frac{\phi'}{f_0 U L}, \quad \hat{b} = \frac{H}{f_0 U L}b',\end{gathered} \tag{5.27}$$

where the hatted variables are nondimensional. The horizontal momentum and hydrostatic equations then become

$$Ro\,\frac{\mathrm{D}\hat{\boldsymbol{u}}}{\mathrm{D}\hat{t}} + \widehat{\boldsymbol{f}} \times \hat{\boldsymbol{u}} = -\nabla\widehat{\phi}, \tag{5.28}$$

and

$$\frac{\partial \widehat{\phi}}{\partial \hat{z}} = \hat{b}. \tag{5.29}$$

The nondimensional mass conservation equation is simply

$$\frac{1}{\tilde{\rho}}\nabla\cdot(\tilde{\rho}\boldsymbol{v}) = \left(\frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} + \frac{1}{\tilde{\rho}}\frac{\partial \tilde{\rho}\widehat{w}}{\partial \hat{z}}\right) = 0, \tag{5.30}$$

and the nondimensional thermodynamic equation is

$$\frac{f_0 U L}{H}\frac{U}{L}\frac{\mathrm{D}\hat{b}}{\mathrm{D}\hat{t}} + \widehat{N}^2 N_0^2 \frac{HU}{L}\widehat{w} = 0, \tag{5.31}$$

or

$$Ro\,\frac{\mathrm{D}\hat{b}}{\mathrm{D}\hat{t}} + \left(\frac{L_d}{L}\right)^2 \widehat{N}^2 \widehat{w} = 0. \tag{5.32}$$

The nondimensional primitive equations are summarized in the box on the following page.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

**Nondimensional Primitive Equations**

Horizontal momentum: $$Ro\,\frac{\mathrm{D}\widehat{\boldsymbol{u}}}{\mathrm{D}\widehat{t}} + \widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}} = -\nabla\widehat{\phi} \qquad \text{(PE.1)}$$

Hydrostatic: $$\frac{\partial\widehat{\phi}}{\partial\widehat{z}} = \widehat{b} \qquad \text{(PE.2)}$$

Mass continuity: $$\left(\frac{\partial\widehat{u}}{\partial\widehat{x}} + \frac{\partial\widehat{v}}{\partial\widehat{y}} + \frac{1}{\widetilde{\rho}}\frac{\partial\widetilde{\rho}\widehat{w}}{\partial\widehat{z}}\right) = 0 \qquad \text{(PE.3)}$$

Thermodynamic: $$Ro\,\frac{\mathrm{D}\widehat{b}}{\mathrm{D}\widehat{t}} + \left(\frac{L_d}{L}\right)^2 \widehat{N}^2\widehat{w} = 0 \qquad \text{(PE.4)}$$

These equations are written for the anelastic equations in a rotating frame of reference. The Boussinesq equations result if we take $\widetilde{\rho} = 1$. The equations in pressure coordinates also have a similar form — see Section 2.6.2.

## 5.2 THE PLANETARY-GEOSTROPHIC EQUATIONS

We now use the low Rossby number scalings above to derive equation sets that are simpler than the original, 'primitive', ones. The planetary-geostrophic equations are probably the simplest such set of equations, and we derive these equations first for the shallow water equations, and then for the stratified primitive equations.

### 5.2.1 Using the Shallow Water Equations

*Informal derivation*

The advection and time derivative terms in the momentum equation (5.10) are order Rossby number smaller than the Coriolis and pressure terms (the term in square brackets is multiplied by $Ro$), and therefore let us neglect them. The momentum equation straightforwardly becomes

$$\widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}} = -\nabla\widehat{\eta}. \tag{5.33}$$

The mass conservation equation (5.12), contains two nondimensional parameters, $Ro = U/(f_0 L)$ (the Rossby number), and $F = (L/L_d)^2$ (the ratio of the length scale of the motion to the deformation scale; $F = Bu^{-1}$) and we must make a choice as to the relationship between these two numbers. We will choose

$$F\,Ro = \mathcal{O}(1), \tag{5.34}$$

which implies

$$L^2 \gg L_d^2 \qquad \text{or equivalently} \qquad F \gg 1, \quad Bu \ll 1. \tag{5.35}$$

That is to say, we suppose that the scales of motion are much larger than the deformation scale. Given this choice, all the terms in the mass conservation equation, (5.12), are of roughly the same size, and we retain them all. Thus, the shallow water planetary geostrophic equations are the full mass continuity equation along with geostrophic balance and a geometric relationship between the height field and the fluid thickness, and in dimensional form these are:

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla\cdot\boldsymbol{u} = 0, \qquad \boldsymbol{f}\times\boldsymbol{u} = -g\nabla\eta, \qquad \eta = h + \eta_b\,. \tag{5.36a,b,c}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

We emphasize that *the planetary-geostrophic equations are only valid for scales of motion much larger than the deformation radius.* The height variations are then as large as the mean height field itself; that is, using (5.8), $\Delta\eta/H = \mathcal{O}(1)$.

### *Formal derivation*

We make the following assumptions:

*(i)* The Rossby number is small. $Ro = U/f_0L \ll 1$.

*(ii)* The scale of the motion is significantly larger than the deformation scale. That is, (5.34) holds or equivalently

$$F = Bu^{-1} = \left(\frac{L}{L_d}\right)^2 \gg 1 \tag{5.37}$$

and in particular

$$F\,Ro = \mathcal{O}(1). \tag{5.38}$$

*(iii)* Time scales advectively, so that $T = L/U$.

The idea is now to expand the nondimensional velocity and height fields in an asymptotic series with the Rossby number as the small parameter, substitute into the equations of motion and derive a simpler set of equations. It is a nearly trivial exercise in this instance, and so illustrates the methodology well. The expansions are

$$\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{u}}_0 + Ro\,\widehat{\boldsymbol{u}}_1 + Ro^2\,\widehat{\boldsymbol{u}}_2 + \dots \qquad \text{and} \qquad \widehat{\eta} = \widehat{\eta}_0 + Ro\,\widehat{\eta}_1 + Ro^2\,\widehat{\eta}_2 + \dots. \tag{5.39a,b}$$

Substituting (5.39) into the momentum equation then gives

$$Ro\left[\frac{\partial \widehat{\boldsymbol{u}}_0}{\partial \hat{t}} + \widehat{\boldsymbol{u}}_0 \cdot \nabla \widehat{\boldsymbol{u}}_0 + \widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}}_1\right] + \widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}}_0 = -\nabla\widehat{\eta}_0 - Ro\left[\nabla\widehat{\eta}_1\right] + \mathcal{O}(Ro^2). \tag{5.40}$$

The Rossby number is an asymptotic ordering parameter; thus, the sum of all the terms at any particular order in Rossby number must vanish. At lowest order we obtain the simple expression

$$\widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}}_0 = -\nabla\widehat{\eta}_0. \tag{5.41}$$

Note that although $f_0$ is a representative value of $f$, we have made no assumptions about the constancy of $f$. In particular, $f$ is allowed to vary by an order one amount, provided that it does not become so small that the Rossby number $U/(fL)$ is not small.

The appropriate height (mass conservation) equation is similarly obtained by substituting (5.39) into the shallow water mass conservation equation. Because $F\,Ro = \mathcal{O}(1)$ at lowest order we simply retain all the terms in the equation to give

$$F\,Ro\left[\frac{\partial \widehat{\eta}_0}{\partial t} + \widehat{\boldsymbol{u}}_0 \cdot \nabla\widehat{\eta}_0\right] + \left[1 + F\,Ro\,\widehat{\eta}\right]\nabla\cdot\widehat{\boldsymbol{u}}_0 = 0. \tag{5.42}$$

Equations (5.41) and (5.42) are a closed set, namely the nondimensional planetary-geostrophic equations. The dimensional forms of these equations are just (5.36).

### *Variation of the Coriolis parameter*

Suppose then that $f$ is a constant ($f_0$). Then, from the curl of (5.41), $\nabla\cdot\boldsymbol{u}_0 = 0$. This means that we can define a streamfunction for the flow and, from geostrophic balance, the height field is just that streamfunction. That is, in dimensional form,

$$\psi = \frac{g}{f_0}\eta, \qquad \boldsymbol{u} = \mathbf{k}\times\nabla\psi, \tag{5.43a,b}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

and (5.42) becomes, in dimensional form,

$$\frac{\partial \eta}{\partial t} + \boldsymbol{u} \cdot \nabla \eta = 0 \qquad \text{or} \qquad \frac{\partial \eta}{\partial t} + J(\psi, \eta) = 0, \tag{5.44}$$

where $J(a, b) \equiv a_x b_y - a_y b_x$. But since $\eta \propto \psi$ the advective term is proportional to $J(\psi, \psi)$, which is zero. Thus, the flow does not evolve at this order. The planetary-geostrophic equations are *uninteresting* if the scale of the motion is such that the Coriolis parameter is not variable. On Earth, the scale of motion on which this parameter regime exists is rather limited, since the planetary-geostrophic equations require that the scale of motion also be larger than the deformation radius. In the Earth's atmosphere, any scale that is larger than the deformation radius will be such that the Coriolis parameter varies significantly over it, and we do not encounter this parameter regime. On the other hand, in the Earth's ocean the deformation radius is relatively small and there exists a small parameter regime (called the frontal geostrophic regime) that has scales larger than the deformation radius but smaller than that on which the Coriolis parameter varies.

### *Potential vorticity*

The shallow water PG equations may be written as an evolution equation for an appropriate potential vorticity. A little manipulation reveals that (5.36) are equivalent to:

$$\begin{gathered} \frac{\mathrm{D}Q}{\mathrm{D}t} = 0, \\ Q = \frac{f}{h}, \quad \boldsymbol{f} \times \boldsymbol{u} = -g \nabla \eta, \quad \eta = h + \eta_b \,. \end{gathered} \tag{5.45}$$

Thus, potential vorticity is a material invariant in the approximate equation set, just as it is in the full equations. The other variables — the free surface height and the velocity — are diagnosed from it, a process known as *potential vorticity inversion*. In the planetary geostrophic approximation, the inversion proceeds using the approximate form $f/h$ rather than the full potential vorticity, $(f + \zeta)/h$. Thus, in a strict sense, we do not approximate potential vorticity, because this is the evolving variable. Rather, we approximate the inversion relations from which we derive the height and velocity fields. The simplest way of all to derive the shallow water PG equations is to *begin* with the conservation of potential vorticity, and to note that at small Rossby number the expression $(\zeta + f)/h$ may be approximated by $f/h$. Then, noting in addition that the flow is geostrophic, (5.45) immediately emerges. *Every* approximate set of equations that we derive in this chapter may be expressed as the evolution of potential vorticity, with the other fields being obtained diagnostically from it.

## 5.2.2 Planetary-Geostrophic Equations for Stratified Flow

To explore the stratified system we will use the inviscid and adiabatic Boussinesq equations of motion with the hydrostatic approximation. The derivation carries through easily enough using the anelastic or pressure-coordinate equations, but as the PG equations have more oceanographic than atmospheric importance, using the incompressible equations is quite appropriate.

### *Simplifying the equations*

The nondimensional equations we begin with are (5.28)–(5.32). As in the shallow water case we expand these in a series in the Rossby number, so that:

$$\widehat{u} = \widehat{u}_0 + Ro\, \widehat{u}_1 + Ro^2\, \widehat{u}_2 + \dots, \qquad \widehat{b} = \widehat{b}_0 + Ro\, \widehat{b}_1 + Ro^2\, \widehat{b}_2 + \dots, \tag{5.46}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

and similarly for $\widehat{v}$, $\widehat{w}$ and $\widehat{\phi}$. Substituting into the nondimensional equations of motion (on page 176) and equating powers of $Ro$ gives the lowest-order momentum, hydrostatic, and mass conservation equations:

$$\widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}}_0 = -\nabla\widehat{\phi}_0, \qquad \frac{\partial\widehat{\phi}_0}{\partial\widehat{z}} = \hat{b}_0, \qquad \nabla\cdot\widehat{\boldsymbol{v}}_0 = 0. \tag{5.47a,b,c}$$

If we also assume that $L_d/L = \mathcal{O}(1)$, then the thermodynamic equation (5.32) becomes

$$\left(\frac{L_d}{L}\right)^2 \widehat{N}^2\widehat{w}_0 = 0. \tag{5.48}$$

Of course we have neglected any diabatic terms in this equation, which would in general provide a non-zero right-hand side. Nevertheless, this is not a useful equation, because the set of the equations we have derived, (5.47) and (5.48), can no longer evolve: all the time derivatives have been scaled away! Thus, although instructive, these equations are not very useful. If instead we assume that the scale of motion is much larger than the deformation scale then the other terms in the thermodynamic equation will become equally important. Thus, we suppose that $L_d^2 \ll L^2$ or, more formally, that $L^2 = \mathcal{O}(Ro^{-1})L_d^2$, and then all the terms in the thermodynamic equation are retained. A closed set of equations is then given by (5.47) and the thermodynamic equation (5.32).

### *Dimensional equations*

Restoring the dimensions, dropping the asymptotic subscripts, and allowing for the possibility of a source term, denoted by $S_{[b']}$, in the thermodynamic equation, the *planetary-geostrophic* equations of motion are:

$$\begin{gathered} \frac{\mathrm{D}b'}{\mathrm{D}t} + wN^2 = S_{[b']}, \\ \boldsymbol{f}\times\boldsymbol{u} = -\nabla\phi', \qquad \frac{\partial\phi'}{\partial z} = b', \qquad \nabla\cdot\boldsymbol{v} = 0. \end{gathered} \tag{5.49}$$

The thermodynamic equation may also be written simply as

$$\frac{\mathrm{D}b}{\mathrm{D}t} = \dot{b}, \tag{5.50}$$

where $b$ now represents the total stratification. The relevant pressure, $\phi$, is then the pressure that is in hydrostatic balance with $b$, so that geostrophic and hydrostatic balance are most usefully written as

$$\boldsymbol{f}\times\boldsymbol{u} = -\nabla\phi, \qquad \frac{\partial\phi}{\partial z} = b. \tag{5.51a,b}$$

### *Potential vorticity*

Manipulation of (5.49) reveals that we can equivalently write the equations as an evolution equation for potential vorticity. Thus, the evolution equations may be written as

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = \dot{Q}, \qquad Q = f\frac{\partial b}{\partial z}, \tag{5.52}$$

where $\dot{Q} = f\partial\dot{b}/\partial z$, and the inversion — i.e., the diagnosis of velocity, pressure and buoyancy — is carried out using the hydrostatic, geostrophic and mass conservation equations.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

*Applicability to the ocean and atmosphere*

In the atmosphere a typical deformation radius $NH/f$ is about 1000 km. The constraint that the scale of motion be significantly larger than the deformation radius is thus hard to satisfy, since one quickly runs out of room on a planet whose equator-to-pole distance is 10 000 km. Only the largest planetary waves can satisfy the planetary-geostrophic scaling in the atmosphere and we should then also write the equations in spherical coordinates. In the ocean the deformation radius is about 100 km, so there is lots of room for the planetary-geostrophic equations to hold, and much of the theory of the large-scale structure of the ocean involves these equations.

## 5.3 THE SHALLOW WATER QUASI-GEOSTROPHIC EQUATIONS

We now derive a set of geostrophic equations that is valid (unlike the PG equations) when the horizontal scale of motion is similar to that of the deformation radius. These equations are called the *quasi-geostrophic* equations, and are perhaps the most widely used set of equations for theoretical studies of the atmosphere and ocean. The specific assumptions we make are as follows:

*(i)* The Rossby number is small, so that the flow is in near-geostrophic balance.

*(ii)* The scale of the motion is not significantly larger than the deformation scale. Specifically, we shall require that

$$Ro\left(\frac{L}{L_d}\right)^2 = \mathcal{O}(Ro). \tag{5.53}$$

For the shallow water equations, this assumption implies, using (5.9), that the variations in fluid depth are small compared to its total depth. For the continuously stratified system it implies, using (5.23), that the variations in stratification are small compared to the background stratification.

*(iii)* Variations in the Coriolis parameter are small; that is, $|\beta L| \ll |f_0|$ where $L$ is the length scale of the motion.

*(iv)* Time scales advectively; that is, the scaling for time is given by $T = L/U$.

The second and third of these differ from the planetary-geostrophic counterparts: we make the second assumption because we wish to explore a different parameter regime, and we then find that the third assumption is necessary to avoid the rather trivial state of $\beta v = 0$ (as we discuss more below). All of the assumptions are the same whether we consider the shallow water equations or a continuously stratified flow, and in this section we consider the former.

### 5.3.1 Single-layer Shallow Water Quasi-Geostrophic Equations

The algorithm is, again, to expand the variables $\widehat{u}, \widehat{v}, \widehat{\eta}$ in an asymptotic series with the Rossby number as the small parameter, substitute into the equations of motion, and derive a simpler set of equations. Thus we let

$$\widehat{u} = \widehat{u}_o + Ro\, \widehat{u}_1 + Ro^2\, \widehat{u}_2 + \dots, \qquad \widehat{v} = \widehat{v}_o + Ro\, \widehat{v}_1 + Ro^2\, \widehat{v}_2 + \dots, \tag{5.54a}$$

$$\widehat{\eta} = \widehat{\eta}_0 + Ro\, \widehat{\eta}_1 + Ro^2\, \widehat{\eta}_2 \dots. \tag{5.54b}$$

We recognize the smallness of $\beta$ compared to $f_0/L$ by letting $\beta = \widehat{\beta} U/L^2$, where $\widehat{\beta}$ is assumed to be a parameter of order unity. Then the expression $f = f_0 + \beta y$ becomes

$$\widehat{f} = f/f_0 = \widehat{f}_0 + Ro\, \widehat{\beta}\widehat{y}, \tag{5.55}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

where $\widehat{f}_0$ is the nondimensional value of $f_0$; its value is unity, but it is helpful to denote it explicitly. Substitute (5.54) into the nondimensional momentum equation (5.10), and equate powers of *Ro*. At lowest order we obtain

$$\widehat{f}_0 \widehat{u}_0 = -\frac{\partial \widehat{\eta}_0}{\partial \widehat{y}}, \qquad \widehat{f}_0 \widehat{v}_0 = \frac{\partial \widehat{\eta}_0}{\partial \widehat{x}}. \tag{5.56}$$

Cross-differentiating gives

$$\nabla \cdot \widehat{\boldsymbol{u}}_0 = 0, \tag{5.57}$$

where, as always, when $\nabla$ operates on a nondimensional variable, the derivatives are taken with respect to the nondimensional coordinates. Evidently the velocity field is divergence-free, with this property arising from the momentum equation rather than the mass conservation equation.

The mass conservation equation is also, at lowest order, $\nabla \cdot \widehat{\boldsymbol{u}}_0 = 0$, and at next order we have

$$F \frac{\partial \widehat{\eta}_0}{\partial \widehat{t}} + F \widehat{\boldsymbol{u}}_0 \cdot \nabla \widehat{\eta}_0 + \nabla \cdot \widehat{\boldsymbol{u}}_1 = 0. \tag{5.58}$$

This equation is not closed, because the evolution of the zeroth-order term involves evaluation of a first-order quantity. For closure, we go to the next order in the momentum equation,

$$\frac{\partial \widehat{\boldsymbol{u}}_0}{\partial \widehat{t}} + (\widehat{\boldsymbol{u}}_0 \cdot \nabla) \widehat{\boldsymbol{u}}_0 + \widehat{\beta} \widehat{y} \mathbf{k} \times \widehat{\boldsymbol{u}}_0 + \widehat{f}_0 \mathbf{k} \times \widehat{\boldsymbol{u}}_1 = -\nabla \widehat{\eta}_1, \tag{5.59}$$

and take its curl to give the vorticity equation:

$$\frac{\partial \widehat{\zeta}_0}{\partial \widehat{t}} + (\widehat{\boldsymbol{u}}_0 \cdot \nabla)(\widehat{\zeta}_0 + \widehat{\beta} \widehat{y}) = -\widehat{f}_0 \nabla \cdot \widehat{\boldsymbol{u}}_1. \tag{5.60}$$

The term on the right-hand side is the *vortex stretching* term. Only vortex stretching by the background or planetary vorticity is present, because the vortex stretching by the relative vorticity is smaller by a factor of the Rossby number. Equation (5.60) is also not closed; however, we may use (5.58) to eliminate the divergence term to give

$$\frac{\partial \widehat{\zeta}_0}{\partial \widehat{t}} + (\widehat{\boldsymbol{u}}_0 \cdot \nabla)(\widehat{\zeta}_0 + \widehat{\beta} \widehat{y}) = \widehat{f}_0 \left( F \frac{\partial \widehat{\eta}_0}{\partial \widehat{t}} + F \widehat{\boldsymbol{u}}_0 \cdot \nabla \widehat{\eta}_0 \right), \tag{5.61}$$

or

$$\frac{\partial}{\partial \widehat{t}} (\widehat{\zeta}_0 - \widehat{f}_0 F \widehat{\eta}_0) + (\widehat{\boldsymbol{u}}_0 \cdot \nabla)(\widehat{\zeta}_0 + \widehat{\beta} \widehat{y} - F \widehat{f}_0 \widehat{\eta}_0) = 0. \tag{5.62}$$

The final step is to note that the lowest-order vorticity and height fields are related through geostrophic balance, so that using (5.56) we can write

$$\widehat{u}_0 = -\frac{\partial \widehat{\psi}_0}{\partial \widehat{y}}, \qquad \widehat{v}_0 = \frac{\partial \widehat{\psi}_0}{\partial \widehat{x}}, \qquad \widehat{\zeta}_0 = \nabla^2 \widehat{\psi}_0, \tag{5.63}$$

where $\widehat{\psi}_0 = \widehat{\eta}_0 / \widehat{f}_0$ is the streamfunction. Equation (5.62) can thus be written as

$$\frac{\partial}{\partial \widehat{t}} (\nabla^2 \widehat{\psi}_0 - \widehat{f}_0^2 F \widehat{\psi}_0) + (\widehat{\boldsymbol{u}}_0 \cdot \nabla)(\widehat{\zeta}_0 + \widehat{\beta} \widehat{y} - \widehat{f}_0^2 F \widehat{\psi}_0) = 0 \tag{5.64}$$

or

$$\frac{\mathrm{D}_0}{\mathrm{D}\widehat{t}} (\nabla^2 \widehat{\psi}_0 + \widehat{\beta} \widehat{y} - \widehat{f}_0^2 F \widehat{\psi}_0) = 0, \tag{5.65}$$

where the subscript '0' on the material derivative indicates that the lowest order velocity, the geostrophic velocity, is the advecting velocity. Restoring the dimensions, (5.65) becomes

$$\boxed{\frac{\mathrm{D}}{\mathrm{D}t} \left( \nabla^2 \psi + \beta y - \frac{1}{L_d^2} \psi \right) = 0,} \tag{5.66}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

where $\psi = (g/f_0)\eta$, $L_d^2 = gH/f_0^2$, and the advective derivative is

$$\frac{\mathrm{D}\,\cdot}{\mathrm{D}t} = \frac{\partial\,\cdot}{\partial t} + u_g\frac{\partial\,\cdot}{\partial x} + v_g\frac{\partial\,\cdot}{\partial y} = \frac{\partial\,\cdot}{\partial t} - \frac{\partial\psi}{\partial y}\frac{\partial\,\cdot}{\partial x} + \frac{\partial\psi}{\partial x}\frac{\partial\,\cdot}{\partial y} = \frac{\partial\,\cdot}{\partial t} + J(\psi,\cdot). \tag{5.67}$$

Another form of (5.66) is

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\zeta + \beta y - \frac{f_0}{H}\eta\right) = 0, \tag{5.68}$$

with $\zeta = (g/f_0)\nabla^2\eta$. Equations (5.66) and (5.68) are forms of the shallow water quasi-geostrophic potential vorticity equation. The quantity

$$\boxed{q \equiv \zeta + \beta y - \frac{f_0}{H}\eta = \nabla^2\psi + \beta y - \frac{1}{L_d^2}\psi} \tag{5.69}$$

is the *shallow water quasi-geostrophic potential vorticity.*

### *Connection to shallow water potential vorticity*

The quantity $q$ given by (5.69) is an approximation (except for dynamically unimportant constant additive and multiplicative factors) to the shallow water potential vorticity. To see the truth of this statement, begin with the expression for the shallow water potential vorticity,

$$Q = \frac{f + \zeta}{h}. \tag{5.70}$$

Now let $h = H(1 + \eta'/H)$, where $\eta'$ is the perturbation of the free-surface height, and assume that $\eta'/H$ is small to obtain

$$Q = \frac{f + \zeta}{H(1 + \eta'/H)} \approx \frac{1}{H}(f + \zeta)\left(1 - \frac{\eta'}{H}\right) \approx \frac{1}{H}\left(f_0 + \beta y + \zeta - f_0\frac{\eta'}{H}\right). \tag{5.71}$$

Because $f_0/H$ is a constant it has no effect in the evolution equation, and the quantity given by

$$q = \beta y + \zeta - f_0\frac{\eta'}{H} \tag{5.72}$$

is materially conserved. Using geostrophic balance we have $\zeta = \nabla^2\psi$ and $\eta' = f_0\psi/g$ so that (5.72) is identical to (5.69). Only the variation in $\eta$ is important in (5.68) or (5.69).

The approximations needed to go from (5.70) to (5.72) are the same as those used in our earlier, more long-winded, derivation of the quasi-geostrophic equations. That is, we assumed that $f$ itself is nearly constant, and that $f_0$ is much larger than $\zeta$, equivalent to a low Rossby number assumption. It was also necessary to assume that $H \gg \eta'$ to enable the expansion of the height field which, using assumption *(ii)* on page 180, is equivalent to requiring that the scale of motion not be significantly larger than the deformation scale. The derivation is completed by noting that the advection of the potential vorticity should be by the geostrophic velocity alone, and we recover (5.66) or (5.68).

### *Two interesting limits*

There are two interesting limits to the quasi-geostrophic potential vorticity equation which, taking $\beta = 0$ for simplicity, are as follows:

(i) *Motion on scales much smaller than the deformation radius.* That is, $L \ll L_d$ and thus $Bu \gg 1$ or $F \ll 1$. Then (5.66) becomes

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = 0, \tag{5.73}$$

where $\zeta = \nabla^2\psi$ and $J(\psi, \zeta) = \psi_x\zeta_y - \psi_y\zeta_x$. Thus, the motion obeys the two-dimensional vorticity equation. Physically, on small length scales the deviations in the height field are very small and may be neglected.

(ii) *Motion on scales much larger than the deformation radius.* Although scales are not allowed to become so large that $Ro(L/L_d)^2$ is of order unity, we may, a posteriori, still have $L \gg L_d$, whence the potential vorticity equation, (5.66), becomes

$$\frac{\partial \psi}{\partial t} + J(\psi, \psi) = 0 \qquad \text{or} \qquad \frac{\partial \eta}{\partial t} + J(\psi, \eta) = 0, \tag{5.74}$$

because $\psi = g\eta/f_0$. The Jacobian term evidently vanishes. Thus, one is left with a trivial equation that implies there is no advective evolution of the height field. There is nothing wrong with our reasoning; the mathematics has indeed pointed out a limit interesting in its uninterestingness. From a physical point of view, however, such a lack of motion is likely to be rare, because on such large scales the Coriolis parameter varies considerably, and we are led to the planetary-geostrophic equations.

In practice, often the most severe restriction of quasi-geostrophy is that variations in layer thickness are small: what does this have to do with geostrophy? If we scale $\eta$ assuming geostrophic balance then $\eta \sim fUL/g$ and $\eta/H \sim Ro(L/L_d)^2$. Thus, if $Ro$ is to remain small, $\eta/H$ can only be of order one if $(L/L_d)^2 \gg 1$. That is, the height variations must occur on a large scale, or we are led to a scaling inconsistency. Put another way, *if there are order-one height variations over a length scale of less than or of the order of the deformation scale, the Rossby number will not be small.* Large height variations are allowed if the scale of motion is large, but this contingency is described by the planetary-geostrophic equations.

### *Another flow regime*

Although perhaps of little terrestrial interest, we can imagine a regime in which the Coriolis parameter varies fully, but the scale of motion remains no larger than the deformation radius. This parameter regime is not quasi-geostrophic, but it gives an interesting result. Because $\eta'/H \sim Ro(L/L_d)^2$ deviations of the height field are at least of order Rossby number smaller than the reference height and $|\eta'| \ll H$. The dominant balance in the height equation is then

$$H\nabla \cdot \boldsymbol{u} = 0, \tag{5.75}$$

presuming that time still scales advectively. This zero horizontal divergence must remain consistent with geostrophic balance,

$$\boldsymbol{f} \times \boldsymbol{u} = -g\nabla\eta, \tag{5.76}$$

where now $f$ is a fully variable Coriolis parameter. Taking the curl of (that is, cross-differentiating) (5.76) gives

$$\beta v + f\nabla \cdot \boldsymbol{u} = 0, \tag{5.77}$$

whence, using (5.75), $v = 0$, and the flow is purely zonal. Although not at all useful as an evolution equation, this illustrates the constraining effect that differential rotation has on meridional velocity. This effect may be the cause of the banded, highly zonal flow on some of the giant planets, and we will revisit this issue in our discussion of geostrophic turbulence.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

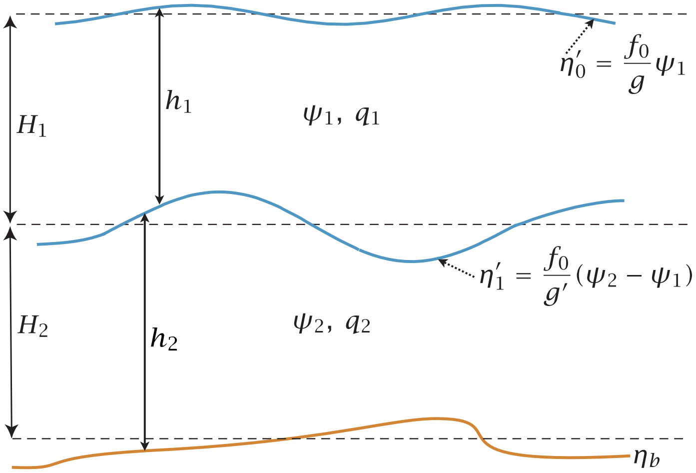


**Fig. 5.1** A quasi-geostrophic fluid system consisting of two immiscible fluids of different density. The quantities $\eta'$ are the interface displacements from the resting basic state, denoted with dashed lines, with $\eta_b$ being the bottom topography.

### 5.3.2 Two-layer and Multi-layer Quasi-Geostrophic Systems

Just as for the one-layer case, the multi-layer shallow water equations simplify to a corresponding quasi-geostrophic system in appropriate circumstances. The assumptions are virtually the same as before, although we assume that the variation in the thickness of *each* layer is small compared to its mean thickness. The basic fluid system for a two-layer case is sketched in Fig. 5.1 (and see also Fig. 3.5), and for the multi-layer case in Fig. 5.2.

Let us proceed directly from the potential vorticity equation for each layer. We will also stay in dimensional variables, foregoing a strict asymptotic approach for the sake of informality and insight, and use the Boussinesq approximation. For each layer the potential vorticity equation is just

$$\frac{\mathrm{D}Q_i}{\mathrm{D}t} = 0, \qquad Q_i = \frac{\zeta_i + f}{h_i}. \tag{5.78}$$

Let $h_i = H_i + h_i'$ where $|h_i'| \ll H_i$. The potential vorticity then becomes

$$Q_i \approx \frac{1}{H_i}(\zeta_i + f)\left(1 - \frac{h_i'}{H_i}\right) \qquad \text{— variations in layer thickness are small,} \tag{5.79a}$$

$$\approx \frac{1}{H_i}\left(f + \zeta_i - f\frac{h_i'}{H_i}\right) \qquad \text{— the Rossby number is small,} \tag{5.79b}$$

$$\approx \frac{1}{H_i}\left(f + \zeta_i - f_0\frac{h_i'}{H_i}\right) \qquad \text{— variations in Coriolis parameter are small .} \tag{5.79c}$$

Now, because Q appears in the equations only as an advected quantity, it is only the *variations* in the Coriolis parameter that are important in the first term on the right-hand side of (5.79c), and given this all three terms are of the same approximate magnitude. Then, because mean layer thicknesses are constant, we can define the quasi-geostrophic potential vorticity in each layer by

$$q_i = \left(\beta y + \zeta_i - f_0\frac{h_i'}{H_i}\right), \tag{5.80}$$

and this will evolve according to $\mathrm{D}q_i/\mathrm{D}t = 0$, where the advective derivative is by the geostrophic wind. As in the one-layer case, the quasi-geostrophic potential vorticity has different dimensions from the full shallow water potential vorticity.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

*Two-layer model*

To obtain a closed set of equations we must obtain an advecting field from the potential vorticity. We use geostrophic balance to do this, and neglecting the advective derivative in (3.49) gives

$$\boldsymbol{f}_0 \times \boldsymbol{u}_1 = -g\nabla\eta_0 = -g\nabla(h_1' + h_2' + \eta_b), \tag{5.81a}$$
$$\boldsymbol{f}_0 \times \boldsymbol{u}_2 = -g\nabla\eta_0 - g'\nabla\eta_1 = -g\nabla(h_1' + h_2' + \eta_b) - g'\nabla(h_2' + \eta_b), \tag{5.81b}$$

where $g' = g(\rho_2 - \rho_1)/\rho_1$ and $\eta_b$ is the height of any bottom topography, and, because variations in the Coriolis parameter are presumptively small, we use a constant value of $f$ (i.e., $f_0$) on the left-hand side. For each layer there is therefore a streamfunction, given by

$$\psi_1 = \frac{g}{f_0}(h_1' + h_2' + \eta_b), \qquad \psi_2 = \frac{g}{f_0}(h_1' + h_2' + \eta_b) + \frac{g'}{f_0}(h_2' + \eta_b), \tag{5.82a,b}$$

and these two equations may be manipulated to give

$$h_1' = \frac{f_0}{g'}(\psi_1 - \psi_2) + \frac{f_0}{g}\psi_1, \qquad h_2' = \frac{f_0}{g'}(\psi_2 - \psi_1) - \eta_b. \tag{5.83a,b}$$

We note as an aside that the interface displacements are given by

$$\eta_0' = \frac{f_0}{g}\psi_1, \qquad \eta_1' = \frac{f_0}{g'}(\psi_2 - \psi_1). \tag{5.84a,b}$$

Using (5.80) and (5.83) the quasi-geostrophic potential vorticity for each layer becomes

$$\begin{aligned} q_1 &= \beta y + \nabla^2\psi_1 + \frac{f_0^2}{g' H_1}(\psi_2 - \psi_1) - \frac{f_0^2}{g H_1}\psi_1, \\ q_2 &= \beta y + \nabla^2\psi_2 + \frac{f_0^2}{g' H_2}(\psi_1 - \psi_2) + f_0\frac{\eta_b}{H_2}. \end{aligned} \tag{5.85a,b}$$

In the rigid-lid approximation the last term in (5.85a) is neglected. The potential vorticity in each layer is advected by the geostrophic velocity, so that the evolution equation for each layer is just

$$\frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = 0, \qquad i = 1, 2. \tag{5.86}$$

*Multi-layer model*

A multi-layer quasi-geostrophic model may be constructed by a straightforward extension of the above two-layer procedure (see Fig. 5.2). The quasi-geostrophic potential vorticity for each layer is still given by (5.80). The pressure field in each layer can be expressed in terms of the thickness of each layer using (3.44) and (3.45), and by geostrophic balance the pressure is proportional to the streamfunction, $\psi_i$, for each layer. Carrying out these steps we obtain, after a little algebra, the following expression for the quasi-geostrophic potential vorticity of an interior layer, in the Boussinesq approximation:

$$q_i = \beta y + \nabla^2\psi_i + \frac{f_0^2}{H_i}\left(\frac{\psi_{i-1} - \psi_i}{g_{i-1}'} - \frac{\psi_i - \psi_{i+1}}{g_i'}\right), \tag{5.87}$$

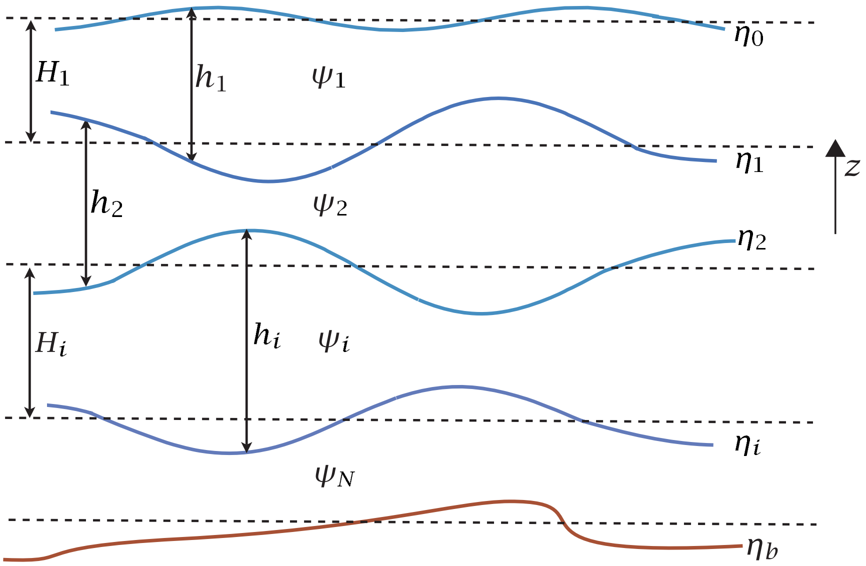


**Fig. 5.2** A multi-layer quasi-geostrophic fluid system. Layers are numbered from the top down, $i$ denotes a general interior layer and $N$ denotes the bottom layer.

and for the top and bottom layers,

$$q_1 = \beta y + \nabla^2\psi_1 + \frac{f_0^2}{H_1}\left(\frac{\psi_2 - \psi_1}{g_1'}\right) - \frac{f_0^2}{gH_1}\psi_1, \tag{5.88a}$$

$$q_N = \beta y + \nabla^2\psi_N + \frac{f_0^2}{H_N}\left(\frac{\psi_{N-1} - \psi_N}{g_{N-1}'}\right) + \frac{f_0}{H_N}\eta_b. \tag{5.88b}$$

In these equations $H_i$ is the basic-state thickness of the $i$th layer, and $g_i' = g(\rho_{i+1} - \rho_i)/\rho_1$. In each layer the evolution equation is (5.86), now for $i = 1 \ldots N$. The displacements of each interface are given, similarly to (5.84), by

$$\eta_0' = \frac{f_0}{g}\psi_1, \qquad \eta_i' = \frac{f_0}{g_i'}(\psi_{i+1} - \psi_i). \tag{5.89a,b}$$

### 5.3.3 † Non-asymptotic and Intermediate Models

The form of the derivation of the previous section suggests that we might be able to improve on the accuracy and the range of applicability of the quasi-geostrophic equations, whilst still filtering gravity waves. For example, a seemingly improved set of geostrophic evolution equations might be

$$\frac{\partial q_i}{\partial t} + \boldsymbol{u}_i \cdot \nabla q_i = 0, \tag{5.90}$$

with

$$q_i = \frac{f + \zeta_i}{h_i}, \qquad \zeta_i = \frac{\partial v_i}{\partial x} - \frac{\partial u_i}{\partial y}, \tag{5.91a,b}$$

and with the velocities given by geostrophic balance, and therefore a function of the layer depths. Thus, the vorticity, height and velocity fields may all be inverted from potential vorticity. Note that the inversion does not involve the linearization of potential vorticity about a resting state — compare (5.91a) with (5.80)] — and we might also choose to keep the full variation of the Coriolis parameter in (5.81). Thus, the model consisting of (5.90) and (5.91) contains both the planetary

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

geostrophic and quasi-geostrophic equations. However, the informality of the derivation hides the fact that this is not an asymptotically consistent set of equations: it mixes asymptotic orders in the same equation, and good conservation properties are not assured. The set above does not, in fact, exactly conserve energy. Models that are either more accurate or more general than the quasi-geostrophic or planetary-geostrophic equations yet that still filter gravity waves are called 'intermediate models'.[4]

A model that is derived asymptotically will, in general, maintain the conservation properties of the original set. To see this, albeit in a rather abstract way, suppose that the original equations (e.g., the primitive equations) may be written in nondimensional form, as

$$\frac{\partial \varphi}{\partial t} = F(\varphi, \epsilon), \tag{5.92}$$

where $\varphi$ is a set of variables, $F$ is some operator and $\epsilon$ is a small parameter, such as the Rossby number. Suppose also that this set of equations has various invariants (such as energy and potential vorticity) that hold for any value of $\epsilon$. The asymptotically derived lowest-order model (such as quasi-geostrophy) is simply a version of this equation set valid in the limit $\epsilon = 0$, and therefore it will preserve the invariants of the original set. These invariants may seem to have a different form in the simplified set: for example, in deriving the hydrostatic primitive equations from the Navier–Stokes equations the small parameter is the aspect ratio, and this multiplies the vertical velocity. Thus, in the limit of zero aspect ratio, and therefore in the primitive equations, the kinetic energy component of the energy invariant has contributions only from the horizontal velocity. In other cases, some invariants may be reduced to trivialities in the simplified set. On the other hand, there is nothing to preclude new invariants emerging that hold only in the limit $\epsilon = 0$, and enstrophy (considered later in this chapter) is one example.

## 5.4 THE CONTINUOUSLY STRATIFIED QUASI-GEOSTROPHIC SYSTEM

We now consider the quasi-geostrophic equations for the continuously stratified hydrostatic system. The primitive equations of motion are given by (5.15), and we extract the mean stratification so that the thermodynamic equation is given by (5.17). We also stay on the $\beta$-plane for simplicity. Readers who wish for a briefer, more informal derivation may peruse the box on page 193; however, it is important to realize that there is a systematic asymptotic derivation of the quasi-geostrophic equations, for it is this that ensures that the resulting equations have good conservation properties, as explained above.

### 5.4.1 Scaling and Assumptions

The scaling assumptions we make are just those we made for the shallow water system on page 180, with a deformation radius now given by $L_d = NH/f_0$. The nondimensionalization and scaling are initially precisely that of Section 5.1.2, and we obtain the following nondimensional equations:

$$\text{horizontal momentum:} \qquad Ro\,\frac{\mathrm{D}\widehat{\boldsymbol{u}}}{\mathrm{D}\hat{t}} + \widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}} = -\nabla_z \widehat{\phi}, \tag{5.93}$$

$$\text{hydrostatic:} \qquad \frac{\partial \widehat{\phi}}{\partial \widehat{z}} = \widehat{b}, \tag{5.94}$$

$$\text{mass continuity:} \qquad \frac{\partial \widehat{u}}{\partial \widehat{x}} + \frac{\partial \widehat{v}}{\partial \widehat{y}} + \frac{1}{\widetilde{\rho}}\frac{\partial \widetilde{\rho}\widehat{w}}{\partial \widehat{z}} = 0, \tag{5.95}$$

$$\text{thermodynamic:} \qquad Ro\,\frac{\mathrm{D}\widehat{b}}{\mathrm{D}\hat{t}} + \left(\frac{L_d}{L}\right)^2 \widehat{N}^2 \widehat{w} = 0. \tag{5.96}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

In Cartesian coordinates we may express the Coriolis parameter as

$$\boldsymbol{f} = \boldsymbol{f}_0 + \beta y\,\mathbf{k}, \tag{5.97}$$

where $\boldsymbol{f}_0 = f_0\,\mathbf{k}$. The variation of the Coriolis parameter is assumed to be small (this is a key difference between the quasi-geostrophic system and the planetary-geostrophic system), and in particular we shall assume that $\beta y$ is approximately the size of the relative vorticity, and so is much smaller than $f_0$ itself.[5] Thus,

$$\beta y \sim \frac{U}{L}, \qquad \beta \sim \frac{U}{L^2}, \tag{5.98}$$

and so we define an $\mathcal{O}(1)$ nondimensional beta parameter by

$$\widehat{\beta} = \frac{\beta L^2}{U} = \frac{\beta L}{Ro\, f_0}. \tag{5.99}$$

From this it follows that if $f = f_0 + \beta y$, the corresponding nondimensional version is

$$\widehat{f} = \widehat{f}_0 + Ro\, \widehat{\beta}\widehat{y}. \tag{5.100}$$

where $\widehat{f} = f/f_0$ and $\widehat{f}_0 = f_0/f_0 = 1$.

### 5.4.2 Asymptotics

We now expand the nondimensional dependent variables in an asymptotic series in Rossby number, and write

$$\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{u}}_0 + Ro\, \widehat{\boldsymbol{u}}_1 + \dots, \qquad \widehat{\phi} = \widehat{\phi}_0 + Ro\, \widehat{\phi}_1 + \dots, \qquad \hat{b} = \hat{b}_0 + Ro\, \hat{b}_1 + \dots. \tag{5.101}$$

Substituting these into the equations of motion, the lowest-order momentum equation is simply geostrophic balance,

$$\widehat{\boldsymbol{f}}_0 \times \widehat{\boldsymbol{u}}_0 = -\nabla\widehat{\phi}_0, \tag{5.102}$$

with a *constant* value of the Coriolis parameter. (Here and for the rest of this chapter we drop the subscript $z$ from the $\nabla$ operator.) From (5.102) it is evident that

$$\nabla\cdot\widehat{\boldsymbol{u}}_0 = 0. \tag{5.103}$$

Thus, the horizontal flow is, to leading order, non-divergent; this is a consequence of geostrophic balance, and is *not* a mass conservation equation. Using (5.103) in the mass conservation equation, (5.95), gives

$$\frac{\partial}{\partial\widehat{z}}(\widetilde{\rho}\widehat{w}_0) = 0, \tag{5.104}$$

which implies that if $w_0$ is zero somewhere (e.g., at a solid surface) then $w_0$ is zero everywhere (essentially the Taylor–Proudman effect). A physical way of saying this is that the scaling estimate $W = UH/L$ is an overestimate of the size of the vertical velocity, because even though $\partial w/\partial z \approx -\nabla\cdot\boldsymbol{u}$, the horizontal divergence of the geostrophic flow is small if $f$ is nearly constant and $|\nabla\cdot\boldsymbol{u}| \ll U/L$. We might have anticipated this from the outset, and scaled $w$ differently, perhaps using the geostrophic vorticity balance estimate, $w \sim \beta UH/f_0 = Ro\, UH/L$, as the scaling factor for $w$, but there is no a-priori guarantee that this would be correct.

At next order the momentum equation is

$$\frac{\mathrm{D}_0\widehat{\boldsymbol{u}}_0}{\mathrm{D}\hat{t}} + \widehat{\beta}\widehat{y}\mathbf{k}\times\widehat{\boldsymbol{u}}_0 + \widehat{\boldsymbol{f}}\times\widehat{\boldsymbol{u}}_1 = -\nabla\widehat{\phi}_1, \tag{5.105}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

where $D_0/Dt = \partial/\partial\hat{t} + (\hat{\boldsymbol{u}}_0 \cdot \nabla)$, and the next order mass conservation equation is

$$\nabla_z \cdot (\tilde{\rho}\hat{\boldsymbol{u}}_1) + \frac{\partial}{\partial z}(\tilde{\rho}\widehat{w}_1) = 0. \tag{5.106}$$

From (5.96), the lowest-order thermodynamic equation is just

$$\left(\frac{L_d}{L}\right)^2 \widehat{N}^2 \widehat{w}_0 = 0, \tag{5.107}$$

provided that, as we have assumed, the scales of motion are not sufficiently large that $Ro(L/L_d)^2 = \mathcal{O}(1)$. (This is a key difference between quasi-geostrophy and planetary geostrophy.) At next order we obtain an evolution equation for the buoyancy, and this is

$$\frac{\mathrm{D}_0 \hat{b}_0}{\mathrm{D}\hat{t}} + \widehat{w}_1 \widehat{N}^2 \left(\frac{L_d}{L}\right)^2 = 0. \tag{5.108}$$

*The potential vorticity equation*

To obtain a single evolution equation for lowest-order quantities we eliminate $w_1$ between the thermodynamic and momentum equations. Cross-differentiating the first-order momentum equation (5.105) gives the vorticity equation,

$$\frac{\partial \widehat{\zeta}_0}{\partial \hat{t}} + (\hat{\boldsymbol{u}}_0 \cdot \nabla)\widehat{\zeta}_0 + \hat{v}_0 \widehat{\beta} = -\widehat{f}_0 \nabla_z \cdot \hat{u}_1. \tag{5.109}$$

(In dimensional terms, the divergence on the right-hand side is small, but is multiplied by the large term $f_0$, and their product is of the same order as the terms on the left-hand side.) Using the mass conservation equation (5.106), (5.109) becomes

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}(\zeta_0 + \widehat{f}) = \frac{\widehat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial z}(w_1 \tilde{\rho}). \tag{5.110}$$

Combining (5.110) and (5.108) gives

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}(\zeta_0 + \widehat{f}) = -\frac{\widehat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial \hat{z}} \left[\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}(F\tilde{\rho}\hat{b}_0)\right], \tag{5.111}$$

where $F \equiv (L/\widehat{N}L_d)^2$. The right-hand side of this equation is

$$\frac{\partial}{\partial \hat{z}}\left(\frac{\mathrm{D}_0 \hat{b}_0}{\mathrm{D}\hat{t}}\right) = \frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}\left(\frac{\partial \hat{b}_0}{\partial \hat{z}}\right) + \frac{\partial \hat{\boldsymbol{u}}_0}{\partial \hat{z}} \cdot \nabla \hat{b}_0. \tag{5.112}$$

The second term on the right-hand side vanishes identically using the thermal wind equation

$$\mathbf{k} \times \frac{\partial \hat{\boldsymbol{u}}_0}{\partial \hat{z}} = -\frac{1}{\widehat{f}_0} \nabla \hat{b}_0, \tag{5.113}$$

and so (5.111) becomes

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}\left[\widehat{\zeta}_0 + \widehat{f} + \frac{\widehat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial \hat{z}} \left(\tilde{\rho} F \hat{b}_0\right)\right] = 0, \tag{5.114}$$

or, after using the hydrostatic equation,

$$\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}}\left[\widehat{\zeta}_0 + \widehat{f} + \frac{\widehat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial \hat{z}} \left(\tilde{\rho} F \frac{\partial \widehat{\phi}_0}{\partial z}\right)\right] = 0. \tag{5.115}$$

Since the lowest-order horizontal velocity is divergence-free, we can define a streamfunction $\widehat{\psi}$ such that

$$\widehat{u}_0 = -\frac{\partial \widehat{\psi}}{\partial \widehat{y}}, \qquad \widehat{v}_0 = \frac{\partial \widehat{\psi}}{\partial \widehat{x}}, \tag{5.116}$$

where also, using (5.102), $\phi_0 = \widehat{f}_0 \widehat{\psi}$. The vorticity is then given by $\widehat{\zeta}_0 = \nabla^2 \widehat{\psi}$ and (5.115) becomes a single equation in a single unknown:

$$\boxed{\frac{\mathrm{D}_0}{\mathrm{D}\hat{t}} \left[ \nabla^2 \widehat{\psi} + \widehat{\beta} \widehat{y} + \frac{\widehat{f}_0^2}{\widetilde{\rho}} \frac{\partial}{\partial \widehat{z}} \left( \widetilde{\rho} F \frac{\partial \widehat{\psi}}{\partial \widehat{z}} \right) \right] = 0,} \tag{5.117}$$

where the material derivative is evaluated using $\widehat{\boldsymbol{u}}_0 = \mathbf{k} \times \nabla \widehat{\psi}$. This is the nondimensional form of the quasi-geostrophic potential vorticity equation, one of the most important equations in dynamical meteorology and oceanography. In deriving it we have reduced the Navier–Stokes equations, which are six coupled nonlinear partial differential equations in six unknowns $(u, v, w, T, p, \rho)$ to a single (albeit nonlinear) first-order partial differential equation in a single unknown.[6]

### *Dimensional equations*

The dimensional version of the quasi-geostrophic potential vorticity equation may be written as

$$\boxed{\frac{\mathrm{D}q}{\mathrm{D}t} = 0, \qquad q = \nabla^2 \psi + f + \frac{f_0^2}{\widetilde{\rho}} \frac{\partial}{\partial z} \left( \frac{\widetilde{\rho}}{N^2} \frac{\partial \psi}{\partial z} \right),} \tag{5.118a,b}$$

where only the variable part of $f$ (e.g., $\beta y$) is relevant in the second term on the right-hand side of the expression for $q$. The quantity $q$ is known as the *quasi-geostrophic potential vorticity.* It is analogous to the exact (Ertel) potential vorticity (see Section 5.5 for more about this), and it is conserved when advected by the *horizontal* geostrophic flow. All the other dynamical variables may be obtained from potential vorticity as follows:

*(i)* Streamfunction, using (5.118b).

*(ii)* Velocity: $\boldsymbol{u} = \mathbf{k} \times \nabla \psi$ $[\equiv \nabla^{\perp} \psi = -\nabla \times (\mathbf{k} \psi)]$.

*(iii)* Relative vorticity: $\zeta = \nabla^2 \psi$ .

*(iv)* Perturbation pressure: $\phi = f_0 \psi$.

*(v)* Perturbation buoyancy: $b' = f_0 \partial \psi / \partial z$.

The length scale, $L_d = NH/f_0$, emerges naturally from the quasi-geostrophic dynamics. It is the scale at which buoyancy and relative vorticity effects contribute equally to the potential vorticity, and is called the *deformation radius;* it is analogous to the quantity $\sqrt{gH}/f_0$ arising in shallow water theory. In the upper ocean, with $N \approx 10^{-2}\,\mathrm{s}^{-1}$, $H \approx 10^3\,\mathrm{m}$ and $f_0 \approx 10^{-4}\,\mathrm{s}^{-1}$, then $L_d \approx 100\,\mathrm{km}$. At high latitudes the ocean is much less stratified and $f$ is somewhat larger, and the deformation radius may be as little as 30 km (see Fig. 12.13 on page 469, where the deformation radius is defined slightly differently). In the atmosphere, with $N \approx 10^{-2}\,\mathrm{s}^{-1}$, $H \approx 10^4\,\mathrm{m}$, then $L_d \approx 1000\,\mathrm{km}$. It is this order of magnitude difference in the deformation scales that accounts for a great deal of the quantitative difference in the dynamics of the ocean and the atmosphere. If we take the limit $L_d \to \infty$ then the stratified quasi-geostrophic equations reduce to

$$\frac{\mathrm{D}q}{\mathrm{D}t} = 0, \qquad q = \nabla^2 \psi + f. \tag{5.119}$$

This is the two-dimensional vorticity equation, identical to (4.67). The high stratification of this limit has suppressed all vertical motion, and variations in the flow become confined to the horizontal plane. Finally, we note that it is typical in quasi-geostrophic applications to omit the prime on the buoyancy perturbations, and write $b = f_0 \partial\psi/\partial z$; however, we will keep the prime in this chapter.

### 5.4.3 Buoyancy Advection at the Surface

The solution of the elliptic equation in (5.118) requires vertical boundary conditions on $\psi$ at the ground and at the top of the atmosphere, and these are given by use of the thermodynamic equation. For a flat, slippery, rigid surface the vertical velocity is zero so that the thermodynamic equation may be written as

$$\frac{\mathrm{D}b'}{\mathrm{D}t} = 0, \qquad b' = f_0 \frac{\partial\psi}{\partial z}. \tag{5.120}$$

We apply this at the ground and at the tropopause, treating the latter as a lid on the lower atmosphere. In the presence of friction and topography the vertical velocity is not zero, but is given by

$$w = r\nabla^2\psi + \boldsymbol{u}\cdot\nabla\eta_b, \tag{5.121}$$

where the first term represents Ekman friction (with the constant $r$ proportional to the thickness of the Ekman layer) and the second term represents topographic forcing. The boundary condition becomes

$$\frac{\partial}{\partial t}\left(f_0\frac{\partial\psi}{\partial z}\right) + \boldsymbol{u}\cdot\nabla\left(f_0\frac{\partial\psi}{\partial z} + N^2\eta_b\right) + N^2 r\nabla^2\psi = 0, \tag{5.122}$$

where all the fields are evaluated at $z = 0$ or at $z = H$, the height of the lid. Thus, the quasi-geostrophic system is characterized by the horizontal advection of potential vorticity in the interior and the advection of buoyancy at the boundary. Instead of a lid at the top, then in a compressible fluid such as the atmosphere we may suppose that all disturbances tend to zero as $z \to \infty$.

#### ✦ *A potential vorticity sheet at the boundary*

Rather than regarding buoyancy advection as providing the boundary condition, it is sometimes useful to think of there being a very thin sheet of potential vorticity just above the ground and another just below the lid, specifically with a vertical distribution proportional to $\delta(z-\epsilon)$ or $\delta(z-H+\epsilon)$, where $\epsilon$ is small. The boundary condition (5.120) or (5.122) can be replaced by this, along with the condition that there are no variations of buoyancy at the boundary and $\partial\psi/\partial z = 0$ at $z = 0$ and $z = H$.[7]

To see this, we first note that the differential of a step function is a delta function. Thus, a discontinuity in $\partial\psi/\partial z$ at a level $z = z_1$ is equivalent to a delta function in potential vorticity there:

$$q(z_1) = \left[\frac{f_0^2}{N^2}\frac{\partial\psi}{\partial z}\right]_{z_1-}^{z_1+}\delta(z-z_1). \tag{5.123}$$

Now, suppose that the lower boundary condition, given by (5.120), has some arbitrary distribution of buoyancy on it. We can replace this condition by the simpler condition $\partial\psi/\partial z = 0$ at $z = 0$, provided we also add to our definition of potential vorticity a term given by (5.123) with $z_1 = \epsilon$. This term is then advected by the horizontal flow, as are the other contributions. A buoyancy source at the boundary must similarly be treated as a sheet of potential vorticity source in the interior. Any flow with buoyancy variations over a horizontal boundary is thus equivalent to a flow with uniform buoyancy at the boundary, but with a spike in potential vorticity adjacent to the boundary. This approach brings notational and conceptual advantages, in that now everything is expressed in terms of potential vorticity and its advection. However, in practice there may be less to be gained, because the boundary terms must still be included in any particular calculation that is to be performed.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

### 5.4.4 Vertical Velocity and the Omega Equation

The vertical velocity is not needed in order to evolve the quasi-geostrophic equations. However, it is not zero and a relatively simple recipe can be found that is of practical use in diagnosing the vertical velocity in weather charts. When deriving the potential vorticity equation, we eliminated vertical velocity from the vorticity equation and thermodynamic equations to give a single evolution equation. Here our approach is complementary: we begin with the same two equations, but eliminate the time derivatives. We will proceed using dimensional variables and write the vorticity and thermodynamic equations as

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = \frac{f_0}{\tilde{\rho}} \frac{\partial (\tilde{\rho} w)}{\partial z} + \mathcal{Z}, \qquad \frac{\partial b}{\partial z} + J(\psi, b) + wN^2 = Q, \tag{5.124a,b}$$

where $b = f_0 \partial\psi/\partial z$ and $\zeta = \nabla^2\psi$, $\mathcal{Z}$ and $Q$ are friction and heating terms that we can leave unspecified, and $\tilde{\rho}$ is a reference density profile. If we take $f_0\partial/\partial z$ of the first equation and $\nabla^2$ of the second we can eliminate time derivatives to find

$$\frac{\partial}{\partial z}\left[\frac{f_0^2}{\tilde{\rho}} \frac{\partial (\tilde{\rho} w)}{\partial z}\right] + N^2\nabla^2 w = f_0 \frac{\partial}{\partial z}[J(\psi, \zeta)] - \nabla^2[J(\psi, b)] - f_0 \frac{\partial \mathcal{Z}}{\partial z} + \nabla^2 Q. \tag{5.125}$$

The equation is called the *omega equation* because omega ($\omega$) is the vertical velocity in pressure coordinates, which was where the equation first appeared. It is an elliptic equation for $w$, and is in fact a Poisson equation if $\tilde{\rho}$ is a constant. It may be easily solved by numerical methods, given the state of the flow at any given time. However, there is rarely a need to solve it exactly, for there is no need to calculate $w$ to step forward the equations. Rather, the equation finds use as an interpretive guide for meteorologists: in the thermodynamic equation both heating itself and warm advection will tend to produce vertical motion, as will the vertical differential of vorticity advection.

### 5.4.5 Quasi-Geostrophy in Pressure Coordinates

The derivation of the quasi-geostrophic system in pressure coordinates is very similar to that in height coordinates, with the main difference coming at the boundaries, and we give only the results. The starting point is the primitive equations in pressure coordinates, (P.1) on page 81. In pressure coordinates, the quasi-geostrophic potential vorticity is found to be

$$q = f + \nabla^2\psi + \frac{\partial}{\partial p}\left(\frac{f_0^2}{S^2}\frac{\partial \psi}{\partial p}\right), \tag{5.126}$$

where $\psi = \Phi/f_0$ is the streamfunction and $\Phi$ the geopotential, and

$$S^2 \equiv -\frac{R}{p}\left(\frac{p}{p_R}\right)^{\kappa} \frac{\mathrm{d}\tilde{\theta}}{\mathrm{d}p} = -\frac{1}{\rho\theta}\frac{\mathrm{d}\tilde{\theta}}{\mathrm{d}p}, \tag{5.127}$$

where $\tilde{\theta}$ is a reference profile and a function of pressure only. In log-pressure coordinates, with $Z = -H \ln p$, the potential vorticity may be written as

$$q = f + \nabla^2\psi + \frac{1}{\rho_*}\frac{\partial}{\partial Z}\left(\frac{\rho_* f_0^2}{N_Z^2}\frac{\partial \psi}{\partial Z}\right), \tag{5.128}$$

where

$$N_Z^2 = S^2\left(\frac{p}{H}\right)^2 = -\left(\frac{R}{H}\right)\left(\frac{p}{p_R}\right)^{\kappa}\frac{\mathrm{d}\tilde{\theta}}{\mathrm{d}Z} \tag{5.129}$$

### Informal Derivation of Stratified QG Equations

We will use the Boussinesq equations, but very similar derivations could be given using the anelastic equations or pressure coordinates. The first ingredient is the vertical component of the vorticity equation, (4.66); in the Boussinesq version there is no baroclinic term and we have

$$\frac{\mathrm{D}_3}{\mathrm{D}t}(\zeta + f) = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \left(\frac{\partial u}{\partial z}\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\frac{\partial w}{\partial x}\right). \tag{QG.1}$$

We now apply the assumptions on page 180. The advection and the vorticity on the left-hand side are geostrophic, but we keep the horizontal divergence (which is small) on the right-hand side where it is multiplied by the big term $f$. Furthermore, because $f$ is nearly constant we replace it with $f_0$ except where it is differentiated. The second term (tilting) on the right-hand side is smaller than the advection terms on the left-hand side by the ratio $[UW/(HL)]/[U^2/L^2] = [W/H]/[U/L] \ll 1$, because $w$ is small ($\partial w/\partial z$ equals the divergence of the ageostrophic velocity). We therefore neglect it, and given all this (QG.1) becomes

$$\frac{\mathrm{D}_g}{\mathrm{D}t}(\zeta_g + f) = -f_0\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = f_0\frac{\partial w}{\partial z}, \tag{QG.2}$$

where the second equality uses mass continuity and $\mathrm{D}_g/\mathrm{D}t = \partial/\partial t + \boldsymbol{u}_g \cdot \nabla$.

The second ingredient is the three-dimensional thermodynamic equation,

$$\frac{\mathrm{D}_3 b}{\mathrm{D}t} = 0. \tag{QG.3}$$

The stratification is assumed to be nearly constant, so we write $b = \tilde{b}(z) + b'(x, y, z, t)$, where $\tilde{b}$ is the basic state buoyancy. Furthermore, because $w$ is small it only advects the basic state, and with $N^2 = \partial\tilde{b}/\partial z$ (QG.3) becomes

$$\frac{\mathrm{D}_g b'}{\mathrm{D}t} + wN^2 = 0. \tag{QG.4}$$

Hydrostatic and geostrophic wind balance enable us to write the geostrophic velocity, vorticity, and buoyancy in terms of streamfunction $\psi$ [$= p/(f_0\rho_0)$]:

$$\boldsymbol{u}_g = \mathbf{k} \times \nabla\psi, \qquad \zeta_g = \nabla^2\psi, \qquad b' = f_0\partial\psi/\partial z. \tag{QG.5}$$

The quasi-geostrophic potential vorticity equation is obtained by eliminating $w$ between (QG.2) and (QG.4), and this gives

$$\frac{\mathrm{D}_g q}{\mathrm{D}t} = 0, \qquad q = \zeta_g + f + \frac{\partial}{\partial z}\left(\frac{f_0 b'}{N^2}\right). \tag{QG.6}$$

This equation is the Boussinesq version of (5.118), and using (QG.5) it may be expressed entirely in terms of the streamfunction, with $\mathrm{D}_g \cdot/\mathrm{D}t = \partial/\partial t + J(\psi, \cdot)$. The vertical boundary conditions, at $z = 0$ and $z = H$ say, are given by (QG.4) with $w = 0$, with straightforward generalizations if topography or friction are present.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

is the buoyancy frequency and $\rho_* = \exp(-z/H)$. Temperature and potential temperature are related to the streamfunction by

$$T = -\frac{f_0 p}{R}\frac{\partial \psi}{\partial p} = \frac{H f_0}{R}\frac{\partial \psi}{\partial Z}, \tag{5.130a}$$

$$\theta = -\left(\frac{p_R}{p}\right)^{\kappa}\left(\frac{f_0 p}{R}\right)\frac{\partial \psi}{\partial p} = \left(\frac{p_R}{p}\right)^{\kappa}\left(\frac{H f_0}{R}\right)\frac{\partial \psi}{\partial Z}. \tag{5.130b}$$

In pressure or log-pressure coordinates, potential vorticity is advected along isobaric surfaces, analogously to the horizontal advection in height coordinates.

The surface boundary condition again is derived from the thermodynamic equation. In log-pressure coordinates this is

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\partial \psi}{\partial Z}\right) + \frac{N_Z^2}{f_0}W = 0, \tag{5.131}$$

where $W = \mathrm{D}Z/\mathrm{D}t$. This is not the real vertical velocity, $w$, but it is related to it by

$$w = \frac{f_0}{g}\frac{\partial \psi}{\partial t} + \frac{RT}{gH}W. \tag{5.132}$$

Thus, choosing $H = RT(0)/g$, we have, at $Z = 0$,

$$\frac{\partial}{\partial t}\left(\frac{\partial \psi}{\partial Z} - \frac{N_Z^2}{g}\psi\right) + \boldsymbol{u}\cdot\nabla\frac{\partial \psi}{\partial Z} = -\frac{N^2}{f_0}w, \tag{5.133}$$

where

$$w = \boldsymbol{u}\cdot\nabla\eta_b + r\nabla^2\psi. \tag{5.134}$$

This differs from the expression in height coordinates only by the second term in the local time derivative. In applications where accuracy is not the main issue the simpler boundary condition $\mathrm{D}(\partial_Z\psi)/\mathrm{D}t = 0$ is sometimes used. Finally, we remark that in pressure coordinates, the equivalent to vertical velocity, $\partial p/\partial t$, is denoted $\omega$ (omega), but it need not be evaluated to solve the equations.

### 5.4.6 The Two-level Quasi-Geostrophic System

The quasi-geostrophic system has, in general, continuous variation in the vertical direction (and horizontal, of course). By finite-differencing the continuous equations we can obtain a *multi-level* model, and a crude but important special case of this is the *two-level* model, also known as the Phillips model.[8] To obtain the equations of motion one way to proceed is to take a crude finite difference of the continuous relation between potential vorticity and streamfunction given in (5.118b). In the Boussinesq case (or in pressure coordinates, with a slight reinterpretation of the meaning of the symbols) the continuous expression for potential vorticity is

$$q = \zeta + f + \frac{\partial}{\partial z}\left(\frac{f_0 b'}{N^2}\right), \tag{5.135}$$

where $b' = f_0\partial\psi/\partial z$. In the case with a flat bottom and rigid lid at the top (and incorporating topography is an easy extension) the boundary condition of $w = 0$ is satisfied by $\mathrm{D}\partial_z\psi/\mathrm{D}t = 0$ at the top and bottom. An obvious finite-differencing of (5.135) in the vertical direction (see Fig. 5.3) then gives

$$q_1 = \zeta_1 + f + \frac{2f_0^2}{N^2 H_1 H}(\psi_2 - \psi_1), \qquad q_2 = \zeta_2 + f + \frac{2f_0^2}{N^2 H_2 H}(\psi_1 - \psi_2). \tag{5.136}$$

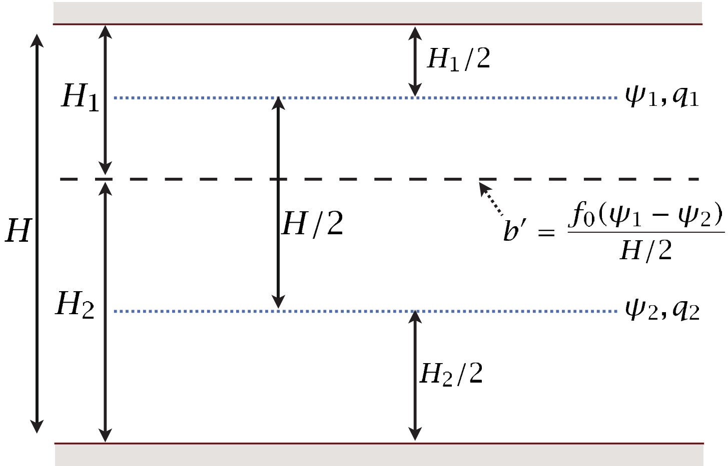


**Fig. 5.3** A two-level quasi-geostrophic system with a flat bottom and rigid lid at which $w = 0$.

In atmospheric problems it is common to choose $H_1 = H_2$, whereas in oceanic problems we might choose to have a thinner upper layer, representing the flow above the main thermocline. Note that the boundary conditions of $w = 0$ at the top and bottom are already taken care of in (5.136): *they are incorporated into the definition of the potential vorticity* — a finite-difference analogue of the delta-function construction of Section 5.4.3. At each level the potential vorticity is advected by the streamfunction so that the evolution equation for each level is:

$$\frac{\mathrm{D}q_i}{\mathrm{D}t} = \frac{\partial q_i}{\partial t} + \boldsymbol{u}_i \cdot q_i = \frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = 0, \qquad i = 1, 2. \tag{5.137}$$

Models with more than two levels can be constructed by extending the finite-differencing procedure in a natural way.

### *Connection to the layered system*

The two-level expressions, (5.136), have an obvious similarity to the *two-layer* expressions, (5.85). Noting that $N^2 = \partial\hat{b}/\partial z$ and that $b = -g\delta\rho/\rho_0$ it is natural to let

$$N^2 = -\frac{g}{\rho_0}\frac{\rho_1 - \rho_2}{H/2} = \frac{g'}{H/2}. \tag{5.138}$$

With this identification we find that (5.136) becomes

$$q_1 = \zeta_1 + f + \frac{f_0^2}{g'H_1}(\psi_2 - \psi_1), \qquad q_2 = \zeta_2 + f + \frac{f_0^2}{g'H_2}(\psi_1 - \psi_2). \tag{5.139}$$

These expressions are identical to (5.85) in the flat-bottomed, rigid lid case. Similarly, a multi-layered system with $n$ layers is equivalent to a finite-difference representation with $n$ levels. It should be said, though, that in the pantheon of quasi-geostrophic models the two-level and two-layer models hold distinguished places.

## 5.5 ✦ QUASI-GEOSTROPHY AND ERTEL POTENTIAL VORTICITY

When using the shallow water equations, quasi-geostrophic theory could be naturally developed beginning with the expression for potential vorticity. Is such an approach possible for the stratified primitive equations? The answer is yes, but with complications.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

### 5.5.1 ✦ Using Height Coordinates

Noting the general expression, (4.117), for potential vorticity in a hydrostatic fluid, the potential vorticity in the Boussinesq hydrostatic equations is given by

$$Q = \left[(v_x - u_y)b_z - v_z b_x + u_z b_y + f b_z\right], \tag{5.140}$$

where the $x, y, z$ subscripts denote derivatives. Without approximation, we write the stratification as $b = \tilde{b}(z) + b'(x, y, z, t)$, and (5.140) becomes

$$Q = [f_0 N^2] + [(\beta y + \zeta)N^2 + f_0 b'_z] + [(\beta y + \zeta)b'_z - (v_z b'_x - u_z b'_y)], \tag{5.141}$$

where, under quasi-geostrophic scaling, the terms in square brackets are in decreasing order of size. Neglecting the third term, and taking the velocity and buoyancy fields to be in geostrophic and thermal wind balance, we can write the potential vorticity as $Q \approx \widetilde{Q} + Q'$, where $\widetilde{Q} = f_0 N^2$ and

$$Q' = (\beta y + \zeta)N^2 + f_0 b'_z = (\beta y + \nabla^2\psi)N^2 + f_0^2 \frac{\partial^2 \psi}{\partial z^2}. \tag{5.142}$$

The potential vorticity evolution equation is then

$$\frac{\mathrm{D}Q'}{\mathrm{D}t} + w\frac{\partial \widetilde{Q}}{\partial z} = 0. \tag{5.143}$$

The vertical advection is important only in advecting the basic state potential vorticity $\widetilde{Q}$ and so, neglecting $w\partial Q'/\partial z$ and dividing by $N^2$, (5.143) becomes

$$\frac{\partial q_*}{\partial t} + \boldsymbol{u}_g \cdot \nabla q_* + \frac{w}{N^2}\frac{\partial \widetilde{Q}}{\partial z} = 0, \tag{5.144}$$

where $\widehat{q}$ is

$$q_* = (\beta y + \zeta) + \frac{f_0}{N^2} b'_z. \tag{5.145}$$

This is the approximation to the (perturbation) Ertel potential vorticity in the quasi-geostrophic limit. However, it is not the same as the expression for the quasi-geostrophic potential vorticity, (5.118b) and, furthermore, (5.144) involves a vertical advection. (Thus, we might refer to the expression in (5.118) as the 'quasi-geostrophic pseudopotential vorticity', but the prefix 'quasi-geostrophic' alone normally suffices.) We can derive (5.118) by eliminating $w$ between (5.144) and the quasi-geostrophic thermodynamic equation $\partial b'/\partial t + \boldsymbol{u}_g \cdot \nabla b' + w\partial\tilde{b}/\partial z = 0$.

### 5.5.2 Using Isentropic Coordinates

An illuminating and somewhat simpler path from Ertel potential vorticity to the quasi-geostrophic equations goes by way of isentropic coordinates.[9] We begin with the isentropic expression for the Ertel potential vorticity of an ideal gas,

$$Q = \frac{f + \zeta}{\sigma}, \tag{5.146}$$

where $\sigma = -\partial p/\partial\theta$ is the thickness density (which we will just call the thickness), and in adiabatic flow the potential vorticity is advected along isopycnals. We now employ quasi-geostrophic scaling to derive an approximate equation set from this. First, assume that variations in thickness are small compared with the reference state, so that

$$\sigma = \widetilde{\sigma}(\theta) + \sigma', \qquad |\sigma'| \ll |\sigma|, \tag{5.147}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

and similarly for pressure and density. Assuming also that the variations in the Coriolis parameter are small, then on the $\beta$–plane (5.146) becomes

$$Q \approx \left[\frac{f_0}{\widetilde{\sigma}}\right] + \left[\frac{1}{\widetilde{\sigma}}(\zeta + \beta y) - \frac{f_0}{\widetilde{\sigma}}\frac{\sigma'}{\widetilde{\sigma}}\right]. \tag{5.148}$$

We now use geostrophic and hydrostatic balance to express the terms on the right-hand side in terms of a single variable, noting that the first term does not vary along isentropic surfaces. Hydrostatic balance is

$$\frac{\partial M}{\partial \theta} = \Pi, \tag{5.149}$$

where $M = c_p T + gz$ and $\Pi = c_p(p/p_R)^\kappa$. Writing $M = \widetilde{M}(\theta) + M'$ and $\Pi = \widetilde{\Pi}(\theta) + \Pi'$, where $\widetilde{M}$ and $\widetilde{\Pi}$ are hydrostatically balanced reference profiles, we obtain

$$\frac{\partial M'}{\partial \theta} = \Pi' \approx \frac{\mathrm{d}\widetilde{\Pi}}{\mathrm{d}p}p' = \frac{1}{\theta\widetilde{\rho}}p', \tag{5.150}$$

where the last equality follows using the equation of state for an ideal gas and $\widetilde{\rho}$ is a reference profile. The perturbation thickness field may then be written as

$$\sigma' = -\frac{\partial}{\partial \theta}\left(\widetilde{\rho}\theta\frac{\partial M'}{\partial \theta}\right). \tag{5.151}$$

Geostrophic balance is $\boldsymbol{f}_0 \times \boldsymbol{u} = -\nabla_\theta M'$ where the velocity, and the horizontal derivatives, are along isentropic surfaces. This enables us to define a flow streamfunction by

$$\psi \equiv \frac{M'}{f_0}, \tag{5.152}$$

and we can then write all the variables in terms of $\psi$:

$$u = -\left(\frac{\partial \psi}{\partial y}\right)_\theta, \qquad v = \left(\frac{\partial \psi}{\partial x}\right)_\theta, \qquad \zeta = \nabla_\theta^2\psi, \qquad \sigma' = -f_0\frac{\partial}{\partial \theta}\left(\widetilde{\rho}\theta\frac{\partial \psi'}{\partial \theta}\right). \tag{5.153}$$

Using (5.148), (5.152) and (5.153), the quasi-geostrophic system in isentropic coordinates may be written

$$\boxed{\frac{\mathrm{D}q}{\mathrm{D}t} = 0, \qquad q = f + \nabla_\theta^2\psi + \frac{f_0^2}{\widetilde{\sigma}}\frac{\partial}{\partial \theta}\left(\widetilde{\rho}\theta\frac{\partial \psi}{\partial \theta}\right),} \tag{5.154a,b}$$

where the advection of potential vorticity is by the geostrophically balanced flow, along isentropes. The variable $q$ is an approximation to the second term in square brackets in (5.148), multiplied by $\widetilde{\sigma}$.

### *Projection back to physical-space coordinates*

We can recover the height or pressure coordinate quasi-geostrophic systems by projecting (5.154) on to the appropriate coordinate. This is straightforward because, by assumption, the isentropes in a quasi-geostrophic system are nearly flat. Recall that, from (2.142), a transformation between vertical coordinates may be effected by

$$\left.\frac{\partial}{\partial x}\right|_\theta = \left.\frac{\partial}{\partial x}\right|_p + \left.\frac{\partial p}{\partial x}\right|_\theta \frac{\partial}{\partial p}, \tag{5.155}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

but the second term is $\mathcal{O}(Ro)$ smaller than the first one because, under quasi-geostrophic scaling, isentropic slopes are small. Thus $\nabla_\theta^2\psi$ in (5.154b) may be replaced by $\nabla_p^2\psi$ or $\nabla_z^2\psi$. The vortex stretching term in (5.154) becomes, in pressure coordinates,

$$\frac{f_0^2}{\widetilde{\sigma}}\frac{\partial}{\partial\theta}\left(\widetilde{\rho}\theta\frac{\partial\psi}{\partial\theta}\right) \approx \frac{f_0^2}{\widetilde{\sigma}}\frac{\mathrm{d}\widetilde{p}}{\mathrm{d}\theta}\frac{\partial}{\partial p}\left(\widetilde{\rho}\theta\frac{\mathrm{d}\widetilde{p}}{\mathrm{d}\theta}\frac{\partial\psi}{\partial p}\right) = \frac{\partial}{\partial p}\left(\frac{f_0^2}{S^2}\frac{\partial\psi}{\partial p}\right), \tag{5.156}$$

where $S^2$ is given by (5.127). The expression for the quasi-geostrophic potential vorticity in isentropic coordinates is thus approximately equal to the quasi-geostrophic potential vorticity in pressure coordinates. This near-equality holds because the isentropic expression, (5.154b), does not contain a component proportional to the mean stratification: the second square-bracketed term on the right-hand side of (5.148) is the only dynamically relevant one, and its evolution along isentropes is mirrored by the evolution along isobaric surfaces of quasi-geostrophic potential vorticity in pressure coordinates.

## 5.6 ✦ ENERGETICS OF QUASI-GEOSTROPHY

If the quasi-geostrophic set of equations is to represent a real fluid system in a physically meaningful way then it should have a consistent set of energetics. In particular, the total energy should be conserved, and there should be analogues of kinetic and potential energy and conversion between the two. We now show that such energetic properties do hold, using the Boussinesq set as an example.

Let us write the governing equations as a potential vorticity equation in the interior,

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\nabla^2\psi + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial\psi}{\partial z}\right)\right] + \beta\frac{\partial\psi}{\partial x} = 0, \qquad 0 < z < 1, \tag{5.157}$$

and buoyancy advection at the boundary,

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{\partial\psi}{\partial z}\right) = 0, \qquad z = 0, 1. \tag{5.158}$$

For lateral boundary conditions we may assume that $\psi$ = constant, or impose periodic conditions. If we multiply (5.157) by $-\psi$ and integrate over the domain, using the boundary conditions, we easily find

$$\frac{\mathrm{d}\widehat{E}}{\mathrm{d}t} = 0, \qquad \widehat{E} = \frac{1}{2}\int_V \left[(\nabla\psi)^2 + \frac{f_0^2}{N^2}\left(\frac{\partial\psi}{\partial z}\right)^2\right]\mathrm{d}V. \tag{5.159a,b}$$

The term involving $\beta$ makes no direct contribution to the energy budget. Equation (5.159) is the fundamental energy equation for quasi-geostrophic motion, and it states that in the absence of viscous or diabatic terms the total energy is conserved. The two terms in (5.159b) can be identified as the kinetic energy (KE) and available potential energy (APE) of the flow, where

$$\mathrm{KE} = \frac{1}{2}\int_V (\nabla\psi)^2\,\mathrm{d}V, \qquad \mathrm{APE} = \frac{1}{2}\int_V \frac{f_0^2}{N^2}\left(\frac{\partial\psi}{\partial z}\right)^2\mathrm{d}V. \tag{5.160a,b}$$

The available potential energy may also be written as

$$\mathrm{APE} = \frac{1}{2}\int_V \frac{H^2}{L_d^2}\left(\frac{\partial\psi}{\partial z}\right)^2\mathrm{d}V, \tag{5.161}$$

where $L_d$ is the deformation radius $NH/f_0$ and we may choose $H$ such that $z \sim H$. At some scale $L$ the ratio of the kinetic energy to the potential energy is thus, roughly,

$$\frac{\mathrm{KE}}{\mathrm{APE}} \sim \frac{L_d^2}{L^2}. \tag{5.162}$$

For scales much larger than $L_d$ the potential energy dominates the kinetic energy, and contrariwise.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

### 5.6.1 Conversion Between APE and KE

Let us return to the vorticity and thermodynamic equations,

$$\frac{\mathrm{D}\zeta}{\mathrm{D}t} = f\frac{\partial w}{\partial z}, \qquad \frac{\mathrm{D}b'}{\mathrm{D}t} + N^2 w = 0 \tag{5.163a,b}$$

where $\zeta = \nabla^2\psi$, and $b' = f_0\partial\psi/\partial z$. From (5.163a) we form a kinetic energy equation, namely

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_V (\nabla\psi)^2\,\mathrm{d}V = -\int_V f_0\frac{\partial w}{\partial z}\psi\,\mathrm{d}V = \int_V f_0 w\frac{\partial\psi}{\partial z}\,\mathrm{d}V. \tag{5.164}$$

From (5.163b) we form a potential energy equation, namely

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\int_V \frac{f_0^2}{N^2}\left(\frac{\partial\psi}{\partial z}\right)^2\,\mathrm{d}V = -\int_V f_0 w\frac{\partial\psi}{\partial z}\,\mathrm{d}V. \tag{5.165}$$

Thus, the *conversion* from APE to KE is represented by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{KE} = -\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{APE} = \int_v f_0 w\frac{\partial\psi}{\partial z}\,\mathrm{d}V. \tag{5.166}$$

Because the buoyancy is proportional to $\partial\psi/\partial z$, when warm fluid rises there is a correlation between $w$ and $\partial\psi/\partial z$ and APE is converted to KE. Whether such a phenomenon occurs depends of course on the dynamics of the flow; however, such a conversion *is,* in fact, a common feature of geophysical flows, and in particular of baroclinic instability, as we shall see in Chapter 9.

### 5.6.2 Energetics of Two-layer Flows

Two-layer or two-level flows are an important special case. For layers of equal thickness let us write the evolution equations as

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\nabla^2\psi_1 - \frac{1}{2}k_d^2(\psi_1 - \psi_2)\right] + \beta\frac{\partial\psi_1}{\partial x} = 0, \tag{5.167a}$$

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\nabla^2\psi_2 + \frac{1}{2}k_d^2(\psi_1 - \psi_2)\right] + \beta\frac{\partial\psi_2}{\partial x} = 0, \tag{5.167b}$$

where $k_d^2/2 = (2f_0/NH)^2$. On multiplying these two equations by $-\psi_1$ and $-\psi_2$, respectively, and integrating over the horizontal domain, the advective term in the material derivatives and the beta term all vanish, and we obtain

$$\int_A \left[\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}(\nabla\psi_1)^2 + \frac{1}{2}k_d^2\psi_1\frac{\mathrm{d}}{\mathrm{d}t}(\psi_1 - \psi_2)\right]\mathrm{d}A = 0, \tag{5.168a}$$

$$\int_A \left[\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}(\nabla\psi_2)^2 - \frac{1}{2}k_d^2\psi_2\frac{\mathrm{d}}{\mathrm{d}t}(\psi_1 - \psi_2)\right]\mathrm{d}A = 0. \tag{5.168b}$$

Adding these gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_A \left[\frac{1}{2}(\nabla\psi_1)^2 + \frac{1}{2}(\nabla\psi_2)^2 + \frac{k_d^2}{4}(\psi_1 - \psi_2)^2\right]\mathrm{d}A = 0. \tag{5.169}$$

This is the energy conservation statement for the two layer model. The first two terms represent the kinetic energy and the last term represents the available potential energy.

*Energy in the baroclinic and barotropic modes*

A useful partitioning of the energy is between the energy in the barotropic and baroclinic modes. The barotropic streamfunction, $\overline{\psi}$, is the vertically averaged streamfunction and the baroclinic mode is the difference between the streamfunctions in the two layers. That is, for equal layer thicknesses,

$$\overline{\psi} \equiv \frac{1}{2}(\psi_1 + \psi_2), \qquad \tau \equiv \frac{1}{2}(\psi_1 - \psi_2). \tag{5.170}$$

Substituting (5.170) into (5.169) reveals that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_A \left[(\nabla\overline{\psi})^2 + (\nabla\tau)^2 + k_d^2\tau^2\right] \mathrm{d}A = 0. \tag{5.171}$$

The energy density in the barotropic mode is thus just $(\nabla\overline{\psi})^2$, and that in the baroclinic mode is $(\nabla\tau)^2 + k_d^2\tau^2$. This partitioning will prove particularly useful when we consider baroclinic turbulence in Chapter 12.

### 5.6.3 Enstrophy Conservation

Potential vorticity is advected only by the horizontal flow, and thus it is materially conserved on the horizontal surface at every height and

$$\frac{\mathrm{D}q}{\mathrm{D}t} = \frac{\partial q}{\partial t} + \boldsymbol{u}\cdot\nabla q = 0. \tag{5.172}$$

Furthermore, the advecting flow is divergence-free so that $\boldsymbol{u}\cdot\nabla q = \nabla\cdot(\boldsymbol{u}q)$. Thus, on multiplying (5.172) by $q$ and integrating over a horizontal domain $A$ we obtain

$$\frac{\mathrm{d}\widehat{Z}}{\mathrm{d}t} = 0, \qquad \widehat{Z} = \frac{1}{2}\int_A q^2\,\mathrm{d}A. \tag{5.173}$$

The result holds in an enclosed domain, with no-normal flow boundary conditions, or in a channel with periodic boundary conditions in $x$ and no-normal flow conditions in $y$. The quantity $\widehat{Z}$ is known as the *enstrophy*, and it is conserved at each height as well as, naturally, over the entire volume. (In a doubly-periodic domain, only the relative enstrophy, $\int \zeta^2\,\mathrm{d}A$, is conserved.)

The enstrophy is just one of an infinity of invariants in quasi-geostrophic flow. Because the potential vorticity of a fluid element is conserved, *any* function of the potential vorticity must be a material invariant and we can immediately write

$$\frac{\mathrm{D}}{\mathrm{D}t}F(q) = 0. \tag{5.174}$$

To verify that this is true, simply note that (5.174) implies that $(\mathrm{d}F/\mathrm{d}q)\mathrm{D}q/\mathrm{D}t = 0$, which is true by virtue of (5.172). (However, by virtue of the material advection, the function $F(q)$ need not be differentiable in order for (5.174) to hold.) Each of the material invariants corresponding to different choices of $F(q)$ has a corresponding integral invariant; that is,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_A F(q)\,\mathrm{d}A = 0, \tag{5.175}$$

with the boundary conditions as before. The enstrophy invariant corresponds to choosing $F(q) = q^2$; it plays a particularly important role because, like energy, it is a quadratic invariant, and its presence profoundly alters the behaviour of two-dimensional and quasi-geostrophic flow compared to three-dimensional flow (see Section 11.3).

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

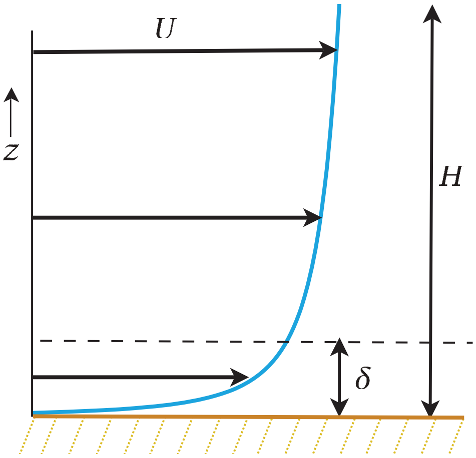


Fig. 5.4 An idealized boundary layer. The values of a field, such as velocity, $U$, may vary rapidly in a boundary in order to satisfy the boundary conditions at a rigid surface. The parameter $\delta$ is a measure of the boundary layer thickness, $H$ is a typical scale of variation away from the boundary, and typically a boundary layer has $\delta \ll H$.

## 5.7 THE EKMAN LAYER

In the final topic of this chapter we consider the effects of friction. The fluid fields in the interior of a domain are often set by different physical processes from those occurring at a boundary, and consequently often change rapidly in a thin *boundary layer,* as in Fig. 5.4. Such boundary layers nearly always involve one or both of viscosity and diffusion, because these appear in the terms of highest differential order in the equations of motion, and so are responsible for the number and type of boundary conditions that the equations must satisfy — for example, the presence of molecular viscosity leads to the condition that the tangential flow, as well as the normal flow, must vanish at a rigid surface.

In many boundary layers in non-rotating flow the dominant balance in the momentum equation is between the advective and viscous terms. In some contrast, in large-scale atmospheric and oceanic flow the effects of rotation are large, and this results in a boundary layer, known as the *Ekman layer,* in which the dominant balance is between Coriolis and frictional or stress terms.[10] Now, the direct effects of molecular viscosity and diffusion are nearly always negligible at distances more than a few millimetres away from a solid boundary, but it is inconceivable that the entire boundary layer between the free atmosphere (or free ocean) and the surface is only a few millimetres thick. Rather, in practice a balance occurs between the Coriolis terms and the forces due to the stress generated by small-scale turbulent motion, and this gives rise to a boundary layer that has a typical depth of a few tens to several hundreds of metres. Because the stress arises from the turbulence we cannot with confidence determine its precise form; thus, we should try to determine what general properties Ekman layers may have that are *independent* of the precise form of the friction, and these properties turn out to be integral ones such as the total mass flux in the Ekman layer.

The atmospheric Ekman layer occurs near the ground, and the stress at the ground itself is due to the surface wind (and its vertical variation). In the ocean the main Ekman layer is near the surface, and the stress at the ocean surface is largely due to the presence of the overlying wind. There is also a weak Ekman layer at the bottom of the ocean, analogous to the atmospheric Ekman layer. To analyze all these layers, let us assume that:

- The Ekman layer is Boussinesq. This is a very good assumption for the ocean, and a reasonable one for the atmosphere if the boundary layer is not too deep.
- The Ekman layer has a finite depth that is less than the total depth of the fluid, this depth being given by the level at which the frictional stresses essentially vanish. Within the Ekman layer, frictional terms are important, whereas geostrophic balance holds beyond it.
- The nonlinear and time-dependent terms in the equations of motion are negligible, hydro-

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

static balance holds in the vertical, and buoyancy is constant, not varying in the horizontal.

- As needed, the friction can be parameterized by a viscous term of the form $\rho_0^{-1}\partial\boldsymbol{\tau}/\partial z = A\,\partial^2\boldsymbol{u}/\partial z^2$, where $A$ is constant and $\boldsymbol{\tau}$ is the stress. (In general, stress is a tensor, $\tau_{ij}$, with an associated force given by $F_i = \partial\tau_{ij}/\partial x_j$, summing over the repeated index. It is common in geophysical fluid dynamics that the vertical derivative dominates, and in this case the force is $\boldsymbol{F} = \partial\boldsymbol{\tau}/\partial z$. We still use the word stress for $\boldsymbol{\tau}$, but it now refers to a vector whose derivative in a particular direction ($z$ in this case) is the force on a fluid.) In laboratory settings $A$ may be the molecular viscosity, whereas in the atmosphere and ocean it is a so-called *eddy viscosity*. In turbulent flows momentum is transferred by the near-random motion of small parcels of fluid and, by analogy with the motion of molecules that produces a molecular viscosity, the associated stress is approximately given by using a turbulent or eddy viscosity that may be orders of magnitude larger than the molecular one.

### 5.7.1 Equations of Motion and Scaling

Frictional–geostrophic balance in the horizontal momentum equation is:

$$\boldsymbol{f} \times \boldsymbol{u} = -\nabla_z \phi + \frac{\partial \widetilde{\boldsymbol{\tau}}}{\partial z}, \tag{5.176}$$

where $\widetilde{\boldsymbol{\tau}} \equiv \boldsymbol{\tau}/\rho_0$ is the kinematic stress and $\boldsymbol{f} = f\mathbf{k}$, where the Coriolis parameter $f$ is allowed to vary with latitude. If we model the stress with an eddy viscosity, (5.176) becomes

$$\boldsymbol{f} \times \boldsymbol{u} = -\nabla_z \phi + A\frac{\partial^2 \boldsymbol{u}}{\partial z^2}. \tag{5.177}$$

The vertical momentum equation is $\partial\phi/\partial z = b$, i.e., hydrostatic balance, and, because buoyancy is constant, we may without loss of generality write this as

$$\frac{\partial \phi}{\partial z} = 0. \tag{5.178}$$

The equation set is completed by the mass continuity equation, $\nabla \cdot \boldsymbol{v} = 0$.

#### *The Ekman number*

We nondimensionalize the equations by setting

$$(u, v) = U(\widehat{u}, \widehat{v}), \quad (x, y) = L(\widehat{x}, \widehat{y}), \quad f = f_0\widehat{f}, \quad z = H\widehat{z}, \quad \phi = \Phi\widehat{\phi}, \tag{5.179}$$

where hatted variables are nondimensional. $H$ is a scaling for the height, and at this stage we will suppose it to be some height scale in the free atmosphere or ocean, not the height of the Ekman layer itself. Geostrophic balance suggests that $\Phi = f_0 UL$. Substituting (5.179) into (5.177) we obtain

$$\widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}} = -\widehat{\nabla}\widehat{\phi} + Ek\frac{\partial^2 \widehat{\boldsymbol{u}}}{\partial \widehat{z}^2}, \tag{5.180}$$

where the parameter

$$\boxed{Ek \equiv \left(\frac{A}{f_0 H^2}\right)}, \tag{5.181}$$

is the *Ekman number*, and it determines the importance of frictional terms in the horizontal momentum equation. If $Ek \ll 1$ then the friction is small in the flow interior where $\widehat{z} = \mathcal{O}(1)$. However, the friction term cannot necessarily be neglected in the boundary layer because it is of the

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

highest differential order in the equation, and so determines the boundary conditions; if $Ek$ is small the vertical scales become small and the second term on the right-hand side of (5.180) remains finite. The case when this term is simply omitted from the equation is therefore a *singular limit*, meaning that it differs from the case with $Ek \to 0$. If $Ek \geq 1$ friction is important everywhere, but it is usually the case that $Ek$ is small for atmospheric and oceanic large-scale flow, and the interior flow is very nearly geostrophic. (In part this is because $A$ itself is only large near a rigid surface where the presence of a shear creates turbulence and a significant eddy viscosity.)

### *Momentum balance in the Ekman layer*

For definiteness, suppose the fluid lies above a rigid surface at $z = 0$. Sufficiently far away from the boundary the velocity field is known, and we suppose this flow to be in geostrophic balance. We then write the velocity field and the pressure field as the sum of the interior geostrophic part, plus a boundary layer correction:

$$\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{u}}_g + \widehat{\boldsymbol{u}}_E, \qquad \widehat{\phi} = \widehat{\phi}_g + \widehat{\phi}_E, \tag{5.182}$$

where the Ekman layer corrections, denoted with a subscript $E$, are negligible away from the boundary layer. Now, in the fluid interior we have, by hydrostatic balance, $\partial\widehat{\phi}_g/\partial\widehat{z} = 0$. In the boundary layer we still have $\partial\widehat{\phi}_g/\partial\widehat{z} = 0$ so that, to satisfy hydrostasy, $\partial\widehat{\phi}_E/\partial\widehat{z} = 0$. But because $\widehat{\phi}_E$ vanishes away from the boundary we have $\widehat{\phi}_E = 0$ everywhere. Thus, *there is no boundary layer in the pressure field.* Note that this is a much stronger result than saying that pressure is continuous, which is nearly always true in fluids; rather, it is a special result for Ekman layers.

Using (5.182) with $\widehat{\phi}_E = 0$, the dimensional horizontal momentum equation (5.176) becomes, in the Ekman layer,

$$\boldsymbol{f} \times \boldsymbol{u}_E = \frac{\partial \widetilde{\boldsymbol{\tau}}}{\partial z}. \tag{5.183}$$

The dominant force balance in the Ekman layer is thus between the Coriolis force and the friction. We can determine the thickness of the Ekman layer if we model the stress with an eddy viscosity so that

$$\boldsymbol{f} \times \boldsymbol{u}_E = A\frac{\partial^2 \boldsymbol{u}_E}{\partial z^2} \qquad \text{or} \qquad \widehat{\boldsymbol{f}} \times \widehat{\boldsymbol{u}}_E = Ek\frac{\partial^2 \widehat{\boldsymbol{u}}_E}{\partial \widehat{z}^2}, \tag{5.184a,b}$$

where the second equation is nondimensional. It is evident that (5.184b) can only be satisfied if $\widehat{z} \neq \mathcal{O}(1)$, implying that $H$ is not a proper scaling for $z$ in the boundary layer. Rather, if the vertical scale in the Ekman layer is $\widehat{\delta}$ (meaning $\widehat{z} \sim \widehat{\delta}$) we must have $\widehat{\delta} \sim Ek^{1/2}$. In dimensional terms this means the thickness of the Ekman layer is

$$\delta = H\widehat{\delta} = HEk^{1/2} \tag{5.185}$$

or, using (5.181),

$$\boxed{\delta = \left(\frac{A}{f_0}\right)^{1/2}.} \tag{5.186}$$

This estimate also emerges directly from (5.184a). Note that (5.185) can be written as

$$Ek = \left(\frac{\delta}{H}\right)^2. \tag{5.187}$$

That is, the Ekman number is equal to the square of the ratio of the depth of the Ekman layer to an interior depth scale of the fluid motion. In laboratory flows where $A$ is the molecular viscosity we can thus estimate the Ekman layer thickness, and if we know the eddy viscosity of the ocean or

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

atmosphere we can estimate their respective Ekman layer thicknesses. We can invert this argument and obtain an estimate of $A$ if we know the Ekman layer depth. In the atmosphere, deviations from geostrophic balance are very small above 1 km, and using this gives $A \approx 10^2\,\mathrm{m^2\,s^{-1}}$. In the ocean Ekman depths are often 50 m or less, and eddy viscosities are about $0.1\,\mathrm{m^2\,s^{-1}}$.

### 5.7.2 Integral Properties of the Ekman Layer

What can we deduce about the Ekman layer without specifying the detailed form of the frictional term? Using dimensional notation we recall frictional–geostrophic balance,

$$\boldsymbol{f} \times \boldsymbol{u} = -\nabla\phi + \frac{1}{\rho_0}\frac{\partial \boldsymbol{\tau}}{\partial z}, \tag{5.188}$$

where $\boldsymbol{\tau}$ is zero at the edge of the Ekman layer. In the Ekman layer itself we have

$$\boldsymbol{f} \times \boldsymbol{u}_E = \frac{1}{\rho_0}\frac{\partial \boldsymbol{\tau}}{\partial z}. \tag{5.189}$$

Consider either a top or bottom Ekman layer, and integrate over its thickness. From (5.189) we obtain

$$\boldsymbol{f} \times \boldsymbol{M}_E = \boldsymbol{\tau}_T - \boldsymbol{\tau}_B\,, \qquad \text{where} \qquad \boldsymbol{M}_E = \int_{Ek} \rho_0 \boldsymbol{u}_E \,\mathrm{d}z. \tag{5.190}$$

Here $\boldsymbol{M}_E$ is the ageostrophic mass transport in the Ekman layer and $\boldsymbol{\tau}_T$ and $\boldsymbol{\tau}_B$ are the respective stresses at the top and the bottom of the Ekman layer at hand. The stress at the top (bottom) will be zero in a bottom (top) Ekman layer and therefore, from (5.190),

$$\begin{aligned} &\text{top Ekman layer:} && \boldsymbol{M}_E = -\frac{1}{f}\mathbf{k} \times \boldsymbol{\tau}_T, \\ &\text{bottom Ekman layer:} && \boldsymbol{M}_E = \frac{1}{f}\mathbf{k} \times \boldsymbol{\tau}_B. \end{aligned} \tag{5.191a,b}$$

The transport is thus at right angles to the stress at the surface, and proportional to the magnitude of the stress. These properties have a simple physical explanation: integrated over the depth of the Ekman layer the surface stress must be balanced by the Coriolis force, which in turn acts at right angles to the mass transport. A consequence of (5.191) is that the mass transports in adjacent oceanic and atmospheric Ekman layers are equal and opposite, because the stress is continuous across the ocean–atmosphere interface. Equation (5.191a) is particularly useful in the ocean, where the stress at the surface is primarily due to the wind, and is largely independent of the interior oceanic flow. In the atmosphere, the surface stress mainly arises as a result of the interior atmospheric flow, and to calculate it we need to parameterize the stress in terms of the flow.

Finally, we obtain an expression for the vertical velocity induced by an Ekman layer. The mass conservation equation is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \tag{5.192}$$

Integrating this over an Ekman layer gives

$$\frac{1}{\rho_0}\nabla \cdot \boldsymbol{M}_{To} = -(w_T - w_B), \tag{5.193}$$

where $\boldsymbol{M}_{To}$ is the total (Ekman plus geostrophic) mass transport in the Ekman layer,

$$\boldsymbol{M}_{To} = \int_{Ek} \rho_0 \boldsymbol{u}\,\mathrm{d}z = \int_{Ek} \rho_0 (\boldsymbol{u}_g + \boldsymbol{u}_E)\,\mathrm{d}z \equiv \boldsymbol{M}_g + \boldsymbol{M}_E, \tag{5.194}$$

and $w_T$ and $w_B$ are the vertical velocities at the top and bottom of the Ekman layer; the former (latter) is zero in a top (bottom) Ekman layer. Equations (5.194) and (5.190) give

$$\mathbf{k} \times (\boldsymbol{M}_{To} - \boldsymbol{M}_g) = \frac{1}{f}(\boldsymbol{\tau}_T - \boldsymbol{\tau}_B). \tag{5.195}$$

Taking the curl of this (i.e., cross-differentiating) gives

$$\nabla \cdot (\boldsymbol{M}_{To} - \boldsymbol{M}_g) = \mathrm{curl}_z[(\boldsymbol{\tau}_T - \boldsymbol{\tau}_B)/f], \tag{5.196}$$

where the $\mathrm{curl}_z$ operator on a vector $\boldsymbol{A}$ is defined by $\mathrm{curl}_z \boldsymbol{A} \equiv \partial_x A_y - \partial_y A_x$. Using (5.193) we obtain, for top and bottom Ekman layers respectively,

$$w_B = \frac{1}{\rho_0}\left(\mathrm{curl}_z \frac{\boldsymbol{\tau}_T}{f} + \nabla \cdot \boldsymbol{M}_g\right), \qquad w_T = \frac{1}{\rho_0}\left(\mathrm{curl}_z \frac{\boldsymbol{\tau}_B}{f} - \nabla \cdot \boldsymbol{M}_g\right), \tag{5.197a,b}$$

where $\nabla \cdot \boldsymbol{M}_g = -(\beta/f)\boldsymbol{M}_g \cdot \mathbf{j}$ is the divergence of the geostrophic transport in the Ekman layer, and this is often small compared to the other terms in these equations. Thus, friction induces a vertical velocity at the edge of the Ekman layer, proportional to the curl of the stress at the surface, and this is perhaps the most used result in Ekman layer theory. Numerical models sometimes do not have the vertical resolution to explicitly resolve an Ekman layer, and (5.197) provides a means of *parameterizing* the layer in terms of resolved or known fields. This is useful for the top Ekman layer in the ocean, where the stress can be regarded as a function of the overlying wind.

### 5.7.3 Explicit Solutions I: a Bottom Boundary Layer

We now assume that the frictional terms can be parameterized as an eddy viscosity and calculate the explicit form of the solution in the boundary layer. The frictional–geostrophic balance may be written as

$$\boldsymbol{f} \times (\boldsymbol{u} - \boldsymbol{u}_g) = A\frac{\partial^2 \boldsymbol{u}}{\partial z^2}, \tag{5.198a}$$

where

$$f(u_g, v_g) = \left(-\frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial x}\right). \tag{5.198b}$$

We continue to assume there are no horizontal gradients of temperature, so that, via thermal wind, $\partial u_g/\partial z = \partial v_g/\partial z = 0$.

#### *Boundary conditions and solution*

Appropriate boundary conditions for a bottom Ekman layer are

$$\text{at } z = 0: \quad u = 0, \quad v = 0 \quad \text{(the no slip condition)} \tag{5.199a}$$

$$\text{as } z \to \infty: \quad u = u_g, \quad v = v_g \quad \text{(a geostrophic interior).} \tag{5.199b}$$

Let us seek solutions to (5.198a) of the form

$$u = u_g + A_0 \mathrm{e}^{\alpha z}, \qquad v = v_g + B_0 \mathrm{e}^{\alpha z}, \tag{5.200}$$

where $A_0$ and $B_0$ are constants. Substituting into (5.198a) gives two homogeneous algebraic equations

$$A_0 f - B_0 A \alpha^2 = 0, \qquad -A_0 A\alpha^2 - B_0 f = 0. \tag{5.201a,b}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

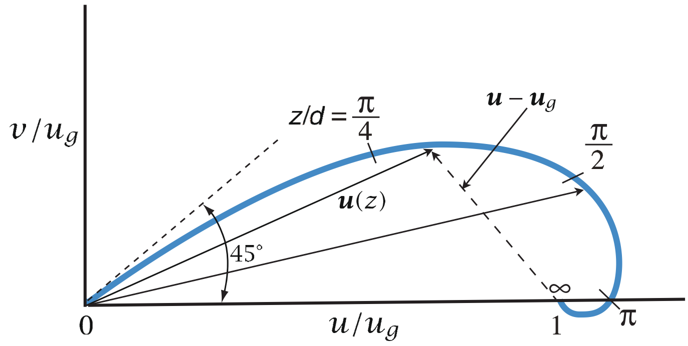


**Fig. 5.5** The idealized Ekman layer solution in the lower atmosphere, plotted as a hodograph of the wind components: the arrows show the velocity vectors at particular heights, and the curve traces out the continuous variation of the velocity. The values on the curve are of the nondimensional variable $z/d$, where $d = (2A/f)^{1/2}$, and $v_g$ is chosen to be zero.

For non-trivial solutions the solvability condition $\alpha^4 = -f^2/A^2$ must hold, from which we find $\alpha = \pm(1 \pm \mathrm{i})\sqrt{f/2A}$. Using the boundary conditions we then obtain the solution

$$u = u_g - \mathrm{e}^{-z/d}\left[u_g \cos(z/d) + v_g \sin(z/d)\right], \tag{5.202a}$$

$$v = v_g + \mathrm{e}^{-z/d}\left[u_g \sin(z/d) - v_g \cos(z/d)\right], \tag{5.202b}$$

where $d = \sqrt{2A/f}$ is, within a constant factor, the depth of the Ekman layer obtained from scaling considerations. The solution decays exponentially from the surface with this e-folding scale, so that $d$ is a good measure of the Ekman layer thickness. Note that the boundary layer correction depends on the interior flow, since the boundary layer serves to bring the flow to zero at the surface.

To illustrate the solution, suppose that the pressure force is directed in the $y$-direction (northwards), so that the geostrophic current is eastwards. Then the solution, the now-famous *Ekman spiral*, is plotted in Figs. 5.5 and 5.6. The wind falls to zero at the surface, and its direction just above the surface is northeastwards; that is, it is rotated by 45° to the left of its direction in the free atmosphere. Although this result is independent of the value of the frictional coefficients, it is dependent on the form of the friction chosen. The force balance in the Ekman layer is between the Coriolis force, the stress, and the pressure force. At the surface the Coriolis force is zero, and the balance is entirely between the northward pressure force and the southward stress force.

### *Transport, force balance and vertical velocity*

The cross-isobaric flow is given by (for $v_g = 0$)

$$V = \int_0^\infty v\,\mathrm{d}z = \int_0^\infty u_g \mathrm{e}^{-z/d}\sin(z/d)\,\mathrm{d}z = \frac{u_g d}{2}. \tag{5.203}$$

For positive $f$, this is to the left of the geostrophic flow — that is, down the pressure gradient. In the general case ($v_g \neq 0$) we obtain

$$V = \int_0^\infty (v - v_g)\,\mathrm{d}z = \frac{d}{2}(u_g - v_g). \tag{5.204}$$

Similarly, the additional zonal transport produced by frictional effects is, for $v_g = 0$,

$$U = \int_0^\infty (u - u_g)\,\mathrm{d}z = -\int_0^\infty \mathrm{e}^{-z/d}\sin(z/d)\,\mathrm{d}z = -\frac{u_g d}{2}, \tag{5.205}$$

and in the general case

$$U = \int_0^\infty (u - u_g)\,\mathrm{d}z = -\frac{d}{2}(u_g + v_g). \tag{5.206}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

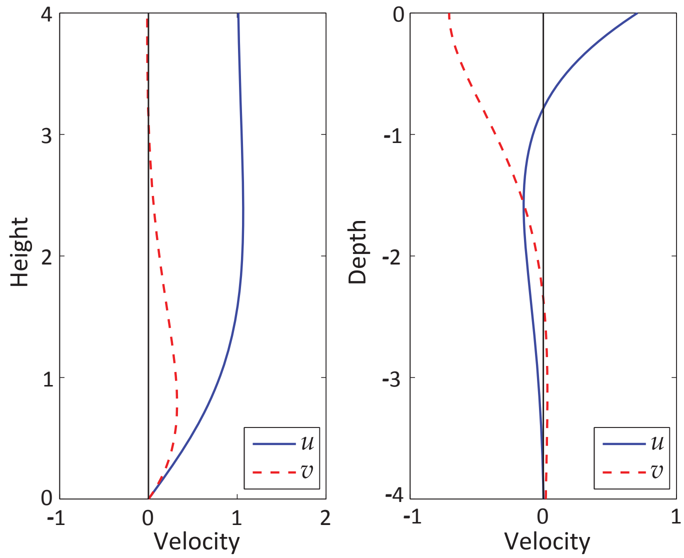


**Fig. 5.6** Solutions for a bottom Ekman layer with a given flow in the fluid interior (left), and for a top Ekman layer with a given surface stress (right), both with $d = 1$. On the left we have $u_g = 1$, $v_g = 0$. On the right we have $u_g = v_g = 0$, $\tilde{\tau}_y = 0$ and $\sqrt{2}\tilde{\tau}_x/(fd) = 1$.

Thus, the total transport caused by frictional forces is

$$\boldsymbol{M}_E = \frac{\rho_0 d}{2}\left[-\mathbf{i}(u_g + v_g) + \mathbf{j}(u_g - v_g)\right]. \tag{5.207}$$

The total stress at the bottom surface $z = 0$ induced by frictional forces is

$$\tilde{\boldsymbol{\tau}}_B = A\frac{\partial \boldsymbol{u}}{\partial z}\bigg|_{z=0} = \frac{A}{d}\left[\mathbf{i}(u_g - v_g) + \mathbf{j}(u_g + v_g)\right], \tag{5.208}$$

using the solution (5.202). Thus, using (5.207), (5.208) and $d^2 = 2A/f$, we see that the total frictionally induced transport in the Ekman layer is related to the stress at the surface by $\boldsymbol{M}_E = (\mathbf{k} \times \boldsymbol{\tau}_B)/f$, reprising the result of the previous more general analysis, (5.197). From (5.208), the stress is at an angle of 45° to the left of the velocity at the surface. (However, this result is not generally true for all forms of stress.) These properties are illustrated in Fig. 5.7.

The vertical velocity at the top of the Ekman layer, $w_E$, is obtained using (5.207) or (5.208). If $f$ is constant we obtain

$$w_E = -\frac{1}{\rho_0}\nabla\cdot\boldsymbol{M}_E = \frac{1}{f_0}\mathrm{curl}_z\tilde{\boldsymbol{\tau}}_B = \frac{d}{2}\zeta_g, \tag{5.209}$$

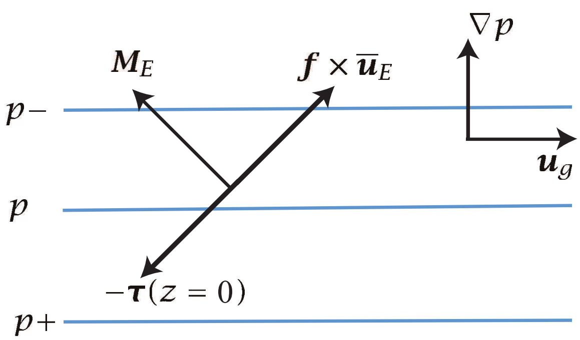


**Fig. 5.7** An Ekman layer generated from an eastward geostrophic flow above with associated pressure levels as shown (blue lines). An overbar denotes a vertical integral over the Ekman layer, so that $-\boldsymbol{f} \times \overline{\boldsymbol{u}}_E$ is the Coriolis force on the vertically integrated Ekman velocity. $\boldsymbol{M}_E$ is the frictionally induced boundary layer transport, and $\boldsymbol{\tau}$ is the stress.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

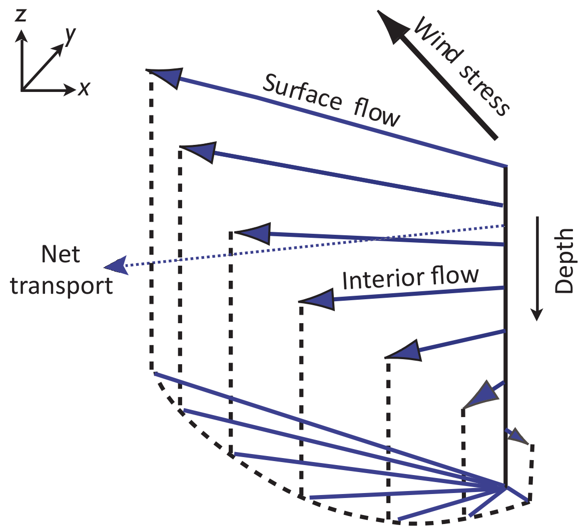


**Fig. 5.8** An idealized Ekman spiral in a southern hemisphere ocean, driven by an imposed wind stress. A northern hemisphere spiral would be the reflection of this in the vertical plane. Such a clean spiral is rarely observed in the real ocean. The net transport is at right angles to the wind, independent of the detailed form of the friction. The angle of the surface flow is 45° to the wind only for a Newtonian viscosity.

where $\zeta_g$ is the vorticity of the geostrophic flow. Thus, the vertical velocity at the top of the Ekman layer, which arises because of the frictionally-induced divergence of the cross-isobaric flow in the Ekman layer, is proportional to the geostrophic vorticity in the free fluid and is proportional to the Ekman layer height $\sqrt{2A/f_0}$.

*Another bottom boundary condition*

In the analysis above we assumed a *no slip* condition at the surface, namely that the velocity tangential to the surface vanishes. This is formally appropriate if $A$ is a molecular viscosity, but in a turbulent flow, where $A$ is to be interpreted as an eddy viscosity, the flow close to the surface may be far from zero. Then, unless we wish to explicitly calculate the flow in an additional very thin viscous boundary layer the no-slip condition may be inappropriate. An alternative, slightly more general boundary condition is to suppose that the stress at the surface is given by

$$\boldsymbol{\tau} = \rho_0 C \boldsymbol{u}, \tag{5.210}$$

where $C$ is a constant. The surface boundary condition is then

$$A\frac{\partial \boldsymbol{u}}{\partial z} = C\boldsymbol{u}. \tag{5.211}$$

If $C$ is infinite we recover the no-slip condition. If $C = 0$, we have a condition of no stress at the surface, also known as a *free slip* condition. For intermediate values of $C$ the boundary condition is known as a 'mixed condition'. Evaluating the solution in these cases is left as an exercise for the reader.

### 5.7.4 Explicit Solutions II: the Upper Ocean

*Boundary conditions and solution*

The wind provides a stress on the upper ocean, and the Ekman layer serves to communicate this to the oceanic interior. Appropriate boundary conditions are thus:

$$\text{at } z = 0: \qquad A\frac{\partial u}{\partial z} = \tilde{\tau}^x, \quad A\frac{\partial v}{\partial z} = \tilde{\tau}^y, \qquad \text{(a given surface stress)} \tag{5.212a}$$

$$\text{as } z \to -\infty: \qquad u = u_g, \qquad v = v_g, \qquad \text{(a geostrophic interior)} \tag{5.212b}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

where $\widetilde{\boldsymbol{\tau}}$ is the given (kinematic) wind stress at the surface. Solutions to (5.198a) with (5.212) are found by the same methods as before, and are

$$u = u_g + \frac{\sqrt{2}}{fd} e^{z/d} \left[\widetilde{\tau}^x \cos(z/d - \pi/4) - \widetilde{\tau}^y \sin(z/d - \pi/4)\right], \tag{5.213}$$

and

$$v = v_g + \frac{\sqrt{2}}{fd} e^{z/d} \left[\widetilde{\tau}^x \sin(z/d - \pi/4) + \widetilde{\tau}^y \cos(z/d - \pi/4)\right]. \tag{5.214}$$

Note that the boundary layer correction depends only on the imposed surface stress, and not the interior flow itself. This is a consequence of the type of boundary conditions chosen, for in the absence of an imposed stress the boundary layer correction is zero — the interior flow already satisfies the gradient boundary condition at the top surface. Similarly to the bottom boundary layer, the velocity vectors of the solution trace a diminishing spiral as they descend into the interior (see Fig. 5.8, which is drawn for the Southern Hemisphere).

*Transport, surface flow and vertical velocity*

The transport induced by the surface stress is obtained by integrating (5.213) and (5.214) from the surface to $-\infty$. We explicitly find

$$U = \int_{-\infty}^{0} (u - u_g)\,dz = \frac{\widetilde{\tau}^y}{f}, \qquad V = \int_{-\infty}^{0} (v - v_g)\,dz = -\frac{\widetilde{\tau}^x}{f}, \tag{5.215}$$

which indicates that the ageostrophic transport is perpendicular to the wind stress, as noted previously from more general considerations. Suppose that the surface wind is eastward; in this case $\widetilde{\tau}^y = 0$ and the solutions immediately give

$$u(0) - u_g = \widetilde{\tau}^x / fd, \qquad v(0) - v_g = -\widetilde{\tau}^x / fd. \tag{5.216}$$

Therefore the magnitudes of the frictional flow in the $x$ and $y$ directions are equal to each other, and the ageostrophic flow is 45° to the right (for $f > 0$) of the wind. This result depends on the form of the frictional parameterization, but not on the size of the viscosity.

At the edge of the Ekman layer the vertical velocity is given by (5.197), and so is proportional to the curl of the wind stress. (The second term on the right-hand side of (5.197) is the vertical velocity due to the divergence of the geostrophic flow, and is usually much smaller than the first term.) The production of a vertical velocity at the edge of the Ekman layer is one of the most important effects of the layer, especially with regard to the large-scale circulation, for it provides an efficient means whereby surface fluxes are communicated to the interior flow (see Fig. 5.9).

### 5.7.5 † Observations of the Ekman Layer

Ekman layers — and in particular the Ekman spiral — are generally quite hard to observe, in either the ocean or atmosphere, both because of the presence of phenomena that are not included in the theory and because of the technical difficulties of actually measuring the vector velocity profile, especially in the ocean. Ekman-layer theory does not take into account the effects of stratification or of inertial and gravity waves (Section 2.10.4 and Chapter 7), nor does it account for the effects of convection or buoyancy-driven turbulence. If gravity waves are present, the instantaneous flow will be non-geostrophic and so time-averaging will be required to extract the geostrophic flow. If strong convection is present, the simple eddy-viscosity parameterizations used to derive the Ekman spiral will be rendered invalid, and the spiral Ekman profile cannot be expected to be observed in either atmosphere or ocean.

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

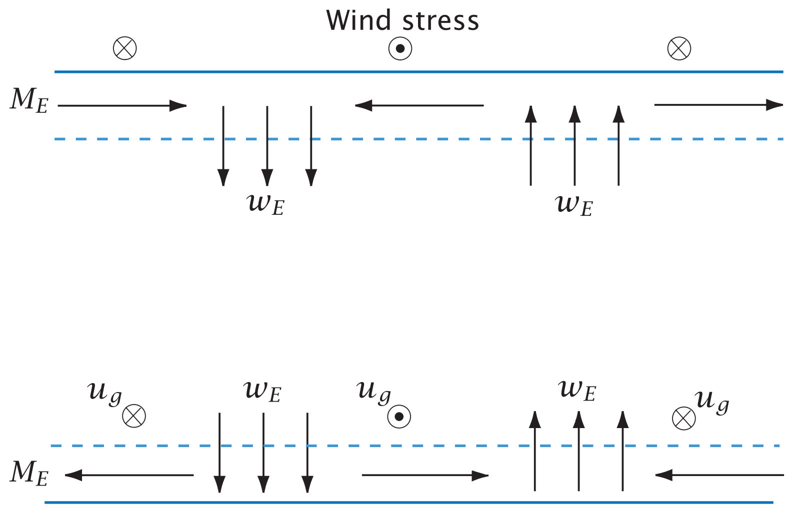


**Fig. 5.9** Upper and lower Ekman layers. The upper Ekman layer in the ocean is primarily driven by an imposed wind stress, whereas the lower Ekman layer in the atmosphere or ocean largely results from the interaction of interior geostrophic velocity and a rigid lower surface. The upper part of the figure shows the vertical Ekman 'pumping' velocities that result from the given wind stress, and the lower part of the figure shows the Ekman pumping velocities given the interior geostrophic flow.

In the atmosphere, in convectively neutral cases, the Ekman profile can sometimes be qualitatively discerned. In convectively unstable situations the Ekman profile is generally not observed, but the flow is nevertheless cross-isobaric, from high pressure to low, consistent with the theory. (For most purposes, it is in any case the integral properties of the Ekman layer that is most important.) In the ocean, from about 1980 onwards improved instruments have made it possible to observe the vector current with depth, and to average that current and correlate it with the overlying wind, and a number of observations generally consistent with Ekman dynamics have emerged.[11] There are some differences between observations and theory, and these can be ascribed to the effects of stratification (which causes a shallowing and flattening of the spiral), and to the interaction of the Ekman spiral with turbulence (and the inadequacy of the eddy-diffusivity parameterization). In spite of these differences, Ekman layer theory remains a remarkable and enduring foundation of geophysical fluid dynamics.

### 5.7.6 ✦ Frictional Parameterization of the Ekman Layer

Suppose that the free atmosphere is described by the quasi-geostrophic vorticity equation,

$$\frac{\mathrm{D}\zeta_g}{\mathrm{D}t} = f_0 \frac{\partial w}{\partial z}, \tag{5.217}$$

where $\zeta_g$ is the geostrophic relative vorticity. Let us further model the atmosphere as a single homogeneous layer of thickness $H$ lying above an Ekman layer of thickness $d \ll H$. If the vertical velocity is negligible at the top of the layer (at $z = H + d$) the equation of motion becomes

$$\frac{\mathrm{D}\zeta_g}{\mathrm{D}t} = \frac{f_0[w(H+d) - w(d)]}{H} = -\frac{f_0 d}{2H}\zeta_g\,, \tag{5.218}$$

https://doi.org/10.1017/9781107588417.006 Published online by Cambridge University Press

using (5.209). This equation shows that the Ekman layer acts as a *linear drag* on the interior flow, with a drag coefficient $r$ equal to $f_0 d/2H$ and with associated time scale $T_{Ek}$ given by

$$T_{Ek} = \frac{2H}{f_0 d} = \frac{2H}{\sqrt{2 f_0 A}}. \tag{5.219}$$

In the oceanic case the corresponding vorticity equation for the interior flow is

$$\frac{\mathrm{D}\zeta_g}{\mathrm{D}t} = \frac{1}{H}\mathrm{curl}_z \tau_s, \tag{5.220}$$

where $\tau_s$ is the surface stress. The surface stress thus acts as if it were a body force on the interior flow, and neither the Coriolis parameter nor the depth of the Ekman layer explicitly appear in this formula.

The Ekman layer is a very efficient way of communicating surface stresses to the interior. To see this, suppose that eddy mixing were the sole mechanism of transferring stress from the surface to the fluid interior, and there were no Ekman layer. Then the time scale of spindown of the fluid would be given by using

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = A\frac{\partial^2 \zeta}{\partial z^2}, \tag{5.221}$$

implying a turbulent spin-down time, $T_{turb}$, of

$$T_{turb} \sim \frac{H^2}{A}, \tag{5.222}$$

where $H$ is the depth over which we require a spin-down. This is much longer than the spin-down of a fluid that has an Ekman layer, for we have

$$\frac{T_{turb}}{T_{Ek}} = \frac{(H^2/A)}{(2H/f_0 d)} = \frac{H}{d} \gg 1, \tag{5.223}$$

using $d = \sqrt{2A/f_0}$. The effects of friction are enhanced because of the presence of a secondary circulation confined to the Ekman layers (as in Fig. 5.9) in which the vertical scales are much smaller than those in the fluid interior and so where viscous effects become significant; these frictional stresses are then communicated to the fluid interior via the induced vertical velocities at the edge of the Ekman layers.

## *Notes*

1 The phrase 'quasi-geostrophic' seems to have been introduced by Durst & Sutcliffe (1938) and the concept used in Sutcliffe's development theory of baroclinic systems (Sutcliffe 1939, 1947). The first systematic derivation of the quasi-geostrophic equations based on scaling theory was given by Charney (1948). The planetary-geostrophic equations were used by Robinson & Stommel (1959) and Welander (1959) in studies of the thermocline (and were first known as the 'thermocline equations'), and were put in the context of other approximate equation sets by Phillips (1963).

2 Carl-Gustav Rossby (1898–1957) played a dominant role in the development of dynamical meteorology in the early and middle parts of the twentieth century, and his work permeates all aspects of dynamical meteorology today. Perhaps the most fundamental nondimensional number in rotating fluid dynamics, the Rossby number, is named for him, as is the perhaps most fundamental wave, the Rossby wave. He also discovered the conservation of potential vorticity (later generalized by Ertel) and contributed important ideas to atmospheric turbulence and the theory of air masses. Swedish born, he studied first with V. Bjerknes before taking a position in Stockholm in 1922 with

the Swedish Meteorological Hydrologic Service and receiving a 'Licentiat' from the University of Stockholm in 1925. Shortly thereafter he moved to the United States, joining the Government Weather Bureau, a precursor of NOAA's National Weather Service. In 1928 he moved to MIT, playing an important role in developing the meteorology department there, while still maintaining connections with the Weather Bureau. In 1940 he moved to the University of Chicago, where he similarly helped develop meteorology there. In 1947 he became Director of the newly formed Institute of Meteorology in Stockholm, and subsequently divided his time between there and the United States. Thus, as well as his scientific contributions, he played an influential role in the institutional development of the field.

3 Burger (1958).

4 Numerical integrations of the potential vorticity equation using (5.91), and performing the inversion without linearizing potential vorticity, do in fact indicate improved accuracy over either the quasi-geostrophic or planetary-geostrophic equations (Mundt *et al.* 1997). In a similar vein, McIntyre & Norton (2000) show how useful potential vorticity inversion can be, and Allen *et al.* (1990a,b) demonstrate the high accuracy of certain intermediate models. Certainly, asymptotic correctness should not be the only criterion used in constructing a filtered model, because the parameter range in which the model is useful may be too limited. Note that there is a difference between extending the parameter range in which a filtered model is useful, as in the inversion of (5.91), and going to higher asymptotic order accuracy in a given parameter regime, as in Allen (1993) and Warn *et al.* (1995). Using Hamiltonian mechanics it is possible to derive equations that both span different asymptotic regimes and that have good conservation properties (Salmon 1983, Allen *et al.* 2002).

5 There is a difference between the *dynamical* demands of the quasi-geostrophic system in requiring $\beta$ to be small, and the *geometric* demands of the Cartesian geometry. On Earth the two demands are similar in practice. But without dynamical inconsistency we may imagine a Cartesian system in which $\beta y \sim f$, and indeed this is common in idealized, planetary geostrophic models of the large-scale ocean circulation.

6 The atmospheric and oceanic sciences are sometimes thought of as not being 'beautiful' in the same way as some branches of theoretical physics. Yet surely quasi-geostrophic theory, and the quasi-geostrophic potential vorticity equation, are quite beautiful, combining austerity of description and richness of behaviour.

7 Bretherton (1966b). Schneider *et al.* (2003) look at the non-QG extension. The equivalence between boundary conditions and delta-function sources is a common feature of elliptic and similar problems, and is analogous to the generation of electromagnetic fields by point charges. It is sometimes exploited in the numerical solution of elliptic equations, both as a simple way to include non-homogeneous boundary conditions and, using the so-called capacitance matrix method, to solve problems in irregular domains (e.g., Hockney 1970).

8 Phillips (1954, 1956) used a two-level model for instability studies and to construct a simple general circulation model of the atmosphere.

9 Charney & Stern (1962). See also Berrisford *et al.* (1993) and Vallis (1996).

10 After Ekman (1905). The problem is said to have been posed to V. W. Ekman (1874–1954), a student of Vilhelm Bjerknes, by Fridtjof Nansen, the polar explorer and statesman, who wanted to understand the motion of pack ice and of his ship, the *Fram*, embedded in the ice.

11 For oceanic observations see Davis *et al.* (1981), Price *et al.* (1987), Rudnick & Weller (1993). For the atmosphere see, e.g., Nicholls (1985).

# Part II

# WAVES, INSTABILITIES AND TURBULENCE

*And the waves sing because they are moving.*
Philip Larkin

# CHAPTER SIX

# Wave Fundamentals

THIS CHAPTER PROVIDES AN INTRODUCTION BOTH TO WAVE MOTION ITSELF and to what is perhaps the most important kind of wave occurring at large scales in the ocean and atmosphere, namely the Rossby wave.[1] The chapter has three main parts to it. In the first, we provide an introduction to wave kinematics, discussing such basic concepts as phase speed and group velocity. The second part, beginning with section 6.4, is a discussion of the dynamics of Rossby waves, and this part may be considered to be the natural follow-on from the previous chapter. Finally, in section 6.8, we return to group velocity in a somewhat more general way and illustrate the results using Poincaré waves. Wave kinematics is a rather formal topic, yet closely tied to wave dynamics: kinematics without a dynamical example is jejune and dry, yet understanding wave dynamics of any sort is hardly possible without appreciating at least some of the formal structure of waves. Readers should flip pages back and forth through the chapter as needed.

Those readers who already have a knowledge of wave motion and who wish to cut to the chase may skip the first few sections and begin at section 6.4, referring back as needed. Other readers may wish to skip the sections on Rossby waves altogether and, after absorbing the sections on the wave theory move on to chapter 7 on gravity waves, returning to Rossby waves (or not) later on. The Rossby wave and gravity wave chapters are largely independent of each other, although they both require that the reader is familiar with the basic ideas of wave analysis such as group velocity and phase speed. Rossby waves and gravity waves can, of course, co-exist and we give an introduction to that topic in section 6.7. Close to the equator the two kinds of waves become more intertwined and we deal with the ensuing waves in more depth in chapter 8. We also extend our discussion of Rossby waves in a global atmospheric context in chapter 16.

## 6.1 FUNDAMENTALS AND FORMALITIES

### 6.1.1 Definitions and kinematics

What is a wave? Rather like turbulence, a wave is more easily recognized than defined. Perhaps a little loosely, a wave may be considered to be a propagating disturbance that has a characteristic relationship between its frequency and size; more formally, a wave is a disturbance that satisfies a *dispersion relation.* In order to see what this means, and what a dispersion relation is, suppose that a disturbance, $\psi(\boldsymbol{x}, t)$ (where $\psi$ might be velocity, streamfunction, pressure, etc), satisfies some equation

$$L(\psi) = 0, \tag{6.1}$$

where $L$ is a linear operator, typically a polynomial in time and space derivatives; an example is $L(\psi) = \partial\nabla^2\psi/\partial t + \beta\partial\psi/\partial x$. (Nonlinear waves exist, but the curious reader must look elsewhere to learn about them.[2]) If (6.1) has constant coefficients (if $\beta$ is constant in this example) then solutions may often be found as a superposition of *plane waves,* each of which satisfy

$$\psi = \mathrm{Re}\,\widetilde{\psi}\mathrm{e}^{\mathrm{i}\theta(x,t)} = \mathrm{Re}\,\widetilde{\psi}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}. \tag{6.2}$$

where $\widetilde{\psi}$ is a complex constant, $\theta$ is the phase, $\omega$ is the wave frequency and $\boldsymbol{k}$ is the vector wavenumber $(k, l, m)$ (also written as $(k^x, k^y, k^z)$ or, in subscript notation, $k_i$). The prefix Re denotes the real part of the expression, but we will drop that notation if there is no ambiguity.

Earlier, we said that waves are characterized by having a particular relationship between the frequency and wavevector known as the *dispersion relation*. This is an equation of the form

$$\omega = \Omega(\boldsymbol{k}) \tag{6.3}$$

where $\Omega(\boldsymbol{k})$ [or $\Omega(k_i)$, and meaning $\Omega(k, l, m)$] is some function determined by the form of $L$ in (6.1) and which thus depends on the particular type of wave — the function is different for sound waves, light waves and the Rossby waves and gravity waves we will encounter in this book (peak ahead to (6.59) and (7.57), and there is more discussion in section 6.1.3). Unless it is necessary to explicitly distinguish the function $\Omega$ from the frequency $\omega$, we will often write $\omega = \omega(\boldsymbol{k})$.

If the medium in which the waves are propagating is inhomogeneous then (6.1) will probably not have constant coefficients (for example, $\beta$ may vary meridionally). Nevertheless, if the medium is varying sufficiently slowly, wave solutions may often still be found with the general form

$$\psi(\boldsymbol{x}, t) = \mathrm{Re}\,a(\boldsymbol{x}, t)\mathrm{e}^{\mathrm{i}\theta(\boldsymbol{x},t)}, \tag{6.4}$$

where $a(\boldsymbol{x}, t)$ varies slowly compared to the variation of the phase, $\theta$. The frequency and wavenumber are then *defined* by

$$\boldsymbol{k} \equiv \nabla\theta, \qquad \omega \equiv -\frac{\partial\theta}{\partial t}. \tag{6.5}$$

The example of (6.2) is clearly just a special case of this. Eq. (6.5) implies the formal relation between $\boldsymbol{k}$ and $\omega$:

$$\frac{\partial\boldsymbol{k}}{\partial t} + \nabla\omega = 0. \tag{6.6}$$

### 6.1.2 Wave propagation and phase speed

An almost universal property of waves is that they propagate through space with some velocity (which in special cases might be zero). Waves in fluids may carry energy and momentum but not normally, at least to a first approximation, fluid parcels themselves. Further, it turns out that the speed at which properties like energy are transported (the group speed) may be different from the speed at which the wave crests themselves move (the phase speed). Let's try to understand this statement, beginning with the phase speed.

*Phase speed*

Let us consider the propagation of monochromatic plane waves, for that is all that is needed to introduce the phase speed. Given (6.2) a wave will propagate in the direction of $\boldsymbol{k}$ (Fig. 6.1). At a given instant and location we can align our coordinate axis along this direction, and we write $\boldsymbol{k} \cdot \boldsymbol{x} = Kx^*$, where $x^*$ increases in the direction of $\boldsymbol{k}$ and $K^2 = |\boldsymbol{k}|^2$ is the magnitude of the wavenumber. With this, we can write (6.2) as

$$\psi = \mathrm{Re}\,\widetilde{\psi} \mathrm{e}^{\mathrm{i}(Kx^* - \omega t)} = \mathrm{Re}\,\widetilde{\psi} \mathrm{e}^{\mathrm{i}K(x^* - ct)}, \tag{6.7}$$

where $c = \omega/K$. From this equation it is evident that the phase of the wave propagates at the speed $c$ in the direction of $\boldsymbol{k}$, and we define the *phase speed* by

$$c_p \equiv \frac{\omega}{K}. \tag{6.8}$$

The wavelength of the wave, $\lambda$, is the distance between two wavecrests — that is, the distance between two locations along the line of travel whose phase differs by $2\pi$ — and evidently this is given by

$$\lambda = \frac{2\pi}{K}. \tag{6.9}$$

In (for simplicity) a two-dimensional wave, and referring to Fig. 6.1(a), the wavelength and wave vectors in the $x$- and $y$-directions are given by,

$$\lambda^x = \frac{\lambda}{\cos\phi}, \quad \lambda^y = \frac{\lambda}{\sin\phi}, \qquad k^x = K\cos\phi, \quad k^y = K\sin\phi. \tag{6.10}$$

In general, lines of constant phase intersect both the coordinate axes and propagate along them. The speed of propagation along these axes is given by

$$c_p^x = c_p\frac{l^x}{l} = \frac{c_p}{\cos\phi} = c_p\frac{K}{k^x} = \frac{\omega}{k^x}, \qquad c_p^y = c_p\frac{l^y}{l} = \frac{c_p}{\sin\phi} = c_p\frac{K}{k^y} = \frac{\omega}{k^y}, \tag{6.11}$$

using (6.8) and (6.10). The speed of phase propagation along any one of the axis is in general *larger* than the phase speed in the primary direction of the wave. The phase speeds are clearly *not* components of a vector: for example, $c_p^x \neq c_p \cos\phi$. Analogously, the wavevector $\boldsymbol{k}$ is a true vector, whereas the wavelength $\lambda$ is not.

To summarize, the phase speed and its components are given by

$$\boxed{c_p = \frac{\omega}{K}, \qquad c_p^x = \frac{\omega}{k^x}, \quad c_p^y = \frac{\omega}{k^y}}. \tag{6.12}$$

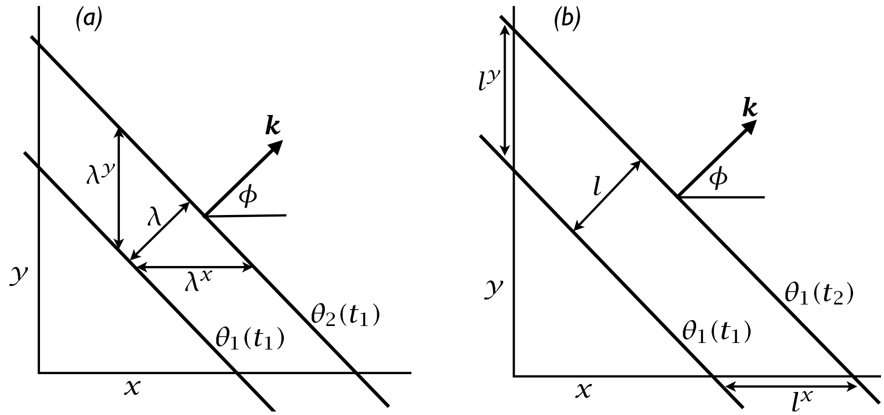


**Fig. 6.1** The propagation of a two-dimensional wave. (a) Two lines of constant phase (e.g., two wavecrests) at a time $t_1$. The wave is propagating in the direction $\boldsymbol{k}$ with wavelength $\lambda$. (b) The same line of constant phase at two successive times. The phase speed is the speed of advancement of the wavecrest in the direction of travel, and so $c_p = l/(t_2 - t_1)$. The phase speed in the $x$-direction is the speed of propagation of the wavecrest along the $x$-axis, and $c_p^x = l^x/(t_2 - t_1) = c_p/\cos\phi$.

*Phase velocity*

Although it is not particularly useful, there is a way of defining a phase speed so that is a true vector, and which might then be called phase velocity. We define the phase velocity to be the velocity that has the magnitude of the phase speed and direction in which wave crests are propagating; that is

$$\boldsymbol{c}_p \equiv \frac{\omega}{K}\frac{\boldsymbol{k}}{|K|} = c_p\frac{\boldsymbol{k}}{|K|}, \tag{6.13}$$

where $\boldsymbol{k}/|K|$ is the unit vector in the direction of wave-crest propagation. The components of the phase velocity in the the $x$- and $y$-directions are then given by

$$c_p^x = c_p\cos\phi, \qquad c_p^y = c_p\sin\phi. \tag{6.14}$$

Defined this way, the quantity given by (6.13) is indeed a true vector velocity. However, the components in the $x$- and $y$-directions are manifestly not the speed at which wave crests propagate in those directions. It is therefore a misnomer to call these quantities phase speeds, although it is helpful to ascribe a direction to the phase speed and so the quantity given by (6.13) can be useful.

### 6.1.3 The dispersion relation

The above description is mostly kinematic and a little abstract, applying to almost any disturbance that has a wavevector and a frequency. The particular *dynamics* of a wave are determined by the relationship between the wavevector and the frequency; that is, by the *dispersion relation*. Once the dispersion relation is known a great many of the properties of the wave follow in a more-or-less straightforward manner, as we will see. Picking up from (6.3), the dispersion relation is a functional relationship between the frequency and the wavevector of the general form

$$\omega = \Omega(\boldsymbol{k}). \tag{6.15}$$

Perhaps the simplest example of a linear operator that gives rise to waves is the one-dimensional equation

$$\frac{\partial \psi}{\partial t} + c\frac{\partial \psi}{\partial x} = 0. \tag{6.16}$$

Substituting a trial solution of the form $\psi = \mathrm{Re}\, A e^{\mathrm{i}(kx-\omega t)}$, where Re denotes the real part, we obtain $(-\mathrm{i}\omega + c\mathrm{i}k)A = 0$, giving the dispersion relation

$$\omega = ck. \tag{6.17}$$

The phase speed of this wave is $c_p = \omega/k = c$. A few other examples of governing equations, dispersion relations and phase speeds are:

$$\frac{\partial \psi}{\partial t} + \boldsymbol{c}\cdot\nabla\psi = 0, \qquad \omega = \boldsymbol{c}\cdot\boldsymbol{k}, \qquad c_p = |\boldsymbol{c}|\cos\theta, \qquad c_p^x = \frac{\boldsymbol{c}\cdot\boldsymbol{k}}{k}, \qquad c_p^y = \frac{\boldsymbol{c}\cdot\boldsymbol{k}}{l} \tag{6.18a}$$

$$\frac{\partial^2 \psi}{\partial t^2} - c^2\nabla^2\psi = 0, \qquad \omega^2 = c^2K^2, \qquad c_p = \pm c, \qquad c_p^x = \pm\frac{cK}{k}, \qquad c_p^y = \pm\frac{cK}{l}, \tag{6.18b}$$

$$\frac{\partial}{\partial t}\nabla^2\psi + \beta\frac{\partial \psi}{\partial x} = 0, \qquad \omega = \frac{-\beta k}{K^2}, \qquad c_p = \frac{\omega}{K}, \qquad c_p^x = -\frac{\beta}{K^2}, \qquad c_p^y = -\frac{\beta k/l}{K^2}. \tag{6.18c}$$

where $K^2 = k^2 + l^2$ and $\theta$ is the angle between $\boldsymbol{c}$ and $\boldsymbol{k}$, and the examples are all two-dimensional, with variation in $x$ and $y$ only.

A wave is said to be *nondispersive* or *dispersionless* if the phase speed is independent of the wavelength. This condition is clearly satisfied for the simple example (6.16) but is manifestly not satisfied for (6.18c), and these waves (Rossby waves, in fact) are *dispersive*. Waves of different wavelengths then travel at different speeds so that a group of waves will spread out — disperse — even if the medium is homogeneous. When a wave is dispersive there is another characteristic speed at which the waves propagate, known as the group velocity, and we come to this in the next section.

Most media are, of course, inhomogeneous, but if the medium varies sufficiently slowly in space and time — and in particular if the variations are slow compared to the wavelength and period — we may still have a *local* dispersion relation between frequency and wavevector,

$$\omega = \Omega(\boldsymbol{k}; \boldsymbol{x}, t). \tag{6.19}$$

Although $\Omega$ is a function of $\boldsymbol{k}$, $\boldsymbol{x}$ and $t$ the semi-colon in (6.19) is used to suggest that $x$ and $t$ are slowly varying parameters of a somewhat different nature than $\boldsymbol{k}$. We'll resume our discussion of this topic in section 6.3, but before that we must introduce the group velocity.

## 6.2 GROUP VELOCITY

Information and energy clearly cannot travel at the phase speed, for as the direction of propagation of the phase line tends to a direction parallel to the $y$-axis, the phase speed in the $x$-direction tends to infinity! Rather, it turns out that most quantities of interest, including energy, propagate at the *group velocity,* a quantity of enormous importance in wave theory.[3]

## Wave Fundamentals

- A wave is a propagating disturbance that has a characteristic relationship between its frequency and size, known as the dispersion relation. Waves typically arise as solutions to a linear problem of the form
$$L(\psi) = 0, \tag{WF.1}$$
where $L$ is, commonly, a linear operator in space and time. Two examples are
$$\frac{\partial^2 \psi}{\partial t^2} - c^2 \nabla^2 \psi = 0 \qquad \text{and} \qquad \frac{\partial}{\partial t} \nabla^2 \psi + \beta \frac{\partial \psi}{\partial x} = 0. \tag{WF.2}$$
The first example is so common in all areas of physics it is sometimes called 'the' wave equation. The second example gives rise to Rossby waves.

- Solutions to the governing equation are often sought in the form of plane waves that have the form
$$\psi = \operatorname{Re} A e^{i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)}, \tag{WF.3}$$
where $A$ is the wave amplitude, $\boldsymbol{k} = (k, l, m)$ is the wavevector, and $\omega$ is the frequency.

- The dispersion relation connects the frequency and wavevector through an equation of the form $\omega = \Omega(\boldsymbol{k})$ where $\Omega$ is some function. The relation is normally derived by substituting a trial solution like (WF.3) into the governing equation (WF.1). For the examples of (WF.2) we obtain $\omega = c^2 K^2$ and $\omega = -\beta k / K^2$ where $K^2 = k^2 + l^2 + m^2$ or, in two dimensions, $K^2 = k^2 + l^2$.

- The phase speed is the speed at which the wave crests move. In the direction of propagation and in the $x$, $y$ and $z$ directions the phase speed is given by, respectively,
$$c_p = \frac{\omega}{K}, \qquad c_p^x = \frac{\omega}{k}, \quad c_p^y = \frac{\omega}{l}, \quad c_p^z = \frac{\omega}{m}. \tag{WF.4}$$
where $K = 2\pi/\lambda$ where $\lambda$ is the wavelength. The wave crests have both a speed ($c_p$) and a direction of propagation (the direction of $\boldsymbol{k}$), like a vector, but the components defined in (WF.4) are not the components of that vector.

- The group velocity is the velocity at which a wave packet or wave group moves. It is a vector and is given by
$$\boldsymbol{c}_g = \frac{\partial \omega}{\partial \boldsymbol{k}} \quad \text{with components} \quad c_g^x = \frac{\partial \omega}{\partial k}, \quad c_g^y = \frac{\partial \omega}{\partial l}, \quad c_g^z = \frac{\partial \omega}{\partial m}. \tag{WF.5}$$
Most physical quantities of interest are transported at the group velocity.

- If the coefficients of the wave equation are not constant (for example if the medium is inhomogeneous) then, if the coefficients are only slowly varying, approximate solutions may sometimes be found in the form
$$\psi = \operatorname{Re} A(\boldsymbol{x}, t) e^{i\theta(\boldsymbol{x}, t)}, \tag{WF.6}$$
where the amplitude $A$ is also slowly varying and the local wavenumber and frequency are related to the phase, $\theta$, by $\boldsymbol{k} = \nabla\theta$ and $\omega = -\partial\theta/\partial t$. The dispersion relation is then a *local* one of the form $\omega = \Omega(\boldsymbol{k}; x, t)$.

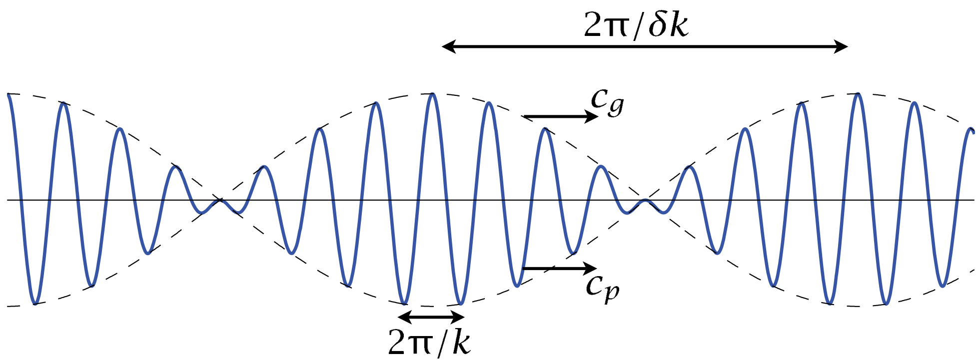


**Fig. 6.2** Superposition of two sinusoidal waves with wavenumbers $k$ and $k + \delta k$, producing a wave (solid line) that is modulated by a slowly varying wave envelope or wave packet (dashed line). The envelope moves at the group velocity, $c_g = \partial\omega/\partial k$ and the phase of the wave moves at the group speed $c_p = \omega/k$.

Roughly speaking, group velocity is the velocity at which a packet or a group of waves will travel, whereas the individual wave crests travel at the phase speed. To introduce the idea we will consider the superposition of plane waves, noting that a monochromatic plane wave already fills space uniformly so that there can be no propagation of energy from place to place. We will restrict attention to waves propagating in one direction, but the argument may be extended to two or three dimensions.

### 6.2.1 Superposition of two waves

Consider the linear superposition of two waves. Limiting attention to the one-dimensional case for simplicity, consider a disturbance represented by

$$\psi = \mathrm{Re}\,\widetilde{\psi}(\mathrm{e}^{\mathrm{i}(k_1 x - \omega_1 t)} + \mathrm{e}^{\mathrm{i}(k_2 x - \omega_2 t)}). \tag{6.20}$$

Let us further suppose that the two waves have similar wavenumbers and frequency, and, in particular, that $k_1 = k + \Delta k$ and $k_2 = k - \Delta k$, and $\omega_1 = \omega + \Delta\omega$ and $\omega_2 = \omega - \Delta\omega$. With this, (6.20) becomes

$$\begin{aligned} \psi &= \mathrm{Re}\,\widetilde{\psi}\mathrm{e}^{\mathrm{i}(kx-\omega t)}[\mathrm{e}^{\mathrm{i}(\Delta k\, x - \Delta\omega\, t)} + \mathrm{e}^{-\mathrm{i}(\Delta k\, x - \Delta\omega\, t)}] \\ &= 2\,\mathrm{Re}\,\widetilde{\psi}\mathrm{e}^{\mathrm{i}(kx-\omega t)}\cos(\Delta k\, x - \Delta\omega\, t). \end{aligned} \tag{6.21}$$

The resulting disturbance, illustrated in Fig. 6.2 has two aspects: a rapidly varying component, with wavenumber $k$ and frequency $\omega$, and a more slowly varying envelope, with wavenumber $\Delta k$ and frequency $\Delta\omega$. The envelope modulates the fast oscillation, and moves with velocity $\Delta\omega/\Delta k$; in the limit $\Delta k \to 0$ and $\Delta\omega \to 0$ this is the *group velocity*, $c_g = \partial\omega/\partial k$. Group velocity is equal to the phase speed, $\omega/k$, only when the frequency is a linear function of wavenumber. The energy in the disturbance must move at the group velocity — note that the node of the envelope moves at the speed of the envelope and no energy can cross the node. These concepts generalize to more than one dimension, and if the wavenumber is the three-dimensional vector

$\boldsymbol{k} = (k, l, m)$ then the three-dimensional envelope propagates at the group velocity given by

$$\boxed{\boldsymbol{c}_g = \frac{\partial \omega}{\partial \mathbf{k}} \equiv \left( \frac{\partial \omega}{\partial k}, \frac{\partial \omega}{\partial l}, \frac{\partial \omega}{\partial m} \right)}. \tag{6.22}$$

The group velocity is also written as $\boldsymbol{c}_g = \nabla_{\boldsymbol{k}} \omega$ or, in subscript notation, $c_{gi} = \partial \Omega / \partial k_i$, with the subscript $i$ denoting the component of a vector.

### 6.2.2 ✦ Superposition of many waves

Now consider a generalization of the above arguments to the case in which many waves are excited. In a homogeneous medium, nearly all wave patterns can be represented as a superposition of an infinite number of plane waves; mathematically the problem is solved by evaluating a Fourier integral. For mathematical simplicity we'll continue to treat only the one-dimensional case but the three dimensional generalization is possible.

A superposition of plane waves, each satisfying some dispersion relation, can be represented by the Fourier integral

$$\psi(x, t) = \int_{-\infty}^{\infty} \widetilde{A}(k) \mathrm{e}^{\mathrm{i}(kx - \omega t)} \, \mathrm{d}k. \tag{6.23a}$$

The function $\widetilde{A}(k)$ is given by the initial conditions:

$$\widetilde{A}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(x, 0) \mathrm{e}^{-\mathrm{i}kx} \, \mathrm{d}x. \tag{6.23b}$$

As an aside, note that if the waves are dispersionless and $\omega = ck$ where $c$ is a constant, then

$$\psi(x, t) = \int_{-\infty}^{+\infty} \widetilde{A}(k) \mathrm{e}^{\mathrm{i}k(x - ct)} \, \mathrm{d}k = \psi(x - ct, 0), \tag{6.24}$$

by comparison with (6.23a) at $t = 0$. That is, the initial condition simply translates at a speed $c$, with no change in structure.

Although the above procedure is quite general it doesn't get us very far: it doesn't provide us with any physical intuition, and the integrals themselves may be hard to evaluate. A physically more revealing case is to consider the case for which the disturbance is a *wave packet* — essentially a nearly plane wave or superposition of waves but confined to a finite region of space. We will consider a case with the initial condition

$$\psi(x, 0) = a(x, 0) \mathrm{e}^{\mathrm{i}k_0 x} \tag{6.25}$$

where $a(x, t)$, rather like the envelope in Fig. 6.3, modulates the amplitude of the wave on a scale much longer than that of the wavelength $2\pi / k_0$, and more slowly than the wave period. That is,

$$\frac{1}{a} \frac{\partial a}{\partial x} \ll k_0, \qquad \frac{1}{a} \frac{\partial a}{\partial t} \ll k_0 c, \tag{6.26a,b}$$

and the disturbance is essentially a slowly modulated plane wave. We suppose that $a(x, 0)$ is peaked around some value $x_0$ and is very small if $|x - x_0| \gg k_0^{-1}$; that is, $a(x, 0)$ is small if we are sufficiently many wavelengths of the plane wave away from the peak, as is the case in Fig. 6.3. We would like to know how such a packet evolves.

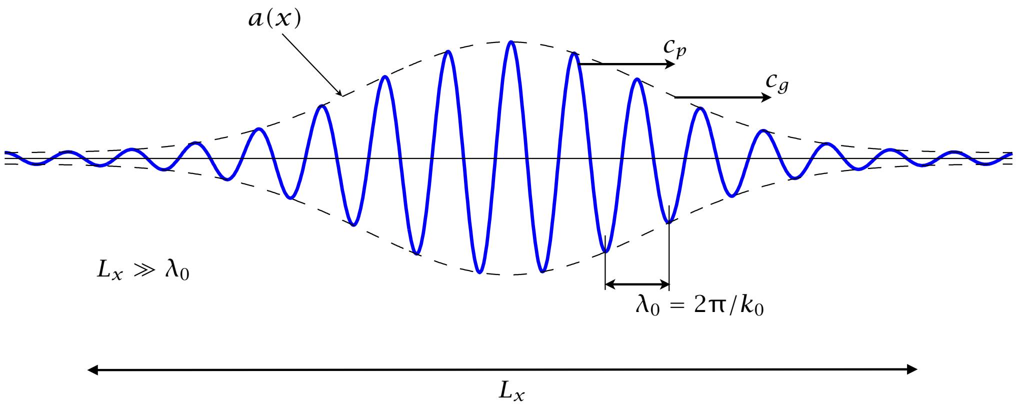


**Fig. 6.3** A generic wave packet. The envelope, $a(x)$, has a scale $L_x$ that is much larger than the wavelength, $\lambda_0$, of the wave embedded within in. The envelope moves at the group velocity, $c_g$, and the phase of the waves at the phase speed, $c_p$.

We can express the envelope as a Fourier integral by first noting that that we can write the initial conditions as a Fourier integral,

$$\psi(x,0) = \int_{-\infty}^{\infty} \widetilde{A}(k) e^{ikx} \, dk \quad \text{where} \quad \widetilde{A}(k) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi(x,0) e^{-ikx} \, dx, \tag{6.27a,b}$$

so that, using (6.25),

$$\widetilde{A}(k) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} a(x,0) e^{i(k_0-k)x} \, dx \quad \text{and} \quad a(x,0) = \int_{-\infty}^{\infty} \widetilde{A}(k) e^{i(k-k_0)x} \, dk. \tag{6.28a,b}$$

We still haven't made much progress beyond (6.23). To do so, we note first that $a(x)$ is confined in space, so that to a good approximation the limits of the integral in (6.28a) can be made finite, $\pm L$ say, provided $L \gg k_0^{-1}$. We then note that when $(k_0 - k)$ is large the integrand in (6.28a) oscillates rapidly; successive intervals in $x$ therefore cancel each other and make a small net contribution to the integral. Thus, the integral is dominated by values of $k$ near $k_0$, and $\widetilde{A}(k)$ is peaked near $k_0$. (Note that the finite spatial extent of $a(x,0)$ is crucial for this argument.)

We can now evaluate how the wave packet evolves. Beginning with (6.23a) we have

$$\psi(x,t) = \int_{-\infty}^{\infty} \widetilde{A}(k) \exp\{i(kx - \omega(k)t)\} \, dk \tag{6.29a}$$

$$\approx \int_{-\infty}^{\infty} \widetilde{A}(k) \exp\left\{ i[k_0 x - \omega(k_0)t] + i(k-k_0)x - i(k-k_0) \left.\frac{\partial \omega}{\partial k}\right|_{k=k_0} t \right\} dk \tag{6.29b}$$

having expanded $\omega(k)$ in a Taylor series about $k_0$ and kept only the first two terms, noting that even though the integral is formally over all wavenumbers, the wavenumber band is effectively limited to a region close to 0. We therefore have

$$\psi(x,t) = \exp\{i[k_0 x - \omega(k_0)t]\} \int \int_{-\infty}^{\infty} \widetilde{A}(k) \exp\left\{ i(k-k_0) \left[ x - \left.\frac{\partial \omega}{\partial k}\right|_{k=k_0} t \right] \right\} dk \tag{6.30a}$$

$$= \exp\{\mathrm{i}[k_0 x - \omega(k_0)t]\}\, a\left(x - c_g t\right), \tag{6.30b}$$

using (6.28b) and where $c_g = \partial\omega/\partial k$ evaluated at $k = k_0$. That is to say, the envelope $a(x,t)$ moves at the group velocity, keeping its initial shape.

The group velocity has a meaning beyond that implied by the derivation above: there is no need to restrict attention to narrow band processes, and it turns out to be a quite general property of waves that energy (and certain other quadratic properties) propagate at the group velocity. This is to be expected, at least in the presence of coherent wave packets, because there is no energy outside of the wave envelope so the energy must propagate with the envelope. Let's now examine this more closely.

### 6.2.3 ✦ The method of stationary phase

We will now relax the assumption that wavenumbers are confined to a narrow band but (since there is no free lunch) we confine ourselves to seeking solutions at large $t$; that is, we will be seeking a description of waves far from their source. Consider a disturbance of the general form

$$\psi(x,t) = \int_{-\infty}^{\infty} \widetilde{A}(k)\mathrm{e}^{\mathrm{i}[kx-\omega(k)t]}\,\mathrm{d}k = \int_{-\infty}^{\infty} \widetilde{A}(k)\mathrm{e}^{\mathrm{i}\Theta(k;x,t)t}\,\mathrm{d}k \tag{6.31}$$

where $\Theta(k;x,t) \equiv kx/t - \omega(k)$. (Here we regard $\Theta$ as a function of $k$ with parameters $x$ and $t$; we will sometimes just write $\Theta(k)$ with $\Theta'(k) = \partial\Theta/\partial k$.) Now, a standard result in mathematics (known as the 'Riemann–Lebesgue lemma') states that

$$I = \lim_{t\to\infty} \int_{-\infty}^{\infty} f(k)\mathrm{e}^{\mathrm{i}kt}\,\mathrm{d}k = 0 \tag{6.32}$$

provided that $f(k)$ is integrable and $\int_{-\infty}^{\infty} f(k)\,\mathrm{d}k$ is finite. Intuitively, as $t$ increases the oscillations in the integral increase and become much faster than any variation in $f(k)$; successive oscillations thus cancel and the integral becomes very small.

Looking at (6.31), with $\widetilde{A}$ playing the role of $f(k)$, the integral will be small if $\Theta$ is everywhere varying with $k$. However, if there is a region where $\Theta$ does not vary with $k$ — that is, if there is a region where the phase is stationary and $\partial\Theta/\partial k = 0$ — then there will be a contribution to the integral from that region. Thus, for large $t$, an observer will predominantly see waves for which $\Theta'(k) = 0$ and so, using the definition of $\Theta$, for which

$$\frac{x}{t} = \frac{\partial\omega}{\partial k}. \tag{6.33}$$

In other words, at some space-time location $(x,t)$ the waves that dominate are those whose group velocity $\partial\omega/\partial k$ is $x/t$. An example is plotted in Fig. 6.4 with a dispersion relation $\omega = -\beta/k$; the wavenumber that dominates, $k_0$ say, is thus given by solving $\beta/k_0^2 = x/t$, which for $x/t = 1$ and $\beta = 400$ gives $k_0 = 20$.

We may actually approximately calculate the contribution to $\psi(x,t)$ from waves moving with the group velocity. Let us expand $\Theta(k)$ around the point, $k_0$, where $\Theta'(k_0) = 0$. We obtain

$$\psi(x,t) = \int_{-\infty}^{\infty} \widetilde{A}(k) \exp\left\{\mathrm{i}t\left[\Theta(k_0) + (k-k_0)\Theta'(k_0) + \tfrac{1}{2}(k-k_0)^2\Theta''(k_0)\ldots\right]\right\}\mathrm{d}k \tag{6.34}$$

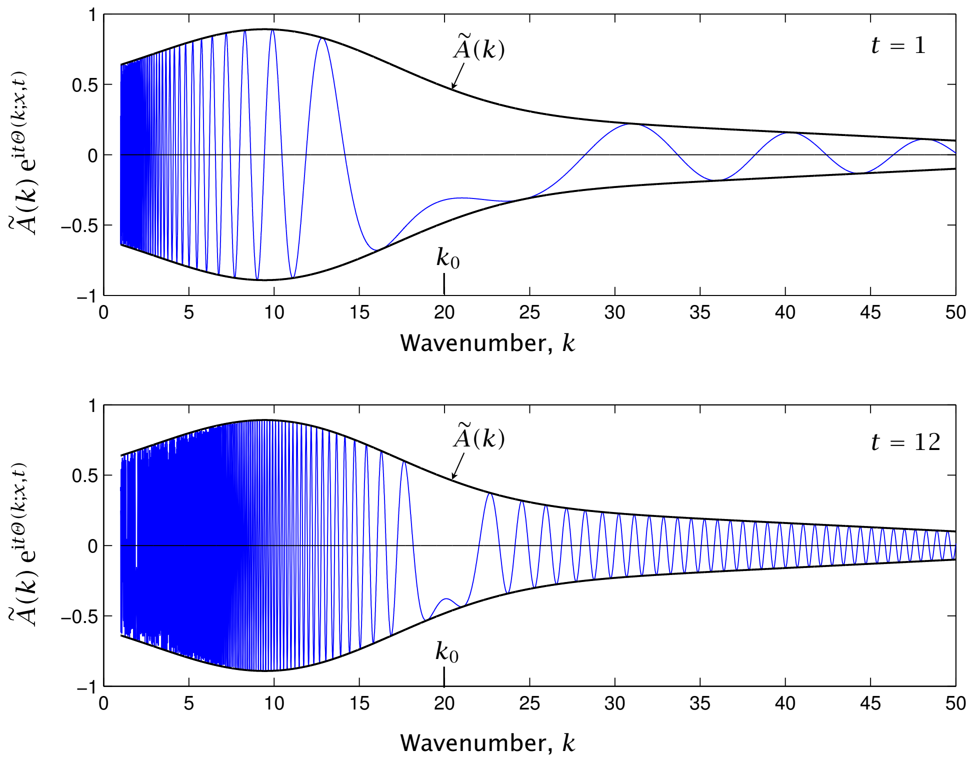


**Fig. 6.4** The integrand of (6.31), namely the function that when integrated over wavenumber gives the wave amplitude at a particular $x$ and $t$. The example shown is for a Rossby wave with $\omega = -\beta/k$, with $\beta = 400$ and $x/t = 1$, and hence $k_0 = 20$, for two times $t = 1$ and $t = 12$. (The amplitude of the envelope, $\widetilde{A}(k)$, diminishes at high wavenumber but is otherwise arbitrary.) At the later time the oscillations are much more rapid in $k$, so that the contribution is more peaked from wavenumbers near to $k_0$.

The higher order terms are small because $k - k_0$ is presumed small (for if it is large the integral vanishes), and the term involving $\Theta'(k_0)$ is zero. The integral becomes

$$\psi(x,t) = \widetilde{A}(k_0)\mathrm{e}^{\mathrm{i}\Theta(k_0)} \int_{-\infty}^{\infty} \exp\left\{\mathrm{i}t\tfrac{1}{2}(k-k_0)^2\Theta''(k_0)\right\}\,\mathrm{d}k. \tag{6.35}$$

We therefore have to evaluate a Gaussian, and because $\int_{-\infty}^{\infty} \mathrm{e}^{-cx^2}\,\mathrm{d}x = \sqrt{\pi/c}$ we obtain

$$\psi(x,t) \approx \widetilde{A}(k_0)\mathrm{e}^{\mathrm{i}\Theta(k_0)}\left[-2\pi/(\mathrm{i}t\theta''(k_0))\right]^{1/2} = \widetilde{A}(k_0)\mathrm{e}^{\mathrm{i}(k_0x-\omega(k_0)t)}\left[2\mathrm{i}\pi/(t\theta''(k_0))\right]^{1/2}. \tag{6.36}$$

The solution is therefore a plane wave, with wavenumber $k_0$ and frequency $\omega(k_0)$, slowly modulated by an envelope determined by the form of $\Theta(k_0; x, t)$, where $k_0$ is the wavenumber such that $x/t = c_g = \partial\omega/\partial k|_{k=k_0}$. [More discussion here, and some relevance to observational data?]

## 6.3 RAY THEORY

Most waves propagate in a media that is inhomogeneous. In the Earth's atmosphere and ocean the stratification varies with altitude and the Coriolis parameter varies with latitude. In these cases it can be hard to obtain the solution of a wave problem by Fourier methods, even approximately. Nonetheless, the ideas of signals propagating at the group velocity is a very robust one, and it turns out that we can often obtain much of the information we want — and in particular the trajectory of a wave — using an approximate recipe known as *ray theory,* using the word theory a little generously.[4]

In an inhomogeneous medium let us suppose that the solution to a particular wave problem is of the form

$$\psi(\boldsymbol{x}, t) = a(\boldsymbol{x}, t)\mathrm{e}^{\mathrm{i}\theta(\boldsymbol{x},t)}, \tag{6.37}$$

where $a$ is the wave amplitude and $\theta$ the phase, and $a$ varies slowly in a sense we will make more precise shortly. The local wavenumber and frequency are defined by,

$$k_i \equiv \frac{\partial \theta}{\partial x_i}, \qquad \omega \equiv -\frac{\partial \theta}{\partial t}. \tag{6.38}$$

where the first expression is equivalent to $\boldsymbol{k} \equiv \nabla\theta$ and so $\nabla \times \boldsymbol{k} = 0$. We suppose that the amplitude $a$ varies slowly over a wavelength and a period; that is $|\Delta a|/|a|$ is small over the length $1/k$ and the period $1/\omega$ or

$$\frac{|\partial a/\partial x|}{a} \ll |k|, \qquad \frac{|\partial a/\partial t|}{a} \ll \omega, \tag{6.39}$$

and similarly in the other directions. We will assume that the wavenumber and frequency as defined by (6.38) are the same as those that would arise if the medium were homogeneous and $a$ were a constant. Thus, we may obtain a local dispersion relation from the governing equation by keeping the spatially (and possibly temporally) varying parameters fixed and obtain

$$\omega = \Omega(k_i; x_i, t), \tag{6.40}$$

and then allow $x_i$ and $t$ to vary, albeit slowly.

Let us now consider how the wavevector and frequency might change with position and time. It follows from their definitions above that the wavenumber and frequency are related by

$$\frac{\partial k_i}{\partial t} + \frac{\partial \omega}{\partial x_i} = 0, \tag{6.41}$$

where we use a subscript notation for vectors and repeated indices are summed. Using (6.41) and (6.40) gives

$$\frac{\partial k_i}{\partial t} + \frac{\partial \Omega}{\partial x_i} + \frac{\partial \Omega}{\partial k_j}\frac{\partial k_j}{\partial x_i} = 0 \qquad \text{or} \qquad \frac{\partial k_i}{\partial t} + \frac{\partial \Omega}{\partial x_i} + \frac{\partial \Omega}{\partial k_j}\frac{\partial k_i}{\partial x_j} = 0, \tag{6.42a,b}$$

where to get (6.42b) we use $\partial k_j/\partial x_i = \partial k_i/\partial x_j$, allowable as $\boldsymbol{k}$ has no curl. Equation (6.42b) may be written as

$$\frac{\partial \boldsymbol{k}}{\partial t} + \boldsymbol{c}_g \cdot \nabla \boldsymbol{k} = -\nabla\Omega \tag{6.43}$$

where

$$c_g = \frac{\partial \Omega}{\partial \boldsymbol{k}} = \left( \frac{\partial \Omega}{\partial k}, \frac{\partial \Omega}{\partial l}, \frac{\partial \Omega}{\partial m} \right) \tag{6.44}$$

is, once more, the group velocity. The left-hand side of (6.43) is similar to an advective derivative, but the velocity is a group velocity not a fluid velocity. Evidently, if the dispersion relation for frequency is not an explicit function of space *the wavevector is propagated at the group velocity.*

The frequency is, in general, a function of space, wavenumber and time, and from the dispersion relation, (6.40), its variation is governed by

$$\frac{\partial \omega}{\partial t} = \frac{\partial \Omega}{\partial t} + \frac{\partial \Omega}{\partial k_i}\frac{\partial k_i}{\partial t} = \frac{\partial \Omega}{\partial t} - \frac{\partial \Omega}{\partial k_i}\frac{\partial \omega}{\partial x_i} \tag{6.45}$$

using (6.41). Using the definition of group velocity, we may write (6.45) as

$$\frac{\partial \omega}{\partial t} + \boldsymbol{c}_g \cdot \nabla \omega = \frac{\partial \Omega}{\partial t}. \tag{6.46}$$

As with (6.43) the left-hand side is like an advective derivative, but with the velocity being a group velocity. Thus, if the dispersion relation is not a function of time, the frequency also propagates at the group velocity.

Motivated by (6.43) and (6.46) we define a *ray* as the trajectory traced by the group velocity, and we see that if the function $\Omega$ is not an explicit function of space or time, then *both the wavevector and the frequency are constant along a ray.*

### 6.3.1 Ray theory in practice

What use is ray theory? The idea is that we may use (6.43) and (6.46) to track a group of waves from one location to another without solving the full wave equations of motion. Indeed, it turns out that we can sometimes solve problems using ordinary differential equations (ODEs) rather than partial differential equations (PDEs).

Suppose that the initial conditions consist of a group of waves at a position $x_0$, for which the amplitude and wavenumber vary only slowly with position. We also suppose that we know the dispersion relation for the waves at hand; that is, we know the functional form of $\Omega(k; x, t)$. Now, the total derivate following the group velocity is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \boldsymbol{c}_g \cdot \nabla, \tag{6.47}$$

so that (6.43) and (6.46) may be written as

$$\frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t} = -\nabla \Omega, \tag{6.48a}$$

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -\frac{\partial \Omega}{\partial t}. \tag{6.48b}$$

These are ordinary differential equations for wavevector and frequency, solvable provided we know the right-hand sides; that is, provided we know the space and time location at which

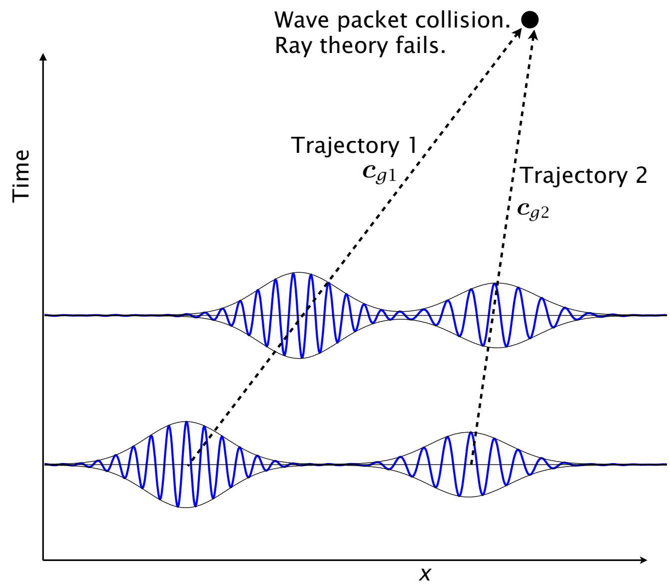


**Fig. 6.5** Schema of the trajectory of two wavepackets, each with a different wavelength and moving with a different group velocity, as might be calculated using ray theory. If the wave packets collide ray theory must fail. Ray theory gives only the trajectory of the wave packet, not the detailed structure of the waves within a packet.

the dispersion relation [i.e., $\Omega(k; x, t)$] is to be evaluated. But the location *is* known because it is moving with the group velocity and so

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{c}_g. \tag{6.48c}$$

where $\boldsymbol{c}_g = \partial\Omega/\partial\boldsymbol{k}|_{\boldsymbol{x},t}$ (i.e., $c_{gi} = \partial\Omega/\partial k_i|_{\boldsymbol{x},t}$). The set (6.48) is a triplet of ordinary differential equations for the wavevector, frequency and position of a wave group. The equations may be solved, albeit sometimes numerically, to give the trajectory of a wave packet or collection of wave packets as schematically illustrated in Fig. 6.5. Of course, if the medium or the wavepacket amplitude is not slowly varying ray theory will fail, and this will perforce happen if two wave packets collide.

The evolution of the amplitude of the wave packet is not given by ray theory. However, the evolution of a quantity related to the amplitude of a wave packet — specifically, the wave activity — may be calculated if the group velocity is known. It may be shown that the wave activity, $A$, satisfies $\partial A/\partial t + \nabla\cdot(\boldsymbol{c}_g A) = 0$; that is, the flux of wave activity is along a ray, but we leave further discussion to chapter 10. Another way to calculate the evolution of a wave and its amplitude in a varying medium is to use 'WKB theory' — see the appendix to chapter 7, with examples in section 7.5 and chapters 16 and 17. Before all that we turn our attention to a specific form of wave — Rossby waves — but the reader whose interest is more in the general properties of waves may skip forward to section 6.8.

## 6.4 ROSSBY WAVES

We now shift gears and consider in some detail a particular wave, namely the Rossby wave in a quasi-geostrophic system. Rossby waves are perhaps the most important large-scale wave in the atmosphere and ocean (although gravity waves, discussed in the next chapter, are arguably as important in some contexts).[5]

### 6.4.1 The linear equation of motion

For most of the rest of this chapter we will be concerned with the quasi-geostrophic equations of motion for which (as discussed in chapter 5) the inviscid, adiabatic potential vorticity equation is

$$\frac{\partial q}{\partial t} + \boldsymbol{u} \cdot \nabla q = 0, \tag{6.49}$$

where $q(x, y, z, t)$ is the potential vorticity and $\boldsymbol{u}(x, y, z, t)$ is the horizontal velocity. The velocity is related to a streamfunction by $u = -\partial\psi/\partial y$, $v = \partial\psi/\partial x$ and the potential vorticity is some function of the streamfunction, which might differ from system to system. Two examples, one applying to a continuously stratified system and the second to a single layer system, are

$$q = f + \zeta + \frac{\partial}{\partial z}\left(S(z)\frac{\partial\psi}{\partial z}\right), \qquad q = \zeta + f - k_d^2\psi. \tag{6.50a,b}$$

where $S(z) = f_0^2/N^2$, $\zeta = \nabla^2\psi$ is the relative vorticity and $k_d = 1/L_d$ is the inverse radius of deformation for a shallow water system. (Note that definitions of $k_d$ and $L_d$ can vary, typically by factors of 2, π, etc.) Boundary conditions may be needed to form a complete system.

We now *linearize* (6.49); that is, we suppose that the flow consists of a time-independent component (the 'basic state') plus a perturbation, with the perturbation being small compared with the mean flow. The basic state must satisfy the time-independent equation of motion, and it is common and useful to linearize about a zonal flow, $\overline{u}(y, z)$. The basic state is then purely a function of $y$ and so we write

$$q = \overline{q}(y, z) + q'(x, y, t), \qquad \psi = \overline{\psi}(y, z) + \psi'(x, y, z, t) \tag{6.51}$$

with a similar notation for the other variables. Note that $\overline{u} = -\partial\overline{\psi}/\partial y$ and $\overline{v} = 0$. Substituting into (6.49) gives, without approximation,

$$\frac{\partial q'}{\partial t} + \overline{\boldsymbol{u}} \cdot \nabla\overline{q} + \overline{\boldsymbol{u}} \cdot \nabla q' + \boldsymbol{u}' \cdot \nabla\overline{q} + \boldsymbol{u}' \cdot \nabla q' = 0. \tag{6.52}$$

The primed quantities are presumptively small so we neglect terms involving their products. Further, we are assuming that we are linearizing about a state that is a solution of the equations of motion, so that $\overline{\boldsymbol{u}} \cdot \nabla\overline{q} = 0$. Finally, since $\overline{v} = 0$ and $\partial\overline{q}/\partial x = 0$ we obtain

$$\boxed{\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial\overline{q}}{\partial y} = 0}\,. \tag{6.53}$$

This equation or one very similar appears very commonly in studies of Rossby waves. To proceed, let us consider the simple example of waves in a single layer.

### 6.4.2 Waves in a single layer

Consider a system obeying (6.49) and (6.50b). The equation could be written in spherical coordinates with $f = 2\Omega \sin\vartheta$, but the dynamics are more easily illustrated on Cartesian $\beta$-plane for which $f = f_0 + \beta y$, and since $f_0$ is a constant it does not appear in our subsequent derivations.

*Infinite deformation radius*

If the scale of motion is much less than the deformation scale then we make the approximation that $k_d = 0$ and the equation of motion may be written as

$$\frac{\partial \zeta}{\partial t} + \boldsymbol{u} \cdot \nabla \zeta + \beta v = 0 \tag{6.54}$$

We linearize about a constant zonal flow, $\overline{u}$, by writing

$$\psi = \overline{\psi}(y) + \psi'(x, y, t), \tag{6.55}$$

where $\overline{\psi} = -\overline{u} y$. Substituting(6.55) into (6.54) and neglecting the nonlinear terms involving products of $\psi'$ to give

$$\frac{\partial}{\partial t}\nabla^2\psi' + \overline{u}\frac{\partial \nabla^2 \psi'}{\partial x} + \beta\frac{\partial \psi'}{\partial x} = 0. \tag{6.56}$$

This equation is just a single-layer version of (6.53), with $\partial \overline{q}/\partial y = \beta$, $q' = \nabla^2\psi'$ and $v' = \partial \psi'/\partial x$.

The coefficients in (6.56) are not functions of $y$ or $z$; this is not a requirement for wave motion to exist but it does enable solutions to be found more easily. Let us seek solutions in the form of a plane wave, namely

$$\psi' = \mathrm{Re}\, \widetilde{\psi} \mathrm{e}^{\mathrm{i}(kx + ly - \omega t)}, \tag{6.57}$$

where $\widetilde{\psi}$ is a complex constant and Re indicates the real part of the function (a notation sometimes omitted if no ambiguity is so-caused). Solutions of this form are valid in a domain with doubly-periodic boundary conditions; solutions in a channel can be obtained using a meridional variation of $\sin ly$, with no essential changes to the dynamics. The amplitude of the oscillation is given by $\widetilde{\psi}$ and the phase by $kx + ly - \omega t$, where $k$ and $l$ are the $x$- and $y$-wavenumbers and $\omega$ is the frequency of the oscillation.

Substituting (6.57) into (6.56) yields

$$[(-\omega + Uk)(-K^2) + \beta k]\widetilde{\psi} = 0, \tag{6.58}$$

where $K^2 = k^2 + l^2$. For non-trivial solutions this implies

$$\boxed{\omega = Uk - \frac{\beta k}{K^2}}\,. \tag{6.59}$$

This is the *dispersion relation* for barotropic Rossby waves, and evidently the velocity $U$ Doppler shifts the frequency. The components of the phase speed and group velocity are given by, respectively,

$$c_p^x \equiv \frac{\omega}{k} = \overline{u} - \frac{\beta}{K^2}, \qquad c_p^y \equiv \frac{\omega}{l} = \overline{u}\frac{k}{l} - \frac{\beta k}{K^2 l}, \tag{6.60a,b}$$

and

$$c_g^x \equiv \frac{\partial \omega}{\partial k} = \overline{u} + \frac{\beta(k^2 - l^2)}{(k^2 + l^2)^2}, \qquad c_g^y \equiv \frac{\partial \omega}{\partial l} = \frac{2\beta k l}{(k^2 + l^2)^2}. \tag{6.61a,b}$$

The phase speed in the absence of a mean flow is *westwards,* with waves of longer wavelengths travelling more quickly, and the eastward current speed required to hold the waves of a particular wavenumber stationary (i.e., $c_p^x = 0$) is $U = \beta/K^2$. The background flow $\overline{u}$ evidently just provides a uniform shift to the phase speed, and could be transformed away by a change of coordinate.

*Finite deformation radius*

For a finite deformation radius the basic state $\Psi = -Uy$ is still a solution of the original equations of motion, but the potential vorticity corresponding to this state is $q = Uyk_d^2 + \beta y$ and its gradient is $\nabla q = (\beta + Uk_d^2)\mathbf{j}$. The linearized equation of motion is thus

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)(\nabla^2\psi' - \psi' k_d^2) + (\beta + \overline{u}k_d^2)\frac{\partial \psi'}{\partial x} = 0. \tag{6.62}$$

Substituting $\psi' = \widetilde{\psi}\mathrm{e}^{\mathrm{i}(kx+ly-\omega t)}$ we obtain the dispersion relation,

$$\boxed{\omega = \frac{k(UK^2 - \beta)}{K^2 + k_d^2} = Uk - k\frac{\beta + Uk_d^2}{K^2 + k_d^2}}. \tag{6.63}$$

The corresponding components of phase speed and group velocity are

$$c_p^x = \overline{u} - \frac{\beta + \overline{u}k_d^2}{K^2 + k_d^2} = \frac{\overline{u}K^2 - \beta}{K^2 + k_d^2}, \qquad c_p^y = \overline{u}\frac{k}{l} - \frac{k}{l}\left(\frac{\overline{u}K^2 - \beta}{K^2 + k_d^2}\right) \tag{6.64a,b}$$

and

$$c_g^x = \overline{u} + \frac{(\beta + \overline{u}k_d^2)(k^2 - l^2 - k_d^2)}{(k^2 + l^2 + k_d^2)^2}, \qquad c_g^y = \frac{2kl(\beta + \overline{u}k_d^2)}{(k^2 + l^2 + k_d^2)^2}. \tag{6.65a,b}$$

The uniform velocity field now no longer provides just a simple Doppler shift of the frequency, nor a uniform addition to the phase speed. From (6.64a) the waves are stationary when $K^2 = \beta/\overline{u} \equiv K_s^2$; that is, the current speed required to hold waves of a particular wavenumber stationary is $\overline{u} = \beta/K^2$. However, this is *not* simply the magnitude of the phase speed of waves of that wavenumber in the absence of a current — this is given by

$$c_p^x = \frac{-\beta}{K_s^2 + k_d^2} = \frac{-\overline{u}}{1 + k_d^2/K_s^2}. \tag{6.66}$$

Why is there a difference? It is because the current does not just provide a uniform translation, but, if $k_d$ is non-zero, it also modifies the basic potential vorticity gradient. The basic state height field $\eta_0$ is sloping; that is $\eta_0 = -(f_0/g)\overline{u}y$, and the ambient potential vorticity field increases with $y$ and $q = (\beta + Uk_d^2)y$. Thus, the basic state defines a preferred frame of reference, and the problem is not Galilean invariant.[6] We also note that, from (6.64b), the group velocity is negative (westward) if the $x$-wavenumber is sufficiently small, compared to the $y$-wavenumber or the deformation wavenumber. That is, said a little loosely, *long waves move information westward and short waves move information eastward,* and this is a common property of Rossby waves. The $x$-component of the phase speed, on the other hand, is always westward relative to the mean flow.

### 6.4.3 The mechanism of Rossby waves

The fundamental mechanism underlying Rossby waves is easily understood. Consider a material line of stationary fluid parcels along a line of constant latitude, and suppose that some disturbance causes their displacement to the line marked $\eta(t = 0)$ in Fig. 6.6. In the displacement,

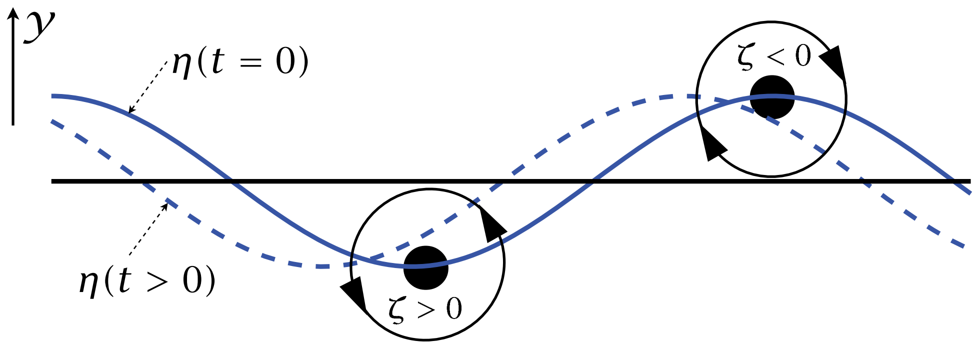


**Fig. 6.6** The mechanism of a two-dimensional ($x$–$y$) Rossby wave. An initial disturbance displaces a material line at constant latitude (the straight horizontal line) to the solid line marked $\eta(t = 0)$. Conservation of potential vorticity, $\beta y + \zeta$, leads to the production of relative vorticity, as shown for two parcels. The associated velocity field (arrows on the circles) then advects the fluid parcels, and the material line evolves into the dashed line. The phase of the wave has propagated westwards.

the potential vorticity of the fluid parcels is conserved, and in the simplest case of barotropic flow on the $\beta$-plane the potential vorticity is the absolute vorticity, $\beta y + \zeta$. Thus, in either hemisphere, a northward displacement leads to the production of negative relative vorticity and a southward displacement leads to the production of positive relative vorticity. The relative vorticity gives rise to a velocity field which, in turn, advects the parcels in material line in the manner shown, and the wave propagates westwards.

In more complicated situations, such as flow in two layers, considered below, or in a continuously stratified fluid, the mechanism is essentially the same. A displaced fluid parcel carries with it its potential vorticity and, in the presence of a potential vorticity gradient in the basic state, a potential vorticity anomaly is produced. The potential vorticity anomaly produces a velocity field (an example of potential vorticity inversion) which further displaces the fluid parcels, leading to the formation of a Rossby wave. The vital ingredient is a basic state potential vorticity gradient, such as that provided by the change of the Coriolis parameter with latitude.

### 6.4.4 Rossby waves in two layers

Now consider the dynamics of the two-layer model, linearized about a state of rest. The two, coupled, linear equations describing the motion in each layer are

$$\frac{\partial}{\partial t}\left[\nabla^2\psi_1' + F_1(\psi_2' - \psi_1')\right] + \beta\frac{\partial \psi_1'}{\partial x} = 0, \tag{6.67a}$$

$$\frac{\partial}{\partial t}\left[\nabla^2\psi_2' + F_2(\psi_1' - \psi_2')\right] + \beta\frac{\partial \psi_2'}{\partial x} = 0, \tag{6.67b}$$

where $F_1 = f_0^2/g'H_1$ and $F_2 = f_0^2/g'H_2$. By inspection (6.67) may be transformed into two uncoupled equations: the first is obtained by multiplying (6.67a) by $F_2$ and (6.67b) by $F_1$ and adding, and the second is the difference of (6.67a) and (6.67b). Then, defining

$$\overline{\psi} = \frac{F_1\psi_2' + F_2\psi_1'}{F_1 + F_2}, \qquad \tau = \frac{1}{2}(\psi_1' - \psi_2'), \tag{6.68a,b}$$

(think '$\tau$ for temperature'), (6.67) become

$$\frac{\partial}{\partial t}\nabla^2\overline{\psi} + \beta\frac{\partial\overline{\psi}}{\partial x} = 0, \tag{6.69a}$$

$$\frac{\partial}{\partial t}\left[(\nabla^2 - k_d^2)\tau\right] + \beta\frac{\partial\tau}{\partial x} = 0, \tag{6.69b}$$

where now $k_d = (F_1 + F_2)^{1/2}$. The internal radius of deformation for this problem is the inverse of this, namely

$$L_d = k_d^{-1} = \frac{1}{f_0}\left(\frac{g' H_1 H_2}{H_1 + H_2}\right)^{1/2}. \tag{6.70}$$

The variables $\overline{\psi}$ and $\tau$ are the *normal modes* for the two-layer model, as they oscillate independently of each other. [For the continuous equations the analogous modes are the eigenfunctions of $\partial_z[(f_0^2/N^2)\partial_z\phi] = \lambda^2\phi$.] The equation for $\overline{\psi}$, *the barotropic mode,* is identical to that of the single-layer, rigid-lid model, namely (6.56) with $U = 0$, and its dispersion relation is just

$$\omega = -\frac{\beta k}{K^2}. \tag{6.71}$$

The barotropic mode corresponds to synchronous, depth-independent, motion in the two layers, with no undulations in the dividing interface.

The displacement of the interface is given by $2f_0\tau/g'$ and so proportional to the amplitude of $\tau$, the *baroclinic mode*. The dispersion relation for the baroclinic mode is

$$\omega = -\frac{\beta k}{K^2 + k_d^2}. \tag{6.72}$$

The mass transport associated with this mode is identically zero, since from (6.68) we have

$$\psi_1 = \overline{\psi} + \frac{2F_1\tau}{F_1 + F_2}, \qquad \psi_2 = \overline{\psi} - \frac{2F_2\tau}{F_1 + F_2}, \tag{6.73a,b}$$

and this implies

$$H_1\psi_1 + H_2\psi_2 = (H_1 + H_2)\overline{\psi}. \tag{6.74}$$

The left-hand side is proportional to the total mass transport, which is evidently associated with the barotropic mode.

The dispersion relation and associated group and phase velocities are plotted in Fig. 6.7. The $x$-component of the phase speed, $\omega/k$, is negative (westwards) for both baroclinic and barotropic Rossby waves. The group velocity of the barotropic waves is always positive (eastwards), but the group velocity of long baroclinic waves may be negative (westwards). For very short waves, $k^2 \gg k_d^2$, the baroclinic and barotropic velocities coincide and their phase and group velocities are equal and opposite. With a deformation radius of 50 km, typical for the mid-latitude ocean, then a non-dimensional frequency of unity in the figure corresponds to a dimensional frequency of $5 \times 10^{-7}\,\mathrm{s^{-1}}$ or a period of about 100 days. In an atmosphere with a deformation radius of 1000 km a non-dimensional frequency of unity corresponds to $1 \times 10^{-5}\,\mathrm{s^{-1}}$ or a period of about 7 days. Non-dimensional velocities of unity correspond to respective dimensional velocities of about $0.25\,\mathrm{m\,s^{-1}}$ (ocean) and $10\,\mathrm{m\,s^{-1}}$ (atmosphere).

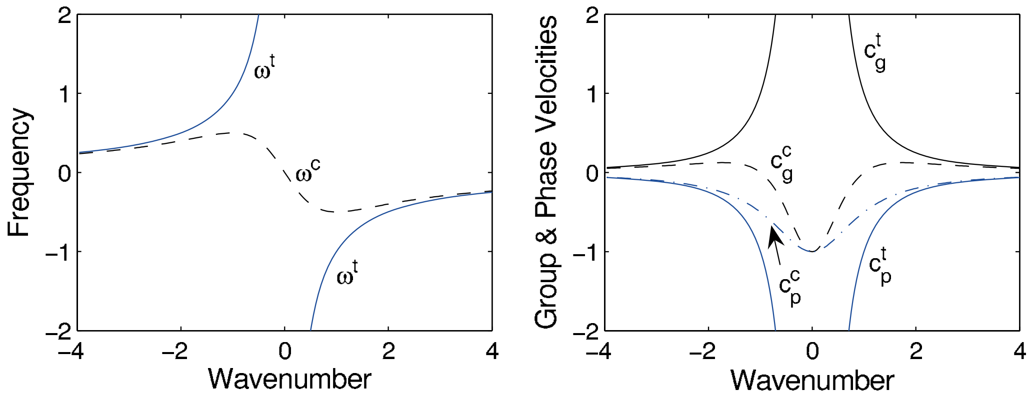


**Fig. 6.7** Left: the dispersion relation for barotropic ($\omega^t$, solid line) and baroclinic ($\omega^c$, dashed line) Rossby waves in the two-layer model, calculated using (6.71) and (6.72) with $k^y = 0$, plotted for both positive and negative zonal wavenumbers and frequencies. The wavenumber is non-dimensionalized by $k_d$, and the frequency is non-dimensionalized by $\beta/k_d$. Right: the corresponding zonal group and phase velocities, $c_g = \partial\omega/\partial k^x$ and $c_p = \omega/k^x$, with superscript 't' or 'c' for the barotropic or baroclinic mode, respectively. The velocities are non-dimensionalized by $\beta/k_d^2$.

The deformation radius only affects the baroclinic mode. For scales much smaller than the deformation radius, $K^2 \gg k_d^2$, we see from (6.69b) that the baroclinic mode obeys the same equation as the barotropic mode so that

$$\frac{\partial}{\partial t}\nabla^2\tau + \beta\frac{\partial \tau}{\partial x} = 0. \tag{6.75}$$

Using this and (6.69a) implies that

$$\frac{\partial}{\partial t}\nabla^2\psi_i + \beta\frac{\partial \psi_i}{\partial x} = 0, \qquad i = 1, 2. \tag{6.76}$$

That is to say, the two layers themselves are uncoupled from each other. At the other extreme, for very long baroclinic waves the relative vorticity is unimportant.

## 6.5 ROSSBY WAVES IN STRATIFIED QUASI-GEOSTROPHIC FLOW

### 6.5.1 Setting up the problem

Let us now consider the dynamics of linear waves in stratified quasi-geostrophic flow on a $\beta$-plane, with a resting basic state. (In chapter 16 we explore the role of Rossby waves in a more realistic setting.) The interior flow is governed by the potential vorticity equation, (5.118), and linearizing this about a state of rest gives

$$\frac{\partial}{\partial t}\left[\nabla^2\psi' + \frac{1}{\tilde{\rho}(z)}\frac{\partial}{\partial z}\left(\tilde{\rho}(z)F(z)\frac{\partial \psi'}{\partial z}\right)\right] + \beta\frac{\partial \psi'}{\partial x} = 0, \tag{6.77}$$

where $\widetilde{\rho}$ is the density profile of the basic state and $F(z) = f_0^2/N^2$. ($F$ is the square of the inverse Prandtl ratio, $N/f_0$.) In the Boussinesq approximation $\widetilde{\rho} = \rho_0$, i.e., a constant. The vertical boundary conditions are determined by the thermodynamic equation, (5.120). If the boundaries are flat, rigid, slippery surfaces then $w = 0$ at the boundaries and if there is no surface buoyancy gradient the linearized thermodynamic equation is

$$\frac{\partial}{\partial t}\left(\frac{\partial \psi'}{\partial z}\right) = 0. \tag{6.78}$$

We apply this at the ground and, with somewhat less justification, at the tropopause: we assume the higher static stability of the stratosphere inhibits vertical motion. If the ground is not flat or if friction provides a vertical velocity by way of an Ekman layer, the boundary condition must be correspondingly modified, but we will stay with the simplest case and apply (6.78) at $z = 0$ and $z = H$.

### 6.5.2 Wave motion

As in the single-layer case, we seek solutions of the form

$$\psi' = \mathrm{Re}\,\widetilde{\psi}(z)\mathrm{e}^{\mathrm{i}(kx+ly-\omega t)}, \tag{6.79}$$

where $\widetilde{\psi}(z)$ will determine the vertical structure of the waves. The case of a sphere is more complicated but introduces no truly new physical phenomena.

Substituting (6.79) into (6.77) gives

$$\omega\left[-K^2\widetilde{\psi}(z) + \frac{1}{\widetilde{\rho}}\frac{\mathrm{d}}{\mathrm{d}z}\left(\widetilde{\rho}F(z)\frac{\mathrm{d}\widetilde{\psi}}{\mathrm{d}z}\right)\right] - \beta k\widetilde{\psi}(z) = 0. \tag{6.80}$$

Now, if $\widetilde{\psi}$ satisfies

$$\frac{1}{\widetilde{\rho}}\frac{\mathrm{d}}{\mathrm{d}z}\left(\widetilde{\rho}F(z)\frac{\mathrm{d}\widetilde{\psi}}{\mathrm{d}z}\right) = -\Gamma\widetilde{\psi}, \tag{6.81}$$

where $\Gamma$ is a constant, then the equation of motion becomes

$$-\,\omega\left[K^2 + \Gamma\right]\widetilde{\psi} - \beta k\widetilde{\psi} = 0, \tag{6.82}$$

and the dispersion relation follows, namely

$$\boxed{\omega = -\frac{\beta k}{K^2 + \Gamma}}\,. \tag{6.83}$$

Equation (6.81) constitutes an eigenvalue problem for the vertical structure; the boundary conditions, derived from (6.78), are $\partial\widetilde{\psi}/\partial z = 0$ at $z = 0$ and $z = H$. The resulting eigenvalues, $\Gamma$ are proportional to the inverse of the squares of the deformation radii for the problem and the eigenfunctions are the vertical structure functions.

*A simple example*

Consider the case in which $F(z)$ and $\widetilde{\rho}$ are constant, and in which the domain is confined between two rigid surfaces at $z = 0$ and $z = H$. Then the eigenvalue problem for the vertical structure is

$$F\frac{\mathrm{d}^2\widetilde{\psi}}{\mathrm{d}z^2} = -\Gamma\widetilde{\psi} \tag{6.84a}$$

with boundary conditions of

$$\frac{\mathrm{d}\widetilde{\psi}}{\mathrm{d}z} = 0, \qquad \text{at } z = 0, H. \tag{6.84b}$$

There is a sequence of solutions to this, namely

$$\widetilde{\psi}_n(z) = \cos(n\pi z/H), \qquad n = 1, 2 \dots \tag{6.85}$$

with corresponding eigenvalues

$$\Gamma_n = n^2\frac{F\pi^2}{H^2} = (n\pi)^2\left(\frac{f_0}{NH}\right)^2, \qquad n = 1, 2 \dots. \tag{6.86}$$

Equation (6.86) may be used to define the deformation radii for this problem, namely

$$L_n \equiv \frac{1}{\sqrt{\Gamma_n}} = \frac{NH}{n\pi f_0}. \tag{6.87}$$

The first deformation radius is the same as the expression obtained by dimensional analysis, namely $NH/f$, except for a factor of $\pi$. (Definitions of the deformation radii both with and without the factor of $\pi$ are common in the literature, and neither is obviously more correct. In the latter case, the first deformation radius in a problem with uniform stratification is given by $NH/f$, equal to $\pi/\sqrt{\Gamma_1}$.) In addition to these baroclinic modes, the case with $n = 0$, that is with $\widetilde{\psi} = 1$, is also a solution of (6.84) for any $F(z)$.

Using (6.83) and (6.86) the dispersion relation becomes

$$\boxed{\omega = -\frac{\beta k}{K^2 + (n\pi)^2(f_0/NH)^2}, \qquad n = 0, 1, 2 \dots} \tag{6.88}$$

and, of course, the horizontal wavenumbers $k$ and $l$ are also quantized in a finite domain. The dynamics of the barotropic mode are independent of height and independent of the stratification of the basic state, and so these Rossby waves are *identical* with the Rossby waves in a homogeneous fluid contained between two flat rigid surfaces. The structure of the baroclinic modes, which in general depends on the structure of the stratification, becomes increasingly complex as the vertical wavenumber $n$ increases. This increasing complexity naturally leads to a certain delicacy, making it rare that they can be unambiguously identified in nature. The eigenproblem for a realistic atmospheric profile is further complicated because of the lack of a rigid lid at the top of the atmosphere.[7]

## Essentials of Rossby Waves

- Rossby waves owe their existence to a gradient of potential vorticity in the fluid. If a fluid parcel is displaced, it conserves its potential vorticity and so its relative vorticity will in general change. The relative vorticity creates a velocity field that displaces neighbouring parcels, whose relative vorticity changes and so on.
- A common source of a potential vorticity gradient is differential rotation, or the $\beta$-effect. In the presence of non-zero $\beta$ the ambient potential vorticity increases northward and the phase of the Rossby waves propagates westward. In general, Rossby waves propagate pseudo-westwards, meaning to the left of the direction of the potential vorticity gradient.
- A common equation of motion for Rossby waves is

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = 0, \tag{RW.1}$$

with an overbar denoting the basic state and a prime a perturbation. In the case of a single layer of fluid with no mean flow this equation becomes

$$\frac{\partial}{\partial t}(\nabla^2 + k_d^2)\psi' + \beta\frac{\partial \psi'}{\partial x} = 0 \tag{RW.2}$$

with dispersion relation

$$\omega = \frac{-\beta k}{k^2 + l^2 + k_d^2}. \tag{RW.3}$$

- The phase speed in the zonal direction ($c_p^x = \omega/k$) is always negative, or westward, and is larger for large waves. For (RW.2) components of the group velocity are given by

$$c_g^x = \frac{\beta(k^2 - l^2 - k_d^2)}{(k^2 + l^2 + k_d^2)^2}, \qquad c_g^y = \frac{2\beta kl}{(k^2 + l^2 + k_d^2)^2}. \tag{RW.4}$$

The group velocity is westward if the zonal wavenumber is sufficiently small, and eastward if the zonal wavenumber is sufficiently large.
- Rossby waves exist in stratified fluids, and have a similar dispersion relation to (RW.3) with an appropriate vertical wavenumber appearing in place of the inverse deformation radius, $k_d$.
- The reflection of such Rossby waves at a wall is specular, meaning that the group velocity of the reflected wave makes the same angle with the wall as the group velocity of the incident wave. The energy flux of the reflected wave is equal and opposite to that of the incoming wave in the direction normal to the wall.

## 6.6 ENERGY PROPAGATION AND REFLECTION OF ROSSBY WAVES

We now consider how energy is fluxed in Rossby waves. To keep matters reasonably simple from an algebraic point of view we will consider waves in a single layer and without a mean flow, but we will allow for a finite radius of deformation. To remind ourselves, the dynamics are governed by the evolution of potential vorticity and the linearized evolution equation is

$$\frac{\partial}{\partial t}\left(\nabla^2 - k_d^2\right)\psi + \beta\frac{\partial \psi}{\partial x} = 0. \tag{6.89}$$

The dispersion relation follows in the usual way and is

$$\omega = \frac{-k\beta}{K^2 + k_d^2}, \tag{6.90}$$

which is a simplification of (6.63), and the group velocities are

$$c_g^x = \frac{\beta(k^2 - l^2 - k_d^2)}{(K^2 + k_d^2)^2}, \qquad c_g^y = \frac{2\beta k l}{(K^2 + k_d^2)^2}, \tag{6.91a,b}$$

which are simplifications of (6.65), and as usual $K^2 = k^2 + l^2$.

To obtain an energy equation multiply (6.89) by $-\psi$ to obtain, after a couple of lines of algebra,

$$\frac{1}{2}\frac{\partial}{\partial t}\left((\nabla\psi)^2 + k_d^2\psi^2\right) - \nabla\cdot\left(\psi\nabla\frac{\partial\psi}{\partial t} + \mathbf{i}\frac{\beta}{2}\psi^2\right) = 0, \tag{6.92}$$

where $\mathbf{i}$ is the unit vector in the $x$ direction. The first group of terms are the energy itself, or more strictly the energy density. (An energy density is an energy per unit mass or per unit volume, depending on the context.) The term $(\nabla\psi)^2/2 = (u^2 + v^2)/2$ is the kinetic energy and $k_d^2\psi^2/2$ is the potential energy, proportional to the displacement of the free surface, squared. The second term is the energy flux, so that we may write

$$\frac{\partial E}{\partial t} + \nabla\cdot\boldsymbol{F} = 0. \tag{6.93}$$

where $E = (\nabla\psi)^2/2 + k_d^2\psi^2$ and $\boldsymbol{F} = -\left(\psi\nabla\partial\psi/\partial t + \mathbf{i}\beta\psi^2\right)$. We haven't yet used the fact that the disturbance has a dispersion relation, and if we do so we may expect, following the derivations of section 6.2, that the energy moves at the group velocity. Let us now demonstrate this explicitly.

We assume a solution of the form

$$\psi = A(x)\cos(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t) = A(x)\cos(kx + ly - \omega t) \tag{6.94}$$

where $A(x)$ is assumed to vary slowly compared to the nearly plane wave. (Note that $\boldsymbol{k}$ is the wave vector, to be distinguished from $\mathbf{k}$, the unit vector in the $z$-direction.) The kinetic energy in a wave is given by

$$KE = \frac{A^2}{2}\left(\psi_x^2 + \psi_y^2\right) \tag{6.95}$$

so that, averaged over a wave period,

$$\overline{KE} = \frac{A^2}{2}(k^2 + l^2)\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\sin^2(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)\,\mathrm{d}t. \tag{6.96}$$

The time-averaging produces a factor of one half, and applying a similar procedure1 to the potential energy we obtain

$$\overline{KE} = \frac{A^2}{4}(k^2 + l^2), \qquad \overline{PE} = \frac{A^2}{4}k_d^2, \tag{6.97}$$

so that the average total energy is

$$\overline{E} = \frac{A^2}{4}(K^2 + k_d^2), \tag{6.98}$$

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

The flux, $\boldsymbol{F}$, is given by

$$\boldsymbol{F} = -\left(\psi\nabla\frac{\partial\psi}{\partial t} + \mathbf{i}\frac{\beta}{2}\psi^2\right) = -A^2\cos^2(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)\left(\boldsymbol{k}\omega - \mathbf{i}\frac{\beta}{2}\right), \tag{6.99}$$

so that evidently the energy flux has a component in the direction of the wavevector, $\boldsymbol{k}$, and a component in the $x$-direction. Averaging over a wave period straightforwardly gives us additional factors of one half:

$$\overline{\boldsymbol{F}} = -\frac{A^2}{2}\left(\boldsymbol{k}\omega + \mathbf{i}\frac{\beta}{2}\right). \tag{6.100}$$

We now use the dispersion relation $\omega = -\beta k/(K^2 + k_d^2)$ to eliminate the frequency, giving

$$\overline{\boldsymbol{F}} = \frac{A^2\beta}{2}\left(\boldsymbol{k}\frac{k}{K^2 + k_d^2} - \mathbf{i}\frac{1}{2}\right), \tag{6.101}$$

and writing this in component form we obtain

$$\overline{\boldsymbol{F}} = \frac{A^2\beta}{4}\left[\mathbf{i}\left(\frac{k^2 - l^2 - k_d^2}{K^2 + k_d^2}\right) + \mathbf{j}\left(\frac{2kl}{K^2 + k_d^2}\right)\right] \tag{6.102}$$

Comparison of (6.102) with (6.91) and (6.98) reveals that

$$\overline{\boldsymbol{F}} = \boldsymbol{c}_g\overline{E} \tag{6.103}$$

so that the energy propagation equation, (6.93), when averaged over a wave, becomes

$$\boxed{\frac{\partial\overline{E}}{\partial t} + \nabla\cdot\boldsymbol{c}_g\overline{E} = 0}. \tag{6.104}$$

It is interesting that the variation of $A$ plays no role in the above manipulations, so that the derivation appears to go through if the amplitude $A(\boldsymbol{x}, t)$ is in fact a constant and the wave is a single plane wave. This seems hard to reconcile with our previous discussion, in which we noted that the group velocity was the velocity of a wave *packet* involving a superposition of plane waves. Indeed, the derivative of the frequency with respect to wavenumber means little if there is only one wavenumber. In fact there is nothing wrong with the above derivation if $A$ is a constant and only a single plane wave is present. The resolution of the paradox arises by noting that a plane wave fills all of space and time; in this case there is no convergence of the energy flux and the energy propagation equation is trivially true.

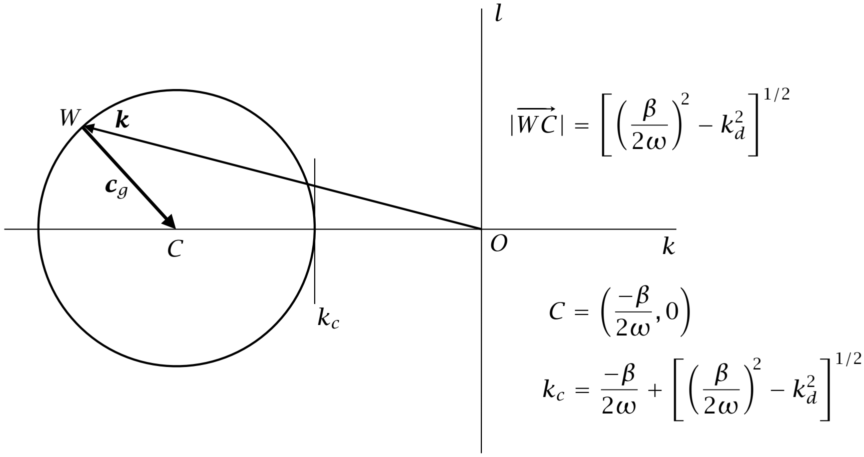


**Fig. 6.8** The energy propagation diagram for Rossby waves. The wavevectors of a given frequency all lie in a circle of radius $[(\beta/2\omega)^2 - k_d^2]^{1/2}$, centered at the point $C$. The closest distance of the circle to the origin is $k_c$, and if the deformation radius is infinite $k_c$ the circle touches the origin. For a given wavenumber $\boldsymbol{k}$, the group velocity is along the line directed from $W$ to $C$.

### 6.6.1 ✦ Rossby wave reflection

We now consider how Rossby waves might be reflected from a solid boundary. The topic has an obvious oceanographic relevance, for the reflection of Rossby waves turns out to be one way of interpreting why intense oceanic boundary currents form on the western sides of ocean basins, not the east. There is also an atmospheric relevance, for meridionally propagating Rossby waves may effectively be reflected as they approach a 'turning latitude' where the meridional wavenumber goes to zero, as considered in chapter 16. As a preliminary, let us give a useful graphic interpretation of Rossby wave propagation.[8]

*The energy propagation diagram*

The dispersion relation for Rossby waves, $\omega = -\beta k/(k^2 + l^2 + k_d^2)$, may be rewritten as

$$\left(k + \beta/2\omega\right)^2 + l^2 = \left(\beta/2\omega\right)^2 - k_d^2. \tag{6.105}$$

This equation is the parametric representation of a circle, meaning that the wavevector $(k, l)$ must lie on a circle centered at the point $(-\beta/2\omega, 0)$ and with radius $[(\beta/2\omega)^2 - k_d^2]^{1/2}$, as illustrated in Fig. 6.8. If $k_d$ is zero the circle touches the origin, and if it is nonzero the distance of the closest point to the circle, $k_c$ say, is given by $k_c = -\beta/2\omega + [(\beta/2\omega)^2 - k_d^2]^{1/2}$. For low frequencies, specifically if $\omega \ll \beta/2k$, then $k_c \approx -\omega k_d^2/\beta$. The radius of the circle is a positive real number only when $\omega < \beta/2k_d$. This is the maximum frequency possible, and it occurs when $l = 0$ and $k = k_d$ and when $c_g^x = c_g^y = 0$.

The group velocity, and hence the energy flux, can be visualized graphically from Fig. 6.8. By direct manipulation of the expressions for group velocity and frequency [equations (RW.3) and (RW.4)] we find that

$$c_g^x = \frac{-2\omega}{K^2 + k_d^2}\left(k + \frac{\beta}{2\omega}\right), \tag{6.106a}$$

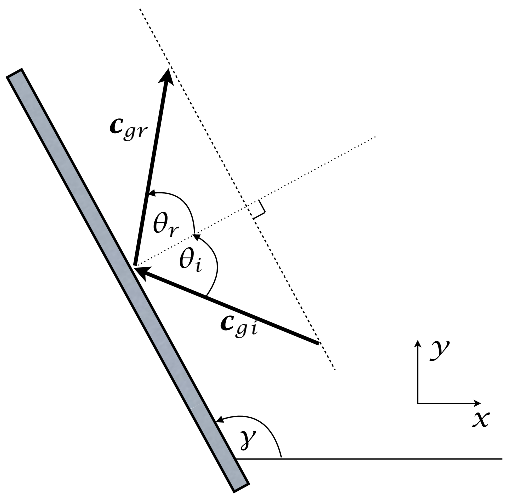


**Fig. 6.9** The reflection of a Rossby wave at a western wall, in physical space. A Rossby wave with a westward group velocity impinges at an angle $\theta_i$ to a wall, inducing a reflected wave moving eastward at an angle $\theta_r$. The reflection is specular, with $\theta_r = \theta_i$, and energy conserving, with $|\boldsymbol{c}_{gr}| = |\boldsymbol{c}_{gi}|$ — see text and Fig. 6.10.

$$c_g^y = \frac{-2\omega}{K^2 + k_d^2}\, l. \tag{6.106b}$$

(To check this, it is easiest to begin with the right-hand sides and use the dispersion relation for $\omega$.) Now, since the center of the circle of wavevectors is at the position $(-\beta/2\omega, 0)$, and referring to Fig. 6.8, we have that

$$\boldsymbol{c}_g = \frac{2\omega}{K^2 + k_d^2}\boldsymbol{R} \tag{6.107}$$

where $\boldsymbol{R} = \overrightarrow{WC}$ is the vector directed from $W$ to $C$, that is from the end of the wavevector itself to the center of the circle around which all the wavevectors lie.

Eq. (6.107) and Fig. 6.8 allow for a useful visualization of the energy and phase. The phase propagates in the direction of the wave vector, and for Rossby waves this is always westward. The group velocity is in the direction of the wave vector to the center of the circle, and this can be either eastward (if $k^2 > l^2 + k_d^2$) or westward ($k^2 < l^2 + k_d^2$). Interestingly, the velocity vector is normal to the wave vector. To see this, consider a purely westward propagating wave for which $l = 0$. Then $v = \partial\psi/\partial x = \mathrm{i}k\widetilde{\psi}$ and $u = -\partial\psi/\partial y = -\mathrm{i}l\widetilde{\psi} = 0$. We now see how some of these properties can help us understand the reflection of Rossby waves.

[Do we need a gray box summarizing some of the properties of reflection? xxx]

### *Reflection at a wall*

Consider Rossby waves incident on wall making an angle $\gamma$ with the $x$-axis, and suppose that somehow these waves are reflected back into the fluid interior. This is a reasonable expectation, for the wall cannot normally simply absorb all the wave energy. We first note a couple of general properties about reflection, namely that the incident and reflected wave will have the same wavenumber component along the wall and their frequencies must be the same To see these properties, consider the case in which the wall is oriented meridionally, along the $y$-axis with $\gamma = 90°$. There is no loss of generality in this choice, because we may simply choose coordinates so that $y$ is parallel to the wall and the $\beta$-effect, which differentiates $x$ from $y$, does not enter the argument. The incident and reflected waves are

$$\psi_i(x, y, t) = A_i e^{i(k_i x + l_i y - \omega_i t)}, \qquad \psi_r(x, y, t) = A_r e^{i(k_r x + l_r y - \omega_r t)}, \tag{6.108}$$

with subscripts $i$ and $r$ denoting incident and reflected. At the wall, which we take to be at $x = 0$, the normal velocity $u = -\partial\psi/\partial y$ must be zero so that

$$A_i l_i e^{i(l_i y - \omega_i t)} + A_r l_r e^{i(l_r y - \omega_r t)} = 0. \tag{6.109}$$

For this equation to hold for all $y$ and all time then we must have

$$l_r = l_i, \qquad \omega_r = \omega_i. \tag{6.110}$$

This result is independent of the detailed dynamics of the waves, requiring only that the velocity is determined from a streamfunction. When we consider Rossby-wave dynamics specifically, the $x$- and $y$-coordinates are not arbitrary and so the wall cannot be taken to be aligned with the $y$-axis; rather, the result means that the *projection* of the incident wavevector, $\boldsymbol{k}_i$ on the wall must equal the *projection* of the reflected wavevector, $\boldsymbol{k}_r$. The magnitude of the wavevector (the wavenumber) is not in general conserved by reflection. Finally, given these results and using (6.109) we see that the incident and reflected amplitudes are related by

$$A_r = -A_i. \tag{6.111}$$

Now let's delve a little deeper into the wave-reflection properties.

Generally, when we consider a wave to be incident on a wall, we are supposing that the *group velocity* is directed toward the wall. Suppose that a wave of given frequency, $\omega$, and wavevector, $\boldsymbol{k}_i$, and with westward group velocity is incident on a predominantly western wall, as in Fig. 6.9. (Similar reasoning, *mutatis mutandis,* can be applied to a wave incident on an eastern wall.) Let us suppose that incident wave, $\boldsymbol{k}_i$ lies at the point $I$ on the wavenumber circle, and the group velocity is found by drawing a line from $I$ to the center of the circle, $C$ (so $\boldsymbol{c}_{gi} \propto \overrightarrow{IC}$), and in this case the vector is directed westward.

The projection of the $\boldsymbol{k}_i$ must be equal to the projection of the reflected wave vector, $\boldsymbol{k}_r$, and both wavevectors must lie in the same wavenumber circle, centered at $-\beta/2\omega$, because the frequencies of the two waves are the same. We may then graphically determines the wavevector of the reflected wave using the construction of Fig. 6.10. Given the wavevector, the group velocity of the reflected wave follows by drawing a line from the wavevector to the center of the circle (the line $\overrightarrow{RC}$). We see from the figure that the reflected group velocity is directed eastward and that it forms the same angle to the wall as does the incident wave; that is, the reflection is *specular.* Since the amplitude of the incoming and reflected wave are the same, the components of the energy flux perpendicular to the wall are equal and opposite. Furthermore, we can see from the figure that the wavenumber of the reflected wave has a larger magnitude than that of the incident wave. For waves reflecting of an easter boundary, the reverse is true. Put simply, at a western boundary incident long waves are reflected as short waves, whereas at an eastern boundary incident short waves are reflected as long waves.

Quantitatively solving for the wavenumbers of the reflected wave is a little tedious in the case when the wall is at angle, but easy enough if the wall is a meridional, along the $y$-axis. We know the frequency, $\omega$, and the $y$-wavenumber, $l$, so that the $x$-wavenumber is may be deduced from the dispersion relation

$$\omega = \frac{-\beta k_i}{k_i^2 + l^2 + k_d^2} = \frac{-\beta k_r}{k_r^2 + l^2 + k_d^2}. \tag{6.112}$$

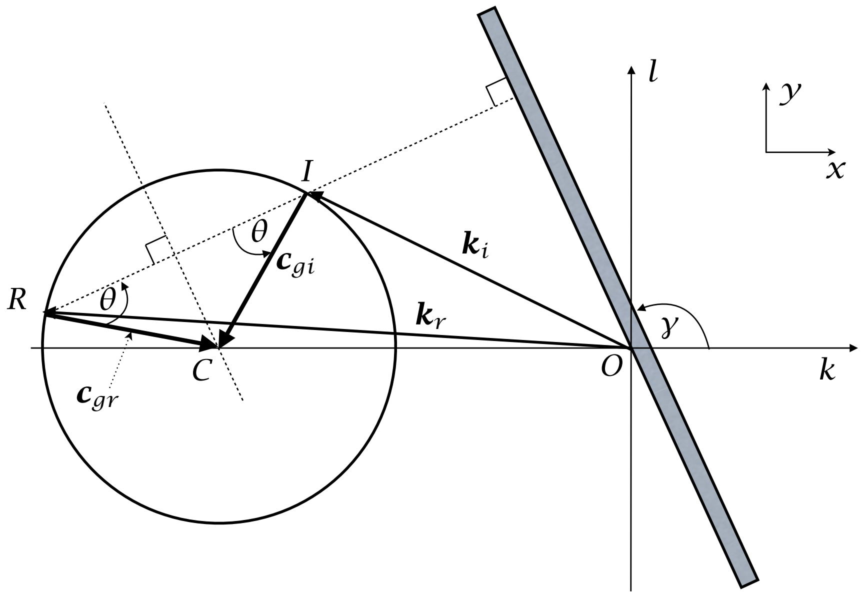


**Fig. 6.10** Graphical representation of the reflection of a Rossby wave at a western wall, in spectral space. The incident wave has wavevector $\boldsymbol{k}_i$, ending at point $I$. Construct the wavevector circle through point $I$ with radius $\sqrt{(\beta/2\omega)^2 - k_d^2}$ and center $C = (-\beta/2\omega, 0)$; the group velocity vector then lies along $\overrightarrow{IC}$ and is directed westward. The reflected wave has a wavevector $\boldsymbol{k}_r$ such that its projection on the wall is equal to that of $\boldsymbol{k}_i$, and this fixes the point $R$. The group velocity of the reflected wave then lies along $\overrightarrow{RC}$, and it can be seen that $\boldsymbol{c}_{gr}$ makes the same angle to the wall as does $\boldsymbol{c}_{gi}$, except that it is directed eastward. The reflection is therefore both specular and is such that the energy flux directed away from the wall is equal to the energy flux directed toward the wall.

We obtain

$$k_i = \frac{-\beta}{2\omega} + \sqrt{\left(\frac{\beta}{2\omega}\right)^2 - (l^2 + k_d^2)}, \qquad k_r = \frac{-\beta}{2\omega} - \sqrt{\left(\frac{\beta}{2\omega}\right)^2 - (l^2 + k_d^2)}. \tag{6.113a,b}$$

The signs of the square-root terms are chosen for reflection at a western boundary, for which, as we noted, the reflected wave has a larger (absolute) wavenumber than the incident wave. For reflection at an eastern boundary, we simply reverse the signs.

### *Oceanographic relevance*

The behaviour of Rossby waves at lateral boundaries is not surprisingly of some oceanographic importance, there being two particularly important examples. One of them concerns the equatorial ocean, and the other the formation of western boundary currents, common in midlatitudes. We only touch on these topics here, deferring a more extensive treatments to later chapters.

Suppose that Rossby waves are generated in the middle of the ocean, for example by the wind or possibly by some fluid dynamical instability in the ocean. Shorter waves will tend to propagate eastward, and be reflected back at the eastern boundary as long waves, and long waves will tend to propagate westward, being reflected back as short waves. The reflection at the western boundary is believed to be particularly important in the dynamics of El Niño although the situation is further complicated because the reflection may also generate eastward moving equatorial Kelvin waves, which we discuss more in the next chapter.

In mid-latitudes the reflection at a western boundary generates Rossby waves that have a short *zonal* length scale (the meridional scale is the same as the incident wave if the wall is meridional), which means that their *meridional* velocity is large. Now, if the zonal wavenumber is much larger than both the meridional wavenumber $l$ and the inverse deformation radius $k_d$ then, using either (6.61) or (6.65) the group velocity in the $x$-direction is given by $c_g^x = \overline{u} + \beta/k^2$, where $\overline{u}$ is the zonal mean flow. If the mean flow is westward, so that $U$ is negative, then very short waves will be unable to escape from the boundary; specifically, if $k > \sqrt{-\beta/U}$ then the waves will be trapped in a western boundary layer. [More here? A schematic figure? xxx]

## 6.7 ROSSBY-GRAVITY WAVES: AN INTRODUCTION

We now consider Rossby waves and shallow water gravity waves together. To keep the treatment tractable we will consider the simplest possible case, namely a single layer of shallow water on the beta plane in which the Coriolis parameter, $f$, is held constant except where it is differentiated, an approximation similar to that made when deriving the quasi-geostrophic equations.[9] A (perforce more complex) treatment of the analogous problem on the equatorial beta-plane, in which we allow both $f$ and $\beta$ to vary fully with latitude, in is given in chapter 8.

Our equations of motion are the shallow water equations in Cartesian coordinates in a rotating frame of reference, namely

$$\frac{\partial u}{\partial t} - fv = -\frac{\partial \phi}{\partial x}, \qquad \frac{\partial v}{\partial t} + fu = -\frac{\partial \phi}{\partial y}, \tag{6.114a,b}$$

$$\frac{\partial \phi}{\partial t} + c^2\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0 \tag{6.114c}$$

where, in terms of possibly more familiar shallow water variables, $\phi = g'\eta$ and $c^2 = g'H$, where $\phi$ is the kinematic pressure, $\eta$ is the free surface height, $H$ is the reference depth of the fluid and $g'$ is the reduced gravity.

After some manipulation (described more fully in section 8.2) we obtain, without additional approximation, a single equation for $v$:

$$\frac{1}{c^2}\frac{\partial^3 v}{\partial t^3} + \frac{f^2}{c^2}\frac{\partial v}{\partial t} - \frac{\partial}{\partial t}\nabla^2 v - \beta\frac{\partial v}{\partial x} = 0. \tag{6.115}$$

In this equation the Coriolis parameter is given by the $\beta$-plane expression $f = f_0 + \beta y$; thus, the equation has a non-constant coefficient, entailing considerable algebraic difficulties. We will address some of these difficulties in chapter 8, but for now we take a simpler approach: we assume that $f$ is constant except where differentiated, an approximation that is reasonable in mid-latitudes provided we are concerned with sufficiently small variations in latitude. Equation (6.115) then has constant coefficients and we may look for plane wave solutions of the form $v = \widetilde{v}\exp\left[\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)\right]$, whence

$$\frac{\omega^2 - f_0^2}{c^2} - (k^2 + l^2) - \frac{\beta k}{\omega} = 0. \tag{6.116a}$$

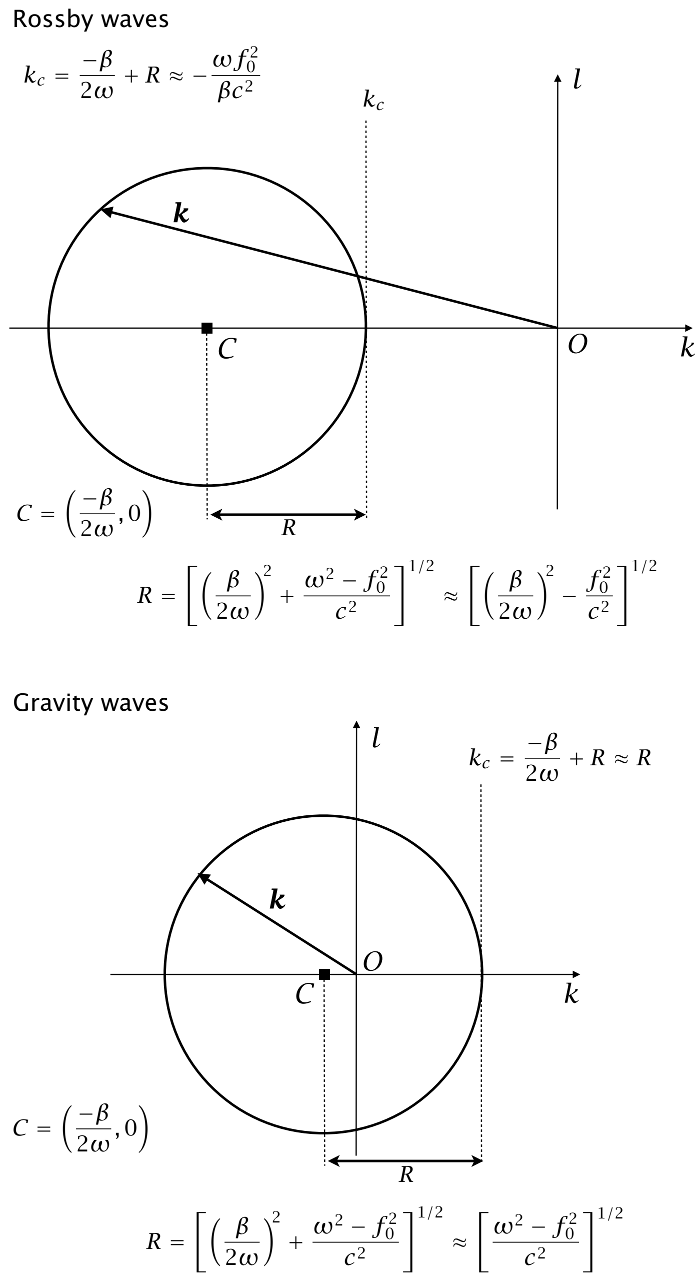


**Fig. 6.11** Wave propagation diagrams for Rossby-gravity waves, obtained using (6.116). The top figure shows the diagram in the low frequency, Rossby wave limit, and the bottom figure shows the high frequency, gravity wave limit. In each case the the locus of wavenumbers for a given frequency is a circle centered at $C = (-\beta/2\omega, 0)$ with a radius $R$ given by (6.117), but the approximate expressions differ significantly at high and low frequency.

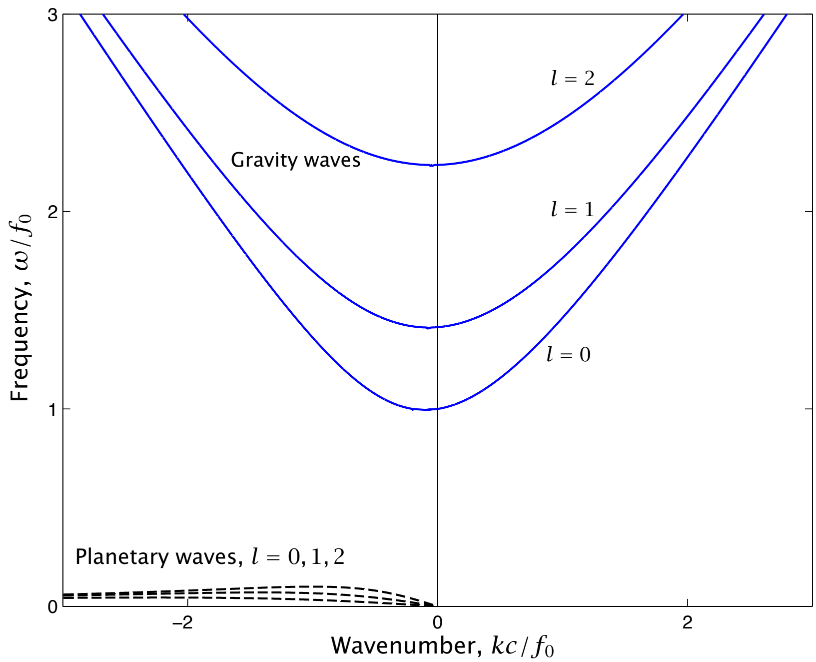


**Fig. 6.12** Dispersion relation for Rossby-gravity waves, obtained from (6.121) with $\widehat{\beta} = 0.2$ for three values of $l$. There a frequency gap between the Rossby or planetary waves and the gravity waves. For the stratified mid-latitude atmosphere or ocean the frequency gap is in reality even larger.

or, written differently,

$$\left(k + \frac{\beta}{2\omega}\right)^2 + l^2 = \left(\frac{\beta}{2\omega}\right)^2 + \frac{\omega^2 - f_0^2}{c^2}. \tag{6.116b}$$

This equation may be compared to (6.105): noting that $k_d^2 = f_0^2/g'H = f_0^2/c^2$, the two equations are identical except for the appearance of a term involving frequency on last term on the right-hand side of (6.116b). The wave propagation diagram is illustrated in Fig. 6.11. The wave vectors a a given frequency all lie on a circle centered at $(-\beta/2\omega, 0)$ and with radius $R$ given by

$$R = \left[\left(\frac{\beta}{2\omega}\right)^2 + \frac{\omega^2 - f_0^2}{c^2}\right]^{1/2}, \tag{6.117}$$

and the radius must be positive in order for the waves to exist. In the low frequency case the diagram is essentially the same as that shown in Fig. 6.8, but is quantitatively significantly different in the high frequency case. These limiting cases are discussed further in section 6.7.1 below.

To plot the full dispersion relation it is useful to nondimensionalize using the following scales for time ($T$), distance ($L$) and velocity ($U$)

$$T = f_0^{-1}, \qquad L = L_d = k_d^{-1} = c/f_0, \qquad U = L/T = c, \tag{6.118a,b}$$

so that, denoting nondimensional quantities with a hat,

$$\omega = \widehat{\omega} f_0, \qquad (k, l) = (\widehat{k}, \widehat{l})k_d, \qquad \beta = \widehat{\beta}\frac{f_0^2}{c} = \widehat{\beta}\frac{f_0}{L_d} = \widehat{\beta} f_0 k_d. \tag{6.119}$$

The dispersion relation (6.116) may then be written as

$$\widehat{\omega}^2 - 1 - (\widehat{k}^2 + \widehat{l}^2) - \widehat{\beta}\frac{\widehat{k}}{\widehat{\omega}} = 0 \tag{6.120}$$

This is a cubic equation in $\omega$, as might be expected given the governing equations (6.114). We may expect that two of the roots correspond to gravity waves and the third to Rossby waves. The only parameter in the dispersion relation is $\widehat{\beta} = \beta c/f_0^2 = \beta L_d/f_0$. In the atmosphere a representative value for $L_d$ is 1000 km, whence $\widehat{\beta} = 0.1$. In the ocean $L_d \sim 100$ km, whence $\widehat{\beta} = 0.01$. If we allow ourselves to consider 'external' Rossby waves (which are of some oceanographic relevance) then $c = \sqrt{gH} = 200\,\mathrm{m\,s^{-1}}$ and $L_d = 2000$ km, whence $\widehat{\beta} = 0.2$.

To actually obtain a solution we regard the equation as a quadratic in $k$ and solve in terms of the frequency, giving

$$\widehat{k} = -\frac{\widehat{\beta}}{2\widehat{\omega}} \pm \frac{1}{2}\left[\frac{\widehat{\beta}^2}{\widehat{\omega}^2} + 4(\widehat{\omega}^2 - \widehat{l}^2 - 1)\right]^{1/2}. \tag{6.121}$$

The solutions are plotted in Fig. 6.12, with $\widehat{\beta} = 0.2$, and we see that the waves fall into two groups, labelled gravity waves and planetary waves in the figure. The gap between the two groups of waves is in fact still larger if a smaller (and generally more relevant) value of $\widehat{\beta}$ is used. To interpret all this let us consider some limiting cases.

### 6.7.1 Special cases and properties of the waves

We now consider a few special cases of the dispersion relation.

*(i) Constant Coriolis parameter*

If $\beta = 0$ then the dispersion relation becomes

$$\omega\left[\omega^2 - f_0^2 - (k^2 + l^2)c^2\right] = 0, \tag{6.122}$$

with the roots

$$\omega = 0, \qquad \omega^2 = f_0^2 + c^2(k^2 + l^2). \tag{6.123a,b}$$

The root $\omega = 0$ corresponds to geostrophic motion (and, since $\beta = 0$, Rossby waves are absent), with the other root corresponding to Poincaré waves, considered in chapter 3. Note that $\omega^2 > f_0^2$.

*(ii) High frequency waves*

If we take the limit of $\omega \gg f_0$ then (6.116a) gives

$$\frac{\omega^2}{c^2} - (k^2 + l^2) - \frac{\beta k}{\omega} = 0. \tag{6.124}$$

## Rossby and Gravity Waves

- Generically speaking, Rossby-gravity waves are waves that arise under the combined effects of a potential vorticity gradient and stratification. Sometimes the definition is restricted to a wave on a single branch of the dispersion curve connecting Rossby and Gravity waves. In mid-latitudes Rossby waves and gravity waves are well separated with distinct physical mechanisms

- The simplest setting in which such waves occur is in the linearized shallow water equations which may be written as a single equation for $v$, namely

$$\frac{1}{c^2}\frac{\partial^3 v}{\partial t^3} + \frac{f^2}{c^2}\frac{\partial v}{\partial t} - \frac{\partial}{\partial t}\nabla^2 v - \beta\frac{\partial v}{\partial x} = 0. \tag{RG.1}$$

- If we take both $f$ and $\beta$ to be constants then the equation above admits of plane-wave solutions with dispersion relation

$$\omega^2 - \frac{\beta k c^2}{\omega} = f_0^2 + c^2(k^2 + l^2). \tag{RG.2}$$

- In Earth's atmosphere and ocean it is common, especially in mid-latitudes, for there to be a frequency separation between two classes of solution. To a good approximation, high frequency waves satisfy

$$\omega^2 = f_0^2 + c^2(k^2 + l^2). \tag{RG.3}$$

  These are gravity waves and which in this context, because of the presence of rotation, are known as Poincaré waves. The low frequency waves satisfy

$$\omega = \frac{-\beta k c^2}{f_0^2 + c^2(k^2 + l^2)} = \frac{-\beta k}{k_d^2 + k^2 + l^2}, \tag{RG.4}$$

  where $k_d^2 = f_0^2/c^2$, and these are called Rossby waves or planetary waves.

- Rossby-gravity waves also exist in the stratified equations. Solutions may be found be decomposing the vertical structure into a series of orthogonal modes, and a sequence of shallow water equations for each mode results, with a different $c$ for each mode. Solutions may also be found if $f$ is allowed to vary in (RG.1), at the price of some algebraic complexity, as discussed in chapter 8.

To be physically realistic we should also now eliminate the $\beta$ term, because if $\omega \gg f_0$ then, from geometric considerations on a sphere, $k^2 \gg \beta k/\omega$. Thus, the dispersion relation is simply $\omega^2 = c^2(k^2 + l^2)$. These waves are just gravity waves uninfluenced by rotation, and are a special case of Poincaré waves.

*(iii) Low frequency waves*

Consider the limit of $\omega \ll f_0$. The dispersion relation reduces to

$$\omega = \frac{-\beta k}{k^2 + l^2 + k_d^2}. \tag{6.125}$$

This is just the dispersion relation for quasi-geostrophic Rossby waves as previously obtained — see (6.63) or (6.90). In this limit, the requirement that the radius of the circle be positive becomes

$$\omega^2 < \frac{\beta^2}{4k_d^2}. \tag{6.126}$$

That is to say, the Rossby waves have a maximum frequency, and directly from (6.125) this occurs when $k = k_d$ and $l = 0$.

*The frequency gap*

The maximum frequency of Rossby waves is usually much less than the frequency of the Poincaré waves: the lowest frequency of the Poincare waves is $f_0$ and the highest frequency of the Rossby waves is $\beta/2k_d$. Thus,

$$\frac{\text{Low gravity wave frequency}}{\text{High Rossby wave frequency}} = \frac{f_0}{\beta/2k_d} = \frac{f_0^2}{2\beta c}. \tag{6.127}$$

If $f_0 = 10^{-4}\,\mathrm{s}^{-1}$, $\beta = 10^{-11}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$ and $k_d = 1/100\,\mathrm{km}^{-1}$ (a representative oceanic baroclinic deformation radius) then $f_0/(\beta/2k_d) = 200$. If $L_d = 1000\,\mathrm{km}$ (an atmospheric baroclinic radius) then the ratio is 20. If we use a barotropic deformation radius of $L_d = 2000\,\mathrm{km}$ then the ratio is 10. Evidently, for most midlatitude applications there is a large gap between the Rossby wave frequency and the gravity wave frequency. Because of this frequency gap, to a good approximation Fig. 6.12 may be obtained by separately plotting (6.123b) for the gravity waves, and (6.125) for the Rossby or planetary waves. The differences between these and the exact results become smaller as $\widehat{\beta}$ gets smaller, virtually indistinguishable in the plots shown.

Finally, we remark that a 'Rossby-gravity wave' is sometimes defined to be the wave on a single branch of the dispersion curve that connects Rossby waves and gravity waves, depending on the value of the wavenumber. The equatorial beta plane does support such a wave — the 'Yanai wave' derived in chapter 8 and shown in Fig. 8.2. However, in the mid-latitude system above there is no such wave; rather, there are Rossby waves and gravity waves, separated by a frequency gap.

### 6.7.2 Planetary geostrophic Rossby waves

A good approximation for the large-scale ocean circulation involves ignoring the time-derivatives and nonlinear terms in the momentum equation, allowing evolution only to occur in the thermodynamic equation. This is the planetary-geostrophic approximation, introduced in section

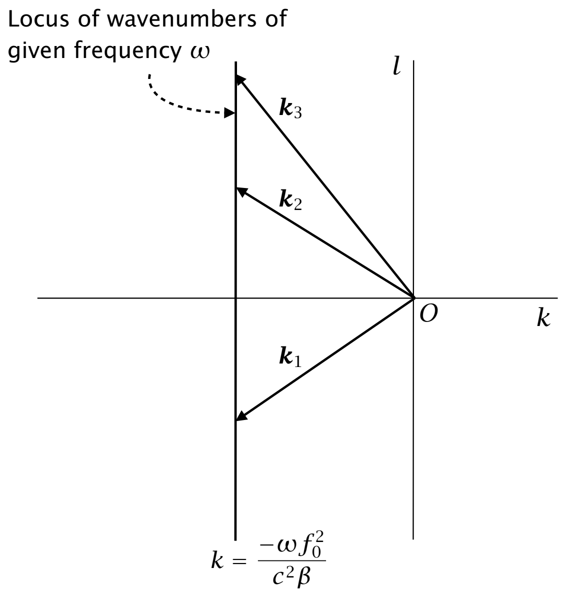


**Fig. 6.13** The locus of points on planetary-geostrophic Rossby waves. Waves of a given frequency all have the same $x$-wavenumber, given by xxx

5.2 204, and it is interesting to see to what extent that system supports Rossby waves.[10] It is easiest just to begin with the linear shallow water equations themselves, and omitting time derivatives in the momentum equation gives

$$-fv = -\frac{\partial \phi}{\partial x}, \qquad fu = -\frac{\partial \phi}{\partial y}, \tag{6.128a,b}$$

$$\frac{\partial \phi}{\partial t} + c^2\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0. \tag{6.128c}$$

From these equations we straightforwardly obtain

$$\frac{\partial \phi}{\partial t} - \frac{c^2\beta}{f^2}\frac{\partial \phi}{\partial x} = 0. \tag{6.129}$$

Again we will treat both $f$ and $\beta$ as constants so that we may look for solutions in the form $\phi = \widetilde{\phi}\exp[\mathrm{i}(\boldsymbol{k}\cdot x - \omega t)]$. The ensuing dispersion relation is

$$\omega = -\frac{c^2\beta}{f_0^2}k = -\frac{\beta k}{k_d^2} \tag{6.130}$$

which is a limiting case of (6.125) with $k^2, l^2 \ll k_d^2$. The waves are a form of Rossby waves with phase and group speeds given by

$$c_p = -\frac{c^2\beta}{f_0^2}, \qquad c_g^x = -\frac{c^2\beta}{f_0^2}. \tag{6.131}$$

That is, the waves are non-dispersive and propagate westward. Eq. (6.120) has the general solution $\phi = G(x + \beta c^2/f^2 t)$, where $G$ is any function, so an initial disturbance will just propagate westward at a speed given by (**??**), without any change in form.

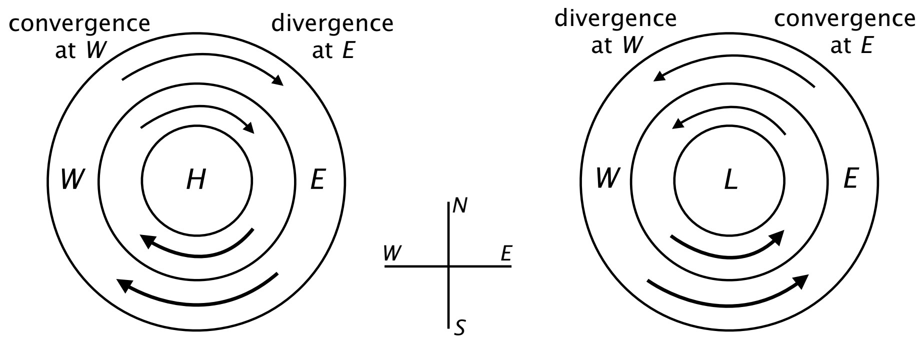


**Fig. 6.14** The westward propagation of planetary-geostrophic Rossby waves. The circular lines are isobars centered around high and low pressure centres. Because of the variation of the Coriolis force, the mass flux between two isobars is greater to the south of a pressure center than it is to the north. Hence, in the left-hand sketch there is convergence to the west of the high pressure and the pattern propagates westward. Similarly, if the pressure centre is a low, as in the right-hand sketch, there is divergence to the west of the pressure centre and the pattern still propagates westward.

Note finally that the locus of wavenumbers in $k$–$l$ space is no longer a circle, as it is for the usual Rossby waves. Rather, since the frequency does not depend on the $y$-wavenumber, the locus is a straight line, parallel to the $y$-axis, as in Fig. 6.13. Waves of a given frequency all have the same $x$-wavenumber, given by $k = -\omega f_0^2/(c^2\beta) = -\omega k_d^2/\beta$, as shown in Fig. 6.13.

*Physical mechanism*

Because the waves *are* a form of Rossby wave their physical mechanism is related to that discussed in section 6.4.3, but with an important difference: relative vorticity is no longer important, but the flow divergence is. Thus, consider flow round a region of high pressure, as illustrated in Fig. 6.14. If the pressure is circularly symmetric as shown, the flow to the south of $H$ in the left-hand sketch, and to the south of $L$ in the right-hand sketch, is larger than that to the north. Hence, in the left sketch the flow converges at $W$ and diverges at $E$, and the flow pattern moves westward. In the flow depicted in the right sketch the low pressure propagates westward in a similar fashion.

## 6.8 ♦ THE GROUP VELOCITY PROPERTY

We now return to a more general discussion of group velocity. Our goal is to show that the group velocity arises in fairly general ways, not just from methods stemming from Fourier analysis or from ray theory. In a purely logical sense this discussion follows most naturally from the end of the section on ray theory (section 6.3), but for most humans it is helpful to have had a concrete introduction to at least one nontrivial form of waves before considering more abstract material. We first give a simple and direct derivation of group velocity that valid in the simple but important special case of a homogeneous medium.[11] Then, in section 6.8.2, we give a rather general derivation of the *group velocity property,* namely that conserved quantities that are quadratic in the wave amplitude — that is,s *wave activities* — are transported at the group velocity.

### 6.8.1 Group velocity in homogeneous media

Consider waves propagating in a homogeneous medium in which the wave equation is a polynomial of the general form

$$L(\psi) = \Lambda\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial x}\right)\psi(x,t) = 0. \tag{6.132}$$

where $\Lambda$ is a polynomial operator in the space and time derivatives. For algebraic simplicity we restrict attention to waves in one dimension, and a simple example is $\Lambda = \partial(\partial_{xx})/\partial t + \beta\partial/\partial x$ so that $L(\psi) = \partial(\partial_{xx}\psi)/\partial t + \beta\partial\psi/\partial x$. We will seek a solution of the form [c.f., (6.4)]

$$\psi(x,t) = A(x,t)\mathrm{e}^{i\theta(x,t)}, \tag{6.133}$$

where $\theta$ is the phase of the disturbance and $A(x,t)$ is the slowly varying amplitude, so that the solution has the form of a wave packet. The phase is such that $k = \partial\theta/\partial x$ and $\omega = -\partial\theta/\partial t$, and the slowly varying nature of the envelope $A(x,t)$ is formalized by demanding that

$$\frac{1}{A}\frac{\partial A}{\partial x} \ll k, \qquad \frac{1}{A}\frac{\partial A}{\partial t} \ll \omega, \tag{6.134}$$

The space and time derivatives of $\psi$ are then given by

$$\frac{\partial\psi}{\partial x} = \left(\frac{\partial A}{\partial x} + \mathrm{i}A\frac{\partial\theta}{\partial x}\right)\mathrm{e}^{i\theta} = \left(\frac{\partial A}{\partial x} + \mathrm{i}Ak\right)\mathrm{e}^{i\theta}, \tag{6.135a}$$

$$\frac{\partial\psi}{\partial t} = \left(\frac{\partial A}{\partial t} + \mathrm{i}A\frac{\partial\theta}{\partial t}\right)\mathrm{e}^{i\theta} = \left(\frac{\partial A}{\partial t} - \mathrm{i}A\omega\right)\mathrm{e}^{i\theta}, \tag{6.135b}$$

so that the wave equation becomes

$$\Lambda\psi = \Lambda\left(\frac{\partial}{\partial t} - \mathrm{i}\omega, \frac{\partial}{\partial x} + \mathrm{i}k\right)A = 0. \tag{6.136}$$

Noting that the space and time derivative of $A$ are small compared to $k$ and $\omega$ we expand the polynomial in a Taylor series about $(\omega, k)$ to obtain

$$\Lambda(-\mathrm{i}\omega, \mathrm{i}k)A + \frac{\partial\Lambda}{\partial(-\mathrm{i}\omega)}\frac{\partial A}{\partial t} + \frac{\partial\Lambda}{\partial(\mathrm{i}k)}\frac{\partial A}{\partial x} = 0. \tag{6.137}$$

The first term is nothing but the linear dispersion relation; that is $\Lambda(-\mathrm{i}\omega, \mathrm{i}k)A = 0$ is the dispersion relation for plane waves. Taking this to be satisfied, (6.137) gives

$$\frac{\partial A}{\partial t} + \frac{\partial\Lambda/\partial k}{\partial\Lambda/\partial\omega}\frac{\partial A}{\partial x} = \frac{\partial A}{\partial t} + \frac{\partial\omega}{\partial k}\frac{\partial A}{\partial x} = 0. \tag{6.138}$$

That is, the envelope moves at the group velocity $\partial\omega/\partial k$.

### 6.8.2 ✦ Group velocity property: a general derivation

In our discussion of Rossby waves in section 6.6 in (6.104) we showed that the energy of the waves is conserved in the sense that

$$\frac{\partial E}{\partial t} + \nabla\cdot\boldsymbol{F} = 0, \tag{6.139}$$

where $E$ is the energy density of the waves and $\boldsymbol{F}$ is its flux. In (6.104) we further showed that, when averaged over a wavelength and a period, the average flux, $\overline{\boldsymbol{F}}$, was related to the average energy, $\overline{E}$, by $\overline{\boldsymbol{F}} = \boldsymbol{c}_g \overline{E}$. This property is called the group velocity property and it is a very general property, not restricted to Rossby waves or even to energy. In the previous section we gave a more general derivation valid in homogeneous media. In fact, the property is still more general and it holds for almost any conserved quantity that is quadratic in the wave amplitude, and we now demonstrate this in a rather general way.[12] A quantity that is quadratic and conserved is known as a *wave activity*. (The corresponding local quantity, such as the wave activity per unit volume, might strictly we called the *wave activity density*.) The group velocity property is useful because if we can determine $\boldsymbol{c}_g$ then we know straightaway how wave activities propagate. Energy itself can be a wave activity but is not always. In a growing baroclinic wave energy is drawn from the background state; however, we will see in chapter 10 that even in a growing baroclinic disturbance it is possible to define a conserved wave activity.

### ✦ *The formal procedure*

The derivation, which is rather formal, will hold for waves and wave activities that satisfy the following three assumptions.

*(i)* The wave activity, $A$, and flux, $\boldsymbol{F}$, obey the general conservation relation

$$\frac{\partial A}{\partial t} + \nabla \cdot \boldsymbol{F} = 0. \tag{6.140}$$

*(ii)* Both the wave activity and the flux are quadratic functions of the wave amplitude.

*(iii)* The waves themselves are of the general form

$$\psi = \widetilde{\psi} \mathrm{e}^{\mathrm{i}\theta(\boldsymbol{x},t)} + \text{c.c.}, \qquad \theta = \boldsymbol{k} \cdot \boldsymbol{x} - \omega t, \qquad \omega = \omega(\boldsymbol{k}), \tag{6.141a,b,c}$$

where (6.141c) is the dispersion relation, and $\psi$ is any wave field. We will carry out the derivation in case in which $\widetilde{\psi}$ is a constant, but the derivation may be extended to the case in which it varies slowly over a wavelength.

Given assumption *(ii)*, the wave activity must have the general form

$$A = b + a\mathrm{e}^{2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} + a^* \mathrm{e}^{-2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}, \tag{6.142a}$$

where the asterisk, $^*$, denotes complex conjugacy, and $b$ is a real constant and $a$ is a complex constant. For example, suppose that $A = \psi^2$ and $\psi = c\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} + c^* \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}$, then we find that (6.142a) is satisfied with $a = c^2$ and $b = 2cc^*$. Similarly, the flux has the general form

$$\boldsymbol{F} = \boldsymbol{g} + \boldsymbol{f}\mathrm{e}^{2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} + \boldsymbol{f}^* \mathrm{e}^{-2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}. \tag{6.142b}$$

where $\boldsymbol{g}$ is a real constant vector and $\boldsymbol{f}$ is a complex constant vector. The mean activity and mean flux are obtained by averaging over a cycle; the oscillating terms vanish on integration and therefore the wave activity and flux are given by

$$\overline{A} = b, \qquad \overline{\boldsymbol{F}} = \boldsymbol{g}, \tag{6.143}$$

where the overbar denotes the mean.

Now formally consider a wave with a slightly different phase, $\theta + \mathrm{i}\,\delta\theta$, where $\delta\theta$ is small compared with $\theta$. Thus, we formally replace $\boldsymbol{k}$ by $\boldsymbol{k} + \mathrm{i}\,\boldsymbol{\delta k}$ and $\omega$ by $\omega + \mathrm{i}\,\delta\omega$ where, to satisfy the dispersion relation, we have

$$\omega + \mathrm{i}\,\delta\omega = \omega(\boldsymbol{k} + \mathrm{i}\,\boldsymbol{\delta k}) \approx \omega(\boldsymbol{k}) + \mathrm{i}\,\boldsymbol{\delta k} \cdot \frac{\partial \omega}{\partial \boldsymbol{k}}, \tag{6.144}$$

and therefore

$$\delta\omega = \boldsymbol{\delta k} \cdot \frac{\partial \omega}{\partial \boldsymbol{k}} = \boldsymbol{\delta k} \cdot \boldsymbol{c}_g, \tag{6.145}$$

where $\boldsymbol{c}_g \equiv \partial\omega/\partial\boldsymbol{k}$ is the group velocity.

The new wave has the general form

$$\psi' = (\widetilde{\psi} + \delta\widetilde{\psi})\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}\mathrm{e}^{-\boldsymbol{\delta k}\cdot\boldsymbol{x}+\delta\omega t} + \text{c.c.}, \tag{6.146}$$

and, analogously to (6.142), the associated wave activity and flux have the forms:

$$A' = \left[b + \delta b + (a + \delta a)\mathrm{e}^{2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} + (a^* + \delta a^*)\mathrm{e}^{-2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}\right]\mathrm{e}^{-2\boldsymbol{\delta k}\cdot\boldsymbol{x}+2\delta\omega t} \tag{6.147a}$$

$$\boldsymbol{F}' = \left[\boldsymbol{g} + \delta\boldsymbol{g} + (\boldsymbol{f} + \delta\boldsymbol{f})\mathrm{e}^{2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} + (\boldsymbol{f}^* + \delta\boldsymbol{f}^*)\mathrm{e}^{-2\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}\right]\mathrm{e}^{-2\boldsymbol{\delta k}\cdot\boldsymbol{x}+2\delta\omega t}, \tag{6.147b}$$

where the $\delta$ quantities are small. If we now demand that $A'$ and $\boldsymbol{F}'$ satisfy assumption (i), then substituting (6.147) into (6.140) gives, after a little algebra,

$$(\boldsymbol{g} + \delta\boldsymbol{g}) \cdot \boldsymbol{\delta k} = (b + \delta b)\delta\omega \tag{6.148}$$

and therefore at first order in $\delta$ quantities, $\boldsymbol{g} \cdot \boldsymbol{\delta k} = b\delta\omega$. Using (6.145) and (6.143) we obtain

$$\boldsymbol{c}_g = \frac{\boldsymbol{g}}{b} = \frac{\overline{\boldsymbol{F}}}{\overline{A}}, \tag{6.149}$$

and using this the conservation law, (6.140), becomes

$$\frac{\partial \overline{A}}{\partial t} + \nabla \cdot (\boldsymbol{c}_g \overline{A}) = 0. \tag{6.150}$$

Thus, for waves satisfying our three assumptions, the flux velocity — that is, the propagation velocity of the wave activity — is equal to the group velocity.

## 6.9 ENERGY PROPAGATION OF POINCARÉ WAVES

In the final section of this chapter we discuss the energetics of Poincaré waves, and show explicitly that the energy propagation occurs at the group velocity. (Poincaré waves were first introduced in section 3.7.2 and the reader may wish to review that section before continuing.) We begin with the one-dimensional problem as this shows the essential aspects and the algebra is a little simpler.

### 6.9.1 Energetics in one dimension

The one-dimensional (i.e., no variations in the $y$-direction), inviscid linear shallow-water equations on the $f$-plane, linearized about a state of rest, are

$$\frac{\partial u}{\partial t} - f_0 v = -g\frac{\partial h}{\partial x}, \qquad \frac{\partial v}{\partial t} + f_0 u = 0, \qquad \frac{\partial \eta}{\partial t} = -H\frac{\partial u}{\partial x}. \tag{6.151a,b,c}$$

To obtain the dispersion relation we differentiate the first equation with respect to $t$ and substitute from the second and third to obtain

$$\frac{\partial^2 u}{\partial t^2} - Hg\frac{\partial u}{\partial x} + f_0^2 u = 0, \tag{6.152}$$

whence, assuming solutions of the form $u = \mathrm{Re}\,\tilde{u}e^{i(kx-\omega t)}$, we obtain the dispersion relation

$$\omega^2 = f^2 + Hgk^2. \tag{6.153}$$

This is immediately recognizable as a special case of the two-dimensional dispersion relation. An interesting property of this equation is obtained by differentiating with respect to $k$, giving $2\omega\partial\omega/\partial k = 2kHg$ or

$$c_g = \frac{Hg}{c_p}, \tag{6.154}$$

where $c_g = \partial\omega/\partial k$ and $c_p = \omega/k$ are the group and phase velocities, respectively. Using (6.153) and (6.154) the ratio of the group and phase velocities is found to be

$$\frac{c_g}{c_p} = \frac{L_d^2 k^2}{1 + L_d^2 k^2}, \tag{6.155}$$

where $L_d = \sqrt{gH}/f$ is the deformation radius. This ratio is always less than unity, tending to zero in the long-wave limit ($kL_d \ll 1$) and to unity for short waves ($kL_d \gg 1$).

The energy equations are obtained by multiplying the three equations of (**??**) by $u, v$ and $\eta$ respectively, and adding, to give

$$\frac{\partial E}{\partial t} + \frac{\partial F}{\partial x} = 0, \tag{6.156a}$$

where

$$E = \frac{1}{2}(Hu^2 + Hv^2 + g\eta^2), \qquad F = gHu\eta, \tag{6.156b}$$

are the energy density and the energy flux, respectively. Note that in the linear approximation the energy is transported only by the pressure term, whereas in the full nonlinear equations there is also an advective transport.

*The group velocity property*

To specialize to the case of propagating *waves* we need to average over a wavelength and use the phase relationships between $u, v$ and $\eta$ implied by the equations of motion. Writing $u = \mathrm{Re}\,\tilde{u}e^{i(kx-\omega t)}$, and similarly for $v$ and $\eta$, we have

$$\tilde{v} = -\mathrm{i} f\frac{\tilde{\eta}}{Hk}, \qquad \tilde{u} = \omega\frac{\tilde{\eta}}{Hk}. \tag{6.157a,b}$$

The kinetic energy, averaged over a wavelength, is then

$$KE = \frac{1}{2}H(\overline{u^2} + \overline{v^2}) = \frac{1}{4}(\omega^2 + f^2)\frac{\widetilde{\eta}^2}{Hk^2} = \frac{1}{4}\frac{\omega^2 + f^2}{\omega^2 - f^2}g\widetilde{\eta}^2 \tag{6.158}$$

using (6.157) and the dispersion relation, with the extra factor of one half arising from the averaging over a wavelength. Similarly, the potential energy of the wave is

$$PE = \frac{1}{2}g\overline{\eta^2} = \frac{1}{4}g\widetilde{\eta}^2 \tag{6.159}$$

Thus, the ratio of kinetic to potential energy is just

$$\frac{KE}{PE} = \frac{\omega^2 + f^2}{\omega^2 - f^2} = 1 + \frac{2}{k^2 L_d^2} \tag{6.160}$$

using the dispersion relation, and where $L_d = \sqrt{gH/f}$ is the deformation radius. Thus, the kinetic energy is always *greater* than the potential energy (there is no equipartition in this problem), with the ratio approaching unity for small scales (large $k$).

The total energy (kinetic plus potential) is then

$$KE + PE = \frac{1}{4}\left(\frac{\omega^2 + f^2}{\omega^2 - f^2} + 1\right)g\widetilde{\eta}^2 = \frac{1}{2}\frac{\omega^2}{k^2 H}\widetilde{\eta}^2 = \frac{1}{2}\frac{c_p^2}{H}\widetilde{\eta}^2, \tag{6.161}$$

again using the dispersion relation. The energy flux, $F$, averaged over a wavelength, is

$$F = gH\overline{u\eta} = \frac{1}{2}\frac{g\omega}{k}\widetilde{\eta}^2 = \frac{1}{2}gc_p\widetilde{\eta}^2. \tag{6.162}$$

From (6.161) an (6.162) the flux and the energy are evidently related by

$$\boxed{F = \frac{Hg}{c_p}E = c_g E}, \tag{6.163}$$

using (6.154). That is, the energy flux is equal to the group velocity times the energy itself. Note that in this problem there is no flux in the $y$ direction, because $v$ and $\eta$ are exactly out of phase from (6.157a).

### 6.9.2 ✦ Energetics in two dimensions

The derivations of the preceding section carry through, *mutatis mutandis,* in the full two-dimensional case. We will give only the key results and allow the reader to fill in the algebra. As derived in section 3.7.2 the dispersion relation is

$$\omega^2 = f_0^2 + gH(k^2 + l^2). \tag{6.164}$$

The relation between the components of the group velocity and the phase speed is very similar to the the one-dimensional case, and in particular we have

$$c_g^x = \frac{\partial \omega}{\partial k} = gH\frac{k}{\omega} = \frac{gH}{c_p^x}, \qquad c_g^y = \frac{\partial \omega}{\partial l} = gH\frac{l}{\omega} = \frac{gH}{c_p^y}. \tag{6.165}$$

The magnitude of the group velocity is $c_g \equiv |\boldsymbol{c}_g| = (c_g^{x2} + c_g^{y2})^{1/2}$. The magnitude of the phase speed, in the direction of travel of the wave crests, is $c_p = \omega/(k^2 + l^2)^{1/2}$ (note that in general this is *smaller* than the phase speed in either the $x$ or $y$ directions, $\omega/k$ or $\omega/l$). Thus, we have

$$c_g^2 = (gH)^2 \frac{k^2 + l^2}{\omega^2} = \frac{(gH)^2}{c_p^2}, \qquad \boldsymbol{c}_g = \left(\frac{gH}{c_p K}\right)\boldsymbol{k}, \tag{6.166}$$

which is analogous to (6.154). The ratio of the magnitudes of the group and phase velocities is, analogously to (6.155),

$$\frac{c_g}{c_p} = \frac{gH}{c_p^2} = \frac{L_d^2 K^2}{1 + L_d^2 K^2}, \tag{6.167}$$

where $K^2 = k^2 + l^2$. As in the one-dimensional case the group velocity is large for short waves, in which rotation plays no role, and small for long waves.

The energy equation is found to be

$$\frac{\partial E}{\partial t} + \nabla \cdot \boldsymbol{F} = 0 \tag{6.168a}$$

with

$$E = \frac{1}{2}(Hu^2 + Hv^2 + g\eta^2), \qquad F = gH(u\mathbf{i} + v\mathbf{j})\eta. \tag{6.168b}$$

From the equations of motion the phase relations between the fields are found to be

$$\widetilde{v} = \frac{\omega l - \mathrm{i}kf}{HK^2}\widetilde{\eta}, \qquad \widetilde{u} = \frac{\omega k - \mathrm{i}lf}{HK^2}\widetilde{\eta}, \tag{6.169}$$

so that the kinetic energy is given by, similar to (6.158),

$$KE = \frac{1}{2}H(\overline{u^2} + \overline{v^2}) = \frac{1}{4}(\omega^2 + f^2)\frac{\widetilde{\eta}^2}{Hk^2} = \frac{1}{4}\frac{\omega^2 + f^2}{\omega^2 - f^2}g\widetilde{\eta}^2, \tag{6.170}$$

and the potential energy by

$$PE = \frac{1}{2}g\overline{\eta^2} = \frac{1}{4}g\widetilde{\eta}^2 \tag{6.171}$$

The ratio of the kinetic and potential energies is given by

$$\frac{KE}{PE} = \frac{\omega^2 + f^2}{\omega^2 - f^2} = 1 + \frac{2}{K^2 L_d^2} \tag{6.172}$$

The total (kinetic plus potential) energy is given by

$$E = KE + PE = \frac{1}{4}\left(\frac{\omega^2 + f^2}{\omega^2 - f^2} + 1\right) g\widetilde{\eta}^2 = \frac{1}{2}\frac{\omega^2}{K^2 H}\widetilde{\eta}^2 = \frac{1}{2}\frac{c_p^2}{H}\widetilde{\eta}^2, \tag{6.173}$$

The energy flux, $\boldsymbol{F}$, averaged over a wavelength, is

$$\boldsymbol{F} = gH\overline{\boldsymbol{u}\eta} = \frac{1}{2}\frac{g\omega}{k^2 + l^2}\widetilde{\eta}^2\boldsymbol{k} = \frac{1}{2}\frac{g\omega}{K^2}\widetilde{\eta}^2\boldsymbol{k}, \tag{6.174}$$

using (6.169) and where $\boldsymbol{k} = k\mathbf{i} + l\mathbf{j}$ is the wavevector of the wave.

From (6.173) and (6.174), and using (6.166), the flux and the energy are found to be related by

$$\boxed{\boldsymbol{F} = \boldsymbol{c}_g E}. \tag{6.175}$$

That is, the energy flux is equal to the group velocity times the energy itself.

## *Notes*

1 A useful introduction to wave motion, from which this chapter has benefited, can be found in unpublished lecture notes by Chapman *et al.* (1989). Other useful material can be found in the 'further reading' section below.

2 For example *Linear and Nonlinear Waves* by G. B. Whitham or *Nonlinear Dispersive Waves* by M. J. Ablowitz.

3 For a review of group velocity, see Lighthill (1965).

4 More detailed treatments of ray theory and related matters are given by Whitham (1974), Lighthill (1978) and LeBlond & Mysak (1980).

5 What are now called Rossby waves were probably first discovered in a theoretical context by Hough (1897, 1898). He considered the linear shallow water equations on a sphere (i.e., Laplace's tidal equations) expanding the solution in powers of the sine of latitude, and obtained two classes of waves: long, rotationally modified, gravity waves and a balanced wave dependent on variations in Coriolis parameter. However, his work was mainly aimed at understanding ocean tides and it was not until the topic was revisited by Rossby (1939) that the meteorological relevance was appreciated. Rossby used the beta-plane approximation in Cartesian co-ordinates, and the simplicity of the presentation along with the meteorological context lead to the work attracting significant notice.

6 This non-Doppler effect also arises quite generally, even in models in height coordinates. See White (1977) and problem 5.5.

7 See Chapman & Lindzen (1970).

8 Following Longuet-Higgins (1964).

9 To read more about this problem, see Paldor *et al.* (2007) and Heifetz & Caballero (2014).

10 Waves of this type seem to have been first deduced by Bjerknes (1937).

11 Following Pedlosky (2003).

12 The form of this derivation was originally given by Hayes (1977) in the context of wave energy. See Vanneste & Shepherd (1998) for generalizations.

## *Further reading*

Majda, A. J., 2003. *Introduction to PDEs and Waves for the Atmosphere and Ocean.*
Provides a compact, somewhat mathematical introduction to various equation sets and their properties, including quasi-geostrophy.

## *Problems*

6.1 Consider the flat-bottomed shallow water potential vorticity equation in the form

$$\frac{\mathrm{D}}{\mathrm{D}t}\frac{\zeta + f}{h} = 0 \tag{P6.1}$$

(a) Suppose that deviations of the height field are small compared to the mean height field, and that the Rossby number is small (so $|\zeta| \ll f$). Further consider flow on a $\beta$-plane such that $f = f_0 + \beta y$ where $|\beta y| \ll f_0$. Show that the evolution equation becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\zeta + \beta y - \frac{f_0 \eta}{H}\right) = 0 \tag{P6.2}$$

where $h = H + \eta$ and $|\eta| \ll H$. Using geostrophic balance in the form $f_0 u = -g\partial\eta/\partial y$, $f_0 v = g\partial\eta/\partial x$, obtain an expression for $\zeta$ in terms of $\eta$.

(b) Linearize (P6.2) about a state of rest, and show that the resulting system supports two-dimensional Rossby waves that are similar to those of the usual two-dimensional barotropic system. Discuss the limits in which the wavelength is much shorter or much longer than the deformation radius.

(c) Linearize (P6.2) about a *geostrophically balanced state* that is translating uniformly eastwards. Note that this means that:

$$u = U + u' \qquad \eta = \eta(y) + \eta',$$

where $\eta(y)$ is in geostrophic balance with $U$. Obtain an expression for the form of $\eta(y)$.

(d) Obtain the dispersion relation for Rossby waves in this system. Show that their speed is different from that obtained by adding a constant $U$ to the speed of Rossby waves in part (b), and discuss why this should be so. (That is, why is the problem not *Galilean invariant*?)

6.2 Obtain solutions to the two-layer Rossby wave problem by seeking solutions of the form

$$\psi_1 = \mathrm{Re}\, \widetilde{\psi}_1 \mathrm{e}^{\mathrm{i}(k_x x + k_y y - \omega t)}, \qquad \psi_2 = \mathrm{Re}\, \widetilde{\psi}_2 \mathrm{e}^{\mathrm{i}(k_x x + k_y y - \omega t)}. \tag{P6.3}$$

Substitute (P6.3) directly into (6.67) to obtain the dispersion relation, and show that the ensuing two roots correspond to the baroclinic and barotropic modes.

6.3 ✦ (Not difficult, but messy.) Obtain the vertical normal modes and the dispersion relationship of the two-layer quasi-geostrophic problem with a free surface, for which the equations of motion linearized about a state of rest are

$$\frac{\partial}{\partial t}\left[\nabla^2\psi_1 + F_1(\psi_2 - \psi_1)\right] + \beta\frac{\partial\psi_1}{\partial x} = 0 \tag{P6.4a}$$

$$\frac{\partial}{\partial t}\left[\nabla^2\psi_2 + F_2(\psi_1 - \psi_2) - F_{ext}\psi_1\right] + \beta\frac{\partial\psi_2}{\partial x} = 0, \tag{P6.4b}$$

where $F_{ext} = f_0/(gH_2)$.

6.4 Given the baroclinic dispersion relation, $\omega = -\beta k^x/(k^{x2} + k_d^2)$, for what value of $k_x$ is the $x$-component of the group velocity the largest (i.e., the most positive), and what is the corresponding value of the group velocity?

6.5 ✦ Show that the non-Doppler effect arises using geometric height as the vertical coordinate, using the modified quasi-geostrophic set of White (1977). In particular, obtain the dispersion relation for stratified quasi-geostrophic flow with a resting basic state. Then obtain the dispersion relation for the equations linearized about a uniformly translating state, paying attention to the lower boundary condition, and note the conditions under which the waves are stationary. Discuss.

6.6 (a) Obtain the dispersion relationship for Rossby waves in the single-layer quasi-geostrophic potential vorticity equation with linear drag.

(b) Obtain the dispersion relation for Rossby waves in the linearized two-layer potential vorticity equation with linear drag in the lowest layer.

(c) ✦ Obtain the dispersion relation for Rossby waves in the continuously stratified quasi-geostrophic equations, with the effects of linear drag appearing in the thermodynamic equation for the lower boundary condition. That is, the boundary condition at $z = 0$ is $\partial_t(\partial_z \psi) + N^2 w = 0$ where $w = \alpha\zeta$ with $\alpha$ being a constant. You may make the Boussinesq approximation and assume $N^2$ is constant if you like.

*It mounts at sea, a concave wall,*
*Down-ribbed with shine,*
*And pushes forward, building tall,*
*Its steep incline.*

Thom Gunn, *From the Wave.*

# CHAPTER SEVEN

# Gravity Waves

In this chapter we consider gravity waves, which are simply waves in a fluid with gravity providing the restoring force. (A completely different form of gravity wave arises in general relativity.) In order for gravity to have an effect the density must vary, as discussed in chapter 2, and this means that the waves must either exist at a fluid interface or that stratification is present (and one might of course regard a fluid interface as being a particularly abrupt form of stratification). It is thus common to think of gravity waves as being either internal waves or surface waves: internal waves occur in the interior of a fluid, often but not always when the density changes are continuous and surface gravity waves, or interfacial waves, are the waves at a fluid interface. Naturally enough there are many similarities between the two classes of waves — indeed surface waves might be considered a limiting form of internal waves, existing when the density of the overlying fluid goes to zero. We have already considered such interfacial waves in the hydrostatic case in chapter 3 (on shallow water systems), and we will first extend that to the nonhydrostatic case. We then consider internal waves in the continuously stratified equations and that constitutes the bulk of the chapter.[1]

In most of the chapter we will restrict attention to the Boussinesq equations, mainly because in making the incompressibility approximation sound waves are eliminated, greatly simplifying the treatment. In the atmosphere the Boussinesq equations are not a quantitatively good approximation except for motions of a small vertical extent; the anelastic equations improve matters in allowing for a vertical variation of the basic state density, an effect that is particularly important when considering the vertical propagation of gravity waves high into the atmosphere. Nevertheless, no truly new types of waves are introduced in this way, though, and so we leave the details for the reader to find in the original literature. If, on the other had, the fluid is truly compressible then sound waves make themselves heard, and we consider the somewhat algebraically complex case of *acoustic-gravity* waves at the end of this chapter. We begin with the simpler case of surface gravity waves at the top of a constant density fluid.

## 7.1 SURFACE GRAVITY WAVES

Let us consider an incompressible fluid with a free surface and a flat bottom that obeys the equations of motion

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla\phi - g\mathbf{k}, \qquad \nabla\cdot\boldsymbol{v} = 0, \tag{7.1}$$

using our standard notation, with $\phi = p/\rho_0$. The above equations are the three-dimensional momentum equation and the mass continuity equation, respectively. We suppose that is a free surface at the top of the fluid, at $z = \eta(x, y, t)$, the mean position of the free surface is at $z = 0$ and the bottom of the fluid, assumed flat, is at $z = -H$ — refer to Fig. 3.1 on on page 124.

In a state of rest the pressure, $\phi_0$ say, is given by hydrostatic balance and so $\phi_0 = -gz$. If we write $\phi = -gz + \phi'$ the momentum equation becomes, without approximation,

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla\phi'. \tag{7.2}$$

Linearizing the equations of motion about such a resting state straightforwardly yields

$$\frac{\partial\boldsymbol{v}'}{\partial t} = -\nabla\phi', \qquad \nabla\cdot\boldsymbol{v}' = 0, \tag{7.3a,b}$$

where a prime denotes a perturbation quantity in the usual way. We now proceed by expressing the problem solely in terms of pressure. (An equivalent alternative is to use a velocity potential, $\xi$ say, such that $\boldsymbol{v} = \nabla\xi$. Such a procedure is possible because, from (7.3a) the flow is irrotational, and solving the problem in this manner is left as an exercise.) Taking the divergence of (7.3a) and using (7.3b) gives us Laplace's equation for the pressure, namely

$$\nabla^2\phi' = 0. \tag{7.4}$$

These has no explicit time dependence, but the boundary conditions are time dependent and that is how we will obtain the dispersion relation.

### 7.1.1 Boundary conditions

Since (7.4) is an equation for pressure we seek boundary conditions on pressure. At the bottom of the fluid ($z = -H$) the condition that $w = 0$ may be turned into a condition on pressure using (7.3b), namely that

$$\frac{\partial\phi'}{\partial z} = 0 \qquad \text{at } z = -H. \tag{7.5}$$

At the top surface, $z = \eta$, the pressure must equal that of the atmosphere above. We will take this to be a constant, and in particular zero, so that $\phi = 0$ at $z = \eta$. Now, the perturbation pressure is given by $\phi = -gz + \phi'$, so that at $z = \eta$ we obtain

$$\phi' = g\eta \qquad \text{at } z = \eta. \tag{7.6}$$

A second boundary condition at the top is the kinematic condition that a fluid parcel in the free surface must remain within it, and therefore that (with full nonlinearity)

$$\frac{\mathrm{D}}{\mathrm{D}t}(z - \eta) = 0. \tag{7.7}$$

If we linearize this and use the definition of $w$ we obtain $w' = \partial\eta/\partial t$ at $z = \eta$, which using (7.6) becomes $w' = g^{-1}\partial\phi'/\partial t$. Using the vertical component of the momentum equation, (7.3a), we obtain the pressure boundary condition

$$\frac{1}{g}\frac{\partial^2\phi'}{\partial t^2} = -\frac{\partial\phi'}{\partial z} \qquad \text{at } z = \eta. \tag{7.8}$$

The value of $\eta$ is in fact unknown without solving the problem itself, and in the general (nonlinear) case we have to solve the whole problem in a self-consistent fashion. However, in the linear problem $\eta$ is presumptively small (remember we are linearizing the free surface about $z = 0$) and we will apply this boundary condition at $z = 0$ rather than at $z = \eta$, for the error will only be second order (see problem 7.**??**).

Having established the equations and the boundary conditions, and noting that we will be dealing exclusively with linear equations in the rest of this section (and in fact for most of this chapter), we'll now drop the primes on perturbation quantities unless ambiguity arises.

### 7.1.2 Wave solutions

We now seek solutions to (7.4) in the form

$$\phi = \mathrm{Re}\, \Phi(z) \exp(\mathrm{i}[\boldsymbol{k}\cdot\boldsymbol{x} - \omega t]) \tag{7.9}$$

where $\boldsymbol{x} = \mathbf{i}x + \mathbf{j}y$ and $\boldsymbol{k} = \mathbf{i}k + \mathbf{j}l$ and Re denotes that the real part is to taken, a notation that we subsequently drop unless it causes ambiguity. We obtain

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2} - K^2\Phi = 0, \tag{7.10}$$

where $K^2 = k^2 + l^2$ and the boundary conditions are that $\mathrm{d}\Phi/\mathrm{d}z = 0$ at $z = -H$ and $\mathrm{d}^2\Phi/\mathrm{d}z^2 = -g\mathrm{d}\Phi/\mathrm{d}z$ at $z = 0$. The bottom boundary condition is satisfied by a solution of the form

$$\Phi = A\cosh K(z + H). \tag{7.11}$$

Substituting into the top boundary condition, (7.8) at $z = 0$, we obtain

$$-\omega^2\cosh KH = gK\sinh KH = 0, \tag{7.12}$$

or

$$\boxed{\omega = \pm\sqrt{gK\tanh KH}}\,. \tag{7.13}$$

This is the dispersion relation for surface gravity waves. The corresponding phase speed is given by

$$c_p = \frac{\omega}{K} = \pm\sqrt{gH}\left(\frac{\tanh KH}{KH}\right)^{1/2}. \tag{7.14}$$

Using (7.9) and (7.11) the full solution for the pressure field is

$$\phi = \mathrm{Re}\, \Phi_0 \cosh K(z + H)\exp(\mathrm{i}[\boldsymbol{k}\cdot\boldsymbol{x} - \omega t]) \tag{7.15a}$$

with $\omega$ given by (7.13) and the amplitude $\Phi_0$ being set by the initial conditions. It is convenient to write the amplitude $\Phi_0$ in terms of the amplitude of the free surface elevation, $\eta_0$, using the upper boundary condition that $\phi = g\eta$ so that $\eta_0 = \phi_0/g$. The other field variables may be found from (7.3a) and are given by

$$u = \eta_0 \frac{k}{\omega} gC \cosh K(z+H), \tag{7.15b}$$

$$v = \eta_0 \frac{l}{\omega} gC \cosh K(z+H), \tag{7.15c}$$

$$w = -i\eta_0 \frac{K}{\omega} gC \sinh K(z+H) \tag{7.15d}$$

where $C = \exp(\mathrm{i}[\boldsymbol{k}\cdot\boldsymbol{x} - \omega t])/\cosh KH$, and as usual it is the real parts of each expression that should be taken. Thus, if we take $\eta_0$ to be real then $u$ and $v$ vary like $\cos(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)$ and $w$ varies as $\sin(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)$.

### 7.1.3 Properties of the solution

Let's now note a few things about the solutions we have obtained. First, from (7.13) we see that for each wavevector amplitude there are two waves propagating in opposite directions, with a frequency and phase speed that depend only on the wavelength $K$ and not the orientation of the wave vector. Second, the waves are *dispersive*. That is, similar to Rossby waves but unlike light waves in a vacuum or shallow water waves, the phase speed is different for waves of different wavelengths. A pattern made up of a superposition of many waves will therefore disperse. Since the frequency is a function only of $K$ (and not of $k$ or $l$ individually) the group velocity is parallel to the wave vector itself and is given by

$$\boldsymbol{c}_g = \nabla_{\boldsymbol{k}}\omega = \frac{\partial \omega}{\partial K}\frac{\boldsymbol{k}}{K}, \tag{7.16}$$

where $\boldsymbol{k} = k\mathbf{i} + l\mathbf{j}$, so that $\boldsymbol{k}/K$ is the unit vector in the direction of propagation. Using the dispersion relation $\omega^2 = gK\tanh KH$ we obtain

$$2\omega\frac{\partial \omega}{\partial K} = g\left(\tanh KH + \frac{KH}{\cosh^2 KH}\right) \tag{7.17}$$

so that

$$\boldsymbol{c}_g = \frac{g}{2c_p K}\left(\tanh KH + \frac{KH}{\cosh^2 KH}\right)\boldsymbol{k} \tag{7.18}$$

and the ratio of the group speed (i.e., the magnitude of the group velocity) to the phase speed is given by

$$\frac{c_g}{c_p} = \frac{1}{2}\left(1 + \frac{2KH}{\sinh 2KH}\right). \tag{7.19}$$

We may note two important limiting cases, as follows.

*(i)* The long wavelength or shallow water limit, $KH \ll 1$. In this limit the wavelength is much greater than the depth of the fluid and the dispersion relation (7.13) reduces to $\omega = K\sqrt{gH}$ (since for small $x$, $\tanh x \to x$) and $c_p = c_g = \sqrt{gH}$ and the waves are

nondispersive. This result is apparent from (7.19) in the limit of $KH \ll 1$. As expected, this is the same dispersion relation as was previously derived *ab initio* for shallow water waves in chapter 3. This limit is appropriate as water waves approach the shore and start feeling the bottom, and for long waves such as tides and tsunamis.

The pressure field in this limit is given, using (7.15a),

$$\phi = \eta_0 g \exp(\mathrm{i}[\boldsymbol{k} \cdot \boldsymbol{x} - \omega t]). \tag{7.20}$$

This is the *perturbation* pressure associated with the wave, and evidently it does not depend on depth. The total pressure at a given point in the fluid is given by the static pressure plus perturbation pressure and this is, including the density $\rho_0$ ,

$$p = -\rho_0 g z + \rho_0 \phi = \rho_0 g(\eta - z). \tag{7.21}$$

Evidently, the pressure in the shallow water limit is hydrostatic. If $1/k > 20H$ the error in this approximation less than 3%.

*(ii)* The short wavelength or deep water limit, $KH \gg 1$. For large $KH$, $\tanh KH \to 1$ so that the dispersion relation becomes $\omega^2 = gK$ and $c_p^2 = g/K$. These waves are dispersive, with long waves travelling faster than short waves. A familiar manifestation of this arises when a rock is thrown into a pool. Initially, waves of all wavelengths are excited (for the initial disturbance is like a delta function), but the long waves propagate away faster than the short waves and reach distance objects first. The group speed in this case is given by

$$c_g = \frac{\partial \omega}{\partial k} = \frac{g}{2\omega} = \frac{1}{2}\sqrt{\frac{g}{K}} = \frac{c_p}{2}. \tag{7.22}$$

This result is also apparent from (7.19) in the limit of short waves, $KH \gg 1$, and it has an interesting consequence for wave packets. Consider a packet of short waves moving in the positive $x$ direction. The envelope moves with the group speed and the individual crests with the phase speed, so that individual crests enter the packet from the rear and travel through the packet, exiting at the front.

### Parcel motion

The trajectories of water parcels is rather interesting in water waves. It turns out that in deep water the parcels make circular orbits with an amplitude diminishing with depth, whereas shallow water waves trace elliptic paths, as illustrated in Fig. 7.1, as we now explain.

We obtain the parcel excursions using the expressions for velocity (7.15b,c,d), taking $v = 0$ without loss of generality. For shallow water waves ($KH \ll 1$) $u$ is depth independent and the velocity and the excursion in the $x$ direction, which we denote as $X$, are given by

$$u = \eta_0 \frac{kg}{\omega} \cos(kx - \omega t), \qquad X = \eta_0 \frac{gk}{\omega^2} \sin(kx - \omega t), \tag{7.23a}$$

and this is independent of $z$. The excursion in the z direction, $Z$, is given by

$$w = \eta_0 \frac{k^2}{\omega}(z + H)\sin(kx - \omega t), \qquad Z = \eta_0 \frac{gk^2}{\omega^2}(z + H)\cos(kx - \omega t). \tag{7.23b}$$

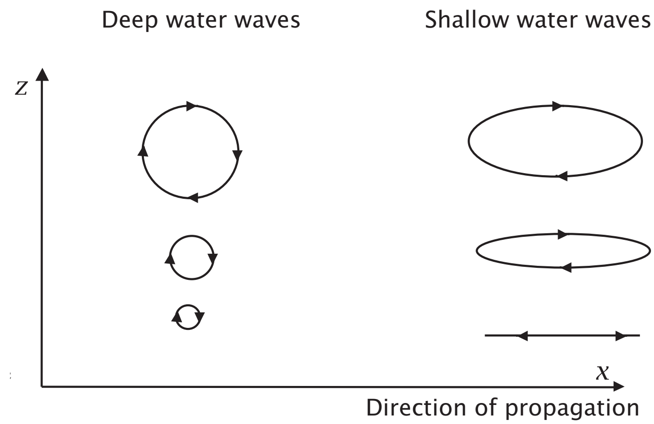


**Fig. 7.1** Schematic of parcel motion for deep and shallow water waves. The motion is circular for deep water waves, with an amplitude that decreases exponentially with depth. The motion is elliptical for shallow water waves, but the horizontal excursion is independent of depth and the vertical excursion decays linearly with depth.

where $\omega = k\sqrt{gH}$. Note that at $z = 0$ $Z = \eta$, as expected. The above expressions for $X$ and $Z$ are, at some fixed location $x$ and $z$, parametric representations of an ellipse. As $z$ varies the horizontal amplitude of the ellipses remains constant whereas the vertical amplitude decreases linearly from the top $z = 0$ to a zero amplitude at the bottom, $z = -H$. The vertical amplitude is also generally much less than the horizontal amplitude, by the ratio

$$\frac{|Z|}{|X|} = \frac{|w|}{|u|} \sim kH \ll 1. \tag{7.24}$$

That it, the fluid motion is mostly horizontal.

In the deep water limit, $kH \gg 1$, the horizontal and vertical velocities and excursions are given by

$$u = \eta_0 \frac{kg}{\omega} \exp kz \cos(kx - \omega t), \qquad X = \eta_0 \frac{kg}{\omega^2} \exp kz \sin(kx - \omega t), \tag{7.25a}$$

$$w = \eta_0 \frac{kg}{\omega} \exp kz \sin(kx - \omega t), \qquad Z = \eta_0 \frac{kg}{\omega^2} \exp kz \cos(kx - \omega t). \tag{7.25b}$$

where $\omega^2 = gk$. [Check signs. xxx] Note again that at $z = 0$, $Z = \eta$. The expressions for $X$ and $Z$, having the same amplitude, are now parametric representations of circles whose amplitudes diminish exponentially with depth. Evidently, all the dynamical variables decrease exponentially with depth, with an e-folding scale of the wavelength itself. In the deep water limit the wave field cannot feel the bottom of the fluid container and all the expressions become independent of depth.

### ✦ *Energy propagation*

For our final discussion on this topic we look at the energy and energy propagation of surface waves. The kinetic energy per unit horizontal area is given by

$$KE = \int_{-H}^{0} \frac{1}{2}\rho_0 \boldsymbol{v}^2 \, \mathrm{d}z. \tag{7.26}$$

The upper limit on the integration is taken to be $z = 0$, rather than $z = \eta$, because using the latter would lead to a term of order $\eta \boldsymbol{v}^2$, which is third order in perturbation quantities. The potential energy per unit horizontal area is

$$PE = \int_{-H}^{\eta} \rho_0 g z \, dz = \frac{\rho_0 g}{2}(\eta^2 - H^2). \tag{7.27}$$

The integral now must be over the complete depth of the fluid in order to calculate the potential energy to quadratic order. The term in $H^2$ is a constant and so is largely irrelevant to the problem of energy propagation. Also, since $\rho_0$ is a constant we will set its value to unity.

The kinetic energy equation is obtained by taking the dot product of the linearized momentum equation, (7.3a) with $\boldsymbol{v}$ and integrating over the depth of the fluid to give

$$\int_{-H}^{0} dz \left[ \frac{\partial}{\partial t}\frac{\boldsymbol{v}^2}{2} + \nabla \cdot (\boldsymbol{u}\phi) + \frac{\partial w\phi}{\partial z} \right] = 0, \tag{7.28}$$

noting that $\boldsymbol{v} = \boldsymbol{u} + w\mathbf{k}$ and $\nabla \cdot \boldsymbol{v} = 0$. The boundary conditions on $w$ are that $w = 0$ at $Z = -H$ and $w = \partial\eta/\partial t$ at $z = 0$. Further, at $z = 0$ $\phi = g\eta$, and using these results (7.28) becomes

$$\int_{-H}^{0} dz \left[ \frac{\partial}{\partial t}\frac{\boldsymbol{v}^2}{2} + \nabla \cdot (\boldsymbol{u}\phi) \right] + g\frac{\partial}{\partial t}\frac{\eta^2}{2} = 0, \tag{7.29}$$

which, using (7.26) and (7.27), is just

$$\boxed{\frac{\partial}{\partial t}(KE + PE) + \nabla \cdot \boldsymbol{F} = 0}. \tag{7.30}$$

where $\boldsymbol{F} = \int_{-H}^{z} \boldsymbol{u}\phi \, dz$ is the energy flux, a vector with only horizontal components. (Thus, the divergence term in (7.30) is just a horizontal divergence.)

Equation (7.30) is an energy conservation equation for the linearized equations. It is fairly general at the moment, for we have not specialized to the case of *wave* motion. Let's do that now, by using the properties of the waves derived above and averaging over a wave period. Without loss of generality we'll assume the waves are propagating in the $x$ direction so that $v = 0$ and $K = k$; nevertheless, the calculation is rather algebraic and the trusting reader may skim it.

The kinetic energy averaged over a wave period, $\overline{KE}$ is given by

$$\begin{aligned}\overline{KE} &= \frac{\omega}{2\pi}\int dt \left( \int \frac{1}{2}\boldsymbol{v}^2 \, dz \right) \\ &= \frac{k^2\eta_0^2 g^2}{2\omega^2\cosh^2 kH}\frac{\omega}{2\pi}\int dt \int dz \times \\ &\qquad \left[\cosh^2 k(z+H)\cos^2(kx-\omega t) + \sinh^2 k(z+H)\sin^2(kx - \omega t)\right].\end{aligned} \tag{7.31}$$

In this expression the time integrals range from 0 to $2\pi/\omega$ and the vertical integrals range from $-H$ to 0. The time averages of $\sin^2$ and $\cos^2$ produce a factor of $1/2$, and noting that $\cosh^2 x + \sinh^2 x = \cosh 2x$ we obtain

$$\overline{KE} = \frac{k^2\eta_0^2 g^2}{2\omega^2\cosh^2 kH}\frac{1}{2}\frac{\sinh(2kH)}{2k}. \tag{7.32}$$

Using the dispersion relation $\omega^2 = gk\tanh kH$ we finally obtain the simple expression

$$\overline{KE} = \frac{g\eta_0^2}{4}. \tag{7.33}$$

The perturbation potential energy is given by

$$\overline{PE} = \frac{\omega}{2\pi}\int \frac{1}{2} g\eta^2 \,\mathrm{d}t = \frac{g\eta_0^2}{2}\frac{\omega}{2\pi}\int \cos^2(kx - \omega t)\,\mathrm{d}t \\ = \frac{g\eta_0^2}{4}. \tag{7.34}$$

Evidently, from (7.33) and (7.34) there is equipartitioning of energy time-averaged potential and kinetic energy components. Such equipartitioning is not, however, a universal property of wave motion.

The time averaged energy flux, which is in the $x$ direction, is given by

$$\overline{F} = \frac{\omega}{2\pi}\int \mathrm{d}t \int u'\phi' \,\mathrm{d}z. \tag{7.35}$$

Using the wave expressions (7.15) we obtain, after a couple of lines of algebra,

$$\overline{F} = \frac{1}{2}\eta_0^2 \frac{g^2}{2c}\frac{1}{\cosh^2 kH}\left[\frac{\sinh 2kH}{2k} + H\right] \tag{7.36}$$

Using (7.18) and the fact that $\sinh 2hK = 2\sinh kH \;\cosh kH$ we obtain

$$\overline{F} = \frac{\eta_0^2 g}{2} c_g = (\overline{KE} + \overline{PE})c_g. \tag{7.37}$$

Thus, using (7.33), (7.34) and (7.37), and generalizing the direction of propagation, we have that

$$\boxed{\frac{\partial \overline{E}}{\partial t} + \nabla\cdot \boldsymbol{c}_g \overline{E} = 0}, \tag{7.38}$$

where $\overline{E} = \overline{KE} + \overline{PE}$. Thus, the flux of energy is equal to the energy times the group velocity, or equivalently the energy in the wave propagates with the group velocity. As we established in chapter 6, this property is a rather general one for wave motion.

## 7.2 SHALLOW WATER WAVES ON FLUID INTERFACES

Let us now generalize our treatment of surface gravity waves to waves that exist on the interface between *two* moving fluids of different densities. The ensuing waves are a simple model of gravity waves that exist in the interior of the atmosphere and, perhaps especially, the ocean, in which we idealize the continuous stratification of the real fluid by supposing that the fluid comprises two (or conceivably more) layers of immiscible fluids of different densities stacked on top of each other. We will consider only the hydrostatic case in which case the layers form a 'stacked shallow water' system. We further limit ourselves to two moving layers; an extension to multiple layers is conceptually if not algebraically straightforward, but it soon becomes easier to treat the continuously stratified case which we do in later sections.

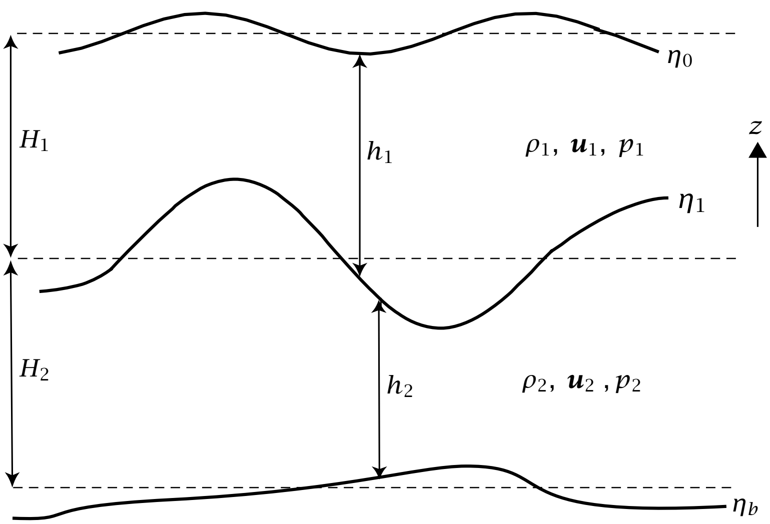


**Fig. 7.2** The two-layer shallow water system. A fluid of density $\rho_1$ lies over a denser fluid of density $\rho_2$.

### 7.2.1 Equations of motion

Consider a two-layer shallow water model as illustrated in Fig. 7.2. From section 3.3 the non-linear momentum equations are, for the upper layer,

$$\frac{\mathrm{D}\boldsymbol{u}_1}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_1 = -g\nabla\eta_0, \tag{7.39a}$$

and in the bottom layer

$$\frac{\mathrm{D}\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_2 = -\frac{\rho_1}{\rho_2}\left(g\nabla\eta_0 + g_1'\nabla\eta_1\right). \tag{7.39b}$$

where $g_1' = g(\rho_2 - \rho_1)/\rho_1$ (we will henceforth drop the subscript 1 and denote this as $g'$), and in the Boussinesq case we take $\rho_1/\rho_2 = 1$. We will only consider the non-rotating case, and after linearization about a resting state we have for the upper and lower layers respectively

$$\frac{\partial \boldsymbol{u}_1'}{\partial t} = -g\nabla\eta_0', \tag{7.40a}$$

$$\frac{\partial \boldsymbol{u}_2'}{\partial t} = -g\nabla\eta_0' - g'\nabla\eta_1'. \tag{7.40b}$$

The equations of motion are completed by the mass continuity equations for each layer, namely

$$\frac{\mathrm{D}}{\mathrm{D}t}(\eta_0 - \eta_1) + h_1\nabla\cdot\boldsymbol{u}_1 = 0 \quad \longrightarrow \quad \frac{\partial}{\partial t}(\eta_0' - \eta_1') + H_1\nabla\cdot\boldsymbol{u}_1' = 0 \tag{7.41a,b}$$

and

$$\frac{\mathrm{D}\eta_1}{\mathrm{D}t} + h_2\nabla\cdot\boldsymbol{u}_2 = 0 \quad \longrightarrow \quad \frac{\partial\eta_1'}{\partial t} + H_2\nabla\cdot\boldsymbol{u}_2' = 0, \tag{7.42a,b}$$

where the two rightmost expressions follow after linearization and we assume that the bottom is flat; that is $\eta_b = 0$. Henceforth we will also omit any primes on the perturbed quantities.

### 7.2.2 Dispersion relation

We first eliminate the velocity from (7.40a) and (7.41b) to give

$$\frac{\partial^2}{\partial t^2}(\eta_0 - \eta_1) - gH_1\nabla^2\eta_0 = 0, \tag{7.43}$$

and similarly for the lower layer:

$$\frac{\partial^2 \eta_1}{\partial t^2} - H_2(g\nabla^2\eta_0 + g'\nabla^2\eta_1) = 0. \tag{7.44}$$

Equations (7.43), and (7.44) form a complete set and in the usual fashion we may look for solutions of the form $\eta_i = \mathrm{Re}\,\tilde{\eta}_i \exp[\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)]$. We obtain

$$(\omega^2 - gH_1K^2)\tilde{\eta}_0 - \omega^2\tilde{\eta}_1 = 0 \tag{7.45a}$$

$$-gH_2K^2\tilde{\eta}_0 + (\omega^2 - g'H_2K^2)\tilde{\eta}_1 = 0. \tag{7.45b}$$

where $K^2 = k^2 + l^2$. For these equations to have non-trivial solutions we must have

$$(\omega^2 - gH_1K^2)(\omega^2 - g'H_2K^2) - \omega^2 gH_2K^2 = 0, \tag{7.46}$$

which, for small $g'/g \ll 1$ gives, after a couple of lines of algebra,

$$\omega^2 = \frac{1}{2}K^2gH \pm \frac{1}{2}K^2gH\sqrt{1 - 4\frac{g'}{g}\frac{H_1H_2}{H^2}} \tag{7.47}$$

$$\approx \frac{1}{2}K^2gH \pm \frac{1}{2}K^2gH\left(1 - 2\frac{g'}{g}\frac{H_1H_2}{H^2}\right). \tag{7.48}$$

where $H = H_1 + H_2$. If $g' = 0$ we recover the familiar single-layer dispersion relation, $\omega = K\sqrt{gH}$ (as well as $\omega = 0$). In the more general case there are two distinct modes:

*(i)* A fast mode with phase speed given by

$$c_p^2 = \left(\frac{\omega}{k}\right)^2 = gH\left(1 - \frac{g'}{g}\frac{H_1H_2}{H^2}\right), \tag{7.49}$$

where, for algebraic simplicity (and, in fact, without loss of generality, since it amounts only to an alignment of our coordinate system), we take $l = 0$. Using (7.45a) we then find that

$$\frac{\eta_0}{\eta_1} \approx \frac{H}{H_2}. \tag{7.50}$$

That is, since $H > H_2$, the displacement of the upper surface is larger than that of the lower. This mode is sometimes called the 'barotropic' mode, for the oscillations are vertically coherent (the phase on the interior surface is the same as that at the surface), and virtually the same oscillation would exist even in the absence of a density jump in the interior.

*(ii)* A slower mode with phase speed given by

$$c_p^2 \approx g' \frac{H_1 H_2}{H}, \tag{7.51}$$

and vertical structure

$$\frac{\eta_0}{\eta_1} \approx \frac{g' H_2}{gH} \ll 1. \tag{7.52}$$

In this case the displacement of the upper surface is smaller than the interior displacement by the ratio of $g'$ to $g$, which in the ocean, where density differences are small, might well be of order 1/100. Furthermore, the internal displacement is *out of phase* with that at the surface. Often, in oceanic situations the interface may be taken as representing the thermocline, in which case $H_2 \gg H_1$ (i.e., the abyss has a greater depth than the thermocline) and $H \approx H_2$. In this case $c_p^2 \approx g' H_1$, and internal waves on the thermocline behave rather like surface waves, but with a weaker restoring force (and consequently a larger amplitude) because the density difference between the two layers of seawater is much smaller than the density difference between the seawater and air above it.

## 7.3 INTERNAL WAVES IN A CONTINUOUSLY STRATIFIED FLUID

We now turn our attention to *internal gravity waves,* namely waves that are internal to a given fluid and that owe their existence to the restoring force of gravity. Interfacial waves are, of course, a model of internal waves with a discontinuous jump in density within the fluid. Surface waves might even be thought of as internal waves if one supposes that part of the fluid has zero density, although this stretches the definition of the word internal somewhat. In this section we will consider the simplest and most fundamental case, that of internal waves in a Boussinesq fluid with constant stratification and no background rotation.

Reprising and extending the material of section 2.9.4, let us consider a continuously stratified Boussinesq fluid, initially at rest, in which the background buoyancy varies only with height and so the buoyancy frequency, $N$, is a function only of $z$. Linearizing the equations of motion about this basic state gives the linear momentum equations,

$$\frac{\partial \boldsymbol{u}'}{\partial t} = -\nabla \phi' \qquad \frac{\partial w'}{\partial t} = -\frac{\partial \phi'}{\partial z} + b', \tag{7.53a,b}$$

the mass continuity and thermodynamic equations,

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0, \qquad \frac{\partial b'}{\partial t} + w' N^2 = 0. \tag{7.54a,b}$$

Our notation is such that $\boldsymbol{u} \equiv u\mathbf{i} + v\mathbf{j}$, $\boldsymbol{v} \equiv u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$, and the gradient operator is two-dimensional unless noted. Thus, $\nabla \equiv \mathbf{i}\,\partial_x + \mathbf{j}\,\partial_y$ and $\nabla_3 \equiv \mathbf{i}\,\partial_x + \mathbf{j}\,\partial_y + \mathbf{k}\partial_z$.

A little algebra gives a single equation for $w'$,

$$\left[\frac{\partial^2}{\partial t^2}\left(\nabla^2 + \frac{\partial^2}{\partial z^2}\right) + N^2 \nabla^2\right] w' = 0. \tag{7.55}$$

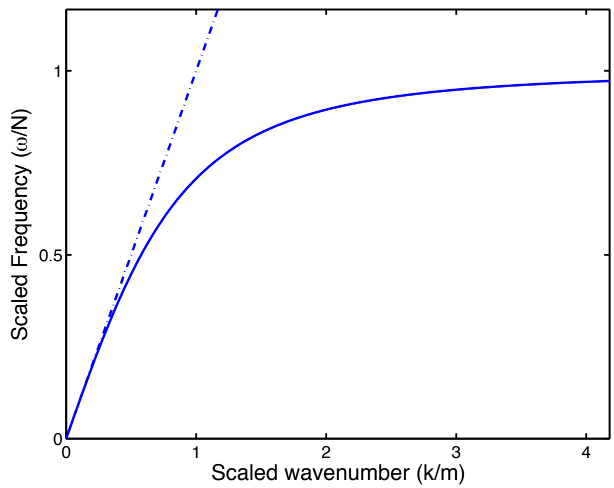


**Fig. 7.3** Scaled frequency, $\omega/N$, plotted as a function of scaled horizontal wavenumber, $k/m$, using the full dispersion relation of (7.57) with $l = 0$ (solid line, asymptoting to unit value for large $k/m$), and with the hydrostatic dispersion relation (7.61) (dashed line, tending to $\infty$ for large $k/m$).

This equation is evidently *not* isotropic. If $N^2$ is a constant — that is, if the background buoyancy varies linearly with $z$ — then the coefficients of each term are constant, and we may then seek solutions of the form

$$w' = \mathrm{Re}\ \widetilde{w} e^{i(kx+ly+mz-\omega t)}, \tag{7.56}$$

where Re denotes the real part, a denotion that will frequently be dropped unless ambiguity arises, and other variables oscillate in a similar fashion. Using (7.56) in (7.55) yields the dispersion relation:

$$\boxed{\omega^2 = \frac{(k^2+l^2)N^2}{k^2+l^2+m^2} = \frac{K^2N^2}{K_3^2}}, \tag{7.57}$$

where $K^2 = k^2 + l^2$ and $K_3^2 = k^2 + l^2 + m^2$. The frequency (see Fig. 7.3) is thus always less than $N$, approaching $N$ for small horizontal scales, $K^2 \gg m^2$. If we neglect pressure perturbations, as in the parcel argument, then the two equations,

$$\frac{\partial w'}{\partial t} = b', \qquad \frac{\partial b'}{\partial t} + w'N^2 = 0, \tag{7.58}$$

form a closed set, and give $\omega^2 = N^2$.

If the basic state density increases with height then $N^2 < 0$ and we expect this state to be unstable. Indeed, the disturbance grows exponentially according to $\exp(\sigma t)$ where

$$\sigma = i\omega = \frac{\pm K\widetilde{N}}{K_3}, \tag{7.59}$$

where $\widetilde{N}^2 \equiv -N^2$. Most convective activity in the ocean and atmosphere is, ultimately, related to an instability of this form, although of course there are many complicating issues — water vapour in the atmosphere, salt in the ocean, the effects of rotation and so forth.

### 7.3.1 Hydrostatic internal waves

Let us now suppose that the fluid satisfies the hydrostatic Boussinesq equations, and for simplicity assume that $l = 0$. The linearized two-dimensional equations of motion become

$$\frac{\partial \boldsymbol{u}'}{\partial t} = -\nabla\phi', \qquad 0 = -\frac{\partial \phi'}{\partial z} + b', \tag{7.60a}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0, \qquad \frac{\partial b'}{\partial t} + w'N^2 = 0, \tag{7.60b}$$

where these are the horizontal and vertical momentum equations, the mass continuity equation and the thermodynamic equation respectively. A little algebra gives the dispersion relation,

$$\omega^2 = \frac{K^2 N^2}{m^2}. \tag{7.61}$$

The frequency and, if $N^2$ is negative, the growth rate, is unbounded as $K^2/m^2 \to \infty$, and the hydrostatic approximation thus has quite unphysical behaviour for small horizontal scales (see also problem 2.11). Many numerical models of the large-scale circulation in the atmosphere and ocean do make the hydrostatic approximation. In these models convection must be *parameterized;* otherwise, it would simply occur at the smallest scale available, namely the size of the numerical grid, and this type of unphysical behaviour should be avoided. Of course in non-hydrostatic models convection must also be parameterized if the horizontal resolution of the model is too coarse to properly resolve the convective scales.

### 7.3.2 Some Properties of Internal Waves

Internal waves have a number of interesting and counter-intuitive properties so let's point a few of them out.

*The dispersion relation*

We can write the dispersion relation, (7.57), as

$$\boxed{\omega = \pm N \cos\vartheta}, \tag{7.62}$$

where $\cos^2\vartheta = K^2/(K^2+m^2)$ so that $\vartheta$ is the angle between the three-dimensional wave-vector, $\boldsymbol{k} = k\mathbf{i} + l\mathbf{j} + m\mathbf{k}$, and the horizontal. The frequency is thus a function only of $N$ and the angle between the vector of propagation, $\boldsymbol{k}_3$ and the horizontal and, if this is given, the frequency is not a function of wavelength. This has some interesting consequences for wave reflection, as we see below.

We can also write the dispersion relation, (7.57), as

$$\frac{\omega^2}{N^2 - \omega^2} = \frac{K^2}{m^2}. \tag{7.63}$$

Thus, and consistently with our first point, given the wave frequency the ratio of the vertical to the horizontal wavenumber is fixed.

*Polarization relations*

If the pressure field is oscillating like $\phi' = \widetilde{\phi}\exp[\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)] = \widetilde{\phi}\exp[\mathrm{i}(kx + ly + mz - \omega t)]$ then, using (7.53a), the horizontal velocity components have the phases

$$(\widetilde{u}, \widetilde{v}) = (k, l)\omega^{-1}\widetilde{\phi} \tag{7.64a}$$

As the frequency is real, the velocities are in phase with the pressure. A little algebra also reveals that the buoyancy perturbation is related to the pressure perturbation by

$$\tilde{b} = \frac{\mathrm{i}mN^2}{N^2 - \omega^2}\widetilde{\phi} = \frac{\mathrm{i}N^2K^2}{m\omega^2}\widetilde{\phi} = \frac{\mathrm{i}K_3^2}{m}\widetilde{\phi}, \tag{7.64b}$$

using the dispersion relation, so that the buoyancy and pressure perturbations are $\pi/2$ out of phase.

The vertical velocity is related to the pressure perturbation by

$$\widetilde{w} = \frac{-\omega m}{N^2 - \omega^2}\widetilde{\phi} = \frac{-K^2}{m\omega}\widetilde{\phi}, \tag{7.64c}$$

where the second expression uses (7.63). The vertical velocity is in phase with the pressure perturbation, and for regions of positive $m$ (and so with upward phase propagation) regions of high relative pressure are associated with downward fluid motion.

The pressure, buoyancy and velocity fields are all real fields and we can write the above phase relationships in terms of sines and cosines as follows.

$$\phi = \Phi_0 \cos(kx + ly + mz - \omega t), \tag{7.65a}$$

$$(u, v) = (k, l)\frac{\Phi_0}{\omega}\cos(kx + ly + mz - \omega t), \tag{7.65b}$$

$$w = \left(\frac{-\omega m}{N^2 - \omega^2} = \frac{-K^2}{m\omega}\right)\Phi_0 \cos(kx + ly + mz - \omega t) \tag{7.65c}$$

$$b = \left(\frac{mN^2}{N^2 - \omega^2} = \frac{N^2K^2}{m\omega^2}\right)\Phi_0 \sin(kx + ly + mz - \omega t), \tag{7.65d}$$

where $\Phi_0$ is a constant. We might equally well have chosen $\phi$ to have a sine dependence in (7.65a), in which case (7.65b,c,d) should be changed appropriately. The relations between pressure, buoyancy and velocity in (7.64) and (7.65) are known as *polarization relations.*

*Relation between wave vector and velocity*

$$\boldsymbol{k}\cdot\widetilde{\boldsymbol{v}} = 0. \tag{7.66}$$

This means that, at any instant, the wave vector is perpendicular to the velocity vector, and the velocity is therefore aligned *along* the direction of the troughs and crests, along which there is no pressure gradient. If the wave vector is purely horizontal (i.e., $m = 0$), then the motion is purely vertical and $\omega = N$.

The vertical and horizontal velocities are related to the wave wavenumbers. Suppose for simplicity (and with little loss of generality) that the motion is all in the $x$-$z$ plane, with $l = 0$ and $v = 0$. Then

$$\frac{\widetilde{u}}{\widetilde{w}} = -\frac{m}{k}. \tag{7.67}$$

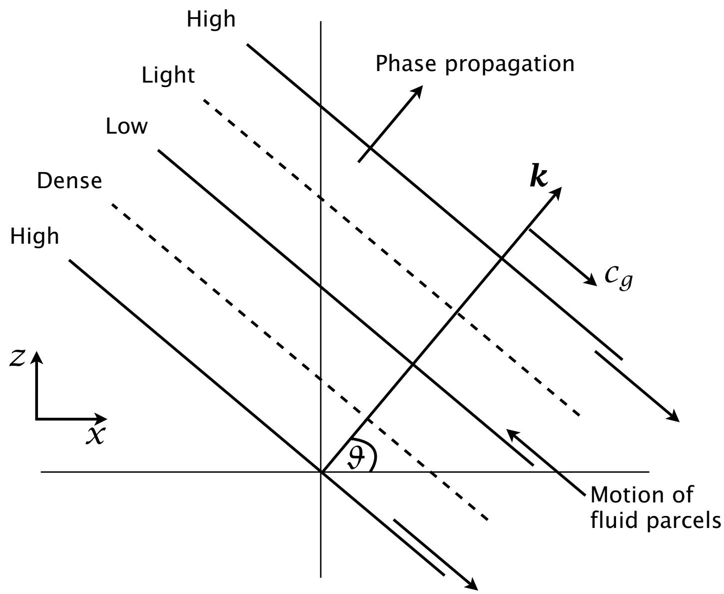


**Fig. 7.4** An internal wave propagating in the direction $\boldsymbol{k}$. If the motion is the the $x$–$z$ plane then both $k$ and $m$ are positive for the wave shown. The solid lines show crests and troughs of constant pressure, and the dashed lines the corresponding crests and troughs of buoyancy (or density). The motion of the fluid parcels in along the lines of constant phase, as shown, and is parallel to the group velocity and perpendicular to the phase speed.

Furthermore, from (7.56) with $l = 0$, at any given instant all of the perturbation quantities in the wave are constant along the lines $kx + mz =$ constant. Thus, *all fluid parcel motions are parallel to the wave fronts.* Now, since the wave frequency is related to the background buoyancy frequency by $\omega = \pm N \cos\vartheta$, it follows that the fluid parcels oscillate along lines that are at an angle $\vartheta = \cos^{-1}(\omega/N)$ to the vertical.

The polarization relations and the group and phase velocities are illustrated in Fig. 7.4. Let us now discuss them, and the figure, in a little more detail.[2]

### 7.3.3 A parcel argument and some physical interpretation

Let us consider first the dispersion relation itself and try to derive it more physically, or at least heuristically. Let us suppose there is a wave propagating in the $(x, z)$ plane at some angle $\vartheta$ to the horizontal, with fluid parcels moving parallel to the troughs and crests, as in Fig. 7.4. In general the restoring force on a parcel is due to both the pressure gradient and gravity, but along the crests there is no pressure gradient. Referring to Fig. 7.5, for a total displacement $\Delta s$ the restoring force in the direction of the particle displacement, $F_{re}$, is

$$\begin{aligned} F_{re} &= g\cos\vartheta \times \Delta\rho = g\cos\vartheta \times \frac{\partial\rho}{\partial z}\Delta z \\ &= g\cos\vartheta \times \frac{\partial\rho}{\partial z}\Delta s\cos\vartheta = \rho_0\frac{\partial b}{\partial z}\cos^2\vartheta\,\Delta s, \end{aligned} \tag{7.68}$$

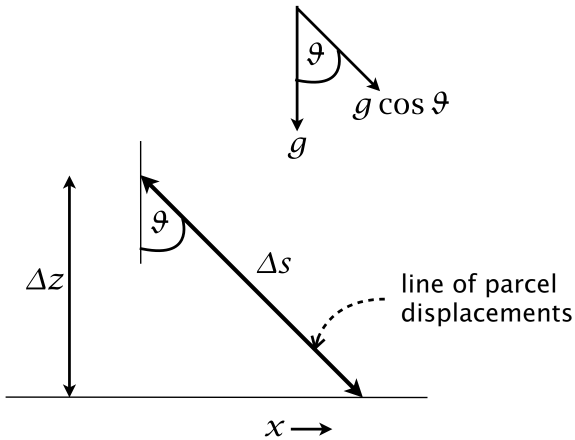


**Fig. 7.5** Parcel displacements and associated forces in a internal gravity wave in which the parcel displacements are occurring at an angle $\vartheta$ to the vertical, as in Fig. 7.4.

noting that $\Delta z = \cos\vartheta\,\Delta s$. The equation of motion of a parcel moving along a trough or crest is therefore

$$\rho_0 \frac{\mathrm{d}^2 \Delta s}{\mathrm{d}t^2} = -\rho_0 N^2 \cos^2\vartheta\,\Delta s, \tag{7.69}$$

which implies a frequency $\omega = N\cos\vartheta$, as in (7.62). One of the $\cos\vartheta$ factors in (7.69) comes from the fact that the parcel displacement is at an angle to the direction of gravity, and the other comes from the fact that the restoring force that a parcel experiences is proportional to $N\cos\vartheta$. (The reader may also wish to refer ahead to Fig. 7.16 and section 7.6.1 for a similar argument.)

Now consider the wave illustrated in Fig. 7.4. For this wave both $k$ and $m$ are positive, and the frequency is assumed positive by convention to avoid duplicative solutions. The slanting solid and dashed lines are lines of constant phase, and from (7.64b) the buoyancy and pressure are 1/4 of a wavelength out of phase. When $k$ and $m$ are both positive the extrema in the buoyancy field lag the extrema in the extrema in the vertical velocity by $\pi/2$, as illustrated. The perturbation velocities are zero along the lines of extreme buoyancy. This follows because the velocities are in phase with the pressure, which as we noted is out of phase with the buoyancy.

Given the direction of the fluid parcel displacement in Fig. 7.4, the direction of the phase propagation $c_p$ up and to the right may be deduced from the following argument. Buoyancy perturbations arise because of vertical advection of the background stratification, $w'\partial b_0/\partial z = w' N^2$. A local maximum in rising motion, and therefore a tendency to increase the fluid density, is present along the 'Low' line 1/4 wavelength upward and to the right of the 'Dense' phase line. Thus, the density of fluid along the 'Low' phase line increases and the 'Dense' phase line moves upward and to the right. If the fluid parcel motion were reversed the pattern of 'High–Dense–Low–Light–High' in Fig. 7.4 would remain the same. However, the downward fluid motion along the 'Low' line would cause the fluid to lose density, and so the phase lines would propagate downward and to the left. Evidently, the wave fronts, or the lines of constant phase, move at right angles to the fluid-parcel trajectories. In the figure we see that the group velocity is denoted as being at right angles to the phase speed, so let's discuss this.

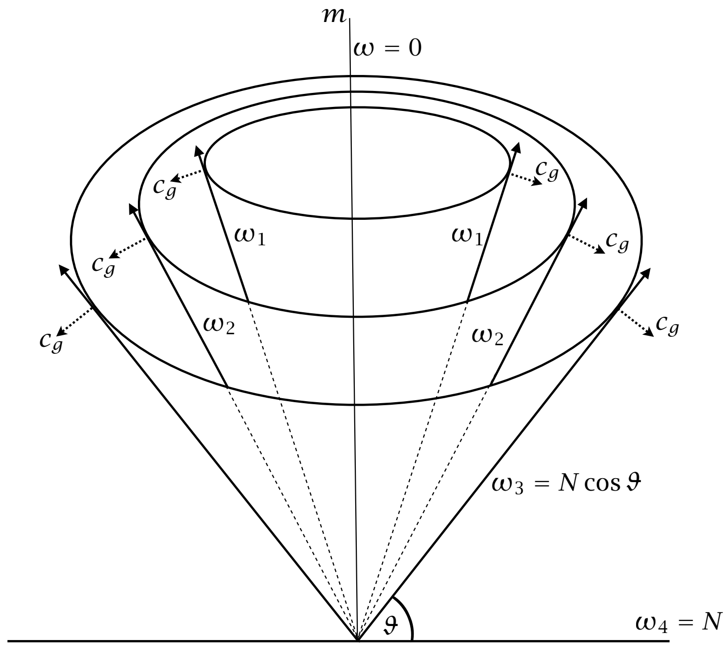


**Fig. 7.6** Internal wave cones. The surfaces of constant frequency are cones, defined by the surface that has a constant angle to the horizontal.The wave vector, and so the phase velocity, points along the cone away from the origin, and the frequency of any wave with a wave vector in the cone is $N \cos \vartheta$. The group velocity is at right angles to the cone and pointed in the direction of increasing frequency, as indicated by the arrows on the dotted lines. In the vertical direction the phase speed and group velocity have opposite signs.

### 7.3.4 Group velocity and phase speed

As we noted above, the frequency of internal waves is given by $\omega = N \cos \vartheta$, where $\vartheta$ is the angle the wave vector makes with the horizontal. This means that the surfaces of constant frequency are *cones,* as illustrated in Fig. 7.6.

To evaluate phase and group velocities in a useful way it is convenient to use spherical polar coordinates, as in Fig. 7.7, in which

$$k = K_3 \cos \vartheta \cos \lambda, \qquad l = K_3 \cos \vartheta \sin \lambda, \qquad m = K_3 \sin \vartheta, \tag{7.70}$$

so that $\boldsymbol{k} = K_3(\cos \vartheta \cos \lambda, \cos \vartheta \sin \lambda, \sin \vartheta)$. The angles are $\vartheta$, the angle of the wave vector with the horizontal and $\lambda$, which determines the orientation in the horizontal plane. (The notation is similar to the spherical coordinates of chapter 2 — see Fig. 2.3 — although here $\vartheta$ is the angle with the horizontal, not the angle with the equatorial plane.) We also note that

$$\sin^2 \vartheta = \frac{m^2}{k^2 + l^2 + m^2}, \qquad \cos^2 \vartheta = \frac{K^2}{K_3^2} = \frac{k^2 + l^2}{k^2 + l^2 + m^2}, \qquad \tan \lambda = \frac{l}{k}. \tag{7.71}$$

In many problems we can align the direction of the wave propagation with the $x$-axis and take $l = 0$ and $\tan \lambda = 0$.

The phase speed of the internal waves in the direction of the wave vector (sometimes referred to as the phase velocity) is given by

$$c_p = \frac{\omega}{K_3} = \frac{N}{K_3}\cos\vartheta = \frac{NK}{K_3^2}. \tag{7.72}$$

The phase speeds (as conventionally-defined) in the $x, y$ and $z$ directions are

$$c_p^x \equiv \frac{\omega}{k} = \frac{N}{k}\cos\vartheta, \qquad c_p^y \equiv \frac{\omega}{l} = \frac{N}{l}\cos\vartheta, \qquad c_p^z \equiv \frac{\omega}{m} = \frac{N}{m}\cos\vartheta. \tag{7.73a,b,c}$$

As noted in section 6.1.2 these quantities are the speed of propagation of the wave crests in the respective directions. In general, each speed is *larger* than the phase speed in the direction perpendicular to the wave crests (that is, in the direction of the wave vector), but no information is transmitted at these speeds.

The group velocity is given by

$$\boldsymbol{c}_g = \left(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial l}, \frac{\partial\omega}{\partial m}\right). \tag{7.74}$$

Using (7.57) we find

$$c_g^x = \frac{\partial\omega}{\partial k} = \frac{Nm}{K_3^2}\frac{km}{KK_3} = \left(\frac{N}{K_3}\sin\vartheta\right)\cos\lambda\sin\vartheta, \tag{7.75a}$$

$$c_g^y = \frac{\partial\omega}{\partial l} = \frac{Nm}{K_3^2}\frac{lm}{KK_3} = \left(\frac{N}{K_3}\sin\vartheta\right)\sin\lambda\sin\vartheta, \tag{7.75b}$$

$$c_g^z = \frac{\partial\omega}{\partial m} = -\frac{Nm}{K_3^2}\frac{K}{K_3} = -\left(\frac{N}{K_3}\sin\vartheta\right)\cos\vartheta. \tag{7.75c}$$

The magnitude of the group velocity is evidently

$$|c_g| = \frac{N}{K_3}\sin\vartheta, \tag{7.76}$$

and the group velocity vector is directed at an angle $\vartheta$ to the vertical, as in Fig. 7.6. This angle is perpendicular to the cone itself; that is, the group velocity is perpendicular to the wave vector, as may be verified by taking the dot product of (7.70) and (7.75) which gives

$$\boldsymbol{k}\cdot\boldsymbol{c}_g = 0. \tag{7.77}$$

The group velocity is therefore parallel to the motion of the fluid parcels, as illustrated in Fig. 7.4. Furthermore, because energy propagates with the group velocity, and the latter is *parallel* to lines of constant phase, energy propagates perpendicular to the direction of phase propagation — very different from the case of acoustic waves or even shallow water waves. In the vertical direction we see from (7.73c) and (7.75c) that

$$\frac{\omega}{m}\frac{\partial\omega}{\partial m} = -\frac{N^2}{K_3^2}\cos^2\vartheta < 0. \tag{7.78}$$

That is, the phase speed and the group velocity have opposite signs, meaning that if the wave crests move downward the group moves upward!

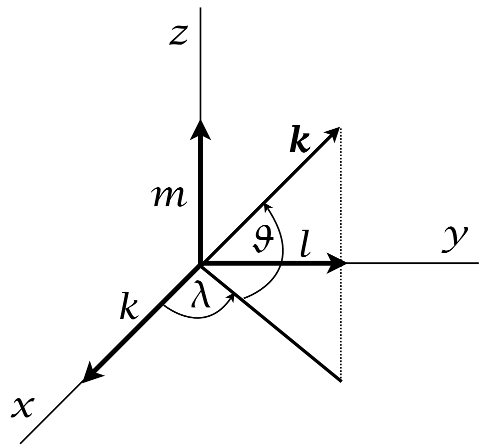


**Fig. 7.7** The spherical coordinates used to describe internal waves, as as in (7.70). The angle $\vartheta$ is the angle of the wave vector with the horizontal, and $\lambda$ determines the orientation in the horizontal plane. The wave vector $\boldsymbol{k}$ is given by $\boldsymbol{k} = (k, l, m)$ in the direction of increasing $(x, y, z)$.

*Effect of a mean flow*

Suppose that there is a mean flow, $U$, in the $x$-direction, as is common in both atmosphere and ocean. The dispersion relation, (7.57), simply becomes

$$(\omega - Uk)^2 = \frac{K^2N^2}{K^2 + m^2}. \tag{7.79}$$

The frequency is Doppler shifted, as expected, but the upward propagation of waves is affected in an interesting way. From (7.79) we find that the vertical component of the group velocity may be written as

$$\frac{\partial \omega}{\partial m} = \frac{-m(\omega - Uk)}{K^2 + m^2} = \frac{-mk(c - U)}{K^2 + m^2}. \tag{7.80}$$

where $c = \omega/k$ is the phase speed in the $x$-direction. If $U$ is not constant but is varying slowly with $z$ then (7.80) still holds, although $m$ itself will also vary slowly with $z$. The point to note is that the group velocity goes to zero at the location where $U = c$, that is at a critical layer and the wave stalls. Of course $m$ itself may become large near a critical layer (as we consider in more detail in section 17.4). In this case — which is essentially the hydrostatic one, with $m^2 \gg K^2$ — we obtain

$$\frac{\partial \omega}{\partial m} = \frac{-k(c - U)}{m} = \frac{-k^2(c - U)^2}{KN}. \tag{7.81}$$

The physical consequence of group velocity going to zero as the wave approaches a critical layer is that any dissipation that may be present has more time to act. That is, we can expect a wave to be preferentially dissipated near a critical layer, giving up its momentum to the mean flow and its energy to create mixing — the former being important in the atmosphere (for this is the mechanism producing the quasi-biennial oscillation) and the latter in the ocean.

### 7.3.5 Energetics of internal waves

In this section we explore the energetics of internal waves, and we first show that the linearized equations conserve a sensible form of energy. Linearized equations do not, of course, automatically conserve energy even if the original nonlinear equations from which they derive do: an unstable wave will draw energy from the background state and grow in amplitude, as we saw in chapter 6 on baroclinic instability.

*Energy Conservation*

To obtain an energy equation we proceed much as in the nonlinear case described in section A2.4.3. From (7.53) we obtain an equation for the evolution of kinetic energy, namely

$$\frac{\partial}{\partial t}\left(\frac{\boldsymbol{v}'^2}{2}\right) = b'w' - \nabla_3 \cdot (\phi' \boldsymbol{v}'), \tag{7.82}$$

where $\boldsymbol{v}'^2 = u'^2 + v'^2 + w'^2$, and from (7.54) we obtain

$$\frac{1}{N^2}\frac{\partial}{\partial t}\frac{b'^2}{2} + w'b' = 0. \tag{7.83}$$

Adding the above two equations gives

$$\boxed{\frac{\partial}{\partial t}\frac{1}{2}\left(\boldsymbol{v}'^2 + \frac{b'^2}{N^2}\right) + \nabla_3 \cdot (\phi' \boldsymbol{v}') = 0}\,. \tag{7.84}$$

This is the linear version of (A2.112). Two differences are apparent: (i) The transport of energy is only by way of the pressure term and the advective transport is absent, as expected in a linear model; (ii) the potential energy term $bz$ of the linear model is replaced by $b'^2/N^2$. It is less obvious why this should be so. However, the quantity

$$A = \frac{1}{2}\int \frac{\overline{b'^2}}{\partial \overline{b}/\partial z}\,\mathrm{d}z\,\mathrm{d}A = \frac{1}{2}\int \frac{\overline{b'^2}}{N^2}\,\mathrm{d}z\,\mathrm{d}A \tag{7.85}$$

is just the *available potential energy* of a Boussinesq fluid in which the isopycnal surfaces vary only slightly from a stable, purely horizontal, resting state (see section A3.10.1).

If we integrate (7.84) over a volume such that the normal component of the velocity vanishes at the boundaries (for example, we integrate over a volume enclosed by rigid walls) then the divergence term vanishes and we obtain the integral conservation statement:

$$\widehat{E} = \frac{1}{2}\int \left(\boldsymbol{v}'^2 + \frac{b'^2}{N^2}\right)\mathrm{d}V, \qquad \frac{\mathrm{d}\widehat{E}}{\mathrm{d}t} = 0. \tag{7.86}$$

The quantity $\widehat{E}$ is an example of a *wave activity:* a conserved quantity that is quadratic in wave amplitude. This conservation statement (7.84) is true whether or not the basic state is stably stratified; that is, whether or not $N^2$ is positive. However, (7.86) only provides a bound on growing perturbations if $N^2$ is positive, in which case all the terms that constitute $\widehat{E}$ are positive definite. If $N^2 < 0$ then both $\boldsymbol{v}'^2$ and $b'^2$ can grow without bound even as $\widehat{E}$ itself remains constant.

Consider now the energy in a *wave*, and we will denote by $\overline{E}$ the energy density, meaning the mean perturbation energy per unit volume, averaged over a wavelength. Thus

$$2\overline{E} = \overline{\boldsymbol{v}'^2} + \frac{\overline{b'^2}}{N^2}. \tag{7.87}$$

If we use the polarization relations of section 7.3.2 then the kinetic and potential energy densities may be written in terms of the pressure amplitude as

$$2\overline{KE} = \left(\frac{k^2}{\omega^2} + \frac{l^2}{\omega^2} + \frac{(k^2+l^2)^2}{m^2\omega^2}\right)|\widetilde{\phi}|^2 = \frac{K^2K_3^2}{m^2\omega^2}|\widetilde{\phi}|^2, \tag{7.88a}$$

$$2\overline{PE} = \frac{N^2K^4}{m^2\omega^4} = \frac{K^2K_3^2}{m^2\omega^2}|\widetilde{\phi}|^2, \tag{7.88b}$$

using also the dispersion relation, $\omega^2K_3^2 = K^2N^2$. Thus, there is *equipartition* between the kinetic and potential energies, a common feature of waves in non-rotating systems (although not a universal feature of waves). The total energy density is thus

$$\overline{E} = \frac{K^2K_3^2}{m^2\omega^2}|\widetilde{\phi}|^2 = \frac{K_3^2}{K^2}|\widetilde{w}|^2 = \frac{|\widetilde{w}|^2}{\cos^2\vartheta}. \tag{7.89}$$

where $\widetilde{w}$ is the amplitude of the vertical component of the velocity perturbation.

### *Energy propagation and the group velocity property*

In section 6.8 we derived, from rather general considerations, the 'group velocity property' for wave activity. We showed that if a wave activity, $A$, and its flux, $\boldsymbol{F}$ obeyed a conservation law of the form $\partial A/\partial t + \nabla\cdot\boldsymbol{F} = 0$, and if the wave activity and its flux were both quadratic functions of the wave amplitude, then the flux is related to the wave activity by $\boldsymbol{F} = \boldsymbol{c}_g A$. The internal wave energy density and its flux do have these properties — see (7.84) — so we should expect the group velocity property to hold, and we now demonstrate that explicitly, albeit briefly.

The energy flux vector for internal waves is

$$\boldsymbol{F} = \overline{\phi'\boldsymbol{v}'} \tag{7.90}$$

and using (7.64a) and (7.64c) this is

$$\boldsymbol{F} = \left(\frac{k}{\omega}, \frac{l}{\omega}, -\frac{K^2}{m\omega}\right)|\widetilde{\phi}|^2. \tag{7.91}$$

Using (7.75) and (7.89) the group velocity times the energy density is

$$c_g^x \times \overline{E} = \left[\frac{Nm^2}{K_3^3}\frac{k}{K}\right]\times\left[\frac{K^2K_3^2}{m^2\omega^2}|\widetilde{\phi}|^2\right] = \frac{k}{\omega}|\widetilde{\phi}|^2, \tag{7.92a}$$

$$c_g^y \times \overline{E} = \left[\frac{Nm^2}{K_3^3}\frac{l}{K}\right]\times\left[\frac{K^2K_3^2}{m^2\omega^2}|\widetilde{\phi}|^2\right] = \frac{l}{\omega}|\widetilde{\phi}|^2, \tag{7.92b}$$

$$c_g^z \times \overline{E} = \left[\frac{NmK}{K_3^3}\right]\times\left[\frac{K^2K_3^2}{m^2\omega^2}|\widetilde{\phi}|^2\right] = -\frac{K^2}{m\omega}|\widetilde{\phi}|^2, \tag{7.92c}$$

which evidently is the same as (7.91), completing our demonstration.

## 7.4 INTERNAL WAVE REFLECTION

Suppose a propagating internal wave encounters a solid boundary — sloping topography, for example. The boundary effectively acts as a source of waves and so the original wave is reflected in some fashion. However, because of the nature of the dispersion relation for internal waves the reflection occurs in a rather peculiar way, as we now discuss.

For algebraic simplicity let us initially suppose that the wave is propagating in the $x$–$z$ plane, and the equation of mass continuity $\partial_x u + \partial_z w = 0$ is then satisfied by introducing a streamfunction $\psi$ such that

$$u = -\frac{\partial \psi}{\partial z}, \qquad w = \frac{\partial \psi}{\partial x}. \tag{7.93}$$

If the incident wave is denoted $\psi_1$ and the reflected wave $\psi_2$ then the total wave field is

$$\psi = \widetilde{\psi}_1 \mathrm{e}^{\mathrm{i}(k_1 x + m_1 z - \omega_1 t)} + \widetilde{\psi}_2 \mathrm{e}^{\mathrm{i}(k_2 x + m_2 z - \omega_2 t)}, \tag{7.94}$$

where as usual a tilde denotes a complex wave amplitude and the real part of the expression is implied. The total streamfunction must be constant *at the boundary* — in fact without loss of generality we may suppose that $\psi = 0$ at the boundary — and this can only be achieved if

$$k_1 x + m_1 z - \omega_1 t = k_2 x + m_2 z - \omega_2 t \tag{7.95}$$

for all $t$ and for all $x$ and $z$ along the boundary. This implies that

$$\omega_1 = \omega_2 \tag{7.96}$$

and

$$k_1 x + m_1 z_b(x) = k_2 x + m_2 z_b(x) \tag{7.97}$$

where $z_b(x)$ parameterizes the height of the reflecting boundary. We can view this another way: suppose that the boundary slopes at an angle $\gamma$ to the horizontal, as in Fig. 7.8 or Fig. 7.9. We then have $z_b = x \tan\gamma$ and a unit vector along the boundary satisfies $\boldsymbol{j}_\gamma = \mathbf{i}\cos\gamma + \mathbf{j}\sin\gamma$. Eq. (7.94) may be written as

$$\psi = \widetilde{\psi}_1 \mathrm{e}^{\mathrm{i}[k_1 + m_1 \tan\gamma)x - \omega_1 t]} + \widetilde{\psi}_2 \mathrm{e}^{\mathrm{i}[(k_2 + m_1 \tan\gamma)x - \omega_2 t]}, \tag{7.98}$$

from which the wavenumber condition that must be satisfied is

$$k_1 + m_1 \tan\gamma = k_2 + m_2 \tan\gamma \tag{7.99}$$

or, and as may also be seen from (7.97),

$$\boldsymbol{k}_1 \cdot \boldsymbol{j} = \boldsymbol{k}_2 \cdot \boldsymbol{j}. \tag{7.100}$$

This means that the components of the wave vector parallel to the boundary for the incoming and outgoing wave are equal to each other. This, and the conservation of frequency expressed by (7.96), are *general* results about wave reflection; they apply to light waves, for example. However, the dispersion relation of internal waves gives rise to rather unintuitive and decidedly non-specular properties of reflection.

### 7.4.1 Properties of internal wave reflection

Suppose an internal wave is incident on a solid boundary, sloping at an angle $\gamma$ to the horizontal, as in Fig. 7.8 or Fig. 7.9. The incident and reflected wave must satisfy the following conditions.

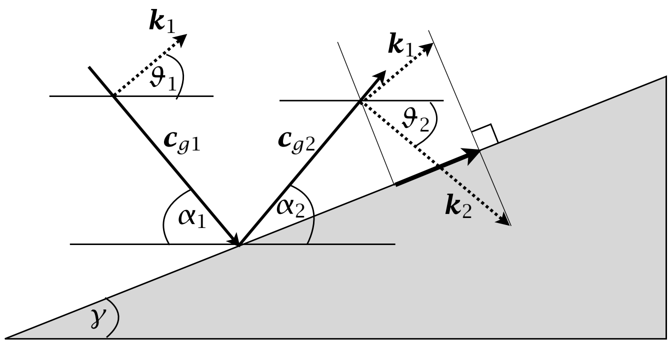


**Fig. 7.8** Internal wave reflection reflection from a shallow sloping boundary. The incoming wave vector, $\boldsymbol{k}_1$, makes an angle $\vartheta_1$ with the horizontal, and the incoming group velocity, $\boldsymbol{c}_{g1}$ makes an angle $\alpha_1 = \pi/2 - \vartheta_1$. The group velocity of the reflected wave, $\boldsymbol{c}_{g2}$ is directed away from the slope, and to satisfy the frequency condition we have $\alpha_2 = \alpha_1$. The projection along the slope of the reflected wave vector, $\boldsymbol{k}_2$ must be equal to that of the incoming wave vector (the projection is the short thick arrow along the slope), and so the magnitude of the reflected wave vector is larger than that incoming wave.

*(i)* The frequency of the reflected wave is equal to that of the incident wave. Because the frequency is given by $\omega = N\cos\vartheta$, the angle of the reflected wave *with respect to the horizontal* is equal to that of the incident wave.

*(ii)* The components of the wave vector along the slope of the reflected wave and incident wave are equal.

*(iii)* The group velocity of the reflected wave must be directed away from the slope.

We did not derive the third of these conditions, but the reflected wave must carry energy and information away from the slope, and these are carried by the group velocity. Similarly, a wave incident on a boundary is one in which the group velocity is directed toward the slope.

Consider a wave approaching a slope as in Fig. 7.8, such that the incoming wave vector makes an angle of $\vartheta_1$ with the horizontal, and the boundary slope is $\gamma$. The condition (7.100) states that the projections along the boundary of the the incoming and outgoing wave vectors are equal to each other, and so

$$\kappa_1 \cos(\vartheta - \gamma) = \kappa_2 \cos(\vartheta + \gamma), \tag{7.101}$$

where $\kappa_1$ and $\kappa_2$ are the magnitudes of the incoming and reflected wave vectors and $\vartheta = \vartheta_1 = \vartheta_2$, because the outgoing wave makes the same angle with the horizontal as does the incoming wave. The group velocity is perpendicular to the wave vector and makes an angle $\alpha = \pi/2 - \vartheta$ to the horizontal, and in terms of this (7.101) may be written, provided $\alpha > \gamma$,

$$\kappa_1 \sin(\alpha + \gamma) = \kappa_2 \sin(\alpha - \gamma). \tag{7.102}$$

The

For a sufficiently steep boundary slope we may have $\alpha < \gamma$, and in this case the wave will be back reflected down the slope, as in Fig. 7.9. A little geometry reveals that the condition (7.102) should be replaced by

$$\kappa_1 \sin(\alpha + \gamma) = \kappa_2 \sin(\gamma - \alpha). \tag{7.103}$$

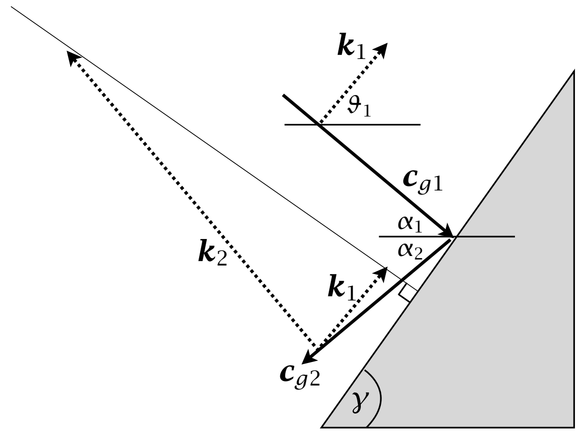


**Fig. 7.9** As for Fig. 7.8, but now showing reflection from a steep slope. The wave is back-reflected down the slope, and in this example the magnitude of the reflected wave is again larger than that of the incoming wave.

The case with $\alpha = \gamma$ is plainly a critical one. In this case the group velocity of the reflected wave is directed along the slope, and the wave vector is perpendicular to the slope. The magnitude of the reflected wave vector is inifinite; that is, the waves are have zero wavelength, and so would in reality be subject to viscous dissipation and diffusion. Reflection of internal waves is in fact an important mechanism leading to mixing in the ocean.

The reflected wave need not, of course, always have a wavenumber that is higher than that of the incident wave: it is a matter of whether the incoming wave vector is more nearly aligned with the slope of the boundary than is the reflected wave, and if it is the reflected wave will have a higher wavenumber, and contrariwise. An example of reflection producing a longer wave is illustrated in Fig. 7.10. Still, the process whereby waves are reflected to produce waves of a shorter wavelength that are then dissipated is an irreversible one, and the net effect of many quasi-random wave reflections is likely to be the dissipation of short waves.

Finally, one might ask why the reflected wave could not simply be back along the track of the incident wave — for example, why could we not have $\boldsymbol{c}_{g1} = -\boldsymbol{c}_{g2}$? If this were so then we would have $\boldsymbol{k}_2 = -\boldsymbol{k}_1$, and it would be impossible for the two wave vectors to project equally on the sloping boundary.

## 7.5 ✦ INTERNAL WAVES IN A FLUID WITH VARYING STRATIFICATION

In most realistic situations the stratification $N^2$ is not constant. In the ocean the stratification is largest in the upper ocean (in the 'pycnocline') diminishing with depth in the weakly stratified abyss. In the atmosphere the stratification tends to be fairly constant in the troposphere but increases fairly abruptly as we pass into the stratosphere. In such circumstances the wave equation (7.55) no longer has constant coefficients and we cannot easily obtain wavelike solutions. However, if the stratification varies *slowly* in the vertical direction, meaning that its variations occur on a larger space scale than the vertical wavelength, while remaining constant in the horizontal direction, then we expect the solution to look locally like plane waves and we can obtain approximate solutions. In this following section we first derive the solution ab initio, using what is essentially WKB theory but without assuming the reader is knowledgeable about the technique. We follow this by a short alternate derivation that directly uses WKB methodology that will be simpler for readers already familiar with the technique or who wish to read

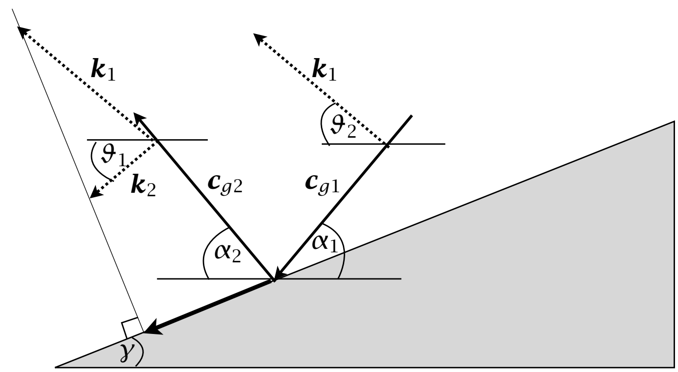


**Fig. 7.10** As for Fig. 7.8, but now showing the production of a reflected wave with a longer wavelength than the incident wave. The wavevector of the reflected wave is more nearly parallel to the sloping boundary than is the wave vector of the incident wave.

the appendix to this chapter.

### 7.5.1 Obtaining the solution

No assumptions are made about the uniformity of $N$ when deriving (7.55), so the equation of motion is again

$$\left[\frac{\partial^2}{\partial t^2}\left(\nabla^2 + \frac{\partial^2}{\partial z^2}\right) + N^2\nabla^2\right] w' = 0. \tag{7.104}$$

where $N^2 = N^2(z)$. We seek solutions in the form

$$w = \mathrm{Re}\, A(z) \mathrm{e}^{i(kx+ly+\chi(z)-\omega t)}, \tag{7.105}$$

where $\chi(z)$ is the vertical phase of the wave and $A(z)$ is its amplitude, which we may take to be real. If $A$ and $\mathrm{d}\chi/\mathrm{d}z$ vary slowly in $z$ (in a manner to be made precise below) then we expect that locally the solution (7.105) will behave like a plane wave, and determining the properties of this wave is our goal.

Substituting (7.105) in (7.104) gives

$$\omega^2\left[K^2 A - \frac{\mathrm{d}^2 A}{\mathrm{d}z^2} + \left(\frac{\partial \chi}{\partial z}\right)^2 A\right] - N^2 K^2 A - \mathrm{i}\omega^2\left[2\frac{\mathrm{d}\chi}{\mathrm{d}z}\frac{\mathrm{d}A}{\mathrm{d}z} + \frac{\mathrm{d}^2\chi}{\mathrm{d}z^2}A\right] = 0 \tag{7.106}$$

or, rearranging,

$$\frac{\mathrm{d}^2 A}{\mathrm{d}z^2} + A\left[\frac{(N^2-\omega^2)K^2}{\omega^2} - m^2\right] - 2\mathrm{i}m^{1/2}\frac{\mathrm{d}}{\mathrm{d}z}\left[m^{1/2}A\right] = 0. \tag{7.107}$$

where $m(z) \equiv \mathrm{d}\chi/\mathrm{d}z$ is the local vertical wavenumber, and the corresponding local vertical wavelength is $2\pi/m$.

Consistent with the small variations of $N^2$, we now assume that $m$ and $A$ vary slowly in the vertical direction, meaning that the vertical scale over which they do vary is much longer than a wavelength itself. For an arbitrary variable $\Phi$ such a condition may be expressed as

$$\left|\frac{1}{\Phi}\frac{\mathrm{d}\Phi}{\mathrm{d}z}\right| \ll m \qquad \text{or} \qquad \left|\frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2}\right| \ll m^2. \tag{7.108a,b}$$

If we apply the second condition to $A$ then the middle term in (7.107) dominates and therefore

$$\boxed{m^2 = \frac{(N^2 - \omega^2)K^2}{\omega^2}}. \tag{7.109}$$

This is an expression for the vertical wavenumber in a medium in which the stratification is varying and the frequency and horizontal wavenumber are known.

A simple rearrangement of (7.109) gives

$$\omega^2 = \frac{N^2K^2}{K^2 + m^2} = N^2 \cos^2 \vartheta(z). \tag{7.110}$$

where $\cos^2\vartheta = K^2/(K^2 + m^2)$ as in (7.71). Equation (7.110) is essentially the same as the dispersion relation for plane waves, as might have been expected given our assumptions. Note that $\omega$ is not a function of $z$, but that $N$ and $\vartheta$ are. Indeed, the expression (7.109) may be thought of as the condition that must be satisfied in order that the frequency satisfy the dispersion relation and be independent of $z$ — as it must be because the medium is time independent (see the discussion in section 6.3). By integrating (7.109) we see that the phase varies according to

$$\chi(z) = \int^z \pm K\left(\frac{N^2 - \omega^2}{\omega^2}\right)^{1/2} \mathrm{d}z'. \tag{7.111}$$

The imaginary part of (7.107) gives

$$\frac{\mathrm{d}}{\mathrm{d}z}(m^{1/2}A) = 0, \tag{7.112}$$

and therefore $A$ varies in the vertical as

$$A(z) = A_0 m^{-1/2}, \tag{7.113}$$

where $A_0$ is a constant. (Equivalently, $A(z) = A(z_0)(m/m_0)^{-1/2}$, where $m_0$ is the wavenumber at $z_0$.) The complete solution thus goes as

$$w = A_0 m^{-1/2} \exp\left(\pm \mathrm{i}\int^z m\,\mathrm{d}z'\right), \tag{7.114}$$

with $m$ given by (7.109).

### *Using WKB theory directly*

The above results can be obtained very quickly if WKB methodology is used from the outset (see the appendix to this chapter). We assume that solutions of (7.104) may be found in the form

$$w' = W(z)\mathrm{e}^{i(kx+ly-\omega t)}, \tag{7.115}$$

whence we obtain the equation of motion

$$\frac{d^2W}{dt^2} + m^2(z)W = 0, \tag{7.116}$$

where $m^2 \equiv (N^2 - \omega^2)K^2/\omega^2$, as in (7.109). The approximate, WKB, solution to (7.116) is (see appendix)

$$W = Cm^{-1/2}\exp\left(\pm i\int m\,dz\right), \tag{7.117}$$

which is equivalent to that of (7.114). We again remark that the phase of the wave is given by $\chi = \int m\,dz$ and $m = d\chi/dz$, so that locally the flow behaves like a plane wave with vertical wavenumber $m$ and with amplitude varying as $m^{-1/2}$.

### 7.5.2 Properties of the solution

The WKB solution above is *almost* that of a plane wave with slowly varying wavenumber. Thus, it seems that the solution (7.114) might be further approximated as

$$w \approx A_0 m^{-1/2}\exp\left(\pm i m(z) z\right) \tag{7.118}$$

where $m(z)$ is given by (7.109). The accuracy of this solution increases as the variation of $m$ diminishes, and in many circumstances (7.118) may be used to infer the qualitative behaviour of a wave. Nonetheless, it is an integral that appears in the phase in the solution (7.114), so the solution is not completely local. The presence of an integral is in fact necessary for the proper interpretation of the wave vector, because the component of the wavevector in the vertical direction ($k^z$ say) is just the vertical derivative of the phase, $\chi$. That is

$$k^z = \frac{d\chi}{dz} = m. \tag{7.119}$$

Thus, the vertical component of the wave vector itself is just $m$ and the wave vector is $(k, l, m)$ as in a plane wave. The solution (7.118), although superficially simpler, does not have this property.

From (7.113) the amplitude $A(z)$ varies with height as $m^{-1/2}$, so that if the stratification ($N^2$) increases $m$ will increase and $A$ will decrease. We have derived this result directly by solving the wave equations of motion, but the result is a consequence of the conservation of energy in internal waves. (Energy is a 'wave activity' — namely a conserved quantity, quadratic in the wave amplitude — in this problem.) As discussed in section 7.3.5, the vertical component of the energy flux, $F^z$, is $c_g^z\overline{E}$, where $\overline{E}$ is the energy density and $c_g^z$ is the vertical component of the group velocity, and for a wave propagating vertically this energy flux must be constant. Now, using (7.75c) and (7.89)

$$c_g^z = -\frac{\omega m}{K_3^2}, \qquad \overline{E} = \left(\frac{A}{\cos\vartheta}\right)^2 \tag{7.120a,b}$$

so that

$$F^z = \frac{A^2\omega m}{K^2} = \text{constant}. \tag{7.121}$$

Thus, because the horizontal wavenumber $K$ is preserved (since there are no inhomogeneities in the horizontal) and the frequency is constant (because the medium itself is not time varying), we must have $A \propto m^{-1/2}$, as in (7.113).

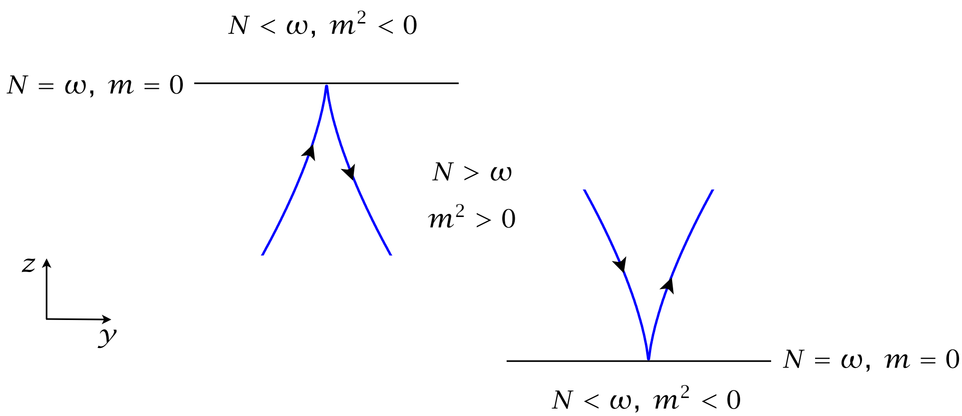


**Fig. 7.11** Trajectories of internal waves approaching a turning height where $N = \omega$. The trajectory makes a cusp, as given by (7.125). If a region of high stratification is sandwiched between two regions of lower stratification then the waves may be vertically confined to a waveguide.

### 7.5.3 Wave trajectories and an idealized example

*Rays*

As we discussed in section 6.3 a wave packet will follow a *ray*, where a ray is simply a trajectory following the group velocity. Restricting attention to two dimensions and using (7.75) the horizontal and vertical components of the group velocity are (for $l > 0$),

$$c_g^y = \frac{Nm^2}{(l^2 + m^2)^{3/2}}, \qquad c_g^z = \frac{-Nlm}{(l^2 + m^2)^{3/2}}. \tag{7.122a,b}$$

The path of a ray may thus be parameterized by the expression

$$\frac{\mathrm{d}z}{\mathrm{d}y} = \frac{c_g^z}{c_g^y} = -\frac{l}{m} = \frac{-\omega}{\sqrt{N^2 - \omega^2}}. \tag{7.123}$$

where the rightmost expression follows from the dispersion relation (7.57) with $k = 0$. The above expressions hold even when $N$ varies in the vertical. Now, for there to be vertical propagation the vertical wavenumber must be positive and the wave frequency must be less than $N$. Suppose a wave is generated in a strongly stratified region and propagates vertically to a more weakly stratified region (with smaller $N$). The vertical wavenumber $m$ becomes smaller and smaller, both the vertical and horizontal components of the group velocity tend to zero and the wave packet will stall. However, $c_g^y$ goes to zero faster than $c_g^z$ and the ray path turns toward the region of lower stratification.

This behaviour may be interpreted in terms of the dispersion relation $\omega = N\cos\vartheta$, where $\vartheta = \cos^{-1}[l^2/(l^2 + m^2)]$ is the angle between the three-dimensional wavevector and the horizontal (see section 7.3.2). If $N$ decreases as we move vertically then $\vartheta$ must decrease until we reach the maximum value of $\cos\vartheta = 1$ and the wave vector is purely horizontal. The group velocity is perpendicular to the wave vector and so is then purely vertical. The wave cannot propagate into the region in which $N^2 < \omega^2$ for then $m$ is imaginary and the disturbance will

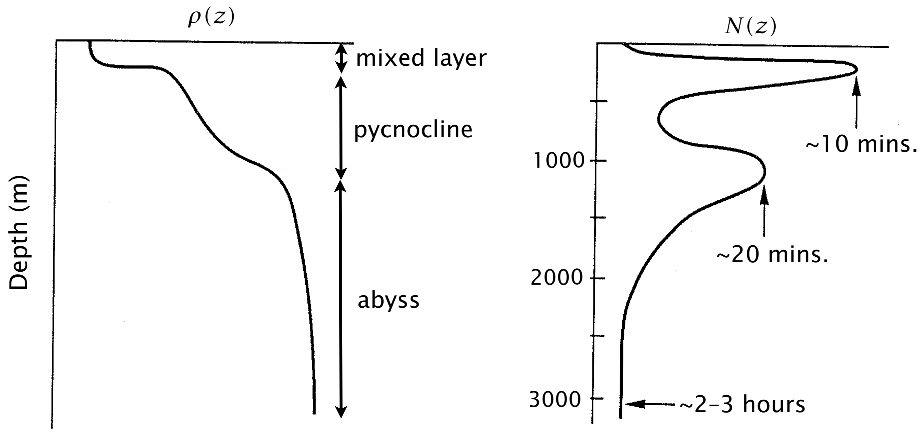


**Fig. 7.12** Schematic of the ocean density, $\rho(z)$ on the left, and corresponding buoyancy frequency $N(z)$ on the right, labelled with approximate buoyancy period. The pycnocline, a region of rapidly changing density, is sandwiched between two weakly stratified, or nearly constant density, regions. The double peak in the buoyancy frequency is exaggerated and seasonal and geographically variable, but the pycnocline is robustly the region of highest frequency internal waves.

decay. Rather, the wave will tend to reflect, and the region where $N = \omega$ is often called a turning level. The trajectory can be obtained analytically in the region of the turning level as follows. Suppose that $N = \omega$ at $z = z^*$ so that, expanding $N^2$ around that point, we have $N^2(z) \approx N^2(z^*) + (z - z^*)\mathrm{d}N^2(z^*)/\mathrm{d}z$. Eq. (7.123) becomes

$$\frac{\mathrm{d}z}{\mathrm{d}y} = \frac{-\omega}{\sqrt{(z - z^*)\mathrm{d}N^2/\mathrm{d}z}}. \tag{7.124}$$

which, upon integrating, yields

$$z - z^* = \frac{\omega(y^* - y)^{2/3}}{\sqrt{\mathrm{d}N^2/\mathrm{d}z}}. \tag{7.125}$$

This cusp-like trajectory is illustrated in Fig. 7.11.

*An idealized oceanic waveguide*

The stratification of the ocean is decidedly nonuniform in the vertical, as schematically illustrated in Fig. 7.12. The density is almost uniform in a layer at the top of the ocean about 50–100 m deep known as the mixed layer. The density then increases fairly rapidly over a region 500-1000 m deep known as the pycnocline, and is then fairly uniform in the abyss. The weak stratification in the abyss and in the mixed layer will inhibit the propagation of internal waves generated in thermocline. For example, consider a wave of frequency $\omega$ propagating downwards from the oceanic thermocline with and into the weakly stratified abyss. As soon as $N(z) < \omega$ the vertical wavenumber becomes imaginary and disturbance will vary like $\mathrm{e}^{\pm mz}$. On physical grounds we must choose the solution that evanesces with depth. Similar behaviour will occur for a wave propagating up from the thermocline into the weakly stratified mixed layer.

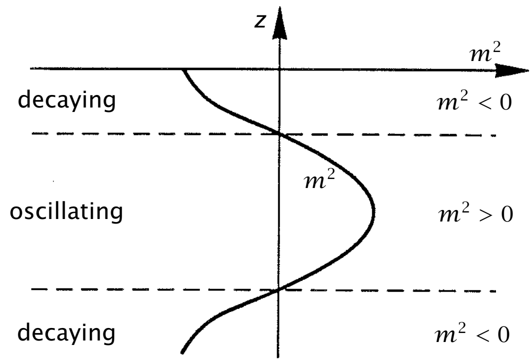


**Fig. 7.13** An oceanic wave guide. A maximum in the vertical density gradient will, using (7.109), give rise to a corresponding maximum $m^2$, as schematically illustrated. Waves generated in the central region will have a positive value of $m^2$, but $m^2$ will become negative in the weakly stratified regions above and below where the disturbance will evanesce, confining the propagating disturbance to the central wave guide.

Thus, waves are trapped in a region where $N^2 > \omega^2$, and this region forms a *wave guide*, as sketched in Fig. 7.13. Essentially the same dynamics are described again in an atmospheric context below.

A specific example is illustrated in Fig. 7.14. The profile of $N^2$ is a simple exponential and the corresponding value of $m^2$ is calculated using (7.109) with $K = \omega = 1$. (the values are nondimensional; the reader is invited to 're-dimensionalize'). The value of $m$ goes to zero near the top and the bottom of the domain, as illustrated. The corresponding group velocities are illustrated in Fig. 7.15, and can be seen to be purely vertical at the two turning heights. The amplitude of a wave becomes very large near the turning heights, but the wave itself need not break because its energy is constant and its vertical wavelength is very large. Rather, the wave will be reflected (following the trajectory illustrated in Fig. 7.11), and the wave is confined in the waveguide.

### 7.5.4 Atmospheric considerations

The atmosphere differs from the ocean in many ways, but for the purposes of internal waves two of these are particularly important: (i) The density diminishes in the vertical and so the Boussinesq approximation is not valid, except for small vertical displacements; (ii) there is no upper surface, so we must consider radiation conditions for large $z$, or require that the solutions

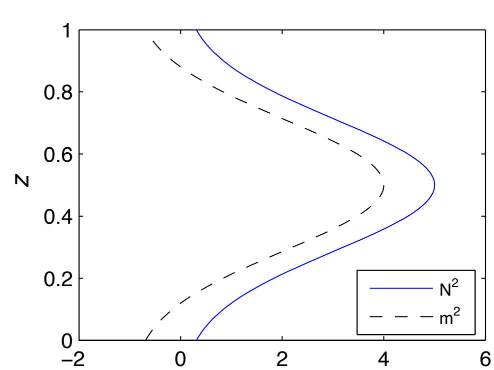


**Fig. 7.14** The value of $N^2$ and $m^2$ giving rise to an idealized oceanic waveguide. The value of $m^2$ is calculated using (7.109)

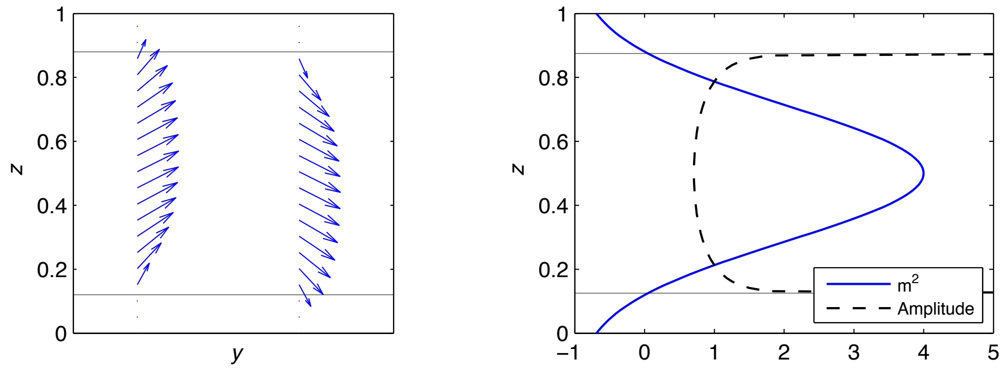


**Fig. 7.15** Left panel: Group velocity vectors for upward and downward propagating gravity waves in a stratification illustrated in Fig. 7.14, calculated using (7.122). Right panel: The values of $m^2$ and the amplitude of the wave, the latter varying as $m^{-1/2}$. The thin horizontal lines in both panels indicate the height at which $m^2 = 0$.

remain bounded for $z \to \infty$, rather than conventional boundary conditions.

There are two common ways to deal with density variations, namely through the use of pressure coordinates or the anelastic equations. In many ways they are equivalent, and we will use pressure coordinates in chapter 17. The anelastic approximation (see section 2.5) differs from the Boussinesq approximation primarily in the mass continuity equation, which is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{\rho_0}\frac{\partial}{\partial z}(w\rho_0) = 0 \tag{7.126}$$

where $\rho_0 = \rho_0(z)$ is a specified profile of density. (Also, for an ideal gas the buoyancy is given by $b = g\delta\theta/\theta_0$ where $\theta_0$ is a constant.) Using (7.126) instead of (7.54a) gives the equation of motion

$$\frac{\partial^2}{\partial t^2}\left(\nabla^2 w' + \frac{\partial}{\partial z}\frac{1}{\rho_0}\frac{\partial \rho_0 w'}{\partial z}\right) + N^2\nabla^2 w' = 0, \tag{7.127}$$

in place of (7.55).[3] Because $\rho_0$ is a function of $z$ we cannot find plane wave solutions without additional approximation — for example unless we assume that $\rho_0$ changes only slowly with $z$. For this reason the Boussinesq approximation is often imposed from the outset in theoretical work, even in atmospheric situations and generally this changes quantitative but often not the qualitative character of the waves.

The second factor (the lack of an upper surface) becomes an issue when considering gravity waves propagating high into the atmosphere, a phenomena we look at in chapter 17. In section 7.7 we consider the generation of internal waves by flow over topography, a phenomena of particular atmospheric importance. To finish this section off, let us consider an atmospheric waveguide. The dynamics are very similar to those of the oceanic waveguide discussed above, but, partly for the sake of variety, we will treat it in a slightly different way.

*An atmospheric waveguide*

Let us suppose the atmosphere to be a semi-infinite region from the ground at $z = 0$ to infinity. If $N^2$ is constant then solutions, as in the bounded case, vary sinusoidally in $z$, for example

$w' \sim \sin mz$, where $m$ is the vertical wavenumber. These solutions remain bounded as $z \to \infty$, although they do not decay. If $N$ varies, then other possibilities exist. Suppose that a region of small stratification, $N_1$ overlies a region of larger stratification, $N_2$; that is

$$N = \begin{cases} N_1 & z > H, \\ N_2 & 0 < z < H. \end{cases} \tag{7.128}$$

where $N_2 > N_1$. (This is *not* a model of the stratosphere overlying the troposphere, because the stratosphere is highly stratified. If anything, it is a model of the mesosphere overlying the stratosphere and troposphere.) The frequency in the two regions must be the same and if $\omega < N_1 < N_2$ then

$$\omega^2 = \frac{N_1^2}{K^2 + m_1^2} = \frac{N_2^2}{K^2 + m_2^2}, \tag{7.129}$$

whence

$$m_1 = m_2 \left( \frac{N_1^2 - \omega^2}{N_2^2 - \omega^2} \right)^{1/2}. \tag{7.130}$$

In contrast, if $N_1 < \omega < N_2$ then wave-like solutions are not allowed in the upper region, because the frequency must always be less than the local value of $N$. Rather, solutions in the upper region evanesce according to

$$w_1' = \widetilde{w}_1 \mathrm{e}^{-\mu z} \mathrm{e}^{\mathrm{i}(kx+ly-\omega t)}, \tag{7.131}$$

where

$$\mu^2 = \frac{\omega^2 - N_1^2}{\omega^2} K^2. \tag{7.132}$$

The solutions still vary sinusoidally in the lower layer, according to

$$w_2' = \widetilde{w}_2 \sin m_1 z \mathrm{e}^{\mathrm{i}(kx+ly-\omega t)}, \tag{7.133}$$

where $m$ now takes on only discrete values in order to satisfy the boundary conditions that $w$ and $\phi$ are continuous $z = H$, and that $w$ vanishes at $z = 0$.

[Expand this discussion a little? xxx]

## 7.6 INTERNAL WAVES IN A ROTATING FRAME OF REFERENCE

In the presence of both a Coriolis force and stratification a displaced fluid will feel two restoring forces — that due to gravity and that due to rotation. The first effect gives rise to gravity waves, as we have discussed, and the second to inertial waves. When the two forces are together simultaneously the resulting waves are called *inertia-gravity waves*. The algebra describing them can be complicated so let us begin with a simpler parcel argument to try to lay bare the basic dynamics; the reader may also wish to first refer back to section 7.3.3.

### 7.6.1 A parcel argument

Consider a parcel that is displaced along a slantwise path in the $x$-$z$ plane, as shown in Fig. 7.16, with a horizontal displacement of $\Delta x$ and a vertical displacement of $\Delta z$. Let us suppose

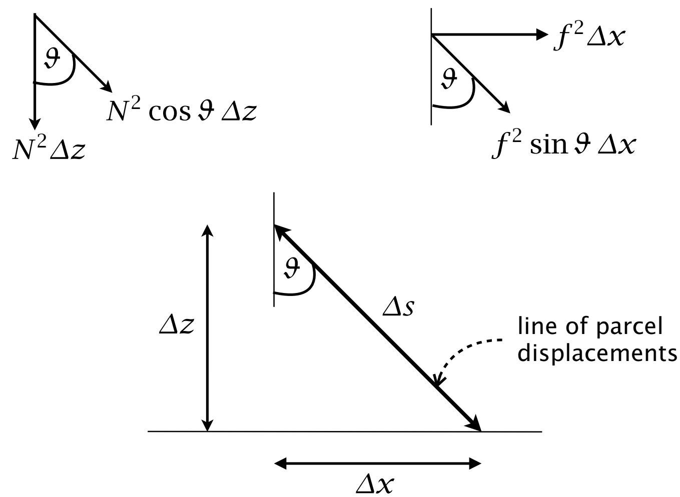


**Fig. 7.16** Parcel displacements and associated forces in an inertia-gravity wave in which the parcel displacements are occurring at an angle $\vartheta$ to the vertical. Both Coriolis and buoyancy forces are present, and $\Delta s = \Delta z/\cos\vartheta = \Delta x/\sin\vartheta$.

that the fluid is Boussinesq and that there is a stable and uniform stratification given by $N^2 = -g\rho_0^{-1}\partial\rho_0/\partial z = \partial b/\partial z$. Referring to (7.68) as needed, the component of the restoring buoyancy force, $F_b$ say, in the direction of the parcel oscillation is given by (7.68)

$$F_b = -N^2 \cos\vartheta\,\Delta z = -N^2\cos^2\vartheta\,\Delta s. \tag{7.134}$$

The parcel will also experience a restoring Coriolis force, $F_C$, and the component of this in the direction of the parcel displacement is

$$F_C = -f^2 \sin\vartheta\,\Delta x = -f^2\sin^2\vartheta\,\Delta s. \tag{7.135}$$

Using (7.134) and (7.135) the (Lagrangian) equation of motion for a displaced parcel is

$$\frac{\mathrm{d}^2\Delta s}{\mathrm{d}t^2} = -(N^2\cos^2\vartheta + f^2\sin^2\vartheta)\Delta s, \tag{7.136}$$

and hence the frequency is given by

$$\omega^2 = N^2\cos^2\vartheta + f^2\sin^2\vartheta. \tag{7.137}$$

Now, nearly everywhere in both atmosphere and ocean, $N^2 > f^2$. From (7.137) we then see that the frequency lies in the interval $N^2 > \omega^2 > f^2$. (To see this, put $N = f$ or $f = N$ in (7.137), and use $\sin^2\vartheta + \cos^2\vartheta = 1$.) If the parcel displacements approach the vertical then the Coriolis force diminishes and $\omega \to N$, and similarly $\omega \to f$ as the displacements become horizontal. The ensuing waves are then pure inertial waves.

We can write (7.137) in terms of wavenumbers if we note that, in the $x$-$z$ plane,

$$\cos^2\vartheta = \frac{k^2}{k^2+m^2}, \qquad \sin^2\vartheta = \frac{m^2}{k^2+m^2} \tag{7.138}$$

where $k$ and $m$ are the horizontal and vertical wavenumbers. The dispersion relation becomes

$$\omega^2 = \frac{N^2k^2 + f^2m^2}{k^2 + m^2}. \tag{7.139}$$

Let's now move on to a discussion using the linearized equations of motion.

### 7.6.2 Equations of motion

In a rotating frame of reference, specifically on an $f$-plane, the linearized equations of motion are the momentum equations

$$\frac{\partial \boldsymbol{u}'}{\partial t} + \boldsymbol{f}_0 \times \boldsymbol{u}' = -\nabla\phi' \qquad \frac{\partial w'}{\partial t} = -\frac{\partial \phi'}{\partial z} + b', \tag{7.140a,b}$$

and the mass continuity and thermodynamic equations,

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0, \qquad \frac{\partial b'}{\partial t} + w'N^2 = 0. \tag{7.140c,d}$$

These are similar to (7.53) and (7.54), with the addition of a Coriolis term in the horizontal momentum equations.

To obtain a single equation for $w'$ we take the horizontal divergence of (7.140a) and use the continuity equation to give

$$\frac{\partial}{\partial t}\left(\frac{\partial w'}{\partial z}\right) + f_0\zeta' = \nabla^2\phi' \tag{7.141}$$

where $\zeta' \equiv (\partial v'/\partial x - \partial u'/\partial y)$ is the vertical component of the vorticity. We may obtain an evolution equation for that vorticity by taking the curl of (7.140a), giving

$$\frac{\partial \zeta'}{\partial t} = f_0\frac{\partial w'}{\partial z}. \tag{7.142}$$

Eliminating vorticity between these equations gives

$$\left(\frac{\partial^2}{\partial t^2} + f_0^2\right)\frac{\partial w'}{\partial z} = \frac{\partial}{\partial t}\nabla^2\phi'. \tag{7.143}$$

We may obtain another equation linking pressure and vertical velocity by eliminating the buoyancy between (7.140b) and (7.140d), so giving

$$\frac{\partial^2 w'}{\partial t^2} + N^2 w' = -\frac{\partial}{\partial t}\frac{\partial \phi'}{\partial z}. \tag{7.144}$$

Eliminating $\phi'$ between (7.143) and (7.144) gives a single equation for $w'$ analogous to (7.55), namely

$$\boxed{\left[\frac{\partial^2}{\partial t^2}\left(\nabla^2 + \frac{\partial^2}{\partial z^2}\right) + f_0^2\frac{\partial^2}{\partial z^2} + N^2\nabla^2\right]w' = 0}. \tag{7.145}$$

If we assume a time dependence of the form $w' = \widehat{w}e^{-i\omega t}$, this equation may be written in the sometimes useful form,

$$\frac{\partial^2\widehat{w}}{\partial z^2} = \left(\frac{N^2 - \omega^2}{\omega^2 - f_0^2}\right)\nabla^2\widehat{w}. \tag{7.146}$$

### 7.6.3 Dispersion Relation

Assuming wave solutions to (7.145) of the form

$$w' = \widetilde{w}e^{i(kx+ly+mz-\omega t)}, \tag{7.147}$$

we readily obtain the dispersion relation

$$\boxed{\omega^2 = \frac{f_0^2 m^2 + (k^2 + l^2)N^2}{k^2 + l^2 + m^2}}, \tag{7.148}$$

We can also write the dispersion relation as

$$\omega^2 = f_0^2 \sin^2 \vartheta + N^2 \cos^2 \vartheta, \tag{7.149}$$

or

$$\omega^2 = f_0^2 + (N^2 - f_0^2)\cos^2 \vartheta, \qquad \text{or} \qquad \omega^2 = N^2 - (N^2 - f_0^2)\sin^2 \vartheta, \tag{7.150}$$

where $\vartheta$ is the angle of the wavevector with the horizontal. The frequency therefore lies between $N$ and $f_0$. The waves satisfying (7.148) are sometimes called inertia-gravity and are analogous to surface gravity waves in a rotating frame — Poincaré waves — discussed in section 3.7.2.

In most atmospheric and oceanic situations $f_0 < N$ (in fact typically $N/f_0 \sim 100$, the main exception being weakly stratified near-surface mixed layers in the ocean) and $f < \omega < N$. From (7.149) the frequency is dependent only on the angle the wavevector makes with the horizontal, and the surfaces of constant frequency again form cones in wavenumber space, although depending on the values of $f$ and $\omega$ the frequency does not necessarily decrease monotonically with $\vartheta$ as in the non-rotating case. For reference, the group velocity is

$$c_g^x = \left[\frac{N^2 - f_0^2}{\omega K_3^3} Km\right] \frac{km}{K\kappa} = \left[\frac{N^2 - f_0^2}{\omega\kappa} \cos\vartheta \sin\vartheta\right] \cos\lambda \sin\vartheta, \tag{7.151a}$$

$$c_g^y = \left[\frac{N^2 - f_0^2}{\omega K_3^3} Km\right] \frac{lm}{K\kappa} = \left[\frac{N^2 - f_0^2}{\omega\kappa} \cos\vartheta \sin\vartheta\right] \sin\lambda \sin\vartheta, \tag{7.151b}$$

$$c_g^z = -\left[\frac{N^2 - f_0^2}{\omega K_3^3} Km\right] \frac{K}{\kappa} = -\left[\frac{N^2 - f_0^2}{\omega\kappa} \cos\vartheta \sin\vartheta\right] \cos\vartheta, \tag{7.151c}$$

[check use of $\kappa$ in above equations xxx]
which reduces to (7.75) if $f = 0$ and in which case $\omega = N\cos\vartheta$. Notice that the directional factors — the terms outside of the square brackets in (7.151) — are the same as those in (7.75). Thus, the group velocity is, as in the non-rotating case, at an angle $\vartheta$ to the vertical, or $\alpha = \pi/2 - \vartheta$ to the horizontal. The magnitude of the group velocity is now given by

$$|\boldsymbol{c}_g| = \frac{N^2 - f_0^2}{\omega K_3^3} Km = \frac{N^2 - f_0^2}{\omega\kappa} \cos\vartheta \sin\vartheta. \tag{7.152}$$

There are a few notable limits:

*(i)* A purely horizontal wave vector. In this case $m = 0$ and $\omega = N$. The waves are then unaffected by the Earth's rotation. This is because the Coriolis force is (in the $f$-plane approximation) due to the product of the Coriolis parameter and the horizontal component of the velocity. If the wave vector is horizontal, the fluid velocities are purely vertical and so the Coriolis force vanishes.

*(ii)* A purely vertical wave vector. In this case $\omega = f_0$. In this case the fluid velocities are horizontal and the fluid parcels do not feel the stratification. The oscillations are then known as *inertial waves*, although they are not inertial in the sense of there being no implied force in an inertial frame of reference.

*(iii)* In the limit $N \to 0$ we have pure inertial waves with a frequency $0 < \omega < f_0$, and specifically $\omega = f_0 \sin\vartheta$. Similarly, as $f_0 \to 0$ we have pure internal waves, as discussed previously, with $\omega = N\cos\vartheta$.

*(iv)* The hydrostatic limit, which we discuss below.

### *The hydrostatic limit*

Hydrostasy occurs in the limit of large horizontal scales, $k, l \ll m$. If we therefore neglect $k^2$ and $l^2$ where they appear with $m^2$ in (7.148) we obtain

$$\omega^2 = f_0^2 + N^2\frac{k^2 + l^2}{m^2} = f_0^2 + N^2\cos^2\vartheta. \tag{7.153}$$

where the rightmost expression arises from (7.149) if we take

$$\sin^2\vartheta = \frac{m^2}{k^2 + l^2 m^2} \to 1, \qquad \cos^2\vartheta = \frac{K^2}{k^2 + m^2} \to \frac{K^2}{m^2} \ll 1, \tag{7.154}$$

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

If we make the hydrostatic approximation from the outset in the rotating, linearized, equations of motion then we have

$$\frac{\partial u'}{\partial t} - f_0 v = -\frac{\partial \phi'}{\partial x}, \qquad \frac{\partial v'}{\partial t} + f_0 u = -\frac{\partial \phi'}{\partial y}, \qquad 0 = -\frac{\partial \phi'}{\partial z} + b', \tag{7.155a}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0, \qquad \frac{\partial b'}{\partial t} + w' N^2 = 0. \tag{7.155b}$$

This reduces to the single equation

$$\left[\frac{\partial^2}{\partial t^2}\frac{\partial^2}{\partial z^2} + f_0^2\frac{\partial^2}{\partial z^2} + N^2\nabla^2\right] w' = 0, \tag{7.156}$$

and corresponding dispersion relation

$$\omega^2 = \frac{f_0^2 m^2 + K^2 N^2}{m^2} = f_0^2 + N^2\alpha'^2, \tag{7.157}$$

so recovering (7.153). This is sometimes known as the *rapidly rotating regime*.

It is notable that the Coriolis parameter $f$ now appears in isolation, and simply provides inertial oscillations that are independent of the wavenumber and the stratification. The group velocity is therefore completely independent of the background rotation.

[the following still needs clarifying xxx]

To make the small aspect ratio limit explicit let us define

$$\alpha' \equiv \frac{\text{vertical scale}}{\text{horizontal scale}} = \frac{K}{m} = \frac{1}{\tan\vartheta} \ll 1, \tag{7.158}$$

and using (7.151) and (7.150) (respectively) this evidently is

$$\alpha' = \frac{c_g^z}{c_g^h} = \frac{\omega^2 - f_0^2}{N^2 - \omega^2}. \tag{7.159}$$

where $c_g^h = (c_g^{x2} + c_g^{y2})^{1/2}$. If $f$ is small, then the hydrostatic limit corresponds to $N^2 \gg \omega^2$ with dispersion relation

$$\omega^2 \approx N^2 \cos^2\vartheta \approx N^2 \frac{K^2}{m^2}. \tag{7.160}$$

This is the same as the dispersion relation in the non-rotating, hydrostatic case derived earlier *ab initio,* giving (7.61) and (A2.252). [Check xxx]

We return to the hydrostatic limit in section 17.2 on gravity waves in the stratosphere.

### 7.6.4 Polarization relations

Just as in the non-rotating case we can derive phase relations between the various fields, useful if we are trying to identify internal waves from observations. As for all waves in an incompressible fluid, the condition $\nabla_3 \cdot \boldsymbol{v} = 0$ gives

$$\boldsymbol{k} \cdot \boldsymbol{v}' = 0, \tag{7.161}$$

so that the fluid motion is in the plane that is perpendicular to the wave vector. The derivations of the other polarization relations are left as excercies for the reader, and the relations are found to be

$$\widetilde{u} = \frac{k\omega + \mathrm{i}lf_0}{\omega^2 - f_0^2}\widetilde{\phi}, \qquad \widetilde{v} = \frac{l\omega - \mathrm{i}kf_0}{\omega^2 - f_0^2}\widetilde{\phi}, \tag{7.162a,b}$$

which should be compared with (7.64a). We also have a relation between buoyancy and pressure,

$$\tilde{b} = \frac{\mathrm{i}mN^2}{N^2 - \omega^2}\widetilde{\phi} \tag{7.163}$$

and one between vertical velocity and pressure,

$$\widetilde{w} = \frac{-m\omega}{N^2 - \omega^2}\widetilde{\phi} = \frac{-\omega K_3^2}{(N^2 - f_0^2)m}\widetilde{\phi}, \tag{7.164}$$

with the second equality following with use of the dispersion relation.

### 7.6.5 Geostrophic motion and vortical modes

If we seek *steady* solutions to (7.140) and (7.140c,d), the equations of motions become

$$-f_0 v = -\frac{\partial\phi'}{\partial x}, \qquad f_0 u = -\frac{\partial\phi'}{\partial y}, \qquad 0 = -\frac{\partial\phi'}{\partial z} + b', \tag{7.165a,b}$$

and

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0, \qquad w'N^2 = 0. \tag{7.166a,b}$$

These are the equations of geostrophic and hydrostatic balance, with zero vertical velocity. What is the dispersion relation corresponding to this?

If instead of eliminating pressure between (7.143) and (7.144) we eliminate vertical velocity we obtain

$$\frac{\partial}{\partial t}\left[\frac{\partial^2}{\partial t^2}\left(\nabla^2+\frac{\partial^2}{\partial z^2}\right)+f_0^2\frac{\partial^2}{\partial z^2}+N^2\nabla^2\right]\phi'=0, \tag{7.167}$$

which is similar to (7.145), except for the extra time derivative, which allows for the possibility of a solution with $\omega = 0$. If $\omega \neq 0$ then

$$\left[\frac{\partial^2}{\partial t^2}\left(\nabla^2+\frac{\partial^2}{\partial z^2}\right)+f_0^2\frac{\partial^2}{\partial z^2}+N^2\nabla^2\right]\phi'=0, \tag{7.168}$$

and the dispersion relation is given by (7.148). If $\omega = 0$, then the quantity in square brackets in (7.167) may not be a function of time; that is

$$\left[\frac{\partial^2}{\partial t^2}\left(\nabla^2+\frac{\partial^2}{\partial z^2}\right)+f_0^2\frac{\partial^2}{\partial z^2}+N^2\nabla^2\right]\phi'=\chi(x,y,z), \tag{7.169}$$

where $\chi$ is a function of space, but not time, and so determined by the initial conditions of $\phi'$. When $\omega \neq 0$, then $\chi = 0$. What is $\chi$? We shall see that it is nothing but the potential vorticity of the flow!

*Potential vorticity*

Recall the vorticity equation and the buoyancy equation, namely

$$\frac{\partial \zeta'}{\partial t}=f_0\frac{\partial w'}{\partial z}, \qquad \frac{\partial b'}{\partial t}+w'N^2=0. \tag{7.170a,b}$$

If we eliminate $w'$ from these equations we obtain

$$\frac{\partial q}{\partial t}=0 \tag{7.171}$$

where

$$q=\left[\zeta'+f_0\frac{\partial}{\partial z}\left(\frac{b'}{N^2}\right)\right] \tag{7.172}$$

is the potential vorticity for this problem. In general, for adiabatic flow potential vorticity is conserved on fluid parcels and $\mathrm{D}Q/\mathrm{D}t = 0$ where for a Boussinesq fluid $Q = \omega_a \cdot \nabla b$. There are two differences between this general case and ours; first, because we have linearized the dynamics the advective term is omitted, and $\partial q/\partial t = 0$. Second, $q$ is not exactly the same as $Q$, but it is an approximation to it valid when the stratification is dominated by its background value, $N^2$. Very informally, we have then, for constant $N$,

$$\begin{aligned} Q=(\boldsymbol{\omega}+\boldsymbol{f}_0)\cdot\nabla b &\approx (\zeta+f_0)\left(N^2+\frac{\partial b'}{\partial z}\right) \\ &\approx f_0N^2+f_0\frac{\partial b'}{\partial z}+\zeta N^2=N^2\left[f_0+\zeta+f_0\frac{\partial}{\partial z}\left(\frac{b'}{N^2}\right)\right]. \end{aligned} \tag{7.173}$$

The first term on the right-hand side of this expression, namely $f_0N^2$, is a constant and so dynamically unimportant, and the remaining terms are equal to $q$ as given by (7.172).

Another way to see that (7.172) is the potential vorticity is to note that the displacement of an isentropic surface, $\eta$ say, is related to the change in buoyancy by

$$\eta \approx -\frac{b'}{\partial \overline{b}/\partial z} = -\frac{b'}{N^2}, \tag{7.174}$$

as illustrated in Fig. 3.12 on page 157. The thickness of an isentropic layer is the difference between the heights of two neighbouring isentropic surfaces, and so is given by

$$h = -\frac{b_1'}{N^2} + \frac{b_2'}{N^2} \approx -H\frac{\partial}{\partial z}\left(\frac{b'}{N^2}\right) \tag{7.175}$$

where $H$ is the mean thickness between the surfaces. Thus, the expression (7.172) may be written

$$q = \left[\zeta' - \frac{f_0 h}{H}\right] \tag{7.176}$$

which is the 'shallow water' expression for the potential vorticity of a fluid layer, linearized about a mean thickness $H$ and a state of rest (so that $|\zeta'| \ll f_0$).

Let us now relate $q$ to $\chi$, and we do this by expressing $\zeta'$ and $b'$ in terms of $\phi'$ and $w'$. From (7.141) and (7.140b) respectively we have,

$$f_0\zeta' = \nabla^2\phi' - w'_{zt}, \tag{7.177a}$$

$$\frac{f_0^2}{N^2}b'_z = \frac{f_0^2}{N^2}w_{zt} + \frac{f_0^2}{N^2}\phi'_{zz}, \tag{7.177b}$$

using subscripts to denote derivatives. Thus, $f_0$ times the potential vorticity is

$$f_0 q = \nabla^2\phi' + \frac{f_0^2}{N^2}\phi'_{zz} + \frac{f_0^2}{N^2}w_{zt} - w_{zt}. \tag{7.178}$$

We now use (7.144) to express the second $w_{zt}$ term in terms of $\phi'$, giving

$$f_0 q = \nabla^2\phi' + \frac{f_0^2}{N^2}\phi'_{zz} + \frac{f_0^2}{N^2}w'_{zt} + \frac{1}{N^2}\left(w'_{zttt} + \phi'_{zztt}\right), \tag{7.179}$$

and we then use (7.143) to eliminate $w'$, giving

$$f_0 q = \nabla^2\phi' + \frac{f_0^2}{N^2}\phi'_{zz} + \frac{1}{N^2}\left(\nabla^2\phi'_{tt} + \phi'_{zztt}\right), \tag{7.180}$$

or, re-arranging,

$$f_0 q = \frac{1}{N^2}\left[\frac{\partial^2}{\partial t^2}\left(\nabla^2\phi' + \frac{\partial^2\phi'}{\partial z^2}\right) + N^2\nabla^2\phi' + f_0^2\frac{\partial^2\phi'}{\partial z^2}\right]. \tag{7.181}$$

Comparing this with (7.169), we can see that

$$\chi = f_0 N^2 q. \tag{7.182}$$

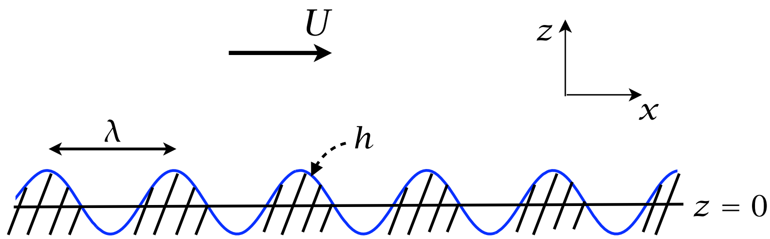


**Fig. 7.17** Uniform flow, $U$, in the $x$-direction flowing over sinusoidal topography, $h$. The vertical co-ordinate is stretched, and in reality $|h| \ll \lambda$.

That is to say, the conserved quantity for motions with $\omega = 0$ is nothing but a constant multiple of the potential vorticity. When $\omega \neq 0$, then $\chi$ and hence the potential vorticity are zero. In other words, *oscillating linear gravity waves, even in a rotating reference frame, have zero potential vorticity.* This is an important result, because large-scale balanced dynamics is characterized by the advection of potential vorticity, so that (in the linear approximation at least) internal waves play not direct role in the potential vorticity budget. However, they *do* play an important role in transporting and dissipating energy, as we will see.

## 7.7 TOPOGRAPHIC GENERATION OF INTERNAL WAVES

How are internal waves generated? One way that is important in both the ocean and atmosphere is by way of a horizontal flow, such as a mean wind or, in the ocean, a tide passing over a topographic feature. This forces the fluid to move up and/or down, so generating an internal wave. In this section we illustrate that with a simple example of steady flow over a sinusoidal topography.[4]

### 7.7.1 Sinusoidal mountain waves

For simplicity we ignore the effects of the Earth's rotation and pose the problem in two dimensions, $x$ and $z$, using the Boussinesq approximation. Our goal is to calculate the response to a steady, uniform flow of magnitude $U$ over a sinusoidally varying boundary $h = \widetilde{h} \cos kx$ at $z = 0$, as in Fig. 7.17 with $k = 2\pi/\lambda$. The topographic variations are assumed small, so allowing the dynamics to be linearized, which would turn enable an arbitrarily shaped boundary to be considered by appropriately summing over Fourier modes. Further, because the problem is linear, the frequency of the response is equal to that of the forcing. Now, suppose we pose the problem in the frame of reference of the mean flow; the topography then has the form

$$h = h_0 \cos[k(x + Ut)]. \tag{7.183}$$

Thus, any resulting internal waves have frequency $\omega = -Uk$, because this is the only time dependence in the problem. This is also a convenient frame in which to work, because there is no mean flow advecting the fields. (An equivalent way to proceed is to stay in the stationary frame and replace each time derivative $\partial/\partial t$ with an advection term $U\partial/\partial x$.)

To proceed it is convenient to write $h = \mathrm{Re}\, h_0 e^{\mathrm{i}(kx-\omega t)}$, and the various dynamical fields in the response, such as the vertical velocity $w$ and the pressure field $\phi(z)$ then have the form

$$w = \mathrm{Re}\, \widetilde{w}(z) \mathrm{e}^{\mathrm{i}(kx-\omega t)}, \qquad \phi = \mathrm{Re}\, \widetilde{\phi}(z) \mathrm{e}^{\mathrm{i}(kx-\omega t)}, \tag{7.184}$$

where $\omega = -Uk$. The problem is to determine the form of $\widetilde{w}(z)$, $\widetilde{\phi}(z)$ and so on.

At the lower boundary the vertical velocity must satisfy the linearized kinematic boundary condition $w = \mathrm{D}h/\mathrm{D}t = \partial h/\partial t + U\partial h/\partial x$. In the moving frame $U = 0$ and therefore we have

$$w = w_0 \mathrm{e}^{\mathrm{i}(kx-\omega t)} = \frac{\partial h}{\partial t} = -\mathrm{i}\omega h_0 \mathrm{e}^{\mathrm{i}(kx-\omega t)}, \qquad \text{at } z = 0, \tag{7.185}$$

where $w_0 = \widetilde{w}(0)$, and so the amplitude of the vertical velocity at the surface is given by

$$w_0 = -\mathrm{i}\omega h_0 = \mathrm{i}Ukh_0. \tag{7.186}$$

The equation of motion to be satisfied by the vertical velocity above the boundary is (7.55), namely

$$\left[\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right) + N^2 \frac{\partial^2}{\partial x^2}\right] w = 0. \tag{7.187}$$

which, given a harmonic dependence in $t$ and $x$ as in (7.184), becomes

$$\frac{\partial^2 \widetilde{w}}{\partial z^2} = \frac{\omega^2 - N^2}{\omega^2} k^2 \widetilde{w}. \tag{7.188}$$

If $N^2$ is constant this equation admits of a solution of the form $\widetilde{w} = w_0 \mathrm{e}^{\mathrm{i}mz}$ where

$$m^2 = \frac{k^2(N^2 - \omega^2)}{\omega^2} = \left(\frac{N}{U}\right)^2 - k^2, \tag{7.189}$$

using $\omega = -Uk$. Equation (7.189) is of course just the dispersion relation for internal gravity waves, but here we are using it to determine the vertical wavenumber since the frequency is given. This solution satisfies the boundary condition at $z = 0$ because the amplitude of the waves is given by (7.186) and the frequency of the waves is given by $\omega = -Uk$. Note that $m^2$ may be negative, and so $m$ imaginary, if $N^2 < k^2U^2$, so evidently there will be a qualitative difference between short waves and long waves.

Given the solution for $w$ we can use the polarization relations of section 7.3.2 with $\omega = -Uk$ to obtain the solutions for perturbation horizontal velocity and pressure. In the stationary frame of reference the solutions are then

$$w = \widetilde{w}(z)\mathrm{e}^{\mathrm{i}kx} = w_0 \mathrm{e}^{\mathrm{i}(kx+mz)} = \mathrm{i}Ukh_0 \mathrm{e}^{\mathrm{i}mz} \mathrm{e}^{\mathrm{i}kx}, \tag{7.190a}$$

$$u = \widetilde{u}(z)\mathrm{e}^{\mathrm{i}kx} = u_0 \mathrm{e}^{\mathrm{i}(kx+mz)} = -\mathrm{i}mUh_0 \mathrm{e}^{\mathrm{i}mz} \mathrm{e}^{\mathrm{i}kx}, \tag{7.190b}$$

$$\phi = \widetilde{\phi}(z)\mathrm{e}^{\mathrm{i}kx} = \phi_0 \mathrm{e}^{\mathrm{i}(kx+mz)} = \mathrm{i}mU^2 h_0 \mathrm{e}^{\mathrm{i}mz} \mathrm{e}^{\mathrm{i}kx}, \tag{7.190c}$$

where $m$ is given by (7.189). In the moving frame of reference we replace $x$ by $x + Ut$; that is, the above solutions are multiplied by $\exp[-\mathrm{i}\omega t]$. The above relationships between $w, u$ and $\phi$ are the de facto polarization relations for this problem.

Having obtained the mathematical form of the solutions let us see what the solutions mean, and if and in what sense the waves propagate away from the mountains.

### 7.7.2 Energy Propagation

The frequency of the mountain waves is just that of internal waves; that is, $\omega = \pm N \cos\vartheta$, where $\vartheta$ is the angle between the wave vector and the horizontal. In our problem, the frequency is determined from the outset by the velocity of the mean flow and the scale of the topography, and thus so is the direction of propagation of the waves crests. The direction of energy propagation is given by the group velocity, given by (7.75), or (7.151) with $f = 0$. The group velocity is at an angle $\vartheta$ to the vertical, and two results that will be useful are that the vertical group velocity and phase speeds are given by

$$c_g^z = \frac{-\omega m}{k^2 + m^2} = \frac{Ukm}{k^2 + m^2}, \qquad c_p^z = -\frac{Uk}{m} \tag{7.191a,b}$$

*Short, trapped waves*

Suppose that the wave frequency is sufficiently high that $\omega^2 > N^2$, which will occur if the undulations on the boundary have a sufficiently short wavelength that $k^2 > (N/U)^2$. From (7.189) $m^2$ is negative and $m$ is pure imaginary. Writing $m = \mathrm{i}s$, so that $s^2 = k^2 - (N/U)^2$, the solutions have the form

$$\widetilde{w} = w_0 \mathrm{e}^{\mathrm{i}(kx-\omega t)-sz}. \tag{7.192}$$

We must choose the solution with $s > 0$ in order that the solution decays away from the mountain, and internal waves are not propagated into the interior. (If there were a rigid lid or a density discontinuity at the top of the fluid (as at the top of the ocean) then the possibility of reflection would arise and we would seek to satisfy the upper boundary condition with a combination of decaying and amplifying modes.) The above result is entirely consistent with the dispersion relation for internal waves, namely $\omega = N\cos\vartheta$: because $\cos\vartheta < 1$ the frequency $\omega$ must be less than $N$ so that if the forcing frequency is higher than $N$ no internal waves will be generated.

Because the waves are trapped waves we do not expect energy to propagate away from the mountains. To verify this, from the polarization relation (7.190) we have

$$\widetilde{w} = \frac{k}{mU}\widetilde{\phi} = \frac{-\mathrm{i}k}{sU}\widetilde{\phi}. \tag{7.193}$$

The pressure and the vertical velocity are therefore out of phase by $\pi/2$, and the vertical energy flux, $\overline{w\phi}$ [see section 7.3.5 and in particular (7.84)] is identically zero. This is consistent with fact that that the energy flux is in the direction of the group velocity; the group velocity is given by (7.191a) and for an imaginary $m$ the real part is zero. A solution is shown in Fig. 7.18 with $s = 1$ ($m = \mathrm{i}$).

*Long, propagating waves*

Suppose now that $k^2 < (N/U)^2$ so that $\omega^2 < N^2$. From (7.189) $m$ is now real and the solution has propagating waves of the form

$$w = w_0 \mathrm{e}^{i(kx+mz-\omega t)}, \qquad m^2 = \left(\frac{N}{U}\right)^2 - k^2 \tag{7.194}$$

Vertical propagation is occurring because the forcing frequency is less than the buoyancy frequency. The angle at which fluid parcel oscillations occur is then slanted off the vertical at an

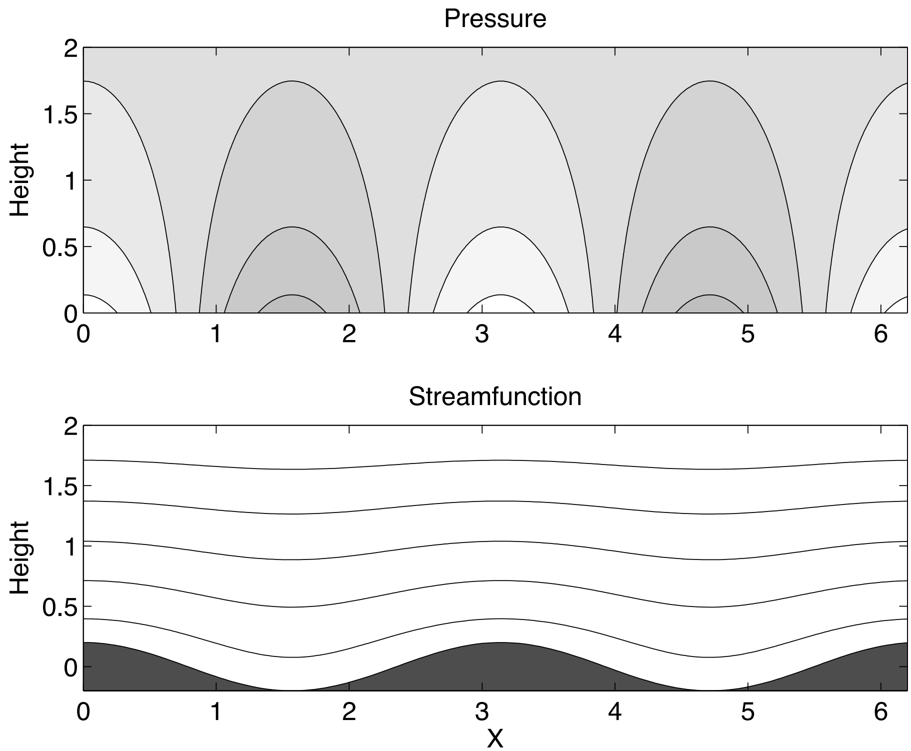


**Fig. 7.18** Solutions for the flow over a sinusoidal ridge, using (7.190), in the short wave limit and with $m = \mathrm{i}$. The top panel shows phase lines of pressure, with darker gray indicating higher pressure. The bottom panel shows contours of the total streamfunction, $\psi - Uz$, with flow coming in from the left, and the topography itself (solid). The perturbation amplitude decreases exponentially with height.

angle $\vartheta$ such that the forcing frequency is equal to the natural frequency of oscillations at that angle, namely

$$\vartheta = \cos^{-1}\left(\frac{Uk}{N}\right). \tag{7.195}$$

The angle $\vartheta$ is also the angle between the wavevector $\boldsymbol{k}$ and the horizontal, as in (7.62), because the wavevector is at right angles to the parcel oscillations. If $Uk = N$ then the fluid parcel oscillations are vertical and, using (7.189), $m = 0$. Thus, although the group velocity is directed vertically, parallel to the fluid parcel oscillations, its magnitude is zero, from (7.191).

Our intuition suggests that if there is vertical propagation there must be an upwards energy flux, since the energy source is at the ground. Let's confirm this. Using the polarization relations (7.190a,c) we obtain

$$w_0 = \frac{k}{mU}\phi_0. \tag{7.196}$$

### Topographically Generated Gravity Waves (Mountain Waves)

- In both atmosphere and ocean an important mechanism for the generation of gravity waves is flow over bottom topography, and the ensuing waves are sometimes called mountain waves. A canonical case is that of a uniform flow over a sinusoidal topography, with constant stratification. If the flow is in the $x$-direction and there is no $y$-variation then the boundary condition is

$$w(x, z = 0) = U\frac{\partial h}{\partial x} = -\mathrm{i}Uk\tilde{h}. \tag{MW.1}$$

  Solutions of the complete problem may be found in the form $w(x, z, t) = w_0 \exp[i(kx + mz - \omega t)]$, where the boundary condition at $z = 0$ is given by (MW.1), the frequency is given by the internal wave dispersion relation, and the other dynamical fields are obtained using the polarization relations.

- One way to easily solve the problem is to transform into a frame moving with the background flow, $U$. The topography then appears to oscillate with a frequency $-Uk$, and this in turn becomes the frequency of the gravity waves.

- Propagating gravity waves can only be supported if the frequency is less than $N$, meaning that $Uk < N$. That is, the waves must be sufficiently long and therefore the topography must be of sufficiently large scale.

- When propagating waves exist, energy is propagated upward away from the topography. The topography also exerts a drag on the background flow.

- If the waves are too short they are evanescent, decaying exponentially with height. That is, they are trapped near the topography

- In the presence of rotation the wave frequency must lie between the buoyancy frequency $N$ and the inertial frequency $f$. That is, waves can radiate upward if

$$f < Uk < N. \tag{MW.2}$$

  Thus, both very long waves and very short waves are evanescent.

and the energy flux in the vertical direction is, from (7.91)

$$F^z = \frac{k}{2mU}|\phi_0|^2 = \frac{mU}{2k}|w_0|^2, \tag{7.197}$$

which is evidently non zero. This energy flux must be upward, away from the source (the topography), and this determines the sign of $m$ that must be chosen by the solution. Specifically, for positive $U$, the group velocity must be positive so from (7.191) $m$ must be positive. If $U$ were negative the sign of $m$ would be negative. Note that if $m = 0$ there is no vertical energy propagation

Because energy is propagating upward and away from the topography there must be a drag at the lower boundary. This stress at the boundary is the rate at which horizontal momentum is transported upwards and so is given by

$$\tau = -\rho_0 \overline{uw}. \tag{7.198}$$

where the overbar denotes averaging over a wavelength. From (7.190)

$$u_0 = -\mathrm{i}mUh_0, \qquad w_0 = \mathrm{i}Ukh_0, \tag{7.199}$$

so that

$$\tau = -\rho_0 \overline{uw} = \frac{1}{2} kmU^2 h_0^2, \tag{7.200}$$

where the factor of 1/2 comes from the averaging, and note that we take the product $u_0 w_0^*$ where $w_0^*$ is the complex conjugate of $w_0$. The sign of the stress depends on the sign of $m$, and thus on the sign of $U$. For positive $U$, $m$ is positive and so the stress is positive at the surface.

Solutions for flow over topography in the long wave limit and $m = 1$ are shown in Fig. 7.19. The flow is coming in from the left, and the phase lines evidently tilt upstream with height. Lines of constant phase follow $kx + mz$ = constant, and in the solution shown both $k$ and $m$ are positive ($k = m = 1$). Thus, the lines slope back at a slope $x/z = -m/k$, and energy propagates up and to the left. The phase propagation is actually downward in this example — see (7.191). The pressure is high on the upstream side of the mountain, and this provides a drag on the flow — a topographic form drag.

### ✦ *Frames of reference, group velocity and critical levels*

For radiating waves, the group velocity seen in the resting and moving frames are different. In the moving frame we have

$$c_g^x = \frac{Um^2}{k^2 + m^2} = -\frac{Nm^2}{(k^2 + m^2)^{3/2}}, \tag{7.201}$$

$$c_g^z = \frac{Ukm}{k^2 + m^2} = \frac{-Nkm}{(k^2 + m^2)^{3/2}}, \tag{7.202}$$

noting that $U = -N/(k^2 + m^2)^{1/2}$ using the dispersion relation and $\omega = -Uk$. We see that $c_g^z / c_g^x = -k/m$, so that the group velocity is, as expected, directed parallel to the phase lines. In the resting frame the horizontal component of the group velocity is shifted by an amount $-U$, so that $c_g^z / c_g^x = -m/k$ and group velocities relative to the ground and air are perpendicular to each other.

If and as the waves propagate upwards it may encounter a *critical level* at which the phase speed of the waves equals the background flow, that is $c = U$. The location of the critical level is not dependent on the choice of frame of reference. At a critical level the wave amplitudes can be expected to become large and linear theory will break down. It is not uncommon for dissipative effects to become important, as we will see explicitly in our discussion of the quasi-biennial oscillation in chapter 17.

### *Atmospheric and oceanic parameters*

A natural question to ask is whether, for typical atmospheric and oceanic parameters, evanescent or propagating gravity waves are more likely to be excited. Consider the atmosphere with a

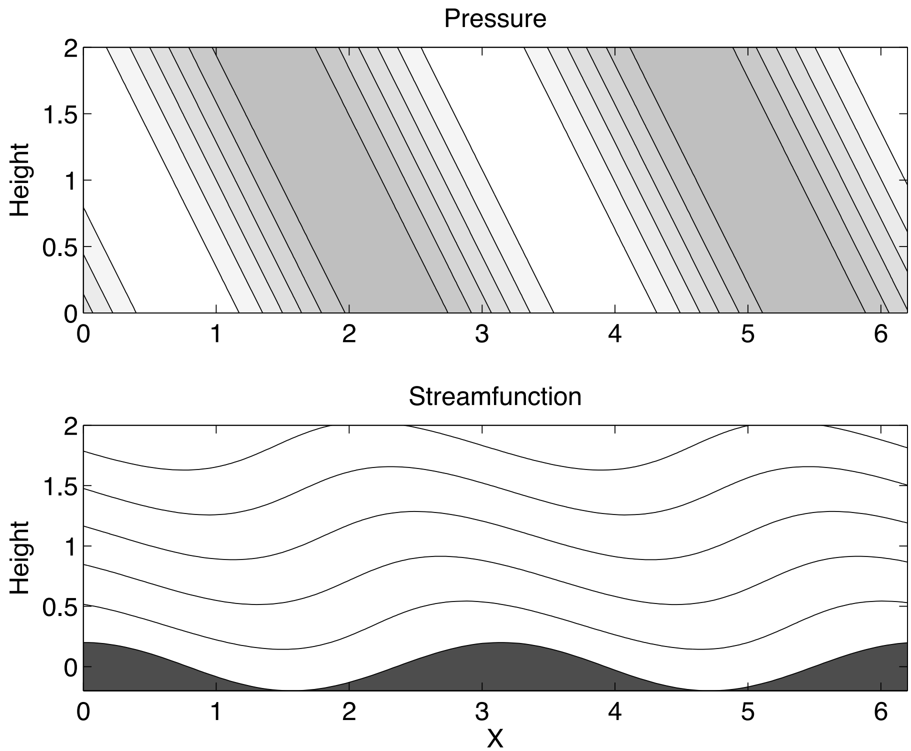


**Fig. 7.19** As for Fig. 7.18, but now showing solutions using (7.190) in the long wave limit with $m = 1$. The top panel shows phase lines of pressure, with gray indicating higher pressure. The bottom panel shows contours of the total streamfunction, $\psi - Uz$, with flow coming in from the left, and the topography itself (solid). Note that pressure is high on the windward side of the topography, and phase lines tilt upstream with height for both pressure and streamfunction.

surface flow of $u = 5\,\mathrm{m\,s^{-1}}$ and $N = 10^{-2}\,\mathrm{s^{-1}}$. Then the critical wavenumber separating evanescent and propagating waves is $k = N/U = 2 \times 10^{-3}$, corresponding to a wavelength of about 3000 m. This is quite large, and almost certainly at that scale rotational effects are also important. Still, large-scale topographic features like the Rockies, Andes and Himalayas do contain such large wavelengths and so we can expect them to excite upward propagating gravity waves.

In the ocean the abyssal stratification is quite weak, typically with $N \approx 10^{-4}\,\mathrm{s^{-1}}$ (although $N$ can be as high as $10^{-3}\,\mathrm{s^{-1}}$) and the velocities are also weak, compared to those of the upper ocean, although they can be of order $1\,\mathrm{cm\,s^{-1}}$ in eddying regions. Using these values we find a critical wavenumber of $k = N/U = 10^{-2}\,\mathrm{m^{-1}}$ with a wavelength of 600 m. Certainly the ocean bathymetry has may scales larger than this (and for smaller values of $u$ the critical scales are correspondingly smaller) meaning that it is relatively easy for abyssal flow to generate gravity waves that propagate upward into the ocean interior. The upper ocean is much more greatly stratified, with $N \approx 10^{-2}\,\mathrm{s^{-1}}$. Gravity waves are no longer generated by flow over topography but by the stirring effects of winds making a turbulent mixed layer. The forcing frequency must still be less than $N$ in order to efficiently generate gravity waves, and using again a velocity of

1 cm s$^{-1}$ we might heuristically estimate that propagating gravity waves can be generated with scales of meters, much smaller than the gravity waves generated in the abyss.

### 7.7.3 Flow over an isolated ridge

Most mountains are of course not perfect sinusoids, but we can construct a solution for any given topography using a superposition of Fourier modes. In this section we will illustrate the solution for a mountain consisting of a single ridge; the actual solution must be obtained numerically and here we will just sketch the method and illustrate the results.

*Sketch of the methodology*

The methodology to compute a solution is as follows. Consider a topographic profile, $h(x)$, and let us suppose that it is periodic in $x$ over some distance $L$. Such profile can (nearly always) be decomposed into a sum of Fourier coefficients, meaning that we can write

$$h(x) = \sum_k \widetilde{h}_k \mathrm{e}^{\mathrm{i}kx} \tag{7.203}$$

where $\widetilde{h}_k$ are the Fourier coefficients. We can obtain the set of $\widetilde{h}_k$ by multiplying (7.203) by $\mathrm{e}^{-ikx}$ and integrating over the domain from $x = 0$ to $x = L$, a procedure known as taking the Fourier transform of $h(x)$, and there are standard computer algorithms for doing this efficiently. Once we have obtained the values of $\widetilde{h}_k$ we essentially solve the problem separately for each $k$ in precisely the same manner as we did in the previous section. Note that for *each* $k$ there will be a vertical wavenumber given by (7.189), so that for each wavenumber we obtain a solution for pressure of the form $\widetilde{\phi}_k(z)$, and similarly for the other variables. Once we have the solution for each wavenumber, then at each level we sum over all the wavenumbers to obtain the solution in real space; that is, we evaluate

$$\phi(x,z) = \sum_k \widetilde{\phi}_z(z)\mathrm{e}^{\mathrm{i}kx}. \tag{7.204}$$

This is known as taking the inverse Fourier transform.

*The solution*

For specificity let us consider the bell-shaped topographic profile

$$h(x) = \frac{h_0 a^2}{a^2 + x^2}, \tag{7.205}$$

sometimes called the Witch of Agnesi.[5] (Results with a Gaussian profile are qualitatively similar.) Such a profile is composed of *many* (in fact an infinite number of) Fourier coefficients of differing amplitudes. If the profile is narrow (meaning $a$ is small in a sense made clearer below) then there will be a great many significant coefficients at high wavenumbers. In fact, in the limiting case of an infinitely thin ridge (a delta function) all wavenumbers are present with equal weight, so there are certainly more large wavenumbers than small wavenumber. However, if $a$ is large, then the contributing wavenumbers will predominantly be small.

In the problem of flow over topography the natural horizontal scale is $U/N$. If $a \gg U/N$ then the dominant wavenumbers are small and the solution will consist of waves propagating

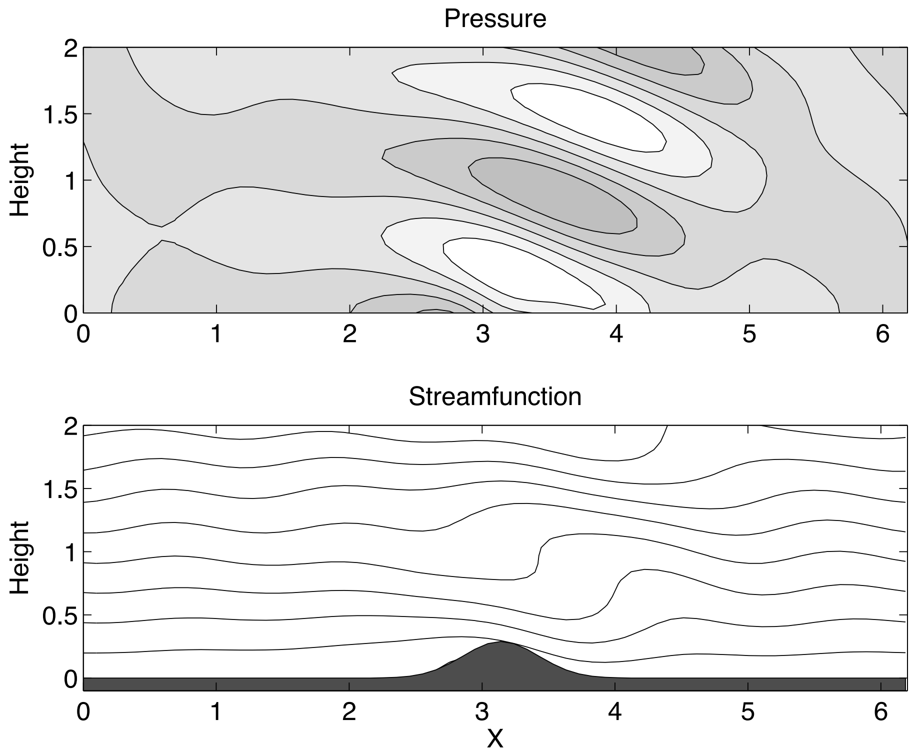


**Fig. 7.20** Solutions for the flow over a bell-shaped ridge (7.205), with $a^2 = 4U^2/N^2$. High pressure is shaded darker, and the flow comes in from the left.

upward with little loss of amplitude and phase lines tilting upstream, as illustrated in Fig. 7.20. If the ridge is sufficiently wide then the solution is essentially hydrostatic, with little dependence of the vertical structure on the horizontal wavenumber. That is, using (7.189) at large scales, $m^2 \approx (N/U)^2$ and the pattern repeats itself in the vertical at intervals of $2\pi U/N$, and so at any given level there can be only one wave crest in the fluid flowing over the ridge.

In the case of a narrow ridge, as illustrated in Fig. 7.21, the perturbation is largely trapped near to the mountain and the perturbation fields largely decay exponentially with height. Nevertheless, because the ridge *does* contain some small wavenumbers some weak, propagating large-scale disturbances are generated. The fluid acts as a low-pass filter, and the perturbation aloft consists only of large scales.

### 7.7.4 Effects of rotation

*General considerations*

We now consider, albeit briefly, the effects of a Coriolis force on mountain waves. The problem is in many ways quite similar to the non-rotating case but the dispersion relation and so the criteria for upward propagation differ accordingly. First, we note that the steady flow must be in geostrophic balance, so that the if the flow is zonal there is a background meridional pressure

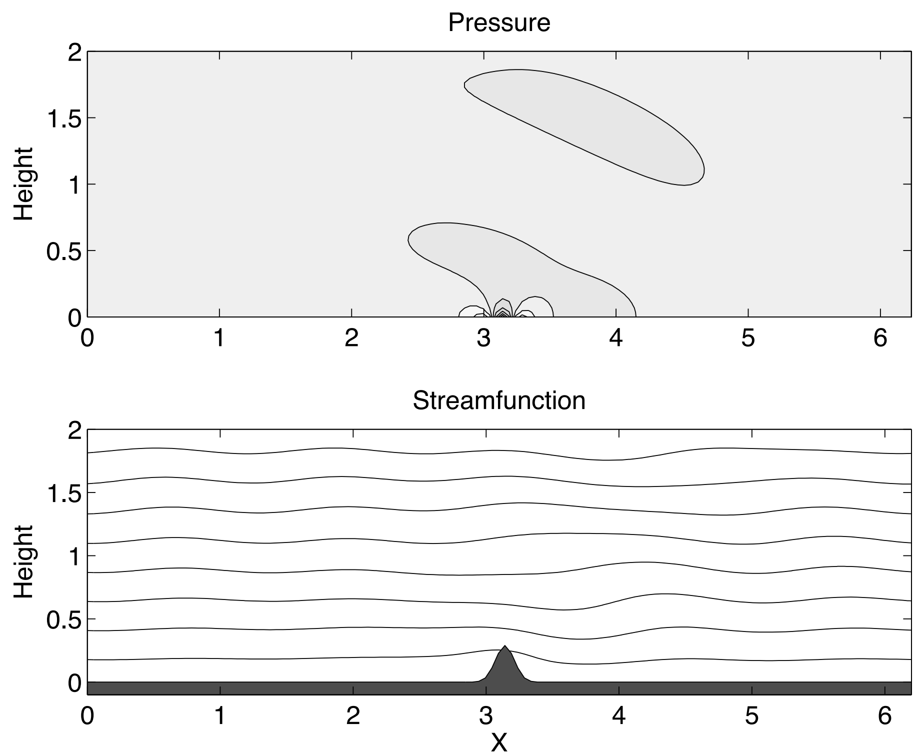


**Fig. 7.21** As for Fig. 7.20 but now for a narrow ridge, with $a^2 = U^2/4N^2$.

gradient that satisfies

$$f_0 U = -\frac{\partial \Phi}{\partial y}. \tag{7.206}$$

The main difference in the solution field arises from the fact that the waves now obey the dispersion relation with rotation, namely (7.148) or, restricting attention to the $x$–$z$ plane,

$$\omega^2 = \frac{f_0^2 m^2 + k^2 N^2}{k^2 + m^2}. \tag{7.207}$$

The frequency of the waves is still given by $\omega = -Uk$ (if we consider the problem in the translating frame), so that the vertical wavenumber is now given by

$$m^2 = \frac{k^2(N^2 - U^2 k^2)}{U^2 k^2 - f_0^2}. \tag{7.208}$$

Evanescent solutions arise when $m$ is imaginary and, as before, such solutions arise for small scales for which $k > N/U$. However, from (7.208), evanescent solutions also arise for very large scales for which $k < f_0/U$. Propagating waves exist in the interval $N/U > k > f_0/U$ with the vertical wavenumber a real number, and these waves have frequencies between $N$ and $f_0$. In the atmosphere with $U = 10\,\mathrm{m\,s^{-1}}$ and $f_0 = 10^{-4}\,\mathrm{s^{-1}}$ the large scale at which evanescence reappears is $L = 2\pi/k = 2\pi U/f \approx 600\,\text{km}$, which of course is not very large at all relative to global scales (and still smaller if we take $U = 5\,\text{m}\,\text{s}^{-1}$). Thus, upward propagating gravity waves exist between scales of a few kilometres (see the calculation on page 329) and several hundred kilometres. For the deep ocean, let us take $N = 10^{-3}\,\text{s}^{-1}$ and $f_0 = 10^{-4}\,\text{s}^{-1}$, and $U = 1\,\text{cm}\,\text{s}^{-1}$. Thus, very roughly, propagating waves exist between scales of a few tens of meters to a few hundred metres. If we use $N = 10^{-4}\,\text{s}^{-1}$ the range of scales is further restricted.

*Wave solutions and energy propagation*

Obtaining a wave solution in the rotating case follows a similar path to the non-rotating case. In the resting frame vertical velocity satisfies the boundary condition $w = U\partial h/\partial x$, and in the moving frame $w = \partial h/\partial t$. Using the polarization relations appropriate for rotation we find analogous relations to (7.190), to wit

$$w = \widetilde{w}(z)e^{ikx} = w_0 e^{i(kx+mz)} = iUkh_0 e^{imz} e^{ikx}, \tag{7.209a}$$

$$u = \widetilde{u}(z)e^{ikx} = u_0 e^{i(kx+mz)} = -imUh_0 e^{imz} e^{ikx}, \tag{7.209b}$$

$$\phi = \widetilde{\phi}(z)e^{ikx} = \phi_0 e^{i(kx+mz)} = \frac{im(U^2k^2 - f_0^2)}{k^2} h_0 e^{imz} e^{ikx}, \tag{7.209c}$$

$$v = \widetilde{v}(z)e^{ikx} = v_0 e^{i(kx+mz)} = -if_0 \frac{m}{k} h_0 e^{imz} e^{ikx}, \tag{7.209d}$$

Of these, the expressions for $w$ and $u$ are no different from the non-rotating case, because $w$ is set by the same boundary condition and $u$ is given by mass continuity, $\partial u/\partial x + \partial w/\partial z = 0$, in both rotating and non-rotating cases. To obtain the expression for the pressure perturbation, (7.209c), we use (7.164). Finally, we note that the solutions now produce a meridional velocity, (7.209d), even when there is no variation in the topography in the $y$-direction. To obtain we use (7.162) with $l = 0$, giving $\widetilde{v} = -i\widetilde{u}f\omega = i\widetilde{u}f_0/Uk$.

As in the non-rotating case, when there are propagating waves there is high pressure on the windward (upstream) side of the topography and low pressure on leeward side, and the phase lines tilt upstream with height. The drag on the flow is equal to the rate of upward momentum transport and using (7.209a,b) we obtain

$$\rho_0 \overline{uw} = -\frac{1}{2}\rho_0 kmU^2 h_0^2 < 0. \tag{7.210}$$

This is just the same as (7.200). It is independent of height, and a momentum flux divergence will only arise in the free atmosphere if the waves break and dissipative effects become important.

The vertical flux of energy density is given by

$$\overline{\phi w} = \frac{1}{2}\rho_0 U \frac{m}{k} h_0^2 (U^2k^2 - f_0^2) > 0 \tag{7.211}$$

Energy is propagating away from the mountain, consistent with the group velocity being directed upward.

## 7.8 ✦ ACOUSTIC-GRAVITY WAVES IN AN IDEAL GAS

In the final section of this chapter we consider wave motion in a stratified, *compressible* fluid such as the Earth's atmosphere. The stratification allows gravity waves to exist, and the compressibility allows sound waves to exist. The resulting problem is, not surprisingly, complicated and arcane and to make it as tractable as possible we will specialize to the case of an isothermal, stationary atmosphere and ignore the effects of rotation and sphericity. The results are not without interest, both in themselves and in illustrating the importance of simplifying the equations of motion from the outset, for example by making the Boussinesq or hydrostatic approximation, in order to isolate phenomena of interest.

In what follows we will denote the unperturbed state with a subscript 0 and the perturbed state with a prime ($'$); we will also omit some of the algebraic details. Because it is at rest, the basic state is in hydrostatic balance,

$$\frac{\partial p_0}{\partial z} = -\rho_0(z) g. \tag{7.212}$$

Ignoring variations in the $y$-direction for algebraic simplicity (and without loss of generality, in fact) the linearized equations of motion are:

$$\begin{aligned}
&u \text{ momentum:} && \rho_0 \frac{\partial u'}{\partial t} = -\frac{\partial p'}{\partial x} && \text{(7.213a)}\\
&w \text{ momentum:} && \rho_0 \frac{\partial w'}{\partial t} = -\frac{\partial p'}{\partial z} - \rho' g && \text{(7.213b)}\\
&\text{mass conservation:} && \frac{\partial \rho'}{\partial t} + w' \frac{\partial \rho_0}{\partial z} = -\rho_0 \left( \frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} \right) && \text{(7.213c)}\\
&\text{thermodynamic:} && \frac{\partial \theta'}{\partial t} + w' \frac{\partial \theta_0}{\partial z} = 0 && \text{(7.213d)}\\
&\text{equation of state:} && \frac{\theta'}{\theta_0} + \frac{\rho'}{\rho_0} = \frac{1}{\gamma} \frac{p'}{p_0}. && \text{(7.213e)}
\end{aligned}$$

For an isothermal basic state we have $p_0 = \rho_0 R T_0$ where $T_0$ is a constant, so that $\rho_0 = \rho_s e^{-z/H}$ and $p_0 = p_s e^{-z/H}$ where $H = RT_0/g$. Further, using $\theta = T(p_s/p)^\kappa$ where $\kappa = R/c_p$, we have $\theta_0 = T_0 e^{\kappa z/H}$ and so $N^2 = \kappa g/H$. It is convenient to use (1.99) on page 23 to rewrite the linear thermodynamic equation in the form

$$\frac{\partial p'}{\partial t} - w' \frac{p_0}{H} = -\gamma p_0 \left( \frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} \right). \tag{7.213f}$$

where $\gamma = c_p/c_v = 1/(1-\kappa)$. The complete set of equations of motion that we use are (7.213a,b,c,f).

Differentiating (7.213a) with respect to time and using (7.213f) leads to

$$\left( \frac{\partial^2}{\partial t^2} - c_s^2 \frac{\partial^2}{\partial x^2} \right) u' = c_s^2 \left( \frac{\partial}{\partial z} - \frac{1}{\gamma H} \right) \frac{\partial}{\partial x} w'. \tag{7.214a}$$

where $c_s^2 = \gamma p_0/\rho_0$ is the square of the speed of sound (equal to $\partial p/\partial \rho$ or $\gamma R T_0$). Similarly, differentiating (7.213b) with respect to time and using (7.213c) and (7.213f) leads to

$$\left( \frac{\partial^2}{\partial t^2} - c_s^2 \left[ \frac{\partial^2}{\partial z^2} - \frac{1}{H} \frac{\partial}{\partial z} \right] \right) w' = c_s^2 \left( \frac{\partial}{\partial z} - \frac{\kappa}{H} \right) \frac{\partial u'}{\partial x}, \tag{7.214b}$$

Equations (7.214a) and (7.214b) combine to give, after some cancellation,

$$\frac{\partial^4 w'}{\partial t^4} - c_s^2 \frac{\partial^2}{\partial t^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{H}\frac{\partial}{\partial z}\right) w' - c_s^2 \frac{\kappa g}{H}\frac{\partial^2 w'}{\partial x^2} = 0. \tag{7.215}$$

If we set $w' = W(x,z,t)e^{z/(2H)}$, so that $W = (\rho_0/\rho_s)^{1/2} w'$, then the term with the single $z$-derivative is eliminated, giving

$$\frac{\partial^4 W}{\partial t^4} - c_s^2 \frac{\partial^2}{\partial t^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{4H^2}\right) W - c_s^2 \frac{\kappa g}{H}\frac{\partial^2 W}{\partial x^2} = 0. \tag{7.216}$$

Although superficially complicated, this equation has constant coefficients and we may seek wave-like solutions of the form

$$W = \mathrm{Re}\, \widetilde{W} e^{i(kx+mz-\omega t)}, \tag{7.217}$$

where $\widetilde{W}$ is the complex wave amplitude. Using (7.217) in (7.216) leads to the dispersion relation for acoustic-gravity waves, namely

$$\omega^4 - c_s^2\omega^2\left(k^2 + m^2 + \frac{1}{4H^2}\right) + c_s^2 N^2 k^2 = 0, \tag{7.218}$$

with solution

$$\boxed{\omega^2 = \frac{1}{2}c_s^2 K^2\left[1 \pm \left(1 - \frac{4N^2k^2}{c_s^2K^4}\right)^{1/2}\right]}, \tag{7.219}$$

where $K^2 = k^2 + m^2 + 1/(4H^2)$. (The factor $[1 - 4N^2k^2/(c_s^2K^4)]$ is always positive — see problem 7.25.) For an isothermal, ideal-gas atmosphere $4N^2H^2/c_s^2 \approx 0.8$ and so this may be written

$$\frac{\omega^2}{N^2} \approx 2.5\widehat{K}^2\left[1 \pm \left(1 - \frac{0.8\widehat{k}^2}{\widehat{K}^4}\right)^{1/2}\right], \tag{7.220}$$

where $\widehat{K}^2 = \widehat{k}^2 + \widehat{m}^2 + 1/4$, and $(\widehat{k}, \widehat{m}) = (kH, mH)$.

### 7.8.1 Interpretation

*Acoustic and gravity waves*

There are two branches of roots in (7.219), corresponding to acoustic waves (using the plus sign in the dispersion relation) and internal gravity waves (using the minus sign). These (and the Lamb wave, described below) are plotted in Fig. 7.22. If $4N^2k^2/c_s^2K^4 \ll 1$ then the two sets of waves are well separated. From (7.220) this is satisfied when

$$\frac{4\kappa}{\gamma}(kH)^2 \approx 0.8(kH)^2 \ll \left[(kH)^2 + (mH)^2 + \frac{1}{4}\right]^2; \tag{7.221}$$

that is, when *either* $mH \gg 1$ or $kH \gg 1$. The two roots of the dispersion relation are then

$$\omega_a^2 \approx c_s^2 K^2 = c_s^2\left(k^2 + m^2 + \frac{1}{4H^2}\right) \tag{7.222}$$

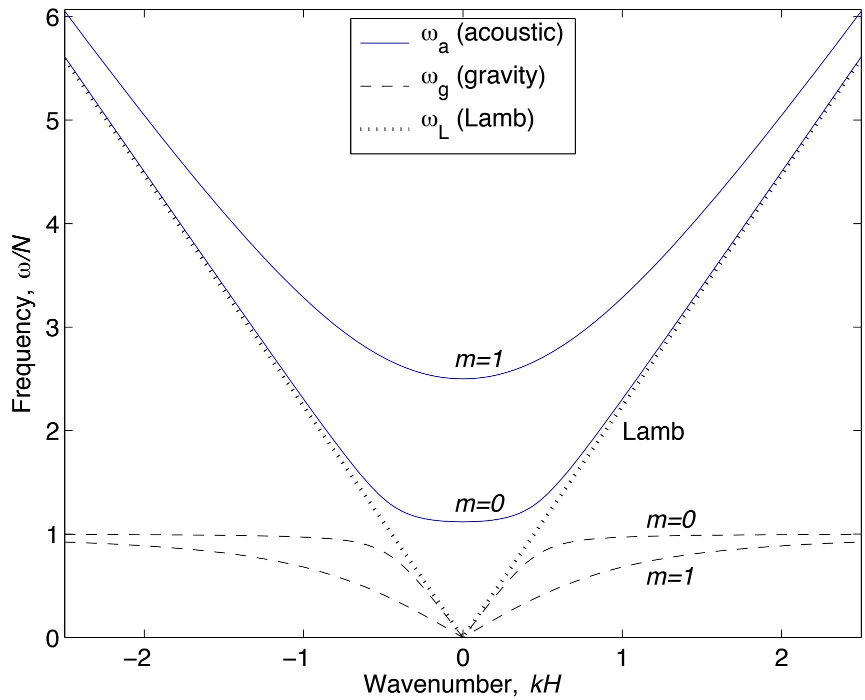


**Fig. 7.22** Dispersion diagram for acoustic gravity waves in an isothermal atmosphere, calculated using (7.220). The frequency is given in units of the the buoyancy frequency $N$, and the wavenumbers are non-dimensionalized by the inverse of the scale height, $H$. The solid curves indicate acoustic waves, whose frequency is always higher than that of the corresponding Lamb wave at the same wavenumber (i.e., $ck$), and of the base acoustic frequency $\approx 1.12N$. The dashed curves indicate internal gravity waves, whose frequency asymptotes to $N$ at small horizontal scales.

and

$$\omega_g^2 \approx \frac{N^2k^2}{k^2 + m^2 + 1/(4H^2)}, \tag{7.223}$$

corresponding to acoustic and gravity waves, respectively. The acoustic waves owe their existence to the presence of compressibility in the fluid, and they have no counterpart in the Boussinesq system. On the other hand, the internal gravity waves are just modified forms of those found in the Boussinesq system, and if we take the limit $(kH, mH) \to \infty$ then the gravity wave branch reduces to $\omega_g^2 = N^2k^2/(k^2 + m^2)$, which is the dispersion relationship for gravity waves in the Boussinesq approximation. We may consider this to be the limit of infinite scale height or (equivalently) the case in which wavelengths of the internal waves are sufficiently small that the fluid is essentially incompressible.

*Vertical structure*

Recall that $w' = W(x, z, t)\mathrm{e}^{z/(2H)}$ and, by inspection of (7.214), $u'$ has the same vertical structure. That is,

$$w' \propto \mathrm{e}^{\,z/(2H)}, \qquad u' \propto \mathrm{e}^{\,z/(2H)}, \tag{7.224}$$

and the amplitude of the velocity field of the internal waves increases with height. The pressure and density perturbation amplitudes fall off with height, varying like

$$p' \propto \mathrm{e}^{-z/(2H)}, \qquad \rho' \propto \mathrm{e}^{-z/(2H)}. \tag{7.225}$$

The kinetic energy of the perturbation, $\rho_0(u'^2 + w'^2)$ is *constant* with height, because $\rho_0 = \rho_s e^{-z/H}$.

*Hydrostatic approximation and Lamb waves*

Equations (7.214) also admit to a solution with $w' = 0$. We then have

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2\frac{\partial^2}{\partial x^2}\right)u' = 0 \qquad \text{and} \qquad \left(\frac{\partial}{\partial z} - \frac{\kappa}{H}\right)\frac{\partial u'}{\partial x} = 0, \tag{7.226}$$

and these have solutions of the form

$$u' = \mathrm{Re}\,\widetilde{U}\mathrm{e}^{\kappa z/H}\mathrm{e}^{\mathrm{i}(kx-\omega t)}, \qquad \omega = ck, \tag{7.227}$$

where $\widetilde{U}$ is the wave amplitude. These are horizontally propagating sound waves, known as *Lamb waves* after the hydrodynamicist Horace Lamb. Their velocity perturbation amplitude increases with height, but the pressure perturbation falls with height; that is

$$u' \propto \mathrm{e}^{\kappa z/H} \approx \mathrm{e}^{2z/(7H)}, \qquad p' \propto \mathrm{e}^{(\kappa-1)z/H} \approx \mathrm{e}^{-5z/(7H)}. \tag{7.228}$$

Their kinetic energy density, $\rho_0 u'^2$, varies as

$$KE \propto \mathrm{e}^{-z/H+2\kappa z/H} = \mathrm{e}^{(2R-c_p)z/(c_pH)]} = \mathrm{e}^{(R-c_v)z/(c_pH)} \approx \mathrm{e}^{-3z/(7H)} \tag{7.229}$$

for an ideal gas. (In a simple ideal gas, $c_v = nR/2$ where $n$ is the number of excited degrees of freedom, 5 for a diatomic molecule.) The kinetic energy density thus falls away exponentially from the surface, and in this sense Lamb waves are an example of edge waves or surface-trapped waves.

Now consider the case in which we make the hydrostatic approximation ab initio, but without restricting the perturbation to have $w' = 0$. The linearized equations are identical to (7.213), except that (7.213b) is replaced by

$$\frac{\partial p'}{\partial z} = -\rho' g. \tag{7.230}$$

The consequence of this is that first term $(\partial^2 w'/\partial t^2)$ in (7.214b) disappears, as do the first two terms in (7.215) [the terms $\partial^4 w'/\partial t^4 - c^2(\partial^2/\partial t^2)(\partial^2 w'/\partial x^2)$]. It is a simple matter to show that the dispersion relation is then

$$\omega^2 = \frac{N^2k^2}{m^2 + 1/(4H^2)}. \tag{7.231}$$

These are long gravity waves, and may be compared with the corresponding Boussinesq result (7.61). Again, the frequency increases without bound as the horizontal wavelength diminishes.

The Lamb wave, of course, still exists in the hydrostatic model, because (7.226) is still a valid solution. Thus, horizontally propagating sound waves still exist in hydrostatic (primitive equation) models, but vertically propagating sound waves do not — essentially because the term $\partial w/\partial t$ is absent from the vertical momentum equation.

## 7.A APPENDIX: THE WKB APPROXIMATION FOR LINEAR WAVES

WKB (Wentzel–Kramers–Brillouin) theory is a way of finding approximate solutions to certain linear differential equations in which the term with the highest derivative is multiplied by a small parameter.[6] The theory for such equations is quite extensive but our interests are modest, being mainly in dispersive waves, and WKB theory can be used to find approximate solutions in cases in which the coefficients of the wave equation vary slowly in space or time. In many cases we find ourselves concerned with finding solutions to an equation of the form

$$\frac{\mathrm{d}^2\xi}{\mathrm{d}z^2} + m^2(z)\xi = 0, \tag{7.232}$$

where $m^2(z)$ is positive for wavelike solutions. If $m$ is constant the solution has the harmonic form

$$\xi = \mathrm{Re}\, A_0 \mathrm{e}^{\mathrm{i}mz} \tag{7.233}$$

where $A_0$ is a complex constant. If $m$ varies only 'slowly' with $z$ — meaning that the variations occur on a scale much longer than $1/m$ — one might reasonably expect that the harmonic solution above would provide a reasonable first approximation. That is, we expect the solution to locally look like a plane wave with local wavenumber $m(z)$, but if so the form cannot be *exactly* like (7.233), because the phase of $\xi$ is $\theta(z) = mz$, so that $\mathrm{d}\theta/\mathrm{d}z = m + z\mathrm{d}m/\mathrm{d}z \neq m$. Thus, in (7.233) $m$ is *not* the wavenumber unless $m$ is constant. Nevertheless, this argument suggests that we seek solutions of a similar form to (7.233), and we find such solutions by way of a perturbation expansion below; readers who are content with a more informal derivation of the solution may skip to section 7.A.2. We note that the condition that variations in $m$, or in the wavelength $m^{-1}$, occur only slowly may be expressed as

$$\frac{m}{|\partial m/\partial z|} \gg m^{-1} \quad \text{or} \quad \left|\frac{\partial m}{\partial z}\right| \ll m^2. \tag{7.234}$$

### 7.A.1 Solution by perturbation expansion

To explicitly recognize the rapid variation of $m$ we rescale the coordinate $z$ with a small parameter $\epsilon$ so that $\widehat{z} = \epsilon z$, whence $\widehat{z}$ varies by $\mathcal{O}(1)$ over the scale on which $m$ varies. Eq. (7.232) becomes

$$\epsilon^2 \frac{\mathrm{d}^2\xi}{\mathrm{d}\widehat{z}^2} + m^2(\widehat{z})\xi = 0, \tag{7.235}$$

and we may now suppose that all variables are $\mathcal{O}(1)$. If $m$ were constant the solution would be of the form $\xi = A\exp(m\widehat{z}/\epsilon)$ and this suggests that we look for a solution to (7.235) of the form

$$\xi(z) = \mathrm{e}^{g(\widehat{z})/\epsilon}, \tag{7.236}$$

where $g(\hat{z})$ is some function. We then have, with primes denoting derivatives,

$$\xi' = \frac{1}{\epsilon} g' \mathrm{e}^{g/\epsilon}, \qquad \xi'' = \left( \frac{1}{\epsilon^2} g'^2 + \frac{1}{\epsilon} g'' \right) \mathrm{e}^{g/\epsilon}. \tag{7.237a,b}$$

Using these expressions in (7.235) yields

$$\epsilon g'' + g'^2 + m^2 = 0, \tag{7.238}$$

and if we let $g = \int h \, \mathrm{d}\hat{z}$ we obtain

$$\epsilon \frac{\mathrm{d}h}{\mathrm{d}\hat{z}} + h^2 + m^2 = 0. \tag{7.239}$$

To obtain a solution of this equation we expand $h$ in powers of the small parameter $\epsilon$,

$$h(\hat{z}; \epsilon) = h_0(\hat{z}) + \epsilon h_1(\hat{z}) + \epsilon^2 h_2(\hat{z}) + \cdots . \tag{7.240}$$

Substituting this in (7.239) and setting successive powers of $\epsilon$ to zero gives, at first and second order,

$$h_0^2 + m^2 = 0, \qquad 2h_0 h_1 + \frac{\mathrm{d}h_0}{\mathrm{d}\hat{z}} = 0. \tag{7.241a,b}$$

The solutions of these equations are

$$h_0 = \pm \mathrm{i} m, \qquad h_1(\hat{z}) = -\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\hat{z}} \ln \frac{m(\hat{z})}{m_0}. \tag{7.242a,b}$$

where $m_0$ is a constant. Now, ignoring higher-order terms, (7.236) may be written in terms of $h_0$ and $h_1$ as

$$\xi(\hat{z}) = \exp\left( \int h_0 \, \mathrm{d}\hat{z}/\epsilon \right) \exp\left( \int h_1 \, \mathrm{d}\hat{z} \right), \tag{7.243}$$

and, using (7.242) and with $z$ in place of $\hat{z}$, we obtain

$$\boxed{\xi(z) = A_0 m^{-1/2} \exp\left( \pm \mathrm{i} \int m \, \mathrm{d}z \right)}. \tag{7.244}$$

where $A_0$ is a constant, and this is the WKB solution to (7.232). In general

$$\xi(z) = B_0 m^{-1/2} \exp\left( \mathrm{i} \textstyle\int m \, \mathrm{d}z \right) + C_0 m^{-1/2} \exp\left( -\mathrm{i} \textstyle\int m \, \mathrm{d}z \right). \tag{7.245}$$

or

$$\xi(z) = D_0 m^{-1/2} \cos\left( \textstyle\int m \, \mathrm{d}z \right) + E_0 m^{-1/2} \sin\left( \textstyle\int m \, \mathrm{d}z \right). \tag{7.246}$$

A property of (7.244) is that the derivative of the phase is just $m$; that is, $m$ is indeed the local wavenumber. Note that a crucial aspect of the derivation is that $m$ varies slowly, so that there is a small parameter, $\epsilon$, in the problem. Having said this, it is often the case that WKB theory can provide qualitative guidance even when there is little scale separation between the variation of the background state and the wavelength. Asymptotics often works when it seemingly shouldn't.

### 7.A.2 Quick derivation

A quick, albeit not obviously well motivated or systematic, way to obtain the same result is to seek solutions of the form

$$\xi = A(z)\mathrm{e}^{\mathrm{i}\theta(z)} \tag{7.247}$$

where $A(z)$ and $\theta(z)$ are both presumptively real. Using (7.247) in (7.232) yields

$$\mathrm{i}\left[2\frac{\mathrm{d}A}{\mathrm{d}z}\frac{\mathrm{d}\theta}{\mathrm{d}z} + A\frac{\mathrm{d}^2\theta}{\mathrm{d}z^2}\right] + \left[A\left(\frac{\mathrm{d}\theta}{\mathrm{d}z}\right)^2 - \frac{\mathrm{d}^2A}{\mathrm{d}z^2} - m^2A\right] = 0. \tag{7.248}$$

The terms in square brackets must each be zero. Given the slow variation of the amplitude we assume that $|A^{-1}\mathrm{d}^2A/\mathrm{d}z^2| \ll m^2$ and consequently ignore the term involving $\mathrm{d}^2A/\mathrm{d}z^2$. The real and imaginary parts of (7.248) become

$$\left(\frac{\mathrm{d}\theta}{\mathrm{d}z}\right)^2 = m^2, \qquad 2\frac{\mathrm{d}A}{\mathrm{d}z}\frac{\mathrm{d}\theta}{\mathrm{d}z} + A\frac{\mathrm{d}^2\theta}{\mathrm{d}z^2} = 0. \tag{7.249a,b}$$

These two equations are very similar to (7.241). The solution of the first equation is

$$\theta = \pm\int m\,\mathrm{d}z, \tag{7.250}$$

and substituting this into the (7.249b) gives

$$2\frac{\mathrm{d}A}{\mathrm{d}z}m + A\frac{\mathrm{d}m}{\mathrm{d}z} = 0, \tag{7.251}$$

the solution of which is

$$A = A_0 m^{-1/2}. \tag{7.252}$$

Using (7.250) and (7.252) in (7.247) recovers (7.244).

*Notes*

1 Treatments of various aspects of internal waves are to be found in Gill (1982), Lighthill (1978), Munk (1981), Pedlosky (2003) and extensively in the book by Sutherland (2010). I am also grateful to have seen unpublished lecture notes kindly provided by S. Legg.

2 Drawing from Durran (1990).

3 Equation (7.127) is slightly different from the corresponding equation in Gill (1982), his (6.4.10), because of our use of the anelastic equation from the outset. Still, Gill goes on to invoke the Boussinesq approximation.

4 We draw from a useful review of mountain waves by Durran (1990).

5 Treatments of this rather canonical profile are given by Queney (1948) and Durran (1990). The profile is named for Maria Agnesi, 1718–1799, an Italian mathematician and later a theologian, who had discussed the properties of the curve.

6 A description of the WKB method, also called the JWKB method, can be found in many books on perturbation methods, for example Simmonds & Mann (1998), Holmes (2013) and Bender & Orszag (1978). As an ironic sign of its importance, developments in perturbation theory have generalized the method to the extent that it does not appear as a separate topic in the

well-known book by Kevorkian & Cole (2011). Wentzel, Kramers and Brillouin separately presented the technique in 1926 as a way to find approximate solutions of the Schrödinger equation. Harold Jeffreys, a mathematical geophysicist, had proposed a similar technique in 1924, and Rayleigh in 1912 had already addressed some aspects of the theory. A general mathematical treatment of the topic was in fact given by Joseph Liouville and George Green in the first half of the nineteenth century. The story thus affirms the hypothesis that methods are often named after the last people to discover them.

## *Problems*

7.1 *Convection and its parameterization*

*(a)* Consider a Boussinesq system in which the vertical momentum equation is modified by the parameter $\alpha$ to read

$$\alpha^2 \frac{\mathrm{D}w}{\mathrm{D}t} = -\frac{\partial \phi}{\partial z} + b, \tag{P7.1}$$

and the other equations are unchanged. (If $\alpha = 0$ the system is hydrostatic, and if $\alpha = 1$ the system is the original one.) Linearize these equations about a state of rest and of constant stratification and obtain the dispersion relation for the system, and plot it for various values of $\alpha$, including 0 and 1. Show that for $\alpha > 1$ the system approaches its limiting frequency more rapidly than with $\alpha = 1$.

*(b)* ✦ Argue that if $N^2 < 0$, convection in a system with $\alpha > 1$ generally occurs at a larger scale than with $\alpha = 1$. Show this explicitly by adding some diffusion or friction to the right-hand sides of the equations of motion and obtaining the dispersion relation. You may do this approximately.

*Changes in latitude, changes in attitude,*
*Nothing remains quite the same.*
Jimmy Buffett, *Changes in latitude.*

# CHAPTER EIGHT

# Linear Dynamics at Low Latitudes

This chapter is an introduction to the dynamics of the atmosphere and ocean at low latitudes, concentrating mainly on the linear dynamics associated with waves. This chapter will likely go in part II of the second edition, which is the 'advanced GFD' part. The chapter is our first real taste of dynamics at low latitudes, and it is a gentle, perhaps even anodyne, introduction although it is somewhat mathematical. In particular, we don't get into the real *phenomenology* of low latitudes: the tropical atmosphere with its humidity, its convection, and its towering cumulonimbus clouds, or the equatorial ocean with its undercurrents and countercurrents. And most certainly we don't get into low latitude atmosphere-ocean interaction and the wonderful phenomenon called El Niño. Rather, this chapter is really just about the linear geophysical fluid dynamics of the shallow water equations at low latitudes, when the beta effect is important and the flow is not completely geostrophically balanced. Still, let us not be too deprecatory about them — they are important both in their own right and as prerequisites for these more complex phenomena that we encounter later.

Why do we talk about the 'tropical' atmosphere but the 'equatorial' ocean? It is because an essential demarcation in the dynamics of the atmosphere lies at the edge of the Hadley Cell, at about 30° latitude, and the dynamics are quite different poleward and equatorward of that line. In some contrast, the dynamics of the ocean do not change their essential character until we approach quite close to the equator. At 10° latitude the ocean dynamics still have many of the characteristics of the mid-latitudes — the Rossby number is still quite small, for example. Only until we get within a very few degrees of the equator does the dynamics change its character in a qualitative way.

In midlatitudes there is a fairly clear separation in the time and space scales between balanced and unbalanced motion, and it is useful to recognize this by explicitly filtering out gravity wave motion and considering purely balanced motion, using for example the quasi-geostrophic equations. In equatorial regions, where the Coriolis parameter can become very small and is zero at the equator, the Rossby number may be order unity or larger and such a separation is less useful. However, even as $f$ becomes small, $\beta$ becomes large and Rossby waves, or their equatorial equivalent, remain as important, or become more even so, than in midlatitudes. The reader may then readily imagine the complications arising even from linear wave problems in equatorial regions: determining the dispersion relation for combined Rossby and gravity waves — Rossby-gravity waves — in a rotating, continuously stratified fluid is an algebraically complex task. The task is greatly simplified by posing the problem in the context of the shallow water equations. The active layer of fluid represents the layer of fluid in and above the main equatorial thermocline, overlying a deep stationary fluid layer of slightly higher density that represents the abyssal ocean. If we accept this physical model we are led to so-called *reduced gravity* equations of motion, as described in chapter 3 of *AOFD*, in which the value of $g$ is replaced by $g' = g\delta\rho/\rho_0$, where $\delta\rho$ is the difference in the value of the density between the two layers of fluid. However, the more accurate equations of motion are the Boussinesq equations. Let us first see how, if the vertical stratification is fixed, the Boussinesq equations may be reduced to the shallow water equations by the device of projecting the equations onto the linear normal modes of the system.

## 8.1 EQUATIONS OF MOTION

### 8.1.1 Vertical Normal Modes of the Linear Equations

In this section we show that the continuously stratified Boussinesq equations have a close correspondence to the shallow water equations. In particular, if the equations are linearized and the flow is stably stratified, then each vertical mode of the continuous equations has the same form as the shallow water equations, with the modes being distinguished by the phase speed of the associated gravity waves.[1] We begin with the hydrostatic Boussinesq equations, linearized about a state of rest and with fixed stratification, $N(z)$.

$$\frac{\partial u}{\partial t} - fv = -\frac{\partial \phi}{\partial x}, \tag{8.1a}$$

$$\frac{\partial v}{\partial t} + fu = -\frac{\partial \phi}{\partial y}, \tag{8.1b}$$

$$0 = -\frac{\partial \phi}{\partial z} + b, \tag{8.1c}$$

$$\nabla \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0, \tag{8.1d}$$

$$\frac{\partial b}{\partial t} + wN^2 = 0. \tag{8.1e}$$

These equations are, respectively, the $u$ and $v$ momentum equations, the hydrostatic equation, the mass continuity equation and the buoyancy or thermodynamic equation, with the $\nabla$ operator being purely horizontal. The stratification, $N^2(z)$ is a time unchanging function of $z$ alone. The boundary conditions on these equations are The 'problem' with these equations is that there are five independent variables in three spatial coordinates so that even the linear problems are algebraically complex, especially when $f$ is variable. The equations are more general than is needed, because it is often observed that the vertical structure of solutions is relatively simple, especially in linear problems. A solution is to project the vertical structure onto appropriate eigenfunctions, and then to retain a very small number — often only one — of these eigenfunctions.

To determine what those eigenfunctions should be, we combine the hydrostatic and buoyancy equations to give

$$\frac{\partial}{\partial t}\left(\frac{\phi_z}{N^2}\right) + w = 0. \tag{8.2}$$

Differentiating with respect to $z$ and using the mass continuity equation gives

$$\frac{\partial}{\partial t}\left(\frac{\phi_z}{N^2}\right)_z - \nabla\cdot\boldsymbol{u} = 0. \tag{8.3}$$

It is this equation that motivates our choice of basis functions: we choose to expand the pressure and horizontal components of velocity in terms of an eigenfunction that satisfies the following Sturm–Liouville problem.

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{N^2}\frac{\mathrm{d}C_m}{\mathrm{d}z}\right) + \frac{1}{c_m^2}C_m = 0, \tag{8.4a}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}C_m(0) = \frac{\mathrm{d}}{\mathrm{d}z}C_m(-H) = 0 \tag{8.4b}$$

The eigenfunctions $C_m$ are complete and orthogonal in the sense that

$$\int_{-H}^{0} C_m C_n \,\mathrm{d}z = \frac{c_m^2}{g}\delta_{mn}. \tag{8.5}$$

where $\delta_{mn} = 0$ unless $m = n$, in which case it equals one. The normalization is somewhat by convention and we include a factor of $g$ for convenience to make the functions themselves nondimensional. There are an infinite number of eigenvalues, $c_m$, namely $c_0, c_1, c_2 \ldots$, normally arranged in descending order, and for each there is a corresponding eigenfunction $C_m$. The pressure and horizontal velocity components are then expressed as

$$[u, v, \phi] = \sum_{0}^{\infty}\left[u_m(x,y,t), v_m(x,y,t), \phi_m(x,y,t)\right] C_m(z). \tag{8.6}$$

A practical advantage of this procedure is that the $z$-derivatives in the equations of motion are replaced by multiplications, and in particular (8.3) becomes

$$\frac{\partial\phi_m}{\partial t} + c_m^2\nabla\cdot\boldsymbol{u}_m = 0. \tag{8.7a}$$

If we define $\eta^* = \phi/g$ then (8.7a) becomes

$$\frac{\partial\eta_m^*}{\partial t} + H_m^*\nabla\cdot\boldsymbol{u}_m = 0. \tag{8.7b}$$

where $H_m^* = c_m^2/g$ is the *equivalent depth* associated with the eigenmode.

Equations (8.7) are evidently of the same form as the familiar linear mass continuity equation in the shallow water equations, which may be written

$$\frac{\partial\eta}{\partial t} + H\nabla\cdot\boldsymbol{u} = 0 \qquad \text{or} \qquad \frac{\partial\hat{\eta}}{\partial t} + c^2\nabla\cdot\boldsymbol{u} = 0 \tag{8.8}$$

where $c = \sqrt{gH}$ and $\widehat{\eta} = g\eta$.

The horizontal momentum equations are simply,

$$\frac{\partial u_m}{\partial t} - f v_m = -\frac{\partial \phi_m}{\partial x}, \qquad \frac{\partial v_m}{\partial t} + f u_m = -\frac{\partial \phi_m}{\partial y}. \tag{8.9a,b}$$

Equations (8.8) and (8.9) are a closed set. If there is a forcing in the momentum equation then the transformed forcing appears on the right-hand sides of (8.9). If there is a source in the buoyancy equation then a corresponding term appears on the right-hand side of (8.8), analogous to a mass source term in the shallow water equations (see problem 8.**??**).

*Eigenfunctions for vertical velocity*

The vertical velocity and the buoyancy do not satisfy the same boundary conditions and so should not be expanded in the same way. Rather, we use eigenfunctions that satisfy the following relations.

$$\frac{1}{N^2}\frac{\mathrm{d}^2 S_m}{\mathrm{d}z^2} + \frac{1}{c_m^2} S_m = 0, \tag{8.10a}$$

$$S_m(0) = S_m(-H) = 0, \tag{8.10b}$$

where $S_m = 0$ if $N = 0$, and with the orthonormalization

$$\int_{-H}^{0} N^2 S_m S_n \,\mathrm{d}z = g\delta_{mn}. \tag{8.11}$$

These functions are related to $C_m$ by

$$C_m = \frac{c_m^2}{g}\frac{\mathrm{d}S_m}{\mathrm{d}z} \qquad N^2 S_m = -g\frac{\mathrm{d}C_m}{\mathrm{d}z}, \tag{8.12}$$

and $S_m = 0$ if $N = 0$.

Given the above, the vertical velocity may be evaluated from the mass continuity equation, $\partial w/\partial z = -\nabla\cdot\boldsymbol{u}$, which becomes

$$w_m \frac{\mathrm{d}S_m}{\mathrm{d}z} = -C_m \nabla\cdot\boldsymbol{u}_m \qquad \Rightarrow \qquad w_m = -\frac{c_m^2}{g}\nabla\cdot\boldsymbol{u}_m. \tag{8.13a,b}$$

*Approximations and interpretation*

The values of $c_m$ can be computed, in general, by solving the eigenvalue problem for the given stratification. In general this is a somewhat complex procedure that must be carried out numerically, but some approximate values can often be computed, especially if the stratification has certain simple forms.

First suppose $N$ is constant. The normal modes are are sines and cosines, and for $m = 1, 2 \ldots$ we have

$$C_m = A_m \cos\frac{m\pi z}{H}, \qquad S_m = B_m \sin\frac{m\pi z}{H}, \qquad c_m = \frac{NH}{m\pi}, \tag{8.14}$$

where, for $m > 0$, $A_m = c_m/\sqrt{gH/2}$ and $B_m = \sqrt{2g/HN^2}$. As an aside we note that the equivalent depth is then given by

$$H_e = \frac{N^2h^2}{m^2\pi^2 g} \sim \frac{g' H}{gm^2\pi^2}, \tag{8.15}$$

where we define $g' \equiv H\partial b/\partial z = -(gH/\rho_0)\partial\rho/\partial z$.

The mode with $m = 0$ is a special one, and is called the *barotropic mode*. With $N$ a constant (is that needed? xxx) we find

$$C_0 = A_0/2, \quad A_0 = \text{constant}, \qquad c_0^2 = gH \tag{8.16}$$

The expression for $c_0$ is particularly important and for an ocean of depth 5 km we find $c_0 \approx 20\,\mathrm{m\,s^{-1}}$, which is much higher than the velocity of fluid parcels or of the higher baroclinic modes (i.e., the modes with $m \geq 1$). For $N = 10^{-2}\,\mathrm{s^{-1}}$ and $H = 1$ km (the scale of depth of the thermocline) we find $c_1 \approx 3\,\mathrm{m\,s^{-1}}$, which is in agreement with numerical calculations that use a more realistic profile of stratification. (It is because the stratification is in reality concentrated in the upper ocean that such a value of $H$ leads to a reasonably realistic answer.)

If the stratification varies sufficiently slowly, WKB methods may be used to approximately evaluate the eigenvalues and eigenfunctions.[2] One may show (see the appendix, but not yet written! xxx) that

$$c_m \approx \frac{1}{m\pi}\int_{-H}^{0} N\,\mathrm{d}z, \qquad\qquad m > 1 \tag{8.17}$$

and

$$S_m \approx S_m^0 \sin\left(\frac{1}{c_m}\int_H^z N(z)\,\mathrm{d}z\right), \quad C_m \approx \left(\frac{c_m N S_m^0}{g}\right)\cos\left(\frac{1}{c_m}\int_{-H}^z N(z)\,\mathrm{d}z\right) \tag{8.18a,b}$$

If $N$ is constant these reduce to the results obtained above; for $N$ non-constant, the eigenfunctions are 'stretched' sines and cosines.

For much of the subsequent development in this chapter we will use the reduced-gravity shallow water form of the equations, rather than the normal mode form, because the notation is more familiar and the physical interpretation is a little simpler. However, this is a somewhat arbitrary choice, and it is always useful to remember that the equations have normal mode analogs that are valid for a continuously stratified ocean. [Need to check and finish this section. Perhaps put it earlier in the book. xxxx]

## 8.2 WAVES ON THE EQUATORIAL BETA PLANE

In this section we derive the dispersion relation and discuss the behaviour of Rossby waves and gravity waves at low latitudes.[3] For small variations in latitude we use, as in section 3.2 of *AOFD*, the $\beta$-plane approximation: we Taylor-expand the Coriolis parameter around a latitude $\vartheta_0$ and obtain

$$f = 2\Omega\sin\vartheta \approx 2\Omega\sin\vartheta_0 + 2\Omega(\vartheta - \vartheta_0)\cos\vartheta_0 = f_0 + \beta y \tag{8.19}$$

where $f_0 = 2\Omega\sin\vartheta_0$, $\beta = 2\Omega\cos\vartheta_0/a$ and $y = a(\vartheta - \vartheta_0)$. For motions at low latitudes we take $\vartheta_0 = 0$, giving the *equatorial beta-plane approximation* in which $\sin\vartheta \approx \vartheta$, $\cos\vartheta \approx 1$ and

$f = 2\Omega\vartheta = \beta y$. The linearized momentum and mass conservation equations are then

$$\frac{\partial u}{\partial t} - fv = -\frac{\partial \phi}{\partial x}, \qquad \frac{\partial v}{\partial t} + fu = -\frac{\partial \phi}{\partial y}, \tag{8.20a,b}$$

$$\frac{\partial \phi}{\partial t} + c^2\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0. \tag{8.20c}$$

To make a connection with the conventional shallow water equations we note that $\phi = g'\eta$, where $g'$ is the reduced gravity and $\eta$ the free surface height, and $c^2 = g'H$ where $H$ is the reference depth of the fluid.

Cross-differentiating (8.20a) and (8.20b) and using (8.20c) to eliminate the divergence we may also derive the linearized potential vorticity equation, namely

$$\frac{\partial}{\partial t}\left(\zeta - \frac{f\phi}{c^2}\right) + \beta v = 0 \tag{8.21}$$

This is the same as the familiar linearized potential vorticity equation on the $f$-plane, with the addition of the term $\mathrm{D}f/\mathrm{D}t = \beta v$. Equation (8.21) is not, of course, independent of (8.20) but it turns out to be convenient to use it. In all of the above equations, $f = \beta y$ and $\beta$ is a constant.

To obtain a single equation for a single unknown, operate on (8.20a) with $(f/c^2)\partial_t$, on (8.20b) with $(1/c^2)\partial_{tt}$, on (8.20c) with $(g'/c^2)\partial_{ty}$) and on (8.21) with $\partial_x$, where $c^2 = g'H$. Using subscripts to denote derivatives the resulting equations are

$$\frac{f}{c^2}u_{tt} - \frac{f^2}{c^2}v_t = -\frac{fg'}{c^2}\eta_{xt}, \tag{8.22a}$$

$$\frac{1}{c^2}v_{ttt} + \frac{f}{c^2}u_{tt} = -\frac{g'}{c^2}\eta_{ytt}, \tag{8.22b}$$

$$\frac{g'}{c^2}\eta_{tty} + (u_{xyt} + v_{yyt}) = 0, \tag{8.22c}$$

$$v_{xxt} - u_{xyt} - \frac{g'f}{c^2}\eta_{xt} + \beta v_x = 0. \tag{8.22d}$$

These equations linearly combine to give a single equation for $v$, namely

$$\boxed{\frac{1}{c^2}\frac{\partial^3 v}{\partial t^3} + \frac{f^2}{c^2}\frac{\partial v}{\partial t} - \frac{\partial}{\partial t}\left(\frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial x^2}\right) - \beta\frac{\partial v}{\partial x} = 0.} \tag{8.23}$$

If $f$ were constant and so $\beta = 0$ we could then straightforwardly obtain the following dispersion relations:

$$\omega = 0, \qquad \omega^2 = f_0^2 + (k^2 + l^2)c^2. \tag{8.24a,b}$$

The first is the dispersion relation for geostrophic waves (the frequency is zero in the absence of a beta effect) and the second is the dispersion relation for Poincaré waves, previously obtained in section 3.7.2. When beta is non-zero the situation is considerably complicated, and we address that below.

We also note one other common approximation, sometimes called the longwave approximation. If zonal scales are much greater than meridional scales then we expect the zonal wind to be in geostrophic balance with the meridional pressure gradient. In this case we replace (8.20b) by

$$fu = -g' \frac{\partial \eta}{\partial y}. \tag{8.25}$$

This equation combines with (8.20a,c) to give

$$\frac{f^2}{c^2}\frac{\partial v}{\partial t} - \frac{\partial}{\partial t}\left(\frac{\partial^2 v}{\partial y^2}\right) - \beta \frac{\partial v}{\partial x} = 0. \tag{8.26}$$

This equation is first order in time and the dispersion relation may be obtained reasonably straightforwardly. This approximation is particularly useful in the forced-dissipative problem as we will see in section 8.4. In the free problem the dispersion equation can in fact be obtained easily enough in the general case, that is from (8.23) as we see below, allowing us to make the longwave approximation at a later stage.

### 8.2.1 Dispersion Relations

In this section we explore the properties of (8.23), in particular obtaining a dispersion relation. Our treatment is initially rather mathematical and formal, but we will follow this by a more physical discussion. [xxx Do we?]

The coefficients of (8.23) vary in the meridional direction but are constant in the zonal direction. We thus search for solutions in the form of a plane wave in the zonal direction only and we let

$$v = \widetilde{v}(y) e^{i(kx - \omega t)}, \tag{8.27}$$

and assume boundary conditions of $\widetilde{v}(y) \to 0$ as $y \to \pm\infty$. Substituting (8.27) into (8.23) gives

$$\frac{d^2 \widetilde{v}}{dy^2} + \left(\frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega} - \frac{\beta^2 y^2}{c^2}\right)\widetilde{v} = 0. \tag{8.28}$$

Given the velocity, $c$ and the presence of the beta effect there is a rather obvious way to nondimensionalize the equations. However, it turns out that by introducing an additional factor of $\sqrt{2}$ into the scaling the mathematics of one of the problems that we address later is simplified. At the risk of wasting a page on a seemingly trivial difference, let's do both. The confident and impatient reader may choose one and skim the other.

#### *Nondimensionalization I*

Let us scale time and distance with the quantities

$$T_{eq} = (c\beta)^{-1/2}, \qquad L_{eq} = (c/\beta)^{1/2} \tag{8.29a,b}$$

where $c \equiv \sqrt{g'H}$. The timescale $T_{eq}$ is related to the lengthscale $L_{eq}$ by $T_{eq} = (L_{eq}\beta)^{-1}$, and the non-dimensional frequency, lengthscale and wavenumber are then given by

$$\widehat{\omega} = \frac{\omega}{(\beta c)^{1/2}}, \qquad \widehat{y} = y\left(\frac{\beta}{c}\right)^{1/2}, \qquad \widehat{k} = k\left(\frac{c}{\beta}\right)^{1/2}. \tag{8.30}$$

The length scale $L_{eq}$ is known as the *equatorial radius of deformation,* and it is a natural scale over which equatorial disturbances decay, as will become apparent very soon. If we take $\delta\rho/\rho_0 = 0.002$, $H = 100\,\mathrm{m}$ and $\beta = 2\Omega/a = 2.3 \times 10^{-11}\,\mathrm{m^{-1}\,s^{-1}}$ then we obtain

$$g' \approx 0.02\,\mathrm{m\,s^{-2}}, \quad c \approx 1.4\,\mathrm{m\,s^{-1}}, \quad L_{eq} \approx 250\,\mathrm{km}, \quad T_\beta = 1.7 \times 10^5\,\mathrm{s} \approx 2\,\mathrm{days}. \tag{8.31}$$

The shallow-water mid-latitude deformation radius, $L_d$ is usually defined as $L_d = c/f$ which differs from (8.29b) most notably in the power of $f$. However, if in mid-latitude expression we take $f = \beta y$, as if near the equator, and $y = L_d$, then $L_d = c/(\beta L_d)$, which is the same as (8.29b). (For the stratified equations we may define the deformation radii as $L_m = c_m/f$, in which case one obtains a sequence of values for each value of $m$.)

Substituting (8.30) into (8.28) gives the slightly simpler-looking equation

$$\frac{\mathrm{d}^2 v}{\mathrm{d}\widehat{y}^2} + \left(\widehat{\omega}^2 - \widehat{k}^2 - \frac{\widehat{k}}{\widehat{\omega}} - \widehat{y}^2\right) v = 0. \tag{8.32}$$

This equation may be put into a standard form[4] by writing

$$v(\widehat{y}) = \Psi(\widehat{y})\mathrm{e}^{-\widehat{y}^2/2}, \tag{8.33}$$

whence (8.32) becomes

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}\widehat{y}^2} - 2\widehat{y}\frac{\mathrm{d}\Psi}{\mathrm{d}\widehat{y}} + \lambda\Psi = 0 \tag{8.34}$$

where $\lambda = \widehat{\omega}^2 - \widehat{k}^2 - \widehat{k}/\widehat{\omega} - 1$. Equation (8.34) is known as *Hermite's equation,* and it is an eigenvalue equation, with solutions if and only if $\lambda = 2m$, for $m = 0, 1, 2, \ldots$. The solutions are Hermite polynomials, $\Psi(\widehat{y}) = H_m(\widehat{y})$, where the first few polynomials are given by

$$H_0 = 1, \qquad H_1 = 2\widehat{y}, \qquad H_2 = 4\widehat{y}^2 - 2, \tag{8.35a}$$

$$H_3 = 8\widehat{y}^3 - 12\widehat{y}, \qquad H_4 = 16\widehat{y}^4 - 48\widehat{y}^2 + 12. \tag{8.35b}$$

A Hermite polynomial is even or odd when $m$ is even or odd, respectively; that is $H_m(-\widehat{y}) = (-1)^m H_m(\widehat{y})$. Hermite polynomials multiplied by a Gaussian are a form of *parabolic cylinder function*,

$$V_m(y) = H_m(y)\exp(-y^2/2). \tag{8.36}$$

These functions are also orthogonal in the interval $[-\infty, +\infty]$; that is

$$\int_{-\infty}^{\infty} V_n V_m \,\mathrm{d}y = \int_{-\infty}^{\infty} H_n(y) H_m(y)\exp(-y^2)\,\mathrm{d}y = \sqrt{\pi}2^n n!\,\delta_{nm}, \tag{8.37}$$

See also the appendix at the end of this chapter for additional details.

Given the Hermite solution for $\Psi$, the solutions for $v$ are given by

$$v(\widehat{y}) = V_m(\widehat{y}) = H_m(\widehat{y})\mathrm{e}^{-\widehat{y}^2/2}, \qquad m = 0,\ 1,\ 2\ldots \tag{8.38}$$

and so decay exponentially as $\widehat{y} \to \pm\infty$ (as we require) with a decay scale of the equatorial deformation radius $\sqrt{c/\beta}$. The functions $V_m$ are plotted in Fig. 8.1 for $m = 0$ to 3.

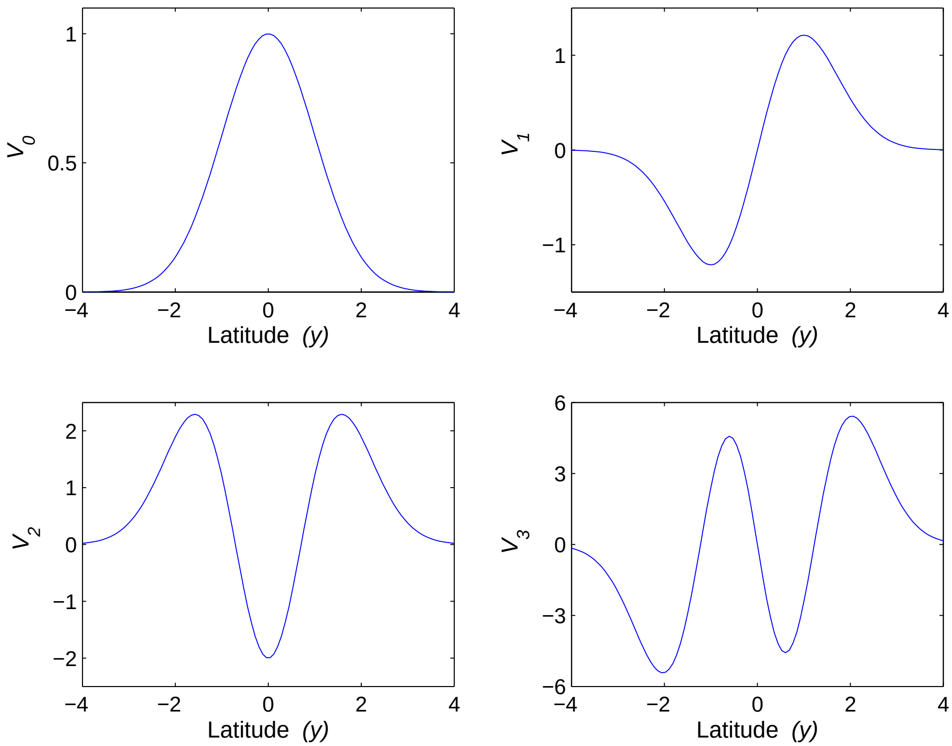


**Fig. 8.1** Latitudinal variation of the wave amplitudes, $V_m(y)$, given by (8.38) as a function of non-dimensional latitude, $\widehat{y}$ for $m = 0, 1, 2, 3$. The parameter $m$ is analogous to a meridional wavenumber. The parabolic cylinder functions given by (8.46) have a similar form.

The dispersion relation follows from the quantization condition $\lambda = 2m$, which implies

$$\widehat{\omega}^2 - \widehat{k}^2 - \frac{\widehat{k}}{\widehat{\omega}} = 2m + 1 \tag{8.39a}$$

or, using (8.30), the dimensional form,

$$\boxed{\omega^2 - c^2k^2 - \beta\frac{kc^2}{\omega} = (2m + 1)\beta c}\,, \tag{8.39b}$$

This is a cubic equation in $\omega$, and although a general solution is possible, it is easier to solve the quadratic equation for the wavenumber in terms of the frequency giving

$$\widehat{k} = -\frac{1}{2\widehat{\omega}} \pm \frac{1}{2}\left[\left(\frac{1}{\widehat{\omega}} - 2\widehat{\omega}\right)^2 - 8m\right]^{1/2}, \tag{8.40a}$$

or, in dimensional form,

$$\boxed{k = -\frac{\beta}{2\omega} \pm \frac{1}{2}\left[\left(\frac{\beta}{\omega} - \frac{2\omega}{c}\right)^2 - \frac{8m\beta}{c}\right]^{1/2}}. \tag{8.40b}$$

*Nondimensionalization II*

We now scale time and distance with the quantities

$$T_{eq} = (2c\beta)^{-1/2}, \qquad L_{eq} = (c/2\beta)^{1/2}. \tag{8.41a,b}$$

Velocity is still nondimensionalized by $c$. The nondimensional version of (8.28) becomes

$$\frac{\mathrm{d}^2\widetilde{v}}{\mathrm{d}\widehat{y}^2} + \left(\widehat{\omega}^2 - \widehat{k}^2 - \frac{\widehat{k}}{2\widehat{\omega}} - \frac{\widehat{y}^2}{4}\right)\widetilde{v} = 0. \tag{8.42}$$

We now make the substitution

$$\widetilde{v}(\widehat{y}) = \Phi \exp(-\widehat{y}^2/4). \tag{8.43}$$

and this leads to

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}\widehat{y}^2} - \widehat{y}\frac{\mathrm{d}\Phi}{\mathrm{d}\widehat{y}} + \gamma\Phi = 0 \tag{8.44}$$

where $\gamma = \widehat{\omega}^2 - \widehat{k}^2 - \widehat{k}/2\widehat{\omega} - 1/2$. Naturally, (8.44) could be transformed into (8.34) by changing to the independent variable $y' = \widehat{y}/\sqrt{2}$, and the dispersion relation then follows in the same way. More directly, solutions of (8.44) are given by the modified Hermite polynomials $\Phi(\widehat{y}) = G_m(\widehat{y})$ where

$$(G_0, G_1, G_2, G_3, G_4) = (1, \widehat{y}, \widehat{y}^2 - 1, \widehat{y}^3 - 3\widehat{y}, \widehat{y}^4 - 6\widehat{y}^2 + 3). \tag{8.45}$$

These are sometimes known as the modified or probabilists' Hermite polynomials, with (8.155) being the physicists' Hermite polynomials, reflecting historical use; the two sets of polynomials are connected by $H_n(y) = 2^{n/2}G_n(y\sqrt{2})$. The corresponding parabolic cylinder functions are given by

$$D_n(\widehat{y}) = G_n(\widehat{y}) \exp(-\widehat{y}^2/4). \tag{8.46}$$

and these functions are solutions of (8.42). The orthonormality condition on the modified polynomials is that

$$\int_{-\infty}^{\infty} D_n(y)D_m(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} G_n(y)\,G_m(y)\exp(-y^2/2)\,\mathrm{d}y = \sqrt{2\pi}n!\,\delta_{nm}, \tag{8.47a}$$

which may be compared to (8.37). The quantization condition on $\gamma$ is that $\gamma = m$, where $m = 0, 1, 2, \dots$. Thus, the nondimensional dispersion relation is

$$\widehat{\omega}^2 - \widehat{k}^2 - \frac{\widehat{k}}{2\widehat{\omega}} - \frac{1}{2} = m \tag{8.48}$$

and restoring the dimensions using (8.41) gives (8.39b). Later on, when dealing with the steady, forced-dissipative problem, the use of the probabilists' polynomials turns out to be more convenient because of the form of certain ladder operators connecting functions of different order.

### 8.2.2 Limiting and special cases

For the wave case we will, for definiteness, stay with the first nondimensionalization, namely (8.29), and with the goal of figuring out what's going on we'll consider various special cases of the dispersion relations (8.39) and (8.40). It is convenient to first partition the waves by frequency, and consider separately high frequency gravity waves and low frequency planetary waves. We need do this only for the case $m \geq 1$ because the $m = 0$ case (mixed Rossby-gravity waves) may be treated exactly. Then finally we look at the so-called $m = -1$ case, namely Kelvin waves.

*High and low frequency waves*

*(i) High frequency waves.* The term $\beta k c^2/\omega$ in (8.39) is small and may be neglected. The dispersion relation becomes

$$\widehat{\omega}^2 = \widehat{k}^2 + 2m + 1 \qquad \text{or} \qquad \omega^2 = c^2 k^2 + \beta c(2m + 1). \tag{8.49a,b}$$

This dispersion relation is similar to that of mid-latitude Poincaré waves, with $\beta c$ replacing $f_0^2$: recall the form of (3.103), namely $\omega^2 = c^2(k^2 + l^2) + f_0^2$. Waves satisfying (8.49) are thus sometimes called equatorially trapped Poincaré waves or equatorially trapped gravity waves.

The approximation requires that $\omega \gg \beta/|k|$, and is somewhat inaccurate for small $k$: note that (8.49) is symmetric around $k = 0$, whereas the full dispersion relation, plotted in Fig. 8.2, is offset. (Formally, the limit is valid for $\widehat{k} \to \infty$, $\widehat{\omega} \to \infty$ and $\widehat{k}/\widehat{\omega} =$ constant.)

For finite $m$ the limiting case at high wavenumber just $\widehat{\omega} = \pm\widehat{k}$, or, in dimensional form, $\omega = \pm ck$. This is just the dispersion relation for familiar conventional shallow water gravity waves, unaffected by rotation and the $\beta$-effect. However, in the rotating case the waves are trapped at the equator and propagate only in the zonal direction, albeit both eastward and westward.

*(ii) Low frequency waves.* For low frequency waves we neglect the term involving $\omega^2$ in (8.39) and the dispersion relation becomes

$$\widehat{\omega} = \frac{-\widehat{k}}{2m + 1 + \widehat{k}^2}, \qquad \omega = \frac{-\beta k}{(2m + 1)\beta/c + k^2}, \tag{8.50}$$

non-dimensionally and dimensionally, respectively. This is recognizable as the dispersion relation for a zonally propagating Rossby wave with large $x$-wavenumber, and these waves are called equatorially trapped Rossby waves, or equatorially trapped planetary waves. We may further consider two limits of these waves, as follows.

*(a) Short, low frequency waves, with* $\widehat{k} \to \infty$, $\widehat{\omega} \to 0$. The dispersion relation becomes

$$\widehat{\omega} = -\frac{1}{\widehat{k}}, \qquad \omega = -\frac{\beta}{k}. \tag{8.51}$$

The phase speed and group velocity in this limit are given by, dimensionally,

$$c_p = -\frac{\beta}{k^2}, \qquad c_g = \frac{\beta}{k^2}, \tag{8.52}$$

Thus, the phase speed is westward but the group velocity, and so the direction of energy propagation, is eastward.

(b) *Long low frequency waves, with $\widehat{k} \to 0$, $\widehat{\omega} \to 0$.* The dispersion relation (8.39) becomes, in nondimensional and dimensional form,

$$\widehat{\omega} = \frac{-\widehat{k}}{2m+1}, \qquad \omega = \frac{-ck}{2m+1}, \tag{8.53}$$

These represent westward propagating waves whose speed is given by $c/(2m+1)$, similar to that of a gravity wave. However, like planetary waves they propagate only westward, and they match with the westward propagating planetary waves derived above as wavenumber increases. They are conveniently nondispersive, and are important near western boundaries where they superpose to create western boundary currents.

The longwave approximation may be made from the outset, and is equivalent to assuming that the zonal flow is in geostrophic balance; that is, (8.20b) is replaced by $fu = -g'\partial\eta/\partial y$. Then, instead of solving (8.23) we solve (8.26). The only difference is in the value of $\lambda$ in (8.34) — we find $\lambda = -\widehat{k}/\widehat{\omega} - 1$ — and so (8.53) immediately emerges. Short waves are filtered out of the system. This approximation will turn out to be particularly important when we consider the steady problem in section 8.5.

There is a distinct gap in frequencies between the minimum frequency of the gravity waves, given by (8.49), and the maximum frequency of the planetary waves, given by (8.52) also with $m$ small. The minimum gravity wave frequency occurs when $k = 0$ and is $\omega_{gmin}^2 = \beta c(2m+1)$. From (8.50) the maximum planetary wave frequency occurs when $k^2 = (2m+1)\beta/c$ and gives $\omega_{pmax}^2 = \beta c/[4(2m+1)]$. The ratio of these two frequencies is

$$\frac{\omega_{gmin}}{\omega_{pmax}} = 2(2m+1), \tag{8.54}$$

giving a value of six for $m = 1$ and two for $m = 0$ (a case we consider more below). Note that this ratio is *independent* of the values of the physical parameters $\beta$ and $c$. Although the gap is distinct, it is not as large as the corresponding gap at midlatitudes, which may be an order of magnitude or more.

### *Special values of m*

In addition to consider limiting cases at low and high frequency, there are two other cases in which we can readily solve the dispersion relation, namely the case with $m = 0$ and the Kelvin wave case, as follows.

(i) *The case with $m = 0$.* The resulting waves are known as *Yanai waves*.[5] From (8.39a) the dispersion relation simplifies to the two cases

$$\widehat{k} = -\widehat{\omega}, \qquad \widehat{k} = -\frac{1}{\widehat{\omega}} + \widehat{\omega}. \tag{8.55a,b}$$

or dimensionally

$$k = -\frac{\omega}{c}, \qquad k = -\frac{\beta}{\omega} + \frac{\omega}{c}. \tag{8.56a,b}$$

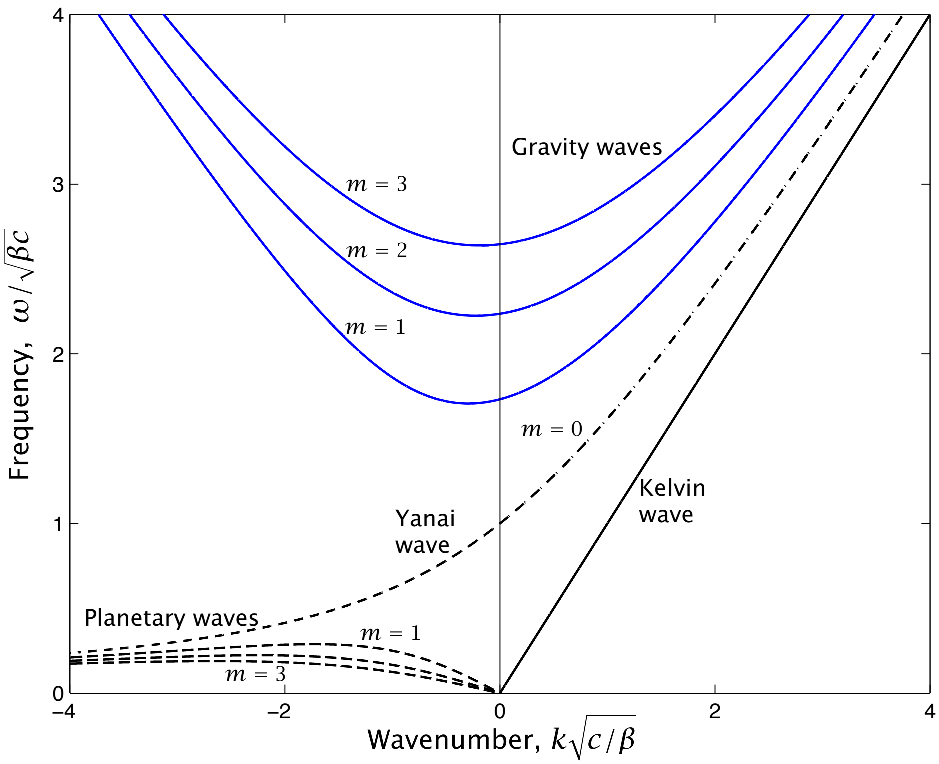


**Fig. 8.2** Dispersion relation for equatorial waves, as given by (8.39), for $m = 0, 1, 2, 3$. The upper group of curves are gravity waves, given approximately by (8.49). The lower group with $k < 0$ are westward propagating planetary waves, given approximately by (8.50). Also shown are the Yanai wave with $m = 0$, satisfying (8.57), and the eastward propagating Kelvin wave (the '$m = -1$' wave) satisfying $\omega = ck$ for $k \geq 0$. [Also need plots of the waves in physical space? xxx]

The case $k = -\omega/c$ is non-physical, for it represents a gravity wave moving westward. Such wave grows without bound as $|y|$ increases away from the equator, as we demonstrate explicitly in the discussion on Kelvin waves below. The physically realizable case, (8.56b) has the explicit dispersion relation

$$\omega = \frac{kc}{2} \pm \frac{1}{2}\sqrt{k^2c^2 + 4\beta c}. \tag{8.57}$$

Again it is useful to consider limiting cases, as follows.

- $k = 0$. In this case (8.57) gives $\omega = \sqrt{\beta c}$ and there is a balance between the two terms on the right-hand side of (8.56b). Note that in Fig. 8.2 the Yanai wave at $k = 0$ intercepts the ordinate at a value of nondimensional frequency of 1.
- $k \to +\infty$. In this case $\omega = ck$, with a balance between the left-hand side and the second term on the right-side of (8.56b). Evidently, this corresponds to eastward propagating gravity waves.

- $k \to -\infty$. In this case, because $\omega$ must be positive, we have $\omega = -\beta/k$, and a balance between the left-hand side and the first term on the right-side of (8.56b). The waves are westward propagating Rossby or planetary waves.

Yanai waves, therefore, are mixed Rossby-gravity waves: the phase of the Rossby wave propagates westward (like all Rossby waves) and has a low frequency, and the gravity wave propagates eastward (and only eastward, unlike conventional gravity waves). The group velocity of Yanai waves is positive in all cases, being given by, from (8.56b),

$$c_g^x \equiv \frac{\partial \omega}{\partial k} = \frac{\omega^2 c}{\beta c + \omega^2}. \tag{8.58}$$

The group velocity of the full problem is, from (8.39),

$$c_g^x = \frac{c^2\omega(\beta + 2\omega k)}{2\omega^3 + \beta k c^2}. \tag{8.59}$$

This may be positive or negative, and vanishes when $\omega = -\beta/(2k)$.

*(ii) Kelvin waves, or the '$m = -1$' case.* In general, Hermite's equation, (8.34), has solutions when $m$ is a positive integer or zero. However, there is a class of waves that also satisfies the dispersion relation (8.39) with $m = -1$, namely equatorial Kelvin waves, as we shall now discover. (This section may be considered to be an extension of section 3.7.3.)

Kelvin waves have identically zero meridional velocity and so their equations of motion are

$$\frac{\partial u}{\partial t} = -g'\frac{\partial \eta}{\partial x}, \qquad fu = -g'\frac{\partial \eta}{\partial y}, \qquad \frac{\partial \eta}{\partial t} + H\frac{\partial u}{\partial x} = 0. \tag{8.60a,b,c}$$

where $f = \beta y$. The zonal velocity is in geostrophic balance with the meridional pressure gradient, and (8.60a) and (8.60c) give the classic wave equation,

$$\frac{\partial^2 u}{\partial t^2} - c^2\frac{\partial^2 u}{\partial x^2} = 0. \tag{8.61}$$

where $c = \sqrt{g'H}$ as before, and so the dispersion relation $\omega = \pm ck$. This is, in fact, a solution of (8.39) with $m = -1$, as may easily be checked.

The solution to (8.61), and the corresponding solution for $\eta$, is

$$u = F_1(x + ct, y) + F_2(x - ct, y), \qquad \eta = \sqrt{\frac{H}{g'}}\,[-F_1(x + ct, y) + F_2(x - ct, y)] \tag{8.62}$$

where $F_1$ and $F_2$ are arbitrary functions, representing waves travelling westwards and eastwards, respectively. We obtain the $y$-dependence of these functions by using (8.60b) giving

$$\beta y F_1 = c\frac{\partial F_1}{\partial y}, \qquad \beta y F_2 = -c\frac{\partial F_2}{\partial y}. \tag{8.63}$$

The solutions of these equations are

$$F_1 = F(x + ct)\exp[y^2/(2L_{eq}^2)], \qquad F_2 = G(x - ct)\exp[-y^2/(2L_{eq}^2)] \tag{8.64a,b}$$

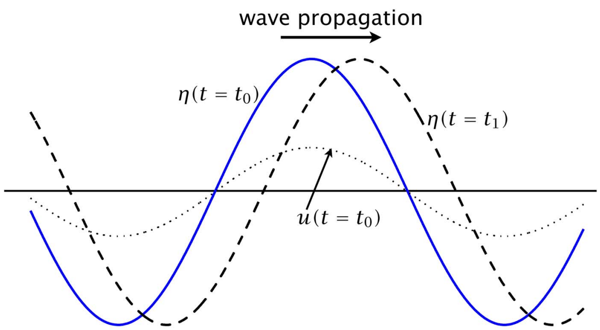


**Fig. 8.3** A shallow water gravity wave, showing the fluid interface at an initial and later time $\eta(t_0)$ and $\eta(t_1)$, and the fluid velocity at the initial time, $u(t_0)$. The fluid flow is in the same direction as the phase speed (i.e., positive in this example) under the fluid crests, and is in the opposite direction under the troughs.

where $F$ and $G$ are the amplitudes at $y = 0$. Evidently, $F_1$ increases without bound away from the equator, and so this solution must be eliminated. The complete solution is thus:

$$u = G(x - ct)\exp[-y^2/(2L_{eq}^2)], \qquad \eta = \frac{H}{c}u, \qquad v = 0. \tag{8.65}$$

with dispersion relation

$$\omega = ck. \tag{8.66}$$

These waves are equatorially trapped Kelvin waves. They propagate eastward only, without dispersion, and their amplitude decays away from the equator in precisely the same way as the other equatorial waves considered above.

### 8.2.3 Why do Kelvin waves have a preferred direction of travel?

Both equatorial and coastal Kelvin waves have a preferred direction of travel: equatorial Kelvin waves move eastward and, consistent with this, coastal Kelvin waves travel such that they have a wall to their right in the Northern Hemisphere and to their left in the Southern Hemisphere. Why is this?

Consider the linearized zonal momentum and mass continuity equations,

$$\frac{\partial u}{\partial t} = -g'\frac{\partial \eta}{\partial x}, \qquad \frac{\partial \eta}{\partial t} = -H\frac{\partial u}{\partial x}. \tag{8.67}$$

Looking for wavelike solutions of the form $(u, h) = (\tilde{u}, \tilde{\eta})e^{i(kx-\omega t)}$ we obtain $\tilde{u} = g'\tilde{\eta}/c$ and $c\tilde{\eta} = H\tilde{u}$. This means that under the crests of fluid (i.e., positive values of $\eta$) $u$ has the same sign as $c$; the parcels of fluid are moving in the same direction as the phase of the wave. This property is also apparent if once considers how the fluid must move in order that the troughs and crests progress in a a particular direction, as illustrated in Fig. 8.3. This property holds for shallow water waves quite generally, and is not restricted to Kelvin waves.

Now we restrict attention to Kelvin waves. In the direction perpendicular to the direction of travel of the wave, the flow is in geostrophic balance:

$$fu = -g'\frac{\partial \eta}{\partial y}. \tag{8.68}$$

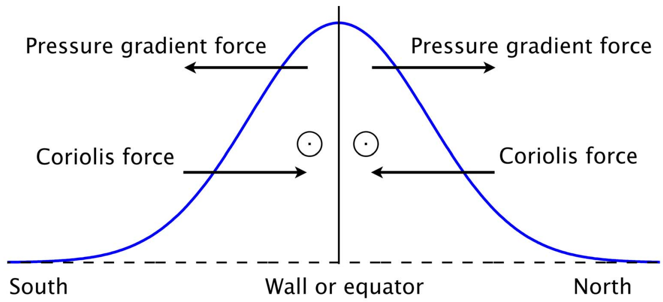


**Fig. 8.4** Balance of forces across a Kelvin wave; the solid line is the fluid surface and the phase speed is directed out of the page. Beneath a crest, as shown, the fluid flow is in the direction of the phase speed and produces Coriolis forces in the directions shown, so balancing the pressure gradient forces. If the wave were travelling in the other direction no such geostrophic balance could be achieved.

Consider the flow under a fluid crest in an equatorial Kelvin wave, as illustrated in Fig. 8.4. The pressure gradient force is directed away from the equator and, if the wave is travelling eastward the pressure force can balanced by the Coriolis force directed toward the equator. Under a trough the fluid is flowing in the opposite direction to the wave itself, and both the pressure gradient force and the Coriolis force are reversed and geostrophic balance still holds. If the wave were to travel westwards, no such balances could be achieved.

Very similar reasoning holds for coastal Kelvin waves, that is, Kelvin waves on an $f$-plane with a cross-wave pressure gradient supported by a wall. Geostrophic balance can now be maintained only if the wall is to the right of the direction of travel in the Northern Hemisphere (where $f > 0$) and to the left in the Southern Hemisphere (where $f < 0$).

### 8.2.4 Potential vorticity dynamics of equatorial Rossby waves

The Rossby waves and Rossby-gravity waves derived above are rather similar to their mid-latitude counterparts, which can be derived from a balanced potential vorticity equation without involving unbalanced dynamics at all. Can we do something similar for equatorial Rossby waves? The answer is yes, as we now show.[6] The method is somewhat ad hoc, but informative. Kelvin waves and inertia-gravity waves are filtered out, but Rossby waves and Rossby-gravity waves are reproduced in a way that transparently illuminates their dynamics.

Let us begin with the unforced linearized potential vorticity equation which to remind ourselves, is

$$\frac{\partial}{\partial t}\left(\zeta - \frac{f\phi}{c^2}\right) + \beta v = 0, \tag{8.69}$$

where, as before, $f = \beta y$. Let us now suppose that the divergence is small and the flow close to geostrophic balance so that the velocity, vorticity and height fields can all be written in terms of a streamfunction,

$$u = -\frac{\partial \psi}{\partial y}, \qquad v = \frac{\partial \psi}{\partial x}, \qquad \zeta = \nabla^2\psi, \qquad \phi = f\psi. \tag{8.70}$$

This is similar to what is done in the quasi-geostrophic approximation, except that here the Coriolis parameter is allowed to vary, with $f = \beta y$. Equation (8.70) is best regarded as an ansatz — by which we mean an approximation or assumption made for convenience — for it has not been rigorously justified by any scaling approximation.

Using (8.70) in (8.69) gives

$$\frac{\partial}{\partial t}\left(\nabla^2\psi - \frac{f^2\psi}{c^2}\right) + \beta\frac{\partial\psi}{\partial x} = 0. \tag{8.71}$$

We can seek wavelike solutions of this in the form

$$\psi = \widetilde{\psi}(y)\mathrm{e}^{\mathrm{i}(kx-\omega t)}, \tag{8.72}$$

and (8.72) becomes

$$\frac{\mathrm{d}^2\widetilde{\psi}}{\mathrm{d}y^2} + \left(-k^2 - \frac{\beta k}{\omega} - \frac{\beta^2 y^2}{c^2}\right)\widetilde{\psi}. \tag{8.73}$$

This is almost the same as (8.28) except for the replacement of $\widetilde{v}$ by $\widetilde{\psi}$ and the absence of the $\omega^2$ term in the bracketed expression. Regarding the first difference, the meridional velocity is just $\partial\psi/\partial x \propto k\psi$, so the meridional velocity obeys the same equation as $\widetilde{\psi}$. The second difference arises because we are, through our use of (8.70), only considering the low-frequency limit. Given that equations (8.73) and (8.28) have the same form, we simply repeat the development following (8.28) and obtain a dispersion relation similar to (8.39b) but without the $\omega^2$ term, to wit

$$\omega = \frac{-\beta k}{(2m+1)\beta/c + k^2}. \tag{8.74}$$

This is the same as the dispersion relation for low frequency waves discussed in section 8.2.2. The balanced system (8.71) thus *exactly* reproduces the Rossby waves and Rossby-gravity waves in the low frequency limit. However, we are not able to recover the behaviour of Kelvin waves by this methodology because such waves are essentially non-balanced: in the meridional direction the Coriolis force balances the height field, as in (8.68), but in the zonal direction there is a balance between the zonal acceleration and the pressure gradient.

## 8.3 RAY TRACING AND EQUATORIAL TRAPPING

We have seen that equatorial waves are trapped near the equator. What then happens to a wave that initially propagates in a direction away from the equator? The waves must either change their character completely, or be refracted back toward the equator. The former can only happen if there exists a class of midlatitude waves with similar frequency and wavenumber; otherwise no such waves can be excited and the waves must, if they are not absorbed, bend back if energy is to be conserved. Let us explore this using some ideas from ray theory (see the appendix to chapter 5 and section 13.2 of *AOFD*).

### 8.3.1 Dispersion relation and ray equations

Consider again the wave equation of motion for the meridional velocity, namely (8.28) or

$$\frac{\mathrm{d}^2\widetilde{v}}{\mathrm{d}y^2} + \left(\frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega} - \frac{\beta^2 y^2}{c^2}\right)\widetilde{v} = 0. \tag{8.75}$$

If the term in brackets is positive then sinusoidal solutions in $y$ are possible, but if the term is negative, which will occur for $y$ larger than some critical value $y_c$, then the physically realizable solutions decay exponentially with $y$; that is, wavelike solutions are trapped between two critical latitudes. Using the dispersion relation (8.39), equation (8.75) becomes

$$\frac{d^2\tilde{v}}{dy^2} + \left(\frac{(2m+1)\beta}{c} - \frac{\beta^2 y^2}{c^2}\right)\tilde{v} = 0. \tag{8.76}$$

and therefore the critical latitudes are given by

$$y_c = \pm\left(\frac{\omega^2}{\beta^2} - \frac{c^2k^2}{\beta^2} - \frac{c^2k}{\beta\omega}\right)^{1/2} = \left((2m+1)\frac{c}{\beta}\right)^{1/2}, \tag{8.77}$$

For $k = 0$, and so for meridionally propagating waves, the critical latitudes are given by $y_c = \omega/\beta$, and so at the critical latitude $\omega = f$. The waves are therefore trapped within their *inertial latitudes,* the latitudes at which their frequency is $f$. For larger $k$ the critical latitudes are correspondingly smaller.

To explore this phenomenon using ray theory we assume that the medium is varying sufficiently slowly that it is possible to find wavelike solutions with spatially varying wavenumbers. Here the medium varies only in the $y$-direction (because of the $\beta$ effect) and so, instead of the more general form (8.27), we seek solutions of (8.23) in the form

$$v = V(y)e^{i(kx+l(y)y-\omega t)}, \tag{8.78}$$

where we assume that $V(y)$ varies slowly enough with $y$ so that its derivatives are small. This procedure is the same as letting $\tilde{v}(y) = V(y)e^{il(y)y}$ in (8.75), [check for consistency with WKB xxx] which then gives an equation for $l$,

$$l^2 = \frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega} - \frac{\beta^2 y^2}{c^2} = \frac{\beta^2}{c^2}(y_c^2 - y^2). \tag{8.79}$$

Writing (8.75) as

$$\frac{d^2V}{dy^2} + l^2(y)V = 0, \tag{8.80}$$

the approximate solution, as given by WKB methods (see the appendix to chapter 7), is

$$V = l^{-1/2}\exp\left(\pm i\int l\, dy\right). \tag{8.81}$$

The formal condition for the validity of this solution is

$$\frac{d^2 l^{-1/2}}{dy^2} \ll l^{3/2}, \tag{8.82}$$

which roughly speaking says that the meridional scale of the wave, $l^{-1}$ should be small compared to the scale over which $l$ varies.

Wave packets travel along rays — paths that are parallel to the direction of the group velocity. That is, their trajectory, $x(t)$, $y(t)$ is defined by

$$\frac{dx}{dt} = c_g^x, \qquad \frac{dy}{dt} = c_g^y, \quad \text{so that} \quad \frac{dy}{dx} = \frac{c_g^y}{c_g^x}. \tag{8.83}$$

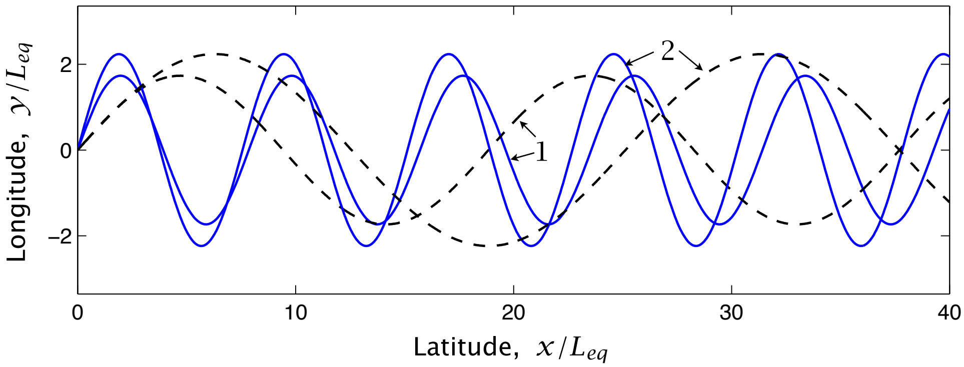


**Fig. 8.5** Rays in the equatorial waveguide calculated using (8.88). The dashed lines show planetary wave trajectories and the solid lines are gravity wave trajectories, with $m = 1, 2$ (numbers marked on the graph) and $\widehat{k} = 1$. The turning latitude for each wave is $(2m+1)^{1/2}L_{eq}$, where $L_{eq} = \sqrt{c/\beta}$.

Using the dispersion relation (8.79) gives

$$\frac{\partial \omega}{\partial l} = \frac{2\omega^2 l c^2}{2\omega^3 + \beta k c^2}, \tag{8.84}$$

and using this and (8.59) gives the slope of the ray in the $x$-$y$ plane,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{c_g^y}{c_g^x} = \frac{l}{k + \beta/(2\omega)}. \tag{8.85}$$

Using the expression for $l$ given by (8.79) we can write this in terms of $y$ instead of $l$, so that

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\beta(y_c^2 - y^2)^{1/2}}{kc + \beta c/(2\omega)}. \tag{8.86}$$

Using the standard result that

$$\int \frac{\mathrm{d}y}{(y_c^2 - y^2)^{1/2}} = \sin^{-1}\frac{y}{y_c}, \tag{8.87}$$

we finally obtain

$$y = y_c \sin\left[\frac{\beta x}{ck + \beta c/(2\omega)}\right], \qquad \widehat{y} = (2m+1)^{1/2}\sin\left[\frac{\widehat{x}}{\widehat{k} + 1/(2\widehat{\omega})}\right]. \tag{8.88}$$

where the second expression is the nondimensional form. The ray path is therefore a sinusoid moving along the equator; the waves are confined to a *waveguide* centered at the equator and with a polewards extent of $y = \pm y_c$. Equation (8.88) holds for both planetary and gravity waves, and for the latter the term $\beta c/(2\omega)$ may be neglected.

### 8.3.2 Discussion

xxx Some observations and implications? Suggestions welcome.

## 8.4 FORCED-DISSIPATIVE WAVELIKE FLOW

[In this section we go back and forth between dimensional and nondimensional variables without changing notation. Maybe that's okay, as it would be clumsy otherwise. xxx]

We now consider linear equatorial dynamics in the presence of forcing. Because there is a forcing we must introduce a damping so that a steady state can be reached, and the simplest form is a linear drag. From a physical perspective the presence of such a drag is the most unsatisfactory aspect of our treatment, for it has no real physical justification especially as, for mathematical reasons, the drag must be the same for momentum and height (implying a frictional spindown time equal to a radiative spindown time). Nevertheless, unresolved small scale processes often do act as some form of damping and a linear damping is the simplest form. In chapter xxx we'll discuss how the equations might be physically justified and in what context the solutions are relevant to the tropical atmosphere and ocean. (It turns out that the equations constitute one of the simplest analytically tractable models of some of the basic features of the large-scale tropical circulation in the atmosphere.) We consider the full problem initially and then special cases.

The linear forced-dissipative equations of motion are

$$\frac{\partial u}{\partial t} + \alpha u - f v + \frac{\partial \phi}{\partial x} = F^x \tag{8.89a}$$

$$\frac{\partial v}{\partial t} + \alpha v + f u + \frac{\partial \phi}{\partial y} = F^y \tag{8.89b}$$

$$\frac{\partial \phi}{\partial t} + \alpha \phi + c^2 \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = -Q. \tag{8.89c}$$

where $F^x$and $F^y$ are the $x$ and $y$ components of the imposed forces, $Q$ is a thermal or mass source and $\alpha$ is a damping coefficient, assumed the same for all three variables. If we interpret $\boldsymbol{F} = (F^x, F^y)$ as wind stress, $\boldsymbol{\tau}$, acting on a layer of fluid we might make the association of $\boldsymbol{F} = \boldsymbol{\tau}/H$. Still, for now we will treat this system simply as a problem in geophysical fluid dynamics. The potential vorticity equation corresponding to (8.89), obtained by cross differentiating (8.89a) and (8.89b), is

$$\left[ \frac{\partial}{\partial t} + \alpha \right] \left( \zeta - \frac{f}{c^2} \phi \right) + \beta v = \mathrm{curl}_z \boldsymbol{F} + \frac{fQ}{c^2}. \tag{8.90}$$

In much the same way as we derived (8.23) we can derive a single partial differential equation for $v$, namely

$$\begin{aligned}\frac{1}{c^2} \left[ \frac{\partial}{\partial t} + \alpha \right]^3 v + \frac{f^2}{c^2} \left[ \frac{\partial}{\partial t} + \alpha \right] v - \left[ \frac{\partial}{\partial t} + \alpha \right] \left( \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial x^2} \right) - \beta \frac{\partial v}{\partial x} \\ = \frac{1}{c^2} \left[ \frac{\partial}{\partial t} + \alpha \right] \frac{\partial Q}{\partial y} - \frac{f}{c^2} \frac{\partial Q}{\partial x} \\ + \frac{1}{c^2} \left[ \frac{\partial}{\partial t} + \alpha \right]^2 F^y - \frac{f}{c^2} \left[ \frac{\partial}{\partial t} + \alpha \right] F^x - \frac{\partial}{\partial x} \left( \frac{\partial F^y}{\partial x} - \frac{\partial F^x}{\partial y} \right).\end{aligned} \tag{8.91}$$

The left-hand side is just a minor variation of that of (8.23). This equation is obviously very complicated and perhaps not very attractive: it certainly does not have the beauty of quasi-geostrophy.

Although the equation might be solved by similar methods to those used on (8.23) (or solved numerically) we will proceed in a slightly different and hopefully more informative way, the differences being twofold.

*(i)* We consider only special cases of (8.89). For example, will often simplify (8.89b) to geostrophic balance, $fu = -\partial\phi/\partial y$, and in section 8.5 we will pay particular attention to the steady version of the equations.

*(ii)* We will change variables from $(u, v, \phi)$ to a set denoted $(q, r, v)$, defined below, that allow an easier connection to be made between $v$ and the variables $u$ and $\phi$.

### 8.4.1 Mathematical Development

As noted, it turns out to be convenient to use a linear combination of $u$ and $\phi$, defined by

$$q = \frac{\phi}{c} + u, \qquad r = \frac{\phi}{c} - u. \tag{8.92}$$

The utility of this will become apparent as we proceed.[7] We may note that $u$ and $\phi$ have the same symmetry across the equator, both tending to be symmetric unless forcing deems otherwise, whereas $v$ tends to be antisymmetric. The equations for $q$ and $r$ become

$$\left(\frac{\partial}{\partial t} + \alpha\right) q + c\frac{\partial q}{\partial x} + c\frac{\partial v}{\partial y} - fv = F^x - Q/c, \tag{8.93a}$$

$$\left(\frac{\partial r}{\partial t} + \alpha\right) r - c\frac{\partial r}{\partial x} + c\frac{\partial v}{\partial y} + fv = -F^x - Q/c \tag{8.93b}$$

and the $v$-momentum equation is

$$\left(\frac{\partial}{\partial t} + \alpha\right) v + \frac{f}{2}(q - r) = -\frac{c}{2}\frac{\partial}{\partial y}(q + r) + F^y. \tag{8.93c}$$

*Nondimensionalization*

We scale velocity by $c$ and time and distance by

$$T_{eq} = (c2\beta)^{-1/2}, \qquad L_{eq} = (c/2\beta)^{1/2}. \tag{8.94a,b}$$

The nondimensional equations of motion are then [Note: we don't use a special notation for nondimensional variables. Consider fixing this, or not. xxx]

$$\left(\frac{\partial}{\partial t} + \alpha\right) q + \frac{\partial q}{\partial x} + \frac{\partial v}{\partial y} - \frac{1}{2}yv = F^x - Q, \tag{8.95a}$$

$$\left(\frac{\partial}{\partial t} + \alpha\right) r - \frac{\partial r}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{2}yv = -F^x - Q \tag{8.95b}$$

$$\left(\frac{\partial}{\partial t} + \alpha\right) v + \frac{y}{4}(q - r) = -\frac{1}{2}\frac{\partial}{\partial y}(q + r) + F^y. \tag{8.95c}$$

The solutions of these equations may be expressed in terms of parabolic cylinder functions, $D_n(y)$. That is, we seek solutions of the form

$$(v, q, r) = \sum_{n=0}^{\infty} \left(v_n(x,t), q_n(x,t), r_n(x,t)\right) D_n(y). \tag{8.96}$$

with the forcing terms expanded in a similar fashion. The parabolic cylinder functions themselves have the form

$$(D_0, D_1, D_2, D_3) = (1, y, y^2 - 1, y^3 - 3y)\exp(-y^2/4), \tag{8.97}$$

and so on. The polynomial terms are just the modified Hermite polynomials $G_m(y)$ given by (8.45). The parabolic cylinder functions obey the ladder properties that

$$\frac{\mathrm{d}D_n}{\mathrm{d}y} + \frac{1}{2} y D_n = n D_{n-1}, \tag{8.98a}$$

$$\frac{\mathrm{d}D_n}{\mathrm{d}y} - \frac{1}{2} y D_n = -D_{n+1}. \tag{8.98b}$$

If we substitute (8.96) into (8.95) we obtain ordinary differential equations for the amplitudes. From the $q$ equation we obtain, after a little algebra,

$$\left(\frac{\partial}{\partial t} + \alpha\right) q_0 + \frac{\partial q_0}{\partial x} = F_0^x - Q_0, \tag{8.99a}$$

$$\left(\frac{\partial}{\partial t} + \alpha\right) q_{n+1} + \frac{\partial q_{n+1}}{\partial x} - v_n = F_{n+1}^x - Q_{n+1}, \qquad n = 0, 1, 2, 3, \dots \tag{8.99b}$$

From the $r$ equation we find

$$\left(\frac{\partial}{\partial t} + \alpha\right) r_{n-1} - \frac{\partial r_{n-1}}{\partial x} + n v_n = -\left(F_{n-1}^x + Q_{n-1}\right), \qquad n = 1, 2, 3, \dots \tag{8.100}$$

and from the $v$ equation we find

$$\left(\frac{\partial}{\partial t} + \alpha\right) v_0 + \frac{q_1}{2} = F_0^y, \tag{8.101a}$$

$$\left(\frac{\partial}{\partial t} + \alpha\right) v_n + \frac{(n+1)}{2} q_{n+1} - \frac{r_{n-1}}{2} = F_n^y, \qquad n = 1, 2, 3, \dots \tag{8.101b}$$

Finally, we note without derivation that these equations may be combined to give

$$\left[\frac{\partial}{\partial t} + \alpha\right]^3 v_n + \left[\frac{\partial}{\partial t} + \alpha\right]\left((2n+1)v_n - \frac{\partial^2 v_n}{\partial x^2}\right) - \frac{\partial v_n}{\partial x} = G \tag{8.102}$$

where $G$ represents the various forcing terms. This equation most easily derived by substituting (8.96) into the nondimensional form of (8.91).

In principle, the above equations provide a means of solving the problem for almost any forcing. The equations have constant coefficients and might be solved by a superposition of harmonic functions in the $x$-direction, in conjunction with the variation in the $y$-direction given by the parabolic cylinder functions. In general, however, this procedure would be tedious and uninformative. Thus, and to avoid being asphyxiated by an avalanche of algebra, we will consider some special cases, but these cases will be the most physically realistic and little of import will be lost. Enthusiasts may continue with the general development by themselves. We may also note that modern geophysical fluid dynamics has advanced by way of using numerical methods to find solutions to complicated equations, in conjunction with using analytic methods to find solutions of simplified cases, or to find more general relations, in order to provide insight and understanding.

The most important problem, or at least the most influential problem, that we will consider we leave until section 8.5. For the rest of this section we content ourselves with some slightly general comments about forced waves.

### 8.4.2 Forced Waves

In this section we will, albeit briefly, consider the problem of forced waves in which we retain some of the forcing terms but neglect the damping.[8] Our purpose is not to give a complete treatment; rather, it is to show what kinds of waves might be excited and to help interpret (8.99)–(8.102). We first note that with $\alpha = 0$ (8.102) becomes, in dimensional form,

$$\frac{\partial}{\partial t}\left[\frac{1}{c^2}\frac{\partial^2 v_n}{\partial t^2} - \frac{\partial^2 v_n}{\partial x^2} + (2n+1)\frac{\beta}{c}v_n\right] - \beta\frac{\partial v_n}{\partial x} = G \tag{8.103}$$

(This equation may be derived directly by using (8.157) or (8.165) in the appendix in the appropriate nondimensional version of (8.23), adding a forcing and redimensionalizing.) From (8.103) the dispersion relation for free waves [that is, (8.39)] follows if we let $G = 0$ and seek harmonic solutions of the form $\exp(\mathrm{i}kx - \mathrm{i}\omega t)$.

Consider now a forcing, $H$ say, that projects only onto the zeroth order parabolic function, $D_0$. Equation (8.99a) becomes, in dimensional form,

$$\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) q_0 = H_0. \tag{8.104}$$

The free solutions of this are Kelvin waves propagating eastwards at speeds $c = \omega/k$ for each $k$ that might be excited; that is $q_0 = \operatorname{Re} C \exp[\mathrm{i}k(x - ct)]$, where $C$ is a constant. Suppose that the forcing is harmonic in $x$ and time,

$$\begin{aligned} G_0 &= \operatorname{Re} A\,\{\exp[\mathrm{i}(k_1 x - \omega_1 t)] + \exp[\mathrm{i}(k_1 x + \omega_1 t)]\} \\ &= A\,[\cos(k_1 x - \omega_1 t) + \cos(k_1 x + \omega_1 t)]\,, \end{aligned} \tag{8.105}$$

if $A$ is real. The solution to (8.104) with this forcing is given by

$$q_0 = -\frac{A\sin(k_1 x - \omega_1 t)}{\omega_1 - ck_1} + \frac{A\sin(k_1 x + \omega_1 t)}{\omega_1 + ck_1}. \tag{8.106}$$

All the parameters in the above equation, $c, k_1, \omega_1$, are positive. If the forcing is just one harmonic then, in general, $c \neq \omega_1/k_1$. However, if the forcing is a superposition of many harmonics then there may be one that is in resonance with the free mode, and this wave, an eastward propagating Kelvin wave represented by an expression like the first term on the right-hand side of (8.106), will be preferentially excited. Similar considerations apply to other waves too; that is, the forcing will excite waves that most resemble the forcing and can resonate with it. Sometimes, a forcing will resemble a delta function in both space and time. For example, a sudden and localized burst of wind over the ocean because of intense storm activity. In these cases, the forcing contains all space and time scales (a true Dirac delta function has equal representation of all Fourier modes). In this case, both eastward propagating Kelvin waves and westward propagating planetary waves will be excited, and to look at some of these it is useful to make a longwave approximation, as we now discuss.

*Planetary waves, revisited*

In the planetary wave, or longwave, approximation the highest time derivative in (8.103) is omitted, leaving

$$\frac{\partial}{\partial t}\left[\frac{\partial^2 v_n}{\partial x^2} - (2n+1)\frac{\beta}{c}v_n\right] + \beta\frac{\partial v_n}{\partial x} = G. \tag{8.107}$$

If $G = 0$ this equation gives the dispersion relation

$$\omega = \frac{-\beta k}{(2m+1)\beta/c + k^2}, \tag{8.108}$$

as in (8.50). Planetary waves will be excited when the forcing itself has a low frequency.

*The longwave approximation, revisited*

Many situations in low latitudes are characterized by having a longer zonal scale than meridional scale; thus, $|\partial\phi/\partial y| \gg |\partial\phi/\partial x|$. When this is the case, geostrophic balance will hold to a good approximation for the zonal flow even in the presence of forcing and dissipation, but not for the meridional flow, and to a good approximation the meridional momentum equation (8.89b) may be replaced by

$$fu = -\frac{\partial\phi}{\partial y}. \tag{8.109}$$

In this limit (8.103) simplifies to

$$\frac{\partial}{\partial t}\left[(2n+1)\frac{\beta}{c}v_n\right] - \beta\frac{\partial v_n}{\partial x} = G, \tag{8.110}$$

from which the dispersion relation,

$$\omega = \frac{-kc}{2n+1}, \tag{8.111}$$

immediately follows. The Kelvin wave is an eastward propagating wave with $n = -1$. When $n \geq 0$ the waves are westward propagating.

The amplitude equations, (8.99)–(8.101) then simplify as follows, also taking $\alpha = 0$. The $q$ equations become

$$\frac{\partial q_0}{\partial t} + \frac{\partial q_0}{\partial x} = F_0^x - Q_0, \tag{8.112a}$$

$$\frac{\partial q_{n+1}}{\partial t} + \frac{\partial q_{n+1}}{\partial x} - v_n = F_{n+1}^x - Q_{n+1}, \qquad n = 0, 1, 2, 3, \ldots \tag{8.112b}$$

The $r$ equation becomes

$$\frac{\partial r_{n-1}}{\partial t} - \frac{\partial r_{n-1}}{\partial x} + n v_n = -\left(F^x_{n-1} + Q_{n-1}\right), \qquad n = 1, 2, 3, \dots \tag{8.113}$$

and from the $v$ equation (geostrophic balance) we find

$$q_1 = 0, \tag{8.114a}$$

$$(n+1) q_{n+1} = r_{n-1}, \qquad n = 1, 2, 3, \dots \tag{8.114b}$$

If we use (8.114b) to eliminate $r_{n-1}$ in (8.113), and then use (8.112b) to eliminate $v_n$ we obtain

$$(2n+1)\frac{\partial q_{n+1}}{\partial t} - \frac{\partial q_{n+1}}{\partial x} = n\left(F^x_{n+1} - Q_{n+1}\right) - \left(F^x_{n-1} + Q_{n-1}\right). \tag{8.115}$$

The above set of equations provide, in principle, a means for studying the response of the system to an imposed forcing, such as winds blowing over the ocean or a diabatic source in the atmosphere. Having neglected dissipation, wavelike solutions of constant amplitude will be found only if the forcing is oscillatory rather than steady. Solutions are found by solving the first-order wave equations (8.112a) and (8.115) for $q_n$, and then using (8.114b) to obtain $r_n$. A simple expression for $v_n$ results if we add (8.112b) and (8.115).

*Waves and adjustment*

The wave described by (8.112a) is a Kelvin wave, moving eastwards with nondimensional speed unity, or dimensional speed $c$. It also follows from the dispersion relation, $\omega = -kc/(2n+1)$, with $n = -1$. In contrast, the waves described by (8.115) are westwards propagating, long, low frequency planetary waves. In dimensional form (8.115) becomes

$$(2n+1)\frac{\partial q_{n+1}}{\partial t} - c\frac{\partial q_{n+1}}{\partial x} = n\left(F^x_{n+1} - Q_{n+1}\right) - \left(F^x_{n-1} + Q_{n-1}\right). \tag{8.116}$$

and hence have a speed $-c/(2n+1)$, just as in (8.53). There are no short low frequency waves in this approximation.

As we noted above, an arbitrary forcing will in general excite both gravity waves and planetary waves and the initial flow will be out of geostrophic balance. In the midlatitude case (discussed in section 3.8 of AOFD) the gravity waves radiate to infinity (at least in the idealized problem) leaving behind an adjusted flow in geostrophic balance, determined by potential vorticity conservation. The process of adjustment is less efficient at low latitudes, because the waves are trapped between their inertial latitudes, as discussed in section 8.3, and in the absence of dissipation the fluid will oscillate endlessly. In the zonal direction both planetary and Kelvin waves propagate. A gravity wave front moves away more quickly, with the eventual adjustment occurring by way of planetary waves.

[More here?]

## 8.5 FORCED, STEADY FLOW: THE MATSUNO–GILL PROBLEM

We now consider the forced, steady version of the equatorial wave problem; that is to say, we seek steady solutions of (8.89), but with a mechanical or thermal forcing on the right-hand side.[9] Because of its importance to the tropical circulation of the atmosphere this problem has become somewhat iconic and some readers may be tempted to begin reading this chapter here. However, the problem is really just the forced, steady version of the wave problems studied in sections 8.2 and 8.4, and the reader should have some familiarity with that material before proceeding. In fact, those readers who have followed the previous sections closely will find the material on the Matsuno–Gill problem a pleasant stroll in the park.

### 8.5.1 Mathematical development

We make one additional approximation from (8.89): we assume that the zonal wind is in geostrophic balance with the meridional pressure gradient. This 'semi-geostrophic' approximation is equivalent to the longwave approximation discussed in previous sections. The equations of motion then become

$$\alpha u - fv + \frac{\partial \phi}{\partial x} = F^x, \tag{8.117a}$$

$$fu + \frac{\partial \phi}{\partial y} = 0, \tag{8.117b}$$

$$\alpha\phi + c^2\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = -Q. \tag{8.117c}$$

From these equations we may derive a single equation for $v$, namely

$$\frac{f^2}{c^2}\alpha v - \alpha\frac{\partial^2 v}{\partial y^2} - \beta\frac{\partial v}{\partial x} = \frac{\alpha}{c^2}\frac{\partial Q}{\partial y} - \frac{f}{c^2}\frac{\partial Q}{\partial x} - \frac{f}{c^2}\alpha F^x + \frac{\partial^2 F^x}{\partial x \partial y}. \tag{8.118}$$

This is just a simplification of (8.91) appropriate for a steady system with the zonal wind in geostrophic balance, obtained by omitting all the time derivatives, the term involving $\alpha^3$, and the $F^y$ term on the right hand side.

From now on we will nondimensionalize all the variables using the length scales

$$T_{eq} = (2c\beta)^{-1/2}, \qquad L_{eq} = (c/2\beta)^{1/2}. \tag{8.119a,b}$$

The equations of motion become

$$\alpha u - \frac{y}{2}v + \frac{\partial \phi}{\partial x} = F^x, \tag{8.120a}$$

$$\frac{y}{2}u + \frac{\partial \phi}{\partial y} = 0, \tag{8.120b}$$

$$\alpha\phi + \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = -Q. \tag{8.120c}$$

and the $v$ equation becomes

$$\frac{y^2}{4}\alpha v - \alpha\frac{\partial^2 v}{\partial y^2} - \frac{1}{2}\frac{\partial v}{\partial x} = \alpha\frac{\partial Q}{\partial y} - \frac{y}{2}\frac{\partial Q}{\partial x} - \frac{\alpha y}{2}\alpha F^x + \frac{\partial^2 F^x}{\partial x \partial y}. \tag{8.121}$$

As before when dealing with wave-like problems it is convenient to change variables to $p$ and $q$ where

$$q = \phi + u, \qquad r = \phi - u. \tag{8.122a,b}$$

The equations of motion become

$$\alpha q + \frac{\partial q}{\partial x} + \frac{\partial v}{\partial y} - \frac{1}{2} y v = F^x - Q, \tag{8.123a}$$

$$\alpha r - \frac{\partial r}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{2} y v = -F^x - Q, \tag{8.123b}$$

$$\frac{y}{4}(q - r) + \frac{1}{2}\frac{\partial}{\partial y}(q + r) = 0. \tag{8.123c}$$

These are special cases of (8.95), with the first two equations being combinations of the $u$-momentum and pressure equations and the last one being the $v$-momentum equation (zonal geostrophic balance).

Again following the general treatment given earlier we expand the variables and the forcing in terms of parabolic cylinder functions. Thus, for example,

$$Q(x) = \sum_{n=0}^{\infty} Q_n(x) D_n(y), \tag{8.124}$$

and similarly for the other variables. The resulting ordinary differential equations are special cases of (8.99)–(8.102), specifically

$$\alpha q_0 + \frac{\partial q_0}{\partial x} = F_0^x - Q_0, \tag{8.125a}$$

$$\alpha q_{n+1} + \frac{\partial q_{n+1}}{\partial x} - v_n = F_{n+1}^x - Q_{n+1}, \qquad n = 0, 1, 2, 3, \ldots \tag{8.125b}$$

$$\alpha r_{n-1} - \frac{\partial r_{n-1}}{\partial x} + n v_n = -\left(F_{n-1}^x + Q_{n-1}\right), \qquad n = 1, 2, 3, \ldots \tag{8.126}$$

$$q_1 = 0, \tag{8.127a}$$

$$(n+1) q_{n+1} - r_{n-1} = 0, \qquad n = 1, 2, 3, \ldots \tag{8.127b}$$

Using (8.125b), (8.126) and (8.127b) we obtain

$$\alpha(2n+1) q_{n+1} - \frac{\partial q_{n+1}}{\partial x} = n\left(F_{n+1}^x - Q_{n+1}\right) - \left(F_{n-1}^x + Q_{n-1}\right) \qquad n = 1, 2, 3, \ldots \tag{8.128}$$

Finally, although we shall not use it, the $v$ equation (8.118) becomes

$$\alpha\left((2n+1) v_n - \frac{\partial^2 v_n}{\partial x^2}\right) - \frac{\partial v_n}{\partial x} = G, \tag{8.129}$$

where $G$ represents the various forcing terms.

As in the wavelike case, the above equations provide, at least in principle, a means of solving for the response for any particular forcing. The procedure is to project the forcing onto parabolic cylinder functions, and then solve the amplitude equations (8.125)–(8.127) for the zonal dependence, and then finally to reconstruct the solutions using the $q_n(x)$, $r_n(x)$ and $v_n(x)$ and the parabolic cylinder functions. Naturally enough, this is easier said than done. We will go through the procedure in detail for one important case, and leave other solutions as exercises for the reader.

### 8.5.2 Symmetric heating

An important canonical case is that in which the system is forced by a heating that is confined in both the $x$- and $y$-directions, and is symmetric across the equator. Confinement in the $y$-direction is easily achieved by supposing that the heating projects solely onto the first parabolic function, so that

$$Q(x) = Q_0(x)D_0(y) = G(x)\exp(-y^2/4), \tag{8.130}$$

and confinement in the $x$-direction may be achieved by supposing that the heating is of the form

$$G(x) = \begin{cases} A\cos kx & |x| < L \\ 0 & |x| > L, \end{cases} \tag{8.131}$$

where $k = \pi/2L$. This may seem an odd form to choose, but the harmonic variation for $|x| < L$ enables an analytic solution to be found in that region, and the absence of any forcing at all in the far field enables solutions to be found there in the form of decaying wavelike disturbances. Although this problem is clearly a special case, we may expect, and certainly hope, that the qualitative form of the solution will transcend its precise details.

*Kelvin wave contribution*

We noted in section 8.4.2 that the equation for $q_0$ represents an eastwards propagating Kelvin wave, and this holds in the damped case also. That is to say, there will be a nonzero solution of (8.125) only in the forced region and eastward of it, where it will be progressively damped. Using this piece of physical insight, we can easily derive the solution in all three regions. First, for $X < -L$, we have

$$q_0 = 0, \qquad x < -L. \tag{8.132a}$$

In the forcing region we have to solve (8.125) with a boundary condition of $q_0 = 0$ at $x = -L$. The solution is

$$q_0 = \frac{-A}{\alpha^2 + k^2}\left\{\alpha\cos kx + k\left[\sin kx + \mathrm{e}^{-\alpha(x+L)}\right]\right\} \qquad |x| < -L. \tag{8.132b}$$

For $x > L$ we solve (8.125), but with a right-hand side equal to zero, with a boundary condition at $x = L$ given by (8.132b), namely $q_0 = -Ak(\alpha^2 + k^2)^{-1}[1 + \exp(-2\alpha L)]$. The solution is

$$q_0 = \frac{-Ak}{\alpha^2 + k^2}\left(1 + \mathrm{e}^{-2\alpha L}\right)\mathrm{e}^{\alpha(L-x)}. \qquad x > L. \tag{8.132c}$$

Because the motion is a decaying Kelvin wave $v = 0$ and the nondimensional $u$ and $\phi$ fields are equal to each other, with $r = 0$. Thus, from (8.122) and (8.132),

$$u = \phi = \frac{1}{2}q_0(x)\exp(-y^2/4), \qquad v = 0. \tag{8.133}$$

This does not mean $r_0$ is zero. Rather, it is associated with the planetary wave solution discussed below. The vertical velocity may be reconstructed from

$$w = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = \alpha\phi + Q, \tag{8.134}$$

whence

$$w = \frac{1}{2}\left[\alpha q_0(x) + Q_0(x)\right]\exp(-y^2/4). \tag{8.135}$$

We'll discuss the meaning of these solutions below, but first we'll complete the solution by finding a planetary wave contribution.

*Planetary wave contribution*

We now find the solution associated with $q_2$ and $r_0$. From (8.128) we have

$$\frac{\mathrm{d}q_2}{\mathrm{d}x} - 3\alpha q_2 = Q_0. \tag{8.136a}$$

From (8.127b) we have

$$r_0 = 2q_2, \tag{8.136b}$$

and from (8.125b) we have

$$v_1 = \alpha q_2 + \frac{\mathrm{d}q_2}{\mathrm{d}x}. \tag{8.136c}$$

These are planetary waves propagating westwards at a dimensional speed of $c/(2n+1) = c/3$. Thus, no signal is transmitted eastwards and we can find a solution to the above equations in an analogous fashion to how we found a solution for the Kelvin wave problem. After just a little algebra, the solution is found to be:

$$q_2 = 0, \qquad x > L \tag{8.137a}$$

$$q_2 = \frac{A}{(3\alpha)^2 + k^2}\left[-3\alpha\cos kx + k\left(\sin kx - \mathrm{e}^{3\alpha(x-L)}\right)\right], \qquad |x| < L \tag{8.137b}$$

$$q_2 = \frac{-Ak}{(3\alpha)^2 + k^2}\left[1 + \mathrm{e}^{-6\alpha L}\right]\mathrm{e}^{3\alpha(x+L)}, \qquad x < -L. \tag{8.137c}$$

The corresponding solutions for the pressure and velocity fields are

$$u = \frac{\mathrm{e}^{-y^2/4}}{2} q_2(x)(y^2 - 3), \qquad v = y\mathrm{e}^{-y^2/4}\left[Q_0(x) + 4\alpha q_2(x)\right], \tag{8.138a,b}$$

$$\phi = \frac{\mathrm{e}^{-y^2/4}}{2} q_2(x)(1 + y^2), \qquad w = \frac{\mathrm{e}^{-y^2/4}}{2}\left[Q_0(x) + \alpha q_2(x)(1 + y^2)\right]. \tag{8.139a,b}$$

The solutions appear complicated (they are complicated!), but they are still amenable the physical interpretation. But first, for the record, we'll combine the Kelvin and planetary wave contributions and restore the dimensions, to give

$$u = \frac{c}{2}\left[q_0(x) + q_2(x)(2\beta y^2/c - 3)\right]\mathrm{e}^{-\beta y^2/2c}, \tag{8.140a}$$

$$v = cy\Big[Q_0(x) + (4\alpha/c)q_2(x)\Big]\mathrm{e}^{-\beta y^2/2c}, \tag{8.140b}$$

$$\phi/c = \frac{c}{2}\left[q_0(x) + q_2(x)(2\beta y^2/c + 1)\right]\mathrm{e}^{-\beta y^2/2c}, \tag{8.140c}$$

$$w = \frac{\mathrm{e}^{-\beta y^2/2c}}{2}\left[2Q_0(x) + \alpha q_0(x) + \alpha q_2(x)(1 + 2\beta y^2/c)\right] \tag{8.140d}$$

The nondimensional forms are recovered by setting $c = 1$ and $\beta = 1/2$, with $\alpha$ taking its nondimensional value.

[Check equation dimensions, especially $w$. xxx]

The solutions above are obviously specific to the form of the forcing function we chose. However, a similar methodology could in principle be applied to forcing of any form, including forcing in the momentum equations, and, because the equations are linear, the solutions could be superposed. The solution above represents the physically important case of a localized heating, but the gross structure (although not the sign) of the far field will often be independent of the details of the forcing: there will be a slowly decaying disturbance west of the forcing and a more rapidly decaying disturbance east of the forcing,

*Interpretation*

Let's now try to figure out what's going on. A solution is illustrated in Fig. 8.6. The heating is confined to a region from $-2 < x < 2$ and exponentially falls away from the equator with an e-folding distance of 2, more-or-less corresponding to the shaded region of vertical velocity in the lower right panel, as intuitively expected and discussed more below.

Consider first the flow in the forcing region. Here the vertical velocity is postive, with the associated horizontal convergence being that of the zonal flow: the meridional flow is polewards, *away* from the maximum of the heating. To understand this, consider the limit $\alpha \to 0$. From (8.140) the vertical velocity field coincides with the heating and the (nondimensional) meridional velocity is given by

$$v = y Q_0 \exp(-y^2/4) = yw. \tag{8.141}$$

Thus, vertical motion is associated with poleward motion. To understand this, consider the inviscid vorticity equation

$$\beta v + f \nabla \cdot \boldsymbol{u} = 0, \qquad \text{or} \qquad \beta v = f w. \tag{8.142a,b}$$

which in nondimensional form is

$$v + y \nabla \cdot \boldsymbol{u} = 0, \qquad \text{or} \qquad v = yw. \tag{8.143a,b}$$

Evidently, (8.141) and (8.143b) are equivalent. Another way to thing about this is to note that the rising motion in the region of the forcing causes vortex stretching, as discussed in chapter 4, and hence the generation of cyclonic vertical vorticity and a polewards migration. From the perspective of potential vorticity, then to the extent that the flow is adiabatic the quantity $Q = (f + \zeta)/h$ is conserved following the flow. The heating increases the value of $h$ (the stretching), so that $f + \zeta$ also tends to increase in magnitude. The flow finds it easier to migrate polewards to increase its value of $f$ than to increase its relative vorticity alone, for the latter would require more energy. If we interpret these equations as the lowest layer of a two-layer system, then the flow in the lower layer is away from the source, and toward the source in the upper layer.

Consider now the flow to the west of the heating, associated with $q_2(x)$. The disturbance here is produced by a decaying westwards propagating Rossby wave — a form of 'Rossby plume' that we will also encounter in chapter 19 (see Fig. 19.12 on page 19.12 and the associated discussion). the vertical velocity is negative, and the horizontal velocity is almost geostrophically balanced: the pressure perturbation is negative everywhere, and so circulating cyclonically around the centers of low pressure just to the west of heating. The flow converges to the equator, producing an eastward flow along the equator, converging in the heating zone. We may be tempted

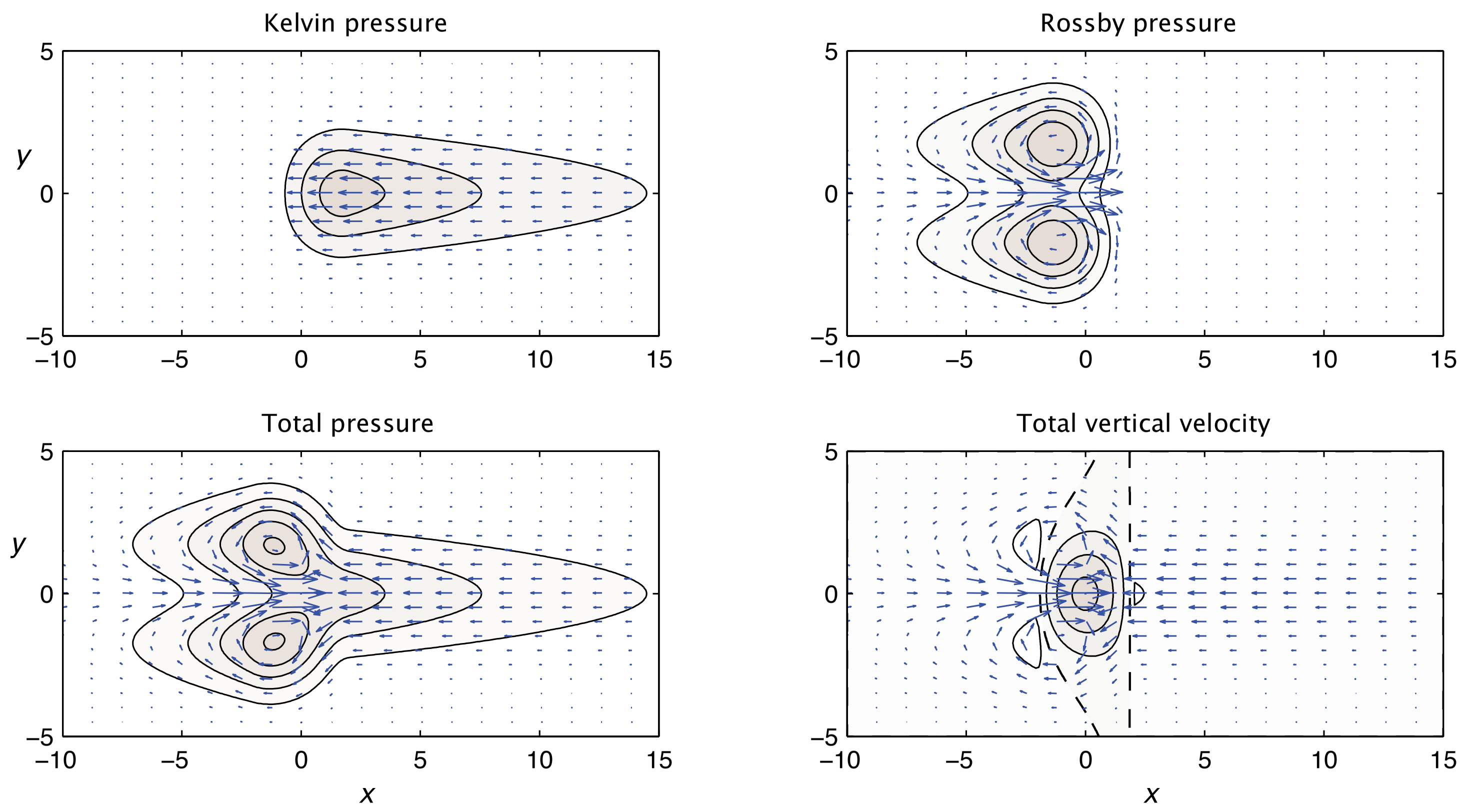


**Fig. 8.6** Nondimensional solutions of the Matsuno–Gill model, with heating given by (8.131) with $L = 2$ and $\alpha = 0.1$. The shaded contours show the field indicated in the titles, and the arrows show the associated horizontal velocities. The 'Kelvin' and 'Rossby' designations indicate that just the Kelvin wave or Rossby (planetary) wave contributions are plotted as given by (8.133)–(8.135) and (8.138)–(8.139), respectively. For the pressure fields the contour interval is 0.3 and all fields are negative, with the zero contour omitted. For vertical velocity the contour interval is 0.3 beginning at -0.1, and so is -0.1, 0.2, 0.5..., with an additional zero contour (dashed) with upward motion within it.

to interpret this in terms of the inviscid vorticity equation, as we did in the forcing region. This would suggest that, away from the forcing region, because the flow is divergent ($\nabla\cdot\boldsymbol{u} > 0,\ w < 0$) then from (8.142) the meridional velocity should be toward the equator in both hemispheres. However, this explanation is at best qualitative, because the vorticity equation above is *not* exactly satisfied by the solution (8.142), because non-zero solutions away from the forcing region depend entirely on the presence of dissipation. [More here?]

The flow east of the forcing region motion is induced by an eastward propagating Kelvin wave, or more precisely the steady, eastward-decaying analogue of such a wave. Evidently, from Fig. 8.6, the pressure field extends further east of the source than west of the source, and this is because Kelvin waves decay more slowly than Rossby waves. Keeping both the time derivative and the damping, the unforced Kelvin wave satisfies, from (8.99),

$$\left[\alpha + \frac{\partial}{\partial t}\right] q_0 + \frac{\partial q_0}{\partial x} = 0, \tag{8.144}$$

whereas the unforced Rossby wave satisfies, from (8.115) and (8.128) for $n = 1$

$$3\left[\alpha + \frac{\partial}{\partial t}\right] q_1 - \frac{\partial q_1}{\partial x} = 0. \tag{8.145}$$

Thus, the effective damping rate on the Rossby wave is three times that on the Kelvin wave. Put another way, the Kelvin wave travels three times as fast as the Rossby wave so that if the damping rate, $\alpha$, is the same the influence of the Kelvin wave spreads three times further east. The horizontal velocity in the Kelvin wave is purely zonal and in this example is directed toward the heating source.

### *Vertical structure*

The zonal structure of the solution is a crude representation of the Walker circulation in the equatorial Pacific. Here, the sea-surface temperature is high in the west, near Indonesia, and low in the east, near South America, because of the upwelling that brings deep, cold water to the surface. This distribution of sea-surface temperature effectively provides a heating in the western Pacific and induces westward winds along the equator, enhancing the westward trade winds that already exist as part of the general circulation. The overturning circulation in the zonal plane is illustrated in Fig. **??**. This solution is obtained by supposing that the fields represent the first vertical mode, as discussed in section 8.1.1. If the stratification is uniform then the modes are just sines and cosines and so we have

$$(u, v, \phi) = (\widetilde{u}, \widetilde{v}, \widetilde{\phi}) \cos(\pi z/D), \qquad w = \widetilde{w} \sin(\pi z/D) \tag{8.146}$$

Now the modal form of the mass continuity equation, (8.13), is $\widetilde{w} = -(c^2/g)\nabla\cdot\widetilde{\boldsymbol{u}}$. If this is to be consistent with the usual form of $\partial w/\partial z = -\nabla\cdot\boldsymbol{u}$ then we make the association $\pi/D = g/c^2 = 1/H_1^*$, where $H_1^*$ is the equivalent depth of the first mode. Given this vertical structure, we integrate the solutions meridionally so enabling a streamfunction to be defined (because $v = 0$ as $y \to \pm\infty$, so $\partial\overline{u}^y/\partial x + \partial\overline{w}^y/\partial z = 0$, with the overbar denoting meridional integration). The expressions for the meridional integrals are given in Appendix B to this chapter, and the meridional structure of the solution is given in Fig. **??**.

[This section still needs cleaning up and figures adding.]

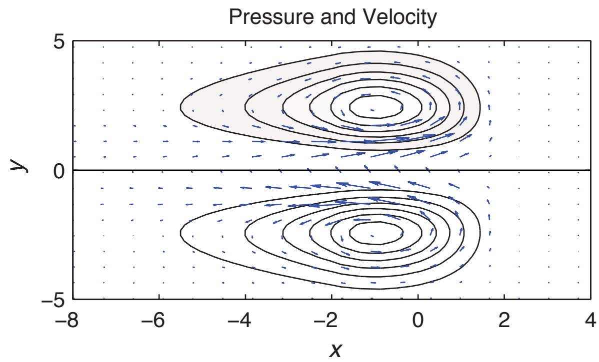


**Fig. 8.7** Solutions to the Matsuno–Gill problem with an asymmetric heating given by (8.147) and decay factor $\alpha = 0.1$. The heating is in the Northern Hemisphere generating a low pressure region (shaded) with inflow and ascent. The contour interval is 0.3, with the zero contour along $y = 0$.

### 8.5.3 Antisymmetric forcing

A solution with asymmetric forcing may be obtained by using a forcing of the form

$$Q(x, y) = Q_1(x)D_1(y) = y \cos kx \exp(-y^2/4), \tag{8.147}$$

using the same form of zonal localization as before. The algebra needed to obtain a solution is somewhat tedious but straightforward, of a very similar nature to that described above. One finds that there are, again, two parts to the response. One part corresponds to a long planetary wave with $n = 0$ and using (8.125)–(8.127) we find

$$q_1 = 0. \qquad v_0 = Q_1. \tag{8.148}$$

There is no response outside the forcing region because long mixed waves have zero propagation velocity. The other part of the solution is obtained, again using (8.125)–(8.127), from

$$v_2 = \frac{dq_3}{dx} + \alpha q_3, \tag{8.149a}$$

$$r_1 = 3q_3, \tag{8.149b}$$

$$\frac{dq_3}{dx} - 5\alpha q_3 = Q_1 \tag{8.149c}$$

The solution of these is left as an exercise for the reader (or by consulting the original literature) and is illustrated in Fig. 8.7. The solutions are zero east of the forcing region because there is no long wave so propagating. West of the forcing region there is eastward inflow into the heating region in the Nortern Hemisphere (which is being heated), as well as a tendency for poleward flow for the reasons described earlier. Thus, there is a cyclone with upward motion somewhat west of the main heating region, and a corresponding anti-cyclone in the cooled region, as illustrated in Fig. 8.7. The zonally averaged solutions (not shown) resemble an asymmetric Hadley Cell, with the air rising in the Norther (summer) hemisphere, moving southwards aloft into the winter hemisphere before sinking.

### 8.5.4 Other forcings

The solution to more general forcings can be constructed by using other forcing coefficients, or a superposition of forcing coefficients, and many solutions of interest to the tropical atmo-

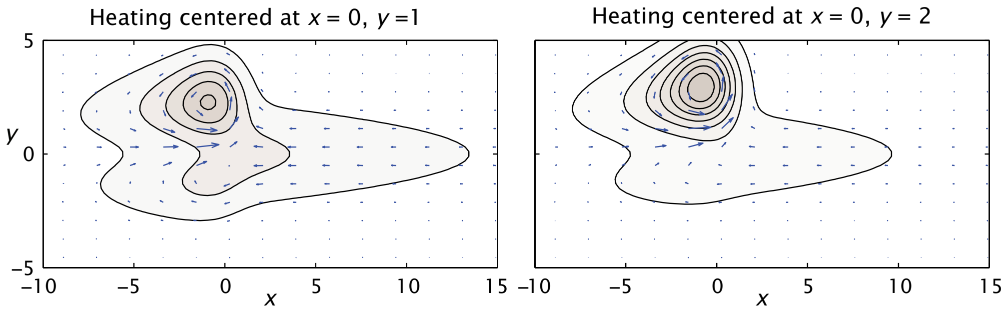


**Fig. 8.8** Solutions of the Matsuno–Gill model with the heating centered off the equator, as labelled, but otherwise similar to that which produces the solutions in Fig. 8.6. The lines are contours of pressure and the arrows are horizontal velocity. As the heating moves to higher latitudes the Kelvin wave response weakens but the magnitude of the local response increases (the contour interval is the same in both panels).

sphere and ocean may be so constructed. Solutions may also be constructed numerically, either by time-stepping the linear shallow water equations to equilibrium or by solving the elliptic equation (8.118) using standard techniques. The solutions we present below were in fact obtained numerically.[10]

We'll present the solutions to two such cases: (i) a heating centered off the equator in the Northern Hemisphere, and (ii) a line source of heating, either centered at the equator or just north of it, roughly mimicking the Inter-Tropical Convergence Zone (ITCZ).

### *Heating off the equator*

A heating off the equator may be constructed by adding the solutions for antisymmetric and symmetric heating presented above. In fact, in Fig. 8.8, we present a solution that has heating of a very similar form to that of the symmetric heating shown in Fig. 8.6, but centered off the equator at $y = 1$ and $y = 2$. The pattern is dominated by a low pressure region just to the west of the heating, with convergence and upward motion within it, and an eastward inflow between the equator and the center of the heating. In the solution with the heating centered at $y = 1$ there is also a response east of the heating region,largest at the equator, produced by the eastward propagating, damped Kelvin wave. As the heating moves further from the equator (in the right panel of Fig. 8.8), the pressure response becomes stronger but the flow around the heating is in near geostrophic balance.

### *A line of heating*

Finally, let us consider the solutions when the heating is independent of $x$, and the solutions themselves are then independent of $x$. Two such solutions are presented in Fig. 8.9 and again, more quantitatively, in Fig. 8.10, for a line of heating at the equator and at $y = 1$. As we noted above, these solutions might be thought of as a rather idealized versions of the ITCZ (although in the real ITCZ the location of the convective region is determined as part of the solution for the overall flow, and not externally imposed).

Consider first the solution with heating at the equator. A low pressure region develops over the heating and the flow converges there, producing equatorward and westward 'trade

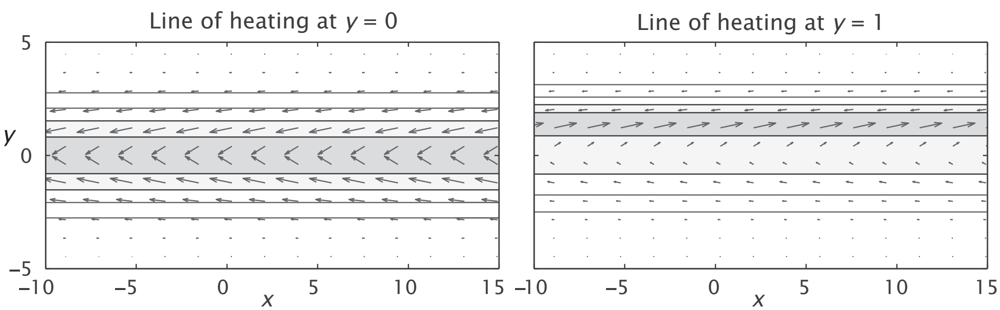


**Fig. 8.9** As for Fig. 8.8 but with a line of heating at the equator (left panel) and at $y = 1$ (right panel). The heating generates a region of low pressure (shaded) where the flow converges. In the right panel the meridional velocity is larger on the equatorward side of the line than on the poleward side. See Fig. 8.10 for a more quantitative picture.

winds' and consequent upward motion at the equator, with the zonal velocity rapidly decreasing actually at the equator.

Now consider the what happens when the heating is off-equator, noting that the real ITCZ is generally situated a little north of the equator, especially in the Pacific Ocean. A low pressure region is formed along the line of the heating and the meridional velocity converges sharply there, with more inflow coming from the equatorial side of the line of heating (as can be seen in the right-hand panels of both Fig. 8.9 and Fig. 8.10). As regards the zonal velocity, there is an *eastward* jet along the line of the heating, with westward flow to either side. That is to say, there is a splitting of the westward trades caused by the line of sharp heating.

## 8.A APPENDIX: NONDIMENSIONALIZATION AND PARABOLIC CYLINDER FUNCTIONS

This appendix provides a brief discussion of the nondimensionalization used to derive the various dispersion relations in this chapter and some of the properties of the associated Hermite polynomials and parabolic cylinder functions. We do not provide any proofs or detailed derivations.[11]

### *Nondimensionalization*

In discussions of equatorial waves and their steady counterparts, one of two slightly different nondimensionalizations are often employed. They lead to the use of parabolic cylinder functions in two slightly different forms; they are essentially equivalent but one may be more convenient than the other depending on the setting. For definiteness, we begin with (8.28), namely

$$\frac{\mathrm{d}^2\tilde{v}}{\mathrm{d}y^2} + \left(\frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega} - \frac{\beta^2 y^2}{c^2}\right)\tilde{v} = 0. \tag{8.150}$$

If, as in the main text, we nondimensionalize time and distance using

$$T_{eq} = (c\beta)^{-1/2}, \qquad L_{eq} = (c/\beta)^{1/2} \tag{8.151a,b}$$

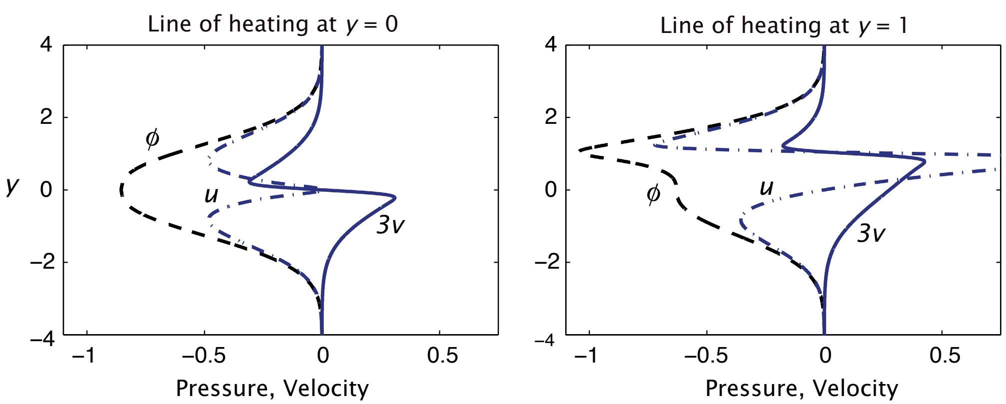


**Fig. 8.10** As for Fig. 8.9, but showing line plots of pressure $\phi$, zonal velocity $u$ and three times the meridional velocity, $3v$. Left panel is for a line of heating at the equator and the right panel for heating at $y = 1$. The heating creates a region of low pressure where the flow converges. Note that in the right panel the meridional velocity is larger on the equatorward side of the line than on the poleward side.

we obtain

$$\frac{\mathrm{d}^2 v}{\mathrm{d}\hat{y}^2} + \left(\hat{\omega}^2 - \hat{k}^2 - \frac{\hat{k}}{\hat{\omega}} - \hat{y}^2\right) v = 0. \tag{8.152}$$

The substitution

$$v(\hat{y}) = \Psi(\hat{y})\mathrm{e}^{-\hat{y}^2/2}, \tag{8.153}$$

leads to

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}\hat{y}^2} - 2\hat{y}\frac{\mathrm{d}\Psi}{\mathrm{d}\hat{y}} + \lambda\Psi = 0 \tag{8.154}$$

where $\lambda = \hat{\omega}^2 - \hat{k}^2 - \hat{k}/\hat{\omega} - 1$. This is Hermite's equation with solutions if and only if $\lambda = 2m$ for $m = 0, 1, 2, \ldots$, and it is this quantization condition that gives the dispersion relation. The solutions are Hermite polynomials; that is, $\Psi(\hat{y}) = H_m(\hat{y})$, where

$$(H_0, H_1, H_2, H_3, H_4) = (1,\ 2\hat{y},\ 4\hat{y}^2 - 2,\ 8\hat{y}^3 - 12\hat{y},\ 16\hat{y}^4 - 48\hat{y}^2 + 12). \tag{8.155}$$

The Hermite polynomial multiplied by a Gaussian is one form of parabolic cylinder functions, $V_m(y)$; that is

$$V_m(y) = H_m(y)\exp(-y^2/2). \tag{8.156}$$

The function $V_m(y)$ satisfies

$$\frac{\mathrm{d}^2 V_m}{\mathrm{d}y^2} + (2m + 1 - y^2)V_m = 0 \tag{8.157}$$

It is often useful to include the normalization coefficient in the definition of the cylinder function, whence

$$P_m = \frac{V_m}{\sqrt{2^m m!\,\sqrt{\pi}}}, \tag{8.158}$$

whence

$$\int_{\infty}^{\infty} P_m P_n = \delta_{mn}. \tag{8.159}$$

As may be verified by direct manipulation, these forms of parabolic cylinder functions obey certain recurrence relations, namely

$$\frac{\mathrm{d}P_m}{\mathrm{d}y} = -\frac{(m+1)^{1/2}}{\sqrt{2}} P_{m+1} + \frac{m^{1/2}}{\sqrt{2}} P_{m-1}, \tag{8.160a}$$

$$yP_m = \frac{m^{1/2}}{\sqrt{2}} P_{m-1} + \frac{(m+1)^{1/2}}{\sqrt{2}} P_{m+1}. \tag{8.160b}$$

and therefore

$$\frac{\mathrm{d}P_m}{\mathrm{d}y} + yP_m = (2m)^{1/2} P_{m-1}, \tag{8.161a}$$

$$\frac{\mathrm{d}P_m}{\mathrm{d}y} - yP_m = -\sqrt{2}(m+1)^{1/2} P_{m+1}. \tag{8.161b}$$

When $m = 0$ the recurrence relations are

$$\frac{\mathrm{d}P_0}{\mathrm{d}y} = \frac{-1}{\sqrt{2}} P_1, \qquad yP_0 = \frac{1}{\sqrt{2}} P_1. \tag{8.162a,b}$$

We don't use these relations in this chapter, although had we developed the forced-dissipative problem using this form of cylinder functions these relations would have been used instead of (8.98). Such a development would have been equivalent, although a little more awkward, to that presented.

*Parabolic cylinder functions*

Parabolic cylinder functions, $D_n(y)$, in the other commonly used form, are the modified Hermite polynomials (8.45) multiplied by a Gaussian; that is

$$D_n(y) = G_n(y) \exp(-y^2/4). \tag{8.163}$$

and these functions are solutions of (8.42) which arises when we use the nondimensionalization

$$T_{eq} = (2c\beta)^{-1/2}, \qquad L_{eq} = (c/2\beta)^{1/2}. \tag{8.164a,b}$$

These parabolic cylinder functions satisfy

$$\frac{\mathrm{d}^2 D_m}{\mathrm{d}y^2} + \frac{1}{2}\left(2m + 1 - \frac{1}{2}y^2\right) D_m = 0, \tag{8.165}$$

which is sometimes called the Weber differential equation. The functions also have the properties that

$$\frac{\mathrm{d}D_n}{\mathrm{d}y} + \frac{1}{2} y D_n = nD_{n-1}, \tag{8.166a}$$

$$\frac{\mathrm{d}D_n}{\mathrm{d}y} - \frac{1}{2}yD_n = -D_{n+1}. \tag{8.166b}$$

These two equations may be combined to give (8.165). The functions also satisfy

$$D_{n+1} - yD_n + nD_{n-1} = 0. \tag{8.167}$$

The simplicity of these particular ladder operators makes this form of the parabolic cylinder functions the most useful in our development of the forced, steady (Matsuno–Gill) problem, although the use of (8.161) is of course possible and, in the end, equivalent.

## 8.B APPENDIX B: SOME MATHEMATICAL RELATIONS IN THE MATSUNO–GILL PROBLEM

In this appendix we provide various zonal and meridional integrals of the solutions given in section 8.5.2.

The zonal integral of the forcing is given by

$$I = \int_{-\infty}^{\infty} Q_0(x)\,\mathrm{d}x = \int_{-L}^{L} \cos(kx)\,\mathrm{d}x = \frac{4L}{\pi}, \tag{8.168}$$

using $k = \pi/2L$.

The zonal integrals of the various $q$, $r$ and $v$ fields are given as follows. Using (8.125) with $F_0 = 0$ we see that

$$\int_{-\infty}^{\infty} \alpha q_0\,\mathrm{d}x = -\Big[q_0\Big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} Q_0(x)\,\mathrm{d}x = -I. \tag{8.169}$$

Using (8.136) we obtain similar results for $q_2, r_0$ and $v_1$, to wit

$$\int_{-\infty}^{\infty} (q_0, q_2, r_0, v_1)\;\mathrm{d}x = \left(-1, -\frac{1}{3}, -\frac{2}{3}, -\frac{\alpha}{3}\right)\frac{I}{\alpha}. \tag{8.170}$$

The zonally integrated pressure and velocity fields are obtained using (8.168), (8.170) and the nondimensional form of (8.140), giving

$$\int_{-\infty}^{\infty} (u, v, w, \phi)\;\mathrm{d}x = \left(\frac{-y^2}{6\alpha}, \frac{-y}{3}, \frac{2-y^2}{6}, \frac{-4-y^2}{6\alpha}\right)\left(\frac{4L}{\pi}\right)\exp(-y^2/4). \tag{8.171}$$

The meridional integrals of the velocity fields may also be calculated. To do this we first note the integrals

$$\int_{-\infty}^{\infty} (1, y, y^2)\exp(-y^2/4)\,\mathrm{d}y = (2, 0, 4)\sqrt{\pi}. \tag{8.172}$$

The first of these is a standard result, the second follows from considerations of symmetry and the third follows on integration by parts. Using (8.172) and the nondimensional form of (8.140) we obtain

$$\int_{-\infty}^{\infty} u\,\mathrm{d}y = \sqrt{\pi}\,[q_0(x) - q_2(x)]\,, \tag{8.173a}$$

$$\int_{-\infty}^{\infty} v\,\mathrm{d}y = 0, \tag{8.173b}$$

$$\int_{-\infty}^{\infty} w \, \mathrm{d}y = \sqrt{\pi}\left[\alpha q_0(x) + 3\alpha q_2(x) + 2Q_0(x)\right], \tag{8.173c}$$

$$\int_{-\infty}^{\infty} \phi \, \mathrm{d}y = \sqrt{\pi}\left[q_0(x) + 3q_2(x)\right]. \tag{8.173d}$$

Equations (8.171) and (8.173) are useful because they allow us to define streamfunctions for the overturning circulation in the zonal and meridional plane, respectively. From (8.171) we see that

$$\overline{w}^x + \frac{\partial \overline{v}^x}{\partial y} = 0, \tag{8.174}$$

and from (8.173), and using (8.125) and (8.136a), we find that

$$\overline{w}^y + \frac{\partial \overline{u}^y}{\partial x} = 0, \tag{8.175}$$

with the overbar denoting a zonal or meridional average, as indicated. These results are to be expected from the mass continuity equation, $w = -(\partial_x u + \partial_y v)$ on zonal and meridional integration, respectively, but the fact that the solutions show it so explicitly is a demonstration of the magical karma of mathematics.

A streamfunction may be constructed by supposing that, in a fluid of depth $H$, the horizontal and vertical velocities vary as

$$(u, v) = (\widetilde{u}, \widetilde{v}) \cos(\pi z/H), \qquad w = \widetilde{w} \sin(\pi z/H). \tag{8.176a,b}$$

Using (8.171) the streamfunction in the meridional plane, $\Psi_M$ is given by

$$\Psi_M(y, z) = \frac{IH}{\pi}\frac{-y}{3}\exp(-y^2/4)\sin \pi z/H. \tag{8.177}$$

Using (8.173) the streamfunction in the zonal plane, $\Psi_Z$, is given by

$$\Psi_Z(x, z) = \frac{\sqrt{\pi}H}{\pi}\left[q_0(x) - q_2(x)\right]\sin \pi z/H. \tag{8.178}$$

*Notes*

1 This correspondence was shown by Matsuno (1966), although there may have been earlier demonstrations.

2 See Chelton *et al.* (1998) for a description of the method and maps of the first deformation radius and related quantities for the world's oceans.

3 The first complete treatment of this seems to have been given by Matsuno (1966). Earlier, Stern (1963) and Bretherton (1964) discussed some special cases. A review of equatorial waves and circulation is provided by McCreary (1985).

4 Standard forms are, of course, in the eye of the beholder.

5 After Yanai. Need some background here. xxx

6 Following Verkley & van der Velde (2010).

7 The notation, and the idea, comes from Gill & Clarke (1974).

8 Gill & Clarke (1974) deal with the problem in more detail.

9 This problem was first considered by Matsuno (1966) and revisited by Gill (1980) in the context of understanding the response of the tropical atmosphere to diabatic heating. It is now commonly referred to as the *Matsuno–Gill* problem. Our mathematical treatment is more similar to that of Gill.

10 That timestepping is a simple way to obtain solutions was pointed out to me by Matthew Barlow. A code that solves the elliptic problem, using Fourier transforms and a tridiagonal inversion, was developed by Chris Bretherton and Adam Sobel, and I am grateful to them for sharing it with me. The stepping code and the elliptic solver give virtually identical results for the solutions shown in Fig. 8.8. The elliptic solver was used to obtain solutions for the line source, shown in Figs. 8.9 and 8.10.

11 For more information about Hermite polynomials and parabolic cylinder functions see, for example, Jeffreys & Jeffreys (1946), Ambramowitz & Stegun (1965) or mathematical software such as Maple™.

*Problems*

8.1 Both easy and fiendishly difficult problems will go here. Please send me some if you have any.

*Another advantage of a mathematical statement is that it is so definite that it might be definitely wrong... Some verbal statements have not this merit.*

L. F. Richardson (1881–1953).

# CHAPTER TEN

# Waves, Mean-Flows and Conservation Properties

Wave–mean-flow interaction is concerned with how some mean flow, perhaps a time or zonal average, interacts with a departure from that mean, and this chapter provides an elementary introduction to a number of topics in this area. It is 'elementary' because our derivations and discussion are obtained by direct and straightforward manipulations of the equations of motion, often in the simplest case that will illustrate the relevant principle. It is implicit in what we do that it is a sensible thing to decompose the fields into a mean plus some departure, and one case when this is so is when the departure is of small amplitude. Departures from the mean — generically called *eddies* — are of course not always small; for example, in the mid-latitude troposphere the eddies are often of similar amplitude to the mean flow, and chapters 12 and 13 will explore this from the standpoint of turbulence. However, in this chapter we will generally assume that eddies are indeed of small amplitude, and, in particular, that eddy–mean-flow interaction is larger than eddy–eddy interaction.

A *wave* is an eddy that satisfies, at least approximately, a dispersion relation. It is the presence of such a dispersion relation that enables a number of results to be obtained that would otherwise be out of our reach, and that gives rise to the appellation '*wave*–mean-flow'. In mid-latitudes the relevant waves are usually Rossby waves, as introduced in chapter 5, although gravity waves also interact with the mean flow. It is implicit in defining waves this way that they are generally of small amplitude, for it is this that allows the equations of motion to be sensibly linearized and a dispersion relation to be obtained (although a monochromatic wave may have finite amplitude and still satisfy a dispersion relation). However, this does not mean that the waves do not interact with each other and with the mean flow; we may expect, or at least hope, that the qualitative nature of such interactions, as calculated by wave–mean-flow interaction theory, will carry over and provide insights into the finite-amplitude problem. Thus, one goal of wave–mean-flow theory is to provide a way of qualitatively understanding more realistic situations, and to suggest diagnostics that might be used to analyze both observations and numerical solutions of the fully nonlinear problem. In this chapter we will almost exclusively concern ourselves with a *zonal* mean, since this is the simplest and often most useful case because of the presence of simple — i.e., periodic — boundary conditions. We will also be mainly concerned with quasi-geostrophic dynamics on a $\beta$-plane. The reader who is anxious for real examples might wish to first look at chapter 17 and then come back to this chapter as needed.

## 10.1 QUASI-GEOSTROPHIC WAVE–MEAN-FLOW INTERACTION

### 10.1.1 Preliminaries

To fix our dynamical system and notation, we write down the quasi-geostrophic potential vorticity equation

$$\frac{\partial q}{\partial t} + J(\psi, q) = D, \tag{10.1}$$

where $D$ represents any non-conservative terms and the potential vorticity in a Boussinesq system is

$$q = \beta y + \zeta + \frac{\partial}{\partial z}\left(\frac{f_0}{N^2} b\right), \tag{10.2}$$

where $\zeta$ is the relative vorticity and $b$ is the buoyancy perturbation from the background state characterized by $N^2$. [In an ideal gas $q = \beta y + \zeta + (f_0/\rho_R)\partial_z(\rho_R b/N^2)$, where $\rho_R$ is a specified density profile, and most of our derivations can be extended to that case.] We will refer to lines of constant $b$ as isentropes. In terms of the streamfunction, the variables are

$$\zeta = \nabla^2\psi, \qquad b = f_0\frac{\partial \psi}{\partial z}, \qquad q = \beta y + \left[\nabla^2 + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\right)\right]\psi. \tag{10.3}$$

where $\nabla^2 \equiv (\partial_x^2 + \partial_y^2)$. The potential vorticity equation holds in the fluid interior; the boundary conditions on (10.3) are provided by the thermodynamic equation

$$\frac{\partial b}{\partial t} + J(\psi, b) + wN^2 = H, \tag{10.4}$$

where $H$ represents heating terms. The vertical velocity at the boundary, $w$, is zero in the absence of topography and Ekman friction, and if $H$ is also zero the boundary condition is just

$$\frac{\partial b}{\partial t} + J(\psi, b) = 0. \tag{10.5}$$

Equations (10.1) and (10.5) are the evolution equations for the system and if both $D$ and $H$ are zero they conserve both the total energy, $\widehat{E}$ and the total enstrophy, $\widehat{Z}$:

$$\begin{aligned} \frac{\mathrm{d}\widehat{E}}{\mathrm{d}t} &= 0, \qquad \widehat{E} = \frac{1}{2}\int_V (\nabla\psi)^2 + \frac{f_0^2}{N^2}\left(\frac{\partial\psi}{\partial z}\right)^2 \mathrm{d}V, \\ \frac{\mathrm{d}\widehat{Z}}{\mathrm{d}t} &= 0, \qquad \widehat{Z} = \frac{1}{2}\int_V q^2\,\mathrm{d}V. \end{aligned} \tag{10.6}$$

where $V$ is a volume bounded by surfaces at which the normal velocity is zero, or that has periodic boundary conditions. The enstrophy is also conserved layerwise; that is, the horizontal integral of $q^2$ is conserved at every level.

### 10.1.2 Potential vorticity flux in the linear equations

Let us decompose the fields into a mean (to be denoted with an overbar) plus a perturbation (denoted with a prime), and let us suppose the perturbation fields are of small amplitude. (In linear problems, such as those considered in chapter 9, we decomposed the flow into a 'basic state' plus a perturbation, with the basic state fixed in time. Our approach here is similar, but soon we will allow the mean state to evolve.) The linearized quasi-geostrophic potential vorticity equation is then

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + u'\frac{\partial \overline{q}}{\partial x} + \overline{v}\frac{\partial q'}{\partial y} + v'\frac{\partial \overline{q}}{\partial y} = D'. \tag{10.7}$$

where $D'$ represents eddy forcing and dissipation and, in terms of streamfunction,

$$\big(u'(x,y,z,t), v'(x,y,z,t)\big) = \left(-\frac{\partial \psi'}{\partial y}, \frac{\partial \psi'}{\partial x}\right), \tag{10.8a}$$

$$q'(x,y,z,t) = \nabla^2\psi' + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \psi'}{\partial z}\right). \tag{10.8b}$$

If the mean is a zonal mean then $\partial\overline{q}/\partial x = 0$ and $\overline{v} = 0$ (because $v$ is purely geostrophic) and (10.7) simplifies to

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = D', \tag{10.9}$$

where

$$\overline{q} = \beta y - \frac{\partial \overline{u}}{\partial y} + \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\overline{b}\right), \quad \text{and} \quad \frac{\partial \overline{q}}{\partial y} = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} - \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \overline{u}}{\partial z}\right). \tag{10.10a,b}$$

using thermal wind, $f_0\partial\overline{u}/\partial z = -\partial b/\partial y$.

Multiplying by $q'$ and zonally averaging gives the enstrophy equation:

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{q'^2} = -\overline{v'q'}\frac{\partial \overline{q}}{\partial y} + \overline{D'q'}. \tag{10.11}$$

The quantity $\overline{v'q'}$ is the meridional flux of potential vorticity; this is downgradient (by definition) when the first term on the right-hand side is positive (i.e., $\overline{v'q'}\partial\overline{q}/\partial y < 0$), and it then acts to increase the variance of the perturbation. (This occurs, for example, when the flux is diffusive so that $\overline{v'q'} = -\kappa\partial\overline{q}/\partial y$, where $\kappa$ may vary but is everywhere positive.) This argument may be inverted: for inviscid flow ($D = 0$), if the waves are growing, as for example in the canonical models of baroclinic instability discussed in chapter 9, then *the potential vorticity flux is downgradient*.

If the second term on the right-hand side of (10.11) is negative, as it will be if $D'$ is a dissipative process (e.g., if $D' = A\nabla^2 q'$ or if $D' = -rq'$, where $A$ and $r$ are positive) then a statistical balance can be achieved between enstrophy production via downgradient transport, and dissipation. If the waves are steady (by which we mean statistically steady, neither growing nor decaying in amplitude) and conservative (i.e., $D' = 0$) then we must have

$$\overline{v'q'} = 0. \tag{10.12}$$

Similar results follow for the buoyancy at the boundary; we start by linearizing the thermodynamic equation (10.5) to give

$$\frac{\partial b'}{\partial t} + \overline{u}\frac{\partial b'}{\partial x} + v'\frac{\partial \overline{b}}{\partial y} = H', \tag{10.13}$$

where $H'$ is a diabatic source term. Multiplying (10.13) by $b'$ and averaging gives

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{b'^2} = -\overline{v'b'}\frac{\partial \overline{b}}{\partial y} + \overline{H'b'}. \tag{10.14}$$

Thus growing adiabatic waves have a downgradient flux of buoyancy at the boundary. In the Eady problem there is no interior gradient of basic-state potential vorticity and all the terms in (10.11) are zero, but the perturbation grows at the boundary. If the waves are steady and adiabatic then, analogously to (10.12),

$$\overline{v'b'} = 0. \tag{10.15}$$

The boundary conditions and fluxes may be absorbed into the interior definition of potential vorticity and its fluxes by way of the delta-function boundary layer construction, described in section 5.4.3. In models with discrete vertical layers or a finite number of levels it is common practice to absorb the boundary conditions into the definition of potential vorticity at top and bottom.

### 10.1.3 Wave–mean-flow interaction

In linear problems we usually suppose that the mean flow is fixed and in that case in the zonal mean terms, $\overline{u}$ and $\overline{q}$, in (10.9),would be functions only of $y$ and $z$. However, we in reality we might expect that the mean flow would change because of momentum and heat flux convergences arising from the eddy-eddy interactions. Thus, if we begin with the potential vorticity equation (10.1) and, in the usual way, express the variables as a zonal mean plus an eddy term, we obtain

$$\frac{\partial \overline{q}}{\partial t} + \nabla\cdot(\overline{\boldsymbol{u}}\,\overline{q}) + \nabla\cdot\overline{(\boldsymbol{u}'q')} = \overline{D}. \tag{10.16}$$

Now, since the mean flow is a zonal mean, and $\overline{v} = 0$, the first term is zero and the mean flow evolves according to

$$\frac{\partial \overline{q}}{\partial t} + \frac{\partial}{\partial y}\overline{v'q'} = \overline{D}. \tag{10.17}$$

Similarly, at the boundary the mean buoyancy evolution equation is

$$\frac{\partial \overline{b}}{\partial t} + \frac{\partial}{\partial y}\overline{v'b'} = \overline{H}. \tag{10.18}$$

To obtain $\overline{u}$ from $\overline{q}$ and $\overline{b}$ we use thermal wind balance to define a streamfunction $\Psi$. That is, since

$$f_0\frac{\partial \overline{u}}{\partial z} = -\frac{\partial \overline{b}}{\partial y}, \quad \text{then} \quad \left(\overline{u}, \frac{1}{f_0}\overline{b}\right) = \left(-\frac{\partial \Psi}{\partial y}, \frac{\partial \Psi}{\partial z}\right) \tag{10.19a,b}$$

whence, using (10.10a), the potential vorticity is

$$\overline{q}(y,z,t) - \beta y = \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \Psi}{\partial z}\right) + \frac{\partial^2 \Psi}{\partial y^2}. \tag{10.20}$$

If $\overline{q}$ is known in the interior from (10.18), and $\overline{b}$ (i.e., $f_0\partial\Psi/\partial z$) is known at the boundaries, then $\overline{u}$ and $\overline{b}$ in the interior may be obtained using (10.20) and (10.19b).

To close the system we suppose that the eddy terms themselves evolve according to (10.9) and (10.13). If in those equations we were to include the eddy-eddy interaction terms we would simply recover the full system, so in neglecting those terms we have constructed an eddy–mean-flow system, commonly called a *wave–mean-flow* system because by eliminating the nonlinear terms in the perturbation equation the eddies will often be wavelike. Non quasi-geostrophic wave–mean-flow systems may be constructed in a similar fashion: for example, we could construct a system using the primitive equations with separate equations for eddy and zonal-mean temperature and velocity fields, and an example involving gravity waves is given in chapter 17.

Note that such systems do differ from linear ones. In constructing linear systems we posit that the eddy terms are small compared to the mean flow and thus neglect the eddy-eddy interaction terms. In a wave–mean-flow problem we similarly suppose the eddy terms are small, and we neglect eddy-eddy interaction terms where they produce another eddy, because the terms involving the mean flow are larger. However, in the mean flow equation, (10.16), there are no larger mean flow terms and we keep the eddy-eddy terms and allow the mean flow to evolve. Still, such a justification is hardly a rigorous one, since if the eddy terms are small then the effects on the mean flow will be small and so one might suppose that the mean flow should be held fixed. The wave–mean flow equations really can only be justified on a case-by-case basis with a detailed examination of the size of the terms and the rate at which they evolve, and that becomes the subject of weakly nonlinear theory. It is also often the case that in applications one allows only a very limited set of wave terms — for example, one might allow the eddies to be represented by just a pair of Fourier modes. In any case, the wave–mean-flow problem can be a useful tool to gain insight into the behaviour of the full system. The equations are summarized in the grey box on page 445.

## 10.2 THE ELIASSEN–PALM FLUX

The eddy flux of potential vorticity may be expressed in terms of vorticity and buoyancy fluxes as

$$v'q' = v'\zeta' + f_0 v'\frac{\partial}{\partial z}\left(\frac{b'}{N^2}\right). \tag{10.21}$$

The second term on the right-hand side can be written as

$$\begin{aligned} f_0 v'\frac{\partial}{\partial z}\left(\frac{b'}{N^2}\right) &= f_0\frac{\partial}{\partial z}\left(\frac{v'b'}{N^2}\right) - f_0\frac{\partial v'}{\partial z}\frac{b'}{N^2} \\ &= f_0\frac{\partial}{\partial z}\left(\frac{v'b'}{N^2}\right) - f_0\frac{\partial}{\partial x}\left(\frac{\partial \psi'}{\partial z}\right)\frac{b'}{N^2} \\ &= f_0\frac{\partial}{\partial z}\left(\frac{v'b'}{N^2}\right) - \frac{f_0^2}{2N^2}\frac{\partial}{\partial x}\left(\frac{\partial \psi'}{\partial z}\right)^2, \end{aligned} \tag{10.22}$$

using $b' = f_0\partial\psi'/\partial z$.

Similarly, the flux of relative vorticity can be written

$$v'\zeta' = -\frac{\partial}{\partial y}(u'v') + \frac{1}{2}\frac{\partial}{\partial x}(v'^2 - u'^2) \tag{10.23}$$

Using (10.22) and (10.23), (10.21) becomes

$$\boxed{v'q' = -\frac{\partial}{\partial y}(u'v') + \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}v'b'\right) + \frac{1}{2}\frac{\partial}{\partial x}\left((v'^2 - u'^2) - \frac{b'^2}{N^2}\right).} \tag{10.24}$$

Thus the meridional potential vorticity flux, in the quasi-geostrophic approximation, can be written as the divergence of a vector: $v'q' = \nabla\cdot\boldsymbol{\mathcal{E}}$ where

$$\boldsymbol{\mathcal{E}} \equiv \frac{1}{2}\left((v'^2 - u'^2) - \frac{b'^2}{N^2}\right)\mathbf{i} - (u'v')\,\mathbf{j} + \left(\frac{f_0}{N^2}v'b'\right)\mathbf{k}. \tag{10.25}$$

A particularly useful form of this arises after zonally averaging, for then (10.24) becomes

$$\boxed{\overline{v'q'} = -\frac{\partial}{\partial y}\overline{u'v'} + \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\overline{v'b'}\right)}. \tag{10.26}$$

The vector defined by

$$\boldsymbol{\mathcal{F}} \equiv -\overline{u'v'}\,\mathbf{j} + \frac{f_0}{N^2}\overline{v'b'}\,\mathbf{k} \tag{10.27}$$

is called the (quasi-geostrophic) *Eliassen–Palm (EP) flux*,[1] and its divergence, given by (10.26), gives the poleward flux of potential vorticity:

$$\overline{v'q'} = \nabla_x\cdot\boldsymbol{\mathcal{F}}, \tag{10.28}$$

where $\nabla_x\cdot \equiv (\partial/\partial y, \partial/\partial z)\cdot$ is the divergence in the meridional plane. Unless the meaning is unclear, the subscript $x$ on the meridional divergence will be dropped.

### 10.2.1 The Eliassen–Palm relation

On dividing by $\partial\overline{q}/\partial y$ and using (10.28), the enstrophy equation (10.11) becomes

$$\boxed{\frac{\partial\mathcal{A}}{\partial t} + \nabla\cdot\boldsymbol{\mathcal{F}} = \mathcal{D}}, \tag{10.29a}$$

where

$$\mathcal{A} = \frac{\overline{q'^2}}{2\partial\overline{q}/\partial y}, \qquad \mathcal{D} = \frac{\overline{D'q'}}{\partial\overline{q}/\partial y}. \tag{10.29b}$$

Equation (10.29a) is known as the *Eliassen–Palm relation*, and it is a conservation law for the *wave activity density* $\mathcal{A}$. If we integrate (10.29b) over a meridional area $A$ bounded by walls where the eddy activity vanishes, and if $\mathcal{D} = 0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_A \mathcal{A}\,\mathrm{d}A = 0. \tag{10.30}$$

The integral is a wave activity — a quantity that is quadratic in the amplitude of the perturbation and that is conserved in the absence of forcing and dissipation. In this case $\mathcal{A}$ is the negative of the *pseudomomentum*, for reasons we will encounter later. Note that neither the perturbation energy nor the perturbation enstrophy are wave activities of the linearized equations, because there can be an exchange of energy or enstrophy between mean and perturbation — indeed, this is how a perturbation grows in baroclinic or barotropic instability! This is already evident from (10.11), or in general take (10.7) with $D' = 0$ and multiply by $q'$ to give the enstrophy equation

$$\frac{1}{2}\frac{\partial q'^2}{\partial t} + \frac{1}{2}\overline{\boldsymbol{u}} \cdot \nabla q'^2 + \boldsymbol{u}' q' \cdot \nabla \overline{q} = 0, \tag{10.31}$$

where here the overbar is an average (although it need not be a zonal average). Integrating this over a volume $V$ gives

$$\frac{\mathrm{d}\hat{Z}'}{\mathrm{d}t} \equiv \frac{\mathrm{d}}{\mathrm{d}t}\int_V \frac{1}{2}q'^2\,\mathrm{d}V = -\int_V \boldsymbol{u}' q' \cdot \nabla \overline{q}\,\mathrm{d}V. \tag{10.32}$$

The right-hand side does not, in general, vanish and so $\hat{Z}'$ is not in general conserved.

### 10.2.2 The group velocity property for Rossby waves

The vector $\boldsymbol{\mathcal{F}}$ describes how the wave activity propagates. We noted in chapter 6 that in the case in which the disturbance is composed of plane or almost plane waves that satisfy a dispersion relation, then $\boldsymbol{\mathcal{F}} = \boldsymbol{c}_g\mathcal{A}$, where $\boldsymbol{c}_g$ is the group velocity and (10.29a) becomes

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot (\mathcal{A}\boldsymbol{c}_g) = 0. \tag{10.33}$$

This is a useful property, because if we can diagnose $\boldsymbol{c}_g$ from observations we can use (10.29a) to determine how wave activity density propagates. Let us demonstrate this explicitly for the pseudomomentum in Rossby waves, that is for (10.29a).

The Boussinesq quasi-geostrophic equation on the $\beta$-plane, linearized around a uniform zonal flow and with constant static stability, is

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = 0, \tag{10.34}$$

where $q' = [\nabla^2 + (f_0^2/N^2)\partial^2/\partial z^2]\psi'$ and, if $\overline{u}$ is constant, $\partial\overline{q}/\partial y = \beta$. Thus we have

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\left[\nabla^2\psi' + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \psi'}{\partial z}\right)\right] + \beta\frac{\partial \psi'}{\partial x} = 0. \tag{10.35}$$

Seeking solutions of the form

$$\psi' = \mathrm{Re}\,\widetilde{\psi}\mathrm{e}^{\mathrm{i}(kx+ly+mz-\omega t)}, \tag{10.36}$$

we find the dispersion relation,

$$\omega = \overline{u}k - \frac{\beta k}{\kappa^2}. \tag{10.37}$$

where $\kappa^2 = (k^2 + l^2 + m^2 f_0^2/N^2)$, and the group velocity components:

$$c_g^y = \frac{2\beta kl}{\kappa^4}, \qquad c_g^z = \frac{2\beta km f_0^2/N^2}{\kappa^4}. \tag{10.38}$$

Also, if $u' = \mathrm{Re}\,\widetilde{u}\exp[\mathrm{i}(kx + ly + mz - \omega t)]$, and similarly for the other fields, then

$$\begin{aligned} \widetilde{u} &= -\mathrm{Re}\ \mathrm{i}l\widetilde{\psi}, \qquad & \widetilde{v} &= \mathrm{Re}\ \mathrm{i}k\widetilde{\psi}, \\ \tilde{b} &= \mathrm{Re}\,\mathrm{i}mf_0\widetilde{\psi}, \qquad & \widetilde{q} &= -\mathrm{Re}\,\kappa^2\widetilde{\psi}, \end{aligned} \tag{10.39}$$

The wave activity density is then

$$\mathcal{A} = \frac{1}{2}\frac{\overline{q'^2}}{\beta} = \frac{\kappa^4}{4\beta}|\widetilde{\psi}^2|, \tag{10.40}$$

where the additional factor of 2 in the denominator arises from the averaging. Using (10.39) the EP flux, (10.27), is

$$\mathcal{F}^y = -\overline{u'v'} = \frac{1}{2}kl|\widetilde{\psi}^2|, \qquad \mathcal{F}^z = \frac{f_0}{N^2}\overline{v'b'} = \frac{f_0^2}{2N^2}km|\widetilde{\psi}^2|. \tag{10.41}$$

Using (10.38), (10.40) and (10.41) we obtain

$$\boxed{\boldsymbol{\mathcal{F}} = (\mathcal{F}^y, \mathcal{F}^z) = \mathbf{c}_g\mathcal{A}}. \tag{10.42}$$

If the properties of the medium are slowly varying, so that a (spatially varying) group velocity can still be defined, then this is a useful expression to estimate how the wave activity propagates in the atmosphere and in numerical simulations.

### 10.2.3 ✦ The orthogonality of modes

It is a direct consequence of the conservation of wave activity that disturbance modes are orthogonal in the 'wave activity norm', defined later on, and thus are a useful measure of the amplitude of a particular mode.[2] To explore this, we start with the linearized potential vorticity equation,

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = 0. \tag{10.43}$$

Let us formally seek solutions of the form $\psi' = \mathrm{Re}\,\Psi\exp(\mathrm{i}kx)$ where $\Psi$ is the sum of *modes*,

$$\Psi = \sum_n \widetilde{\psi}_n(y, z)\mathrm{e}^{-\mathrm{i}kc_nt}, \tag{10.44}$$

where $n$ is an identifier of the modes. The modes satisfy

$$(\overline{u}\Delta_k^2 + \overline{q}_y)\widetilde{\psi}_n = c_n\Delta_k^2\widetilde{\psi}_n, \tag{10.45}$$

where

$$\Delta_k^2 = \frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\right) - k^2. \tag{10.46}$$

The upper and lower boundary conditions (at $z = 0, -H$) are given by the thermodynamic equation

$$\frac{\partial b'}{\partial t} + \overline{u}\frac{\partial b'}{\partial x} + v'\frac{\partial \overline{b}}{\partial y} = 0, \tag{10.47}$$

and if we simplify further by supposing $\partial \overline{u}/\partial z = 0$ then the boundary condition becomes

$$\frac{\partial \psi'_z}{\partial t} + \overline{u}\frac{\partial \psi'_z}{\partial x} = 0. \tag{10.48}$$

There are no meridional buoyancy fluxes at the boundary. If $N^2$ is a constant (a simplifying but not essential assumption) then we can let $\widetilde{\psi}_n(y,z) = \psi_n(y)\cos pz$, with $p = j\pi/H$ where $j$ is an integer and the mode $n$ now labels only the meridional modes. The corresponding potential vorticity modes are given by

$$q_n = \Delta^2_{k,m}\psi_n, \qquad \Delta^2_{k,m} = \frac{\partial^2}{\partial y^2} - (f_0^2/N^2)m^2 - k^2, \tag{10.49}$$

and the boundary conditions are then built in to any solution we construct from (10.45) and (10.49).[3] We may then consider a single zonal and a single vertical wavenumber. (If there is no horizontal variation of the shear, the meridional modes are harmonic functions, for example $\psi_n \propto \sin(n\pi y/L)$ for a channel of width $L$.)

For a given basic state we may imagine solving (10.45), numerically or analytically, and determining the modes. However, these modes are not orthogonal in the sense of either energy or enstrophy. That is, denoting the inner product by

$$\langle a, b\rangle \equiv (2L)^{-1}\int_L ab\,\mathrm{d}y, \tag{10.50}$$

then, in general,

$$I_E = \langle \psi_n, q_m\rangle \neq 0, \qquad I_Z = \langle q_n, q_m\rangle \neq 0, \tag{10.51a,b}$$

for $n \neq m$, where $q_n = \Delta^2_{k,p}\psi_n$. Perturbation energy and enstrophy are thus not wave activities of the linearized equations, and it is not meaningful to talk about the energy or enstrophy of a particular mode. However, by the same token we may expect orthogonality in the wave activity norm. To prove this and understand what it means, suppose that at $t = 0$ the disturbance consists of two modes, $n$ and $m$, so that at a later time $q = (q_n \mathrm{e}^{-\mathrm{i}kc_n t} + q_m \mathrm{e}^{-\mathrm{i}kc_m t} + \text{c.c.})$, where $c_m \neq c_n$ and we assume that both are real. The wave activity is

$$P \equiv \int \mathcal{A}\,\mathrm{d}y\,\mathrm{d}z = \left\langle q_n, q_m^*/\overline{q}_y\right\rangle \mathrm{e}^{-\mathrm{i}k(c_n - c_m)t} + \left\langle q_m, q_m^*/\overline{q}_y\right\rangle + \left\langle q_n, q_n^*/\overline{q}_y\right\rangle + \text{c.c.} \tag{10.52}$$

The second and third terms on the right-hand side are the wave activities of each mode, and these are constants (to see this, consider the case when the disturbance is just a single mode). Now, because $\mathrm{d}P/\mathrm{d}t = 0$ the first term must vanish if $c_n \neq c_m$, implying the modes are orthogonal and, in particular,

$$\mathrm{Re}\int \frac{1}{\overline{q}_y} q_n q_m^*\,\mathrm{d}y = 0, \tag{10.53}$$

for $n \neq m$. The inner product weighted by $1/\overline{q}_y$ defines the wave activity norm. Readers who would prefer a more direct derivation of the orthogonality condition directly from the eigenvalue equation (10.45) should see problem 10.3. Orthogonality is a useful result, for it means that the wave activity is a proper measure of the amplitude of a given mode unlike, for example, energy. The conservation of wave activity will lead to a particularly straightforward derivation of the necessary conditions for stability, given in section 10.7.

## 10.3 THE TRANSFORMED EULERIAN MEAN

The so-called *transformed Eulerian mean*, or TEM, is a transformation of the equations of motion that provides a useful framework for discussing eddy effects under a wide range of conditions.[4] It is useful because, as we shall see, it is equivalent to a very natural form of averaging the equations that serves to eliminate eddy fluxes in the thermodynamic equation and collect them together, in a simple form, in the momentum equation and in so doing it highlights the role of potential vorticity fluxes. The TEM also provides a natural separation between diabatic and adiabatic effects or between advective and diffusive fluxes and, in the case in which the flow is adiabatic, a pleasing simplification of the equations. In later chapters we will use the TEM to better understand the mid-latitude troposphere and the dynamics of the Antarctic Circumpolar Current, and as a framework for the parameterization of eddy fluxes. Of course, there being no free lunch, the TEM brings with it its own difficulties, and in particular the implementation of boundary conditions can cause difficulties, especially in the actual numerical integration of the equations.

### 10.3.1 Quasi-geostrophic form

For simplicity we will use the Boussinesq equations on the beta-plane, and the zonally averaged Eulerian mean equations for the zonally averaged zonal velocity and the buoyancy may then be written as (see section 2.2.6)

$$\frac{\partial \overline{u}}{\partial t} - (f + \overline{\zeta})\overline{v} + \overline{w}\frac{\partial \overline{u}}{\partial z} = -\frac{\partial}{\partial y}\overline{u'v'} - \frac{\partial}{\partial z}\overline{u'w'} + \overline{F}, \tag{10.54a}$$

$$\frac{\partial \overline{b}}{\partial t} + \overline{v}\frac{\partial \overline{b}}{\partial y} + \overline{w}\frac{\partial \overline{b}}{\partial z} = -\frac{\partial}{\partial y}\overline{v'b'} - \frac{\partial}{\partial z}\overline{w'b'} + \overline{S}, \tag{10.54b}$$

where $\overline{F}$ and $\overline{S}$ represent frictional and heating terms, respectively. Note that the meridional velocity, $\overline{v}$, is purely ageostrophic. Using quasi-geostrophic scaling we neglect the vertical eddy flux divergences and all ageostrophic velocities except when multiplied by $f_0$ or $N^2$. The above equations then become

$$\frac{\partial \overline{u}}{\partial t} = f_0\overline{v} - \frac{\partial}{\partial y}\overline{u'v'} + \overline{F}, \tag{10.55a}$$

$$\frac{\partial \overline{b}}{\partial t} = -N^2\overline{w} - \frac{\partial}{\partial y}\overline{v'b'} + \overline{S}. \tag{10.55b}$$

These two equations are connected by the thermal wind relation,

$$f_0\frac{\partial \overline{u}}{\partial z} = -\frac{\partial \overline{b}}{\partial y}, \tag{10.56}$$

which is a combination of the geostrophic $v$-momentum equation ($f_0\overline{u} = -\partial\overline{\phi}/\partial y$) and hydrostasy ($\partial\overline{\phi}/\partial z = \overline{b}$). One less than ideal aspect of (10.55) is that in the extratropics the dominant balance is usually between the first two terms on the right-hand sides of each equation, even in time-dependent cases. Thus, the Coriolis force closely balances the divergence of the eddy momentum fluxes, and the advection of the mean stratification ($N^2w$, or 'adiabatic cooling') often balances the divergence of eddy heat flux, with heating being a small residual. This may lead to an underestimation of the importance of diabatic heating, as this is ultimately responsible for the mean meridional circulation. Furthermore, the link between $\overline{u}$ and $\overline{b}$ via thermal wind dynamically couples buoyancy and momentum, and obscures the understanding of how the eddy fluxes influence these fields — is it through the eddy heat fluxes or momentum fluxes, or some combination?

To address this issue we combine the terms $N^2w$ and the eddy flux in (10.55b) into a single total or *residual* (so recognizing the cancellation between the mean and eddy terms) heat transport term that in a steady state is balanced by the diabatic term $\overline{S}$. To do this, we first note that because $\overline{v}$ and $\overline{w}$ are related by mass conservation we can define a mean meridional streamfunction $\psi_m$ such that

$$(\overline{v}, \overline{w}) = \left(-\frac{\partial\psi_m}{\partial z}, \frac{\partial\psi_m}{\partial y}\right). \tag{10.57}$$

The velocities then satisfy $\partial\overline{v}/\partial y + \partial\overline{w}/\partial z = 0$ automatically. If we define a *residual streamfunction* by

$$\psi^* \equiv \psi_m + \frac{1}{N^2}\overline{v'b'}, \tag{10.58a}$$

the components of the *residual mean meridional circulation* are then given by

$$(\overline{v}^*, \overline{w}^*) = \left(-\frac{\partial\psi^*}{\partial z}, \frac{\partial\psi^*}{\partial y}\right), \tag{10.58b}$$

and

$$\overline{v}^* = \overline{v} - \frac{\partial}{\partial z}\left(\frac{1}{N^2}\overline{v'b'}\right), \qquad \overline{w}^* = \overline{w} + \frac{\partial}{\partial y}\left(\frac{1}{N^2}\overline{v'b'}\right). \tag{10.59}$$

Note that by construction, the residual overturning circulation satisfies

$$\frac{\partial\overline{v}^*}{\partial y} + \frac{\partial\overline{w}^*}{\partial z} = 0. \tag{10.60}$$

Substituting (10.59) into (10.55a) and (10.55b) the zonal momentum and buoyancy equations then take the simple forms

$$\boxed{\begin{aligned} \frac{\partial\overline{u}}{\partial t} &= f_0\overline{v}^* + \overline{v'q'} + \overline{F} \\ \frac{\partial\overline{b}}{\partial t} &= -N^2\overline{w}^* + \overline{S} \end{aligned}}, \tag{10.61a,b}$$

which are known as the (quasi-geostrophic) *transformed Eulerian mean equations*, or TEM equations. The potential vorticity flux, $\overline{v'q'}$, is given in terms of the heat and vorticity fluxes by (10.26), and is equal to the divergence of the Eliassen–Palm flux as in (10.28).

The TEM equations make it apparent that we may consider the potential vorticity fluxes, rather than the separate contributions of the vorticity and heat fluxes, to force the circulation. If we know the potential vorticity flux as well as $\overline{F}$ and $\overline{S}$, then (10.60) and (10.61), along with thermal wind balance

$$f_0 \frac{\partial \overline{u}}{\partial z} = -\frac{\partial \overline{b}}{\partial y} \tag{10.62}$$

form a complete set. The meridional overturning circulation is obtained by eliminating time derivatives from (10.61) using (10.62), giving

$$f_0^2 \frac{\partial^2 \psi^*}{\partial z^2} + N^2 \frac{\partial^2 \psi^*}{\partial y^2} = f_0 \frac{\partial}{\partial z} \overline{v'q'} + f_0 \frac{\partial \overline{F}}{\partial z} + \frac{\partial \overline{S}}{\partial y}. \tag{10.63}$$

Thus, the residual or net overturning circulation is driven by the (vertical derivative of the) potential vorticity fluxes and the diabatic terms — driven in the sense that if we know those terms we can calculate the overturning circulation, although of course the fluxes themselves depend on the circulation. Note that this equation applies at every instant, even if the equations are not in a steady state.

Use of the TEM equations in TEM form is particularly advantageous when the eddy potential vorticity flux arises from wave activity, for example from Rossby waves. The potential vorticity flux is the convergence of the EP flux $\boldsymbol{\mathcal{F}}$, as in (10.28), and if the eddies satisfy a dispersion relation the components of the EP flux are equal to the group velocity multiplied by the wave activity density $\mathcal{A}$, as in (10.42). Thus, knowing the group velocity tells us a great deal about how momentum is transported by waves. We'll use the TEM to deduce the mean flow acceleration in sections 10.5, 10.6 and, in particular, in section 17.4.

### *Connection to potential vorticity and wave–mean-flow interaction*

If we take the curl of (10.61) — that is, cross differentiate its components — then, after using the residual mass continuity equation (10.60), we recover the zonally averaged potential vorticity equation namely

$$\frac{\partial \overline{q}}{\partial t} = -\frac{\partial}{\partial y} \overline{v'q'} - \frac{\partial \overline{F}}{\partial y} \tag{10.64a}$$

where

$$\overline{q}(y,t) = \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\overline{b}\right) - \frac{\partial \overline{u}}{\partial y}, \tag{10.64b}$$

which is essentially the same as (10.18) and (10.20), noting that we may add $\beta y$ to the definition of zonally-averaged potential vorticity with no effect.

The corresponding equation for the evolution of eddy potential vorticity is, in its inviscid form,

$$\left(\frac{\partial}{\partial t} + \overline{u}(y,t)\frac{\partial}{\partial x}\right) q' + v' \frac{\partial \overline{q}}{\partial y} = 0. \tag{10.65}$$

as in (10.7). Eqs. (10.64) and (10.65) are a closed set of quasi-linear equations, and we have recovered the wave–mean-flow system described in section 10.1.3.

## Quasi-Geostrophic Wave–mean flow Interaction

The inviscid and unforced Boussinesq quasi-geostrophic set of wave–mean-flow equations is

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = 0, \tag{WMF.1a}$$

$$\frac{\partial \overline{q}}{\partial t} + \frac{\partial}{\partial y}\overline{v'q'} = 0. \tag{WMF.1b}$$

along with similar equations as needed for buoyancy at the boundary (see main text). The eddy terms are

$$q' = \left[\nabla^2 + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\right)\right]\psi', \qquad (u', v') = \left(-\frac{\partial \psi'}{\partial y}, \frac{\partial \psi'}{\partial x}\right). \tag{WMF.2a,b}$$

The mean flow terms are

$$\overline{q}(y,t) = \beta y - \frac{\partial \overline{u}}{\partial y} + \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\overline{b}\right). \tag{WMF.3}$$

and

$$\frac{\partial \overline{q}}{\partial y} = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} - \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\frac{\partial b}{\partial y}\right) = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} - \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \overline{u}}{\partial z}\right), \tag{WMF.4}$$

using thermal wind. To solve for the mean-flow we may define a streamfunction $\Psi$ such that

$$\left(\overline{u}, \frac{1}{f_0}\overline{b}\right) = \left(-\frac{\partial \Psi}{\partial y}, \frac{\partial \Psi}{\partial z}\right) \tag{WMF.5}$$

whence

$$\overline{q}(y,t) - \beta y = \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \Psi}{\partial z}\right) + \frac{\partial^2 \Psi}{\partial y^2}. \tag{WMF.6}$$

Given $\overline{q}$ from (WMF.1b) we solve (WMF.6) to give $\overline{u}$ and $\overline{b}$. Equivalently, we may derive a single equation for the zonal wind by differentiating (WMF.1b) with respect to $y$ and, using (WMF.4), we obtain

$$\left[\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\right)\right]\frac{\partial \overline{u}}{\partial t} = \frac{\partial^2}{\partial y^2}\overline{v'q'}. \tag{WMF.7}$$

The evolution of the mean flow may also usefully be written in TEM form as

$$\frac{\partial \overline{u}}{\partial t} - f_0\overline{v}^* + \overline{v'q'} = 0, \tag{WMF.8a}$$

$$\frac{\partial \overline{b}}{\partial t} + N^2\overline{w}^* = 0, \tag{WMF.8b}$$

where $\overline{v}^*$ and $\overline{w}^*$ are found by solving the elliptic equation (10.63), and the value of $\partial\overline{q}/\partial y$ [for use in (WMF.1a)] is obtained using (WMF.4).

### 10.3.2 The TEM in isentropic coordinates

The residual circulation has an illuminating interpretation if we think of the fluid as comprising multiple layers of shallow water, or equivalently if we cast the problem in isentropic coordinates (section 3.9). Using the notation of a shallow water system, the momentum and mass conservation equation can then be written as

$$\frac{\partial u}{\partial t} + \boldsymbol{u} \cdot \nabla u - f v = F, \qquad \frac{\partial h}{\partial t} + \nabla \cdot (h\boldsymbol{u}) = S. \tag{10.66a,b}$$

The quantity $h$ is the thickness between two isentropic surfaces and $S$ is a thickness source term. (The field $h$ plays the same role as $\sigma$ in section 3.9.) With quasi-geostrophic scaling, so that variations in Coriolis parameter and layer thickness are small, zonally averaging in a conventional way gives

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v} = \overline{v'\zeta'} + \overline{F}, \qquad \frac{\partial \overline{h}}{\partial t} + H\frac{\partial \overline{v}}{\partial y} = -\frac{\partial}{\partial y}\overline{v'h'} + \overline{S}. \tag{10.67a,b}$$

The overbars in these equations denote averages taken along isentropes — i.e., they are averages for a given layer — but are otherwise conventional, and the meridional velocity is purely ageostrophic. By analogy to (10.59), we define the residual circulation by

$$\overline{v}^* \equiv \overline{v} + \frac{1}{H}\overline{v'h'}, \tag{10.68}$$

where $H$ is the mean thickness of the layer. Using (10.68) in (10.67) gives

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \overline{v'q'} + \overline{F}, \qquad \frac{\partial \overline{h}}{\partial t} + H\frac{\partial \overline{v}^*}{\partial y} = \overline{S}, \tag{10.69a,b}$$

where

$$\overline{v'q'} = \overline{v'\zeta'} - \frac{f_0}{H}\overline{v'h'}, \tag{10.70}$$

is the meridional potential vorticity flux in a shallow water system. From (10.68) we see that the residual velocity is a measure of the *total meridional thickness flux*, eddy plus mean, in an isentropic layer. This is often a more useful quantity than the Eulerian velocity $\overline{v}$ because it is generally the former, not the latter, that is constrained by the external forcing. What we have done, of course, is to effectively use a thickness-weighted mean in (10.66b); to see this, define the thickness-weighted mean by

$$\overline{v}_* \equiv \frac{\overline{hv}}{\overline{h}}. \tag{10.71}$$

(We use $\overline{v}_*$ to denote a thickness- or mass-weighted mean, and $\overline{v}^*$ to denote a residual velocity; the quantities are closely related, as we will see.) From (10.71) we have

$$\overline{v}_* = \overline{v} + \frac{1}{\overline{h}}\overline{v'h'}, \tag{10.72}$$

then the zonal average of (10.66b) is just

$$\frac{\partial \overline{h}}{\partial t} + \frac{\partial}{\partial y}(\overline{h}\overline{v}_*) = \overline{S}, \tag{10.73}$$

## Aspects of the TEM Formulation

*Properties and features*

⋆ The residual mean circulation is equivalent to the total mass-weighted (eddy plus Eulerian mean) circulation, and it is this circulation that is driven by the diabatic forcing.

⋆ There are no explicit eddy fluxes in the buoyancy budget; the only eddy term is the flux of potential vorticity, and this is divergence of the Eliassen–Palm flux; that is $\overline{v'q'} = \nabla_x \cdot \boldsymbol{\mathcal{F}}$.

⋆ The residual circulation, $\overline{v}^*$, becomes part of the solution, just as $\overline{v}$ is part of the solution in an Eulerian mean formulation.

*But note*

⋆ The TEM formulation does not solve the parameterization problem, and eddy fluxes are still present in the equations.

⋆ The theory and practice are well developed for a zonal average, but less so for three-dimensional, non-zonal flow. This is because the geometry enforces simple boundary conditions in the zonal mean case.[5]

⋆ The boundary conditions on the residual circulation are neither necessarily simple nor easily determined; for example, at a horizontal boundary $\overline{w}^*$ is not zero if there are horizontal buoyancy fluxes.

Examples of the use of the TEM and its relatives in the general circulation of the atmosphere and ocean arise in sections 15.2, 15.4, 17.4, 17.8 and 21.6.

which is the same as (10.69b) if we take $H = \overline{h}$. Similarly, if we use the thickness weighted velocity (10.72) in the momentum equation (10.67a) we obtain (10.69a).

Evidently, if the mass-weighted meridional velocity is used in the momentum and thickness equations then the eddy mass flux does not enter the equations explicitly: the only eddy flux in (10.69) is that of potential vorticity. That is, in isentropic coordinates the equations in TEM form are equivalent to the equations that arise from a particular form of averaging — thickness weighted averaging — rather than the conventional Eulerian averaging. A similar correspondence occurs in height coordinates, as we now see.

### 10.3.3 Connection between the residual and thickness-weighted circulation

It is evident from the above arguments that, in a shallow water system or in isentropic coordinates, the residual velocity is a measure of the total (i.e., mean plus eddy) thickness transport. In height coordinates, the definition of residual velocity, (10.58) does not lend itself so easily to such an interpretation. However, the residual velocity in height coordinates is, in fact, also a measure of the total thickness transport, or equivalently of the mass transport between two isentropic surfaces, as we now discover. Specifically, we show that averaging the total transport in isentropic layers is equivalent to the mass transport evaluated by the TEM formalism in

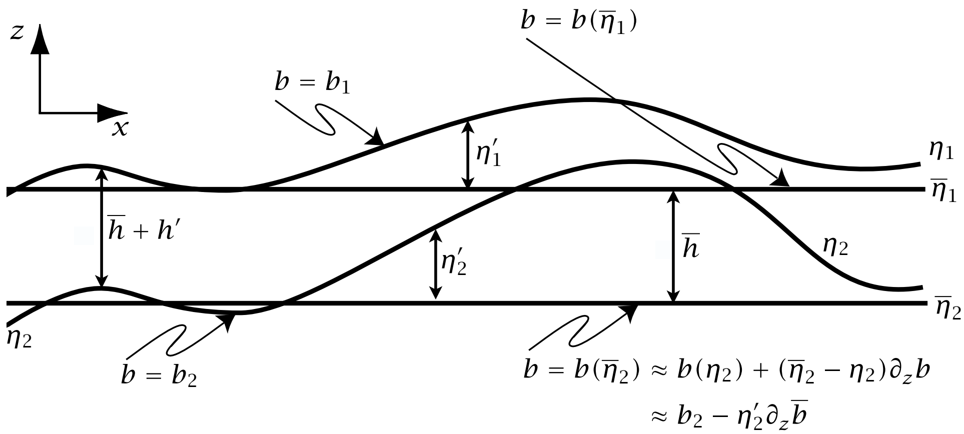


**Fig. 10.1** Two isentropic surfaces, $\eta_1$ and $\eta_2$, and their mean positions, $\overline{\eta}_1$ and $\overline{\eta}_2$. The departure of an isentrope from its mean position is proportional to the temperature perturbation at the mean position of the isentrope, and the variations in thickness ($h'$) of the isentropic layer are proportional to the vertical derivative of this.

height coordinates, and specifically that the thickness-weighted mean, $\overline{v}_*$ is equivalent to the residual velocity, $\overline{v}^*$ in height coordinates. Our demonstration is for a Boussinesq system, but the extension to a compressible gas is reasonably straightforward.[6]

Consider two isentropic surfaces, $\eta_1$ and $\eta_2$ with mean positions $\overline{\eta}_1$ and $\overline{\eta}_2$, as in Fig. 10.1. (We use $z$ to denote the vertical coordinate, and $\eta$ to denote the location of isentropic surfaces.) The meridional transport between these surfaces is given by

$$T = \int_{\eta_2}^{\eta_1} v \, \mathrm{d}z. \tag{10.74}$$

If the velocity does not vary with height within the layer (and in the limit of layer thickness going to zero this is the case) then $T = vh$ where $h = \eta_1 - \eta_2$ is the thickness of the isentropic layer. The zonally averaged transport is then given by

$$\overline{T} = \frac{1}{L}\int_L T \, \mathrm{d}x = \frac{1}{L}\int_L \left( \int_{\eta_2}^{\eta_1} v \, \mathrm{d}z \right) \mathrm{d}x = \overline{\int_{\eta_2}^{\eta_1} v \, \mathrm{d}z} = \overline{vh} = \overline{v}\overline{h} + \overline{v'h'}, \tag{10.75}$$

with obvious notation, and with an overbar denoting a zonal average. Letting the distance between isentropes shrink to zero this result allows us to write

$$\overline{v}_* \equiv \frac{\overline{v\sigma}^b}{\overline{\sigma}} = \overline{v}^b + \frac{\overline{v'\sigma'}^b}{\overline{\sigma}}, \tag{10.76}$$

where $\overline{(\cdot)}^b$ denotes an average along an isentrope and $\overline{\sigma} = \overline{\partial z/\partial b}$ is the thickness density, a measure of the thickness between two isentropes. Equation (10.76) is analogous to (10.72), for a continuously stratified system. The averaged quantity $\overline{v}_*$ is not proportional to the average of the velocity at constant height, or even to the average along an isentrope; rather, it is the *thickness-weighted* zonal average of the velocity *between* two isentropic surfaces, $\Delta b$ apart, of mean separation proportional to $\overline{\sigma}\Delta b$. Our goal is to express this transport in terms of Eulerian-averaged quantities, at a constant height $z$.

Let us first connect an average along an isentrope of some variable $\chi$ to its average at constant height by writing, for small isentropic displacements,

$$\overline{\chi}^b = \overline{\chi(z+\eta')}^z \approx \overline{\chi(z) + \eta' \partial\chi/\partial z}^z, \tag{10.77}$$

where the superscript explicitly denotes how the zonal average is taken, and $\eta'$ is the displacement of the isotherm from its mean position. This can be expressed in terms of the temperature perturbation at the location of the mean isentrope by Taylor expanding $b$ around its value on that mean isentrope. That is,

$$b(\eta) = b(\overline{\eta}) + \left(\frac{\partial b}{\partial z}\right)_{z=\overline{\eta}} (\eta - \overline{\eta}) + \cdots, \tag{10.78}$$

where $\overline{\eta} = \overline{\eta}(z)$, giving

$$\eta' \approx \frac{-b'}{\partial_z b(\overline{\eta})} \approx -\frac{b'}{\partial_z \overline{b}^z}, \tag{10.79}$$

where $\eta' = \eta - \overline{\eta}$ and $b' = b(\overline{\eta}) - b(\eta)$. Using (10.79) in (10.77) (and omitting the superscript $z$ on $\partial_z \overline{b}$) we obtain, with $\chi = v$,

$$\overline{v}^b = \overline{v}^z - \frac{\overline{b'\partial_z v'}^z}{\partial_z \overline{b}}. \tag{10.80}$$

Note that if $v$ is in thermal wind balance with $b$ then the second term vanishes identically, but we will not invoke this.

We now transform the second term on the right-hand side (10.76) to an average at constant $z$. The variations in thickness of an isothermal layer are given by

$$\sigma' \approx \overline{\sigma}\frac{\partial \eta'}{\partial z} = -\overline{\sigma}\frac{\partial}{\partial z}\left(\frac{b'}{\partial_z \overline{b}}\right), \tag{10.81}$$

using (10.79). Thus, neglecting terms that are third-order in amplitude,

$$\overline{v'\sigma'}^b = -\overline{\sigma}\,\overline{v'\frac{\partial}{\partial z}\left(\frac{b'}{\partial_z \overline{b}}\right)}^z. \tag{10.82}$$

Using both (10.80) and (10.82), (10.76) becomes

$$\overline{v}_* = \overline{v}^z - \frac{\overline{b'\partial_z v'}^z}{\partial_z \overline{b}} - \overline{v'\frac{\partial}{\partial z}\left(\frac{b'}{\partial_z \overline{b}}\right)}^z = \overline{v}^z - \overline{\frac{\partial}{\partial z}\left(\frac{v'b'}{\partial_z \overline{b}}\right)}^z. \tag{10.83}$$

The right-hand side of the last equation is the TEM form of the residual velocity; thus, we have shown that

$$\boxed{\overline{v}_* \equiv \frac{\overline{v\sigma}}{\overline{\sigma}} = \overline{v}^b + \frac{\overline{v'\sigma'}^b}{\overline{\sigma}} \approx \overline{v}^z - \frac{\partial}{\partial z}\left(\frac{\overline{v'b'}^z}{\partial_z \overline{b}}\right) \equiv \overline{v}^*}. \tag{10.84}$$

We see the equivalence of the thickness-weighted mean velocity on the left-hand side and the residual velocity on the right-hand side. In the quasi-geostrophic limit $N^2 = \partial_z \overline{b}$ and $\overline{\sigma}$ is a reference thickness.

## 10.4 ✦ THE TEM IN THE PRIMITIVE EQUATIONS

[This section has been removed for further editing]

## 10.5 THE NON-ACCELERATION RESULT

For the rest of this chapter we return to quasi-geostrophic dynamics, and consider further the interpretation and application of the potential vorticity flux and its relatives. We first consider an important result in wave–mean-flow dynamics, the non-acceleration result.[7] This result shows that under certain conditions, to be made explicit below, waves have no net effect on the zonally averaged flow, an important and somewhat counter-intuitive result.

### 10.5.1 A derivation from the potential vorticity equation

Consider how the potential vorticity fluxes affect the mean fields. The unforced and inviscid zonally averaged potential vorticity equation is

$$\frac{\partial \overline{q}}{\partial t} + \frac{\partial \overline{v'q'}}{\partial y} = \overline{F}_q. \tag{10.85}$$

Now, in quasi-geostrophic theory the geostrophically balanced velocity and buoyancy can be determined from the potential vorticity via an elliptic equation, and in particular

$$\overline{q} - \beta y = \frac{\partial^2 \overline{\psi}}{\partial y^2} + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \overline{\psi}}{\partial z}\right), \tag{10.86}$$

where $\overline{\psi}$ is such that $(\overline{u}, \overline{b}/f_0) = (-\partial\overline{\psi}/\partial y, \partial\overline{\psi}/\partial z)$. Differentiating (10.85) with respect to $y$ we obtain

$$\left[\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\right)\right]\frac{\partial \overline{u}}{\partial t} = (\nabla\cdot\boldsymbol{\mathcal{F}})_{yy}. \tag{10.87}$$

where $\nabla\cdot\boldsymbol{\mathcal{F}} = \overline{v'q'}$ is the divergence of the EP flux. This is determined using the wave activity equation which, reprising (10.29a), is

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla_2\cdot\boldsymbol{\mathcal{F}} = \mathcal{D}. \tag{10.88}$$

If the waves are statistically steady (i.e., $\partial\mathcal{A}/\partial t = 0$) and have no dissipation ($\mathcal{D} = 0$) then evidently $\nabla\cdot\boldsymbol{\mathcal{F}} = 0$. If there is no acceleration at the boundaries then the solution of (10.87) is

$$\frac{\partial \overline{u}}{\partial t} = 0. \tag{10.89}$$

This is a *non-acceleration result*. That is to say, under certain conditions the tendency of the mean fields, and in particular of the zonally-averaged zonal flow, are independent of the waves. To be explicit, those conditions are the following.

*(i)* The waves are steady (so that, using the wave activity equation $\mathcal{A}$ does not vary).

*(ii)* The waves are conservative [i.e., $\mathcal{D} = 0$ in (10.29a)]. Given this and item *(i)*, the Eliassen–Palm relation implies that $\nabla\cdot\boldsymbol{\mathcal{F}} = 0$; that is, the potential vorticity flux is zero.

*(iii)* The waves are of small amplitude (all of our analysis has neglected terms that are cubic in perturbation amplitude).
*(iv)* The waves do not affect the boundary conditions (so there are no boundary contributions to the acceleration).

The result applies to the buoyancy and velocity fields that are directly invertible from potential vorticity, and not to the ageostrophic velocities. Given the way we have derived it, it does not seem a surprising result; however, it can be powerful and counter-intuitive, for it means that steady waves (i.e., those whose amplitude does not vary) do not affect the zonal flow. However, they do affect the meridional overturning circulation, and the relative vorticity flux may also be non-zero. In fact, the non-acceleration theorem is telling us that the changes in the vorticity flux are exactly compensated for by changes in the meridional circulation, and there is no net effect on the zonally averaged zonal flow. It is *irreversibility*, often manifested by the breaking of waves, that leads to permanent changes in the mean flow.

The derivation of this result by way of the momentum equation, which one might expect to be more natural, is rather awkward because one must consider momentum and buoyancy fluxes separately. Furthermore, the zonally averaged meridional circulation comes into play: for example, meridional velocity, $\overline{v}$, is, although small because it is purely ageostrophic, not zero and we cannot neglect it because it is multiplied by the Coriolis parameter, which is large. Thus, the eddy vorticity fluxes can affect both the meridional circulation and the acceleration of the zonal mean flow, and it might seem impossible to disentangle the two effects without completely solving the equations of motion. However, we can proceed by way of the momentum and buoyancy equations if we use the transformed Eulerian mean and this provides a useful alternate derivation, as follows.

### 10.5.2 Using TEM to give the non-acceleration result

We may use the TEM formalism to obtain the non-acceleration result. The explanation is largely equivalent to that given above, but the explication may be useful.

#### *A two-dimensional case*

Consider two-dimensional incompressible flow on the $\beta$-plane, for which there is no buoyancy flux. The linearized vorticity equation is

$$\frac{\partial \zeta'}{\partial t} + \overline{u}\frac{\partial \zeta'}{\partial x} + v'\frac{\partial \overline{\zeta}}{\partial y} = D', \tag{10.90}$$

from which we derive, analogously to (10.29a), the Eliassen–Palm relation

$$\frac{\partial \mathcal{A}}{\partial t} + \frac{\partial \mathcal{F}}{\partial y} = \mathcal{D}, \tag{10.91}$$

where $\mathcal{F} = -\overline{u'v'}$, $\mathcal{D}$ represents non-conservative forces, and

$$\mathcal{A} = \frac{\overline{\zeta'^2}}{2\partial_y\overline{\zeta}} = \frac{1}{2}\overline{\eta'^2}\frac{\partial \overline{\zeta}}{\partial y}. \tag{10.92}$$

The quantity $\eta' \equiv -\zeta'/\partial_y\overline{\zeta}$ is proportional to the meridional particle displacement in a disturbance. Now consider the $x$-momentum equation

$$\frac{\partial u}{\partial t} = -\frac{\partial u^2}{\partial x} - \frac{\partial uv}{\partial y} - \frac{\partial \phi}{\partial x} + fv. \tag{10.93}$$

Zonally averaging, noting that $\overline{v} = 0$, gives

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial \overline{uv}}{\partial y} = \overline{v'\zeta'} = \frac{\partial \mathcal{F}}{\partial y}. \tag{10.94}$$

Finally, combining (10.91) and (10.94) gives

$$\frac{\partial}{\partial t}(\overline{u} + \mathcal{A}) = \mathcal{D}. \tag{10.95}$$

In the absence of non-conservative terms (i.e., if $\mathcal{D} = 0$) the quantity $\overline{u}+\mathcal{A}$ is constant.[8] Further, if the waves are steady and conservative then $\mathcal{A}$ is constant and, therefore, so is $\overline{u}$.

*The stratified case*

In the stratified case we can use the TEM form of the momentum equation to derive a similar result. The unforced zonally averaged zonal momentum equation can be written as

$$\frac{\partial \overline{u}}{\partial t} - f_0\overline{v}^* = \nabla \cdot \boldsymbol{\mathcal{F}}, \tag{10.96}$$

and using the Eliassen–Palm relation this may be written as

$$\frac{\partial}{\partial t}(\overline{u} + \mathcal{A}) - f_0\overline{v}^* = \mathcal{D}, \tag{10.97}$$

and so again $\mathcal{A}$ is related to the momentum of the flow. If, furthermore, the waves are steady ($\partial\mathcal{A}/\partial t = 0$) and conservative ($\mathcal{D} = 0$), then $\partial\overline{u}/\partial t - f_0\overline{v}^* = 0$. However, under these same conditions the residual circulation will also be zero. This is because the residual meridional circulation $(\overline{v}^*, \overline{w}^*)$ arises via the necessity to keep the temperature and velocity fields in thermal wind balance, and is thus determined by an elliptic equation, namely (10.63). If the waves are steady and adiabatic then, since $\overline{v'q'} = 0$, the right-hand side of the equation is zero and it becomes

$$f_0^2\frac{\partial^2\psi^*}{\partial z^2} + N^2\frac{\partial^2\psi^*}{\partial y^2} = 0. \tag{10.98}$$

If $\psi^* = 0$ at the boundaries, then the unique solution of this is $\psi^* = 0$ everywhere. At the meridional boundaries we may certainly suppose that $\psi^*$ vanishes if these are quiescent latitudes, and at the horizontal boundaries the buoyancy flux will vanish if the waves there are steady, because from (10.14) we have

$$\overline{v'b'}\frac{\partial\overline{b}}{\partial y} = -\frac{1}{2}\frac{\partial}{\partial t}\overline{b'^2} = 0. \tag{10.99}$$

Under these circumstances, then, the residual meridional circulation vanishes in the interior and, from (10.96), the mean flow is steady, thus reprising the non-acceleration result.

Compare (10.96) with the momentum equation in conventional Eulerian form, namely

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v} = \overline{v'\zeta'}. \tag{10.100}$$

There is no reason that the vorticity flux should vanish when waves are present, even if they are steady. However, such a flux is (under non-acceleration conditions) precisely compensated by the meridional circulation $f_0\overline{v}$, something that is hard to infer or intuit directly from (10.100); even when non-acceleration conditions do not apply there will be a significant cancellation between the Coriolis and eddy terms. The difficulty boils down to the fact that, in contrast to $\overline{v'q'}$, $\overline{v'\zeta'}$ is not the flux of a wave activity.

Unlike the proof of the non-acceleration result given in section 10.5.1, the above argument does not use the invertibility property of potential vorticity directly, suggesting an extension to the primitive equations, but we do not pursue that here.[9] Various results regarding the TEM and non-acceleration are summarized in the shaded box on the following page.

### 10.5.3 The EP flux and form drag

It may seem a little magical that the zonal flow is driven by the Eliassen–Palm flux via (10.96). The poleward vorticity flux is clearly related to the momentum flux convergence, but why should a poleward buoyancy flux affect the momentum? The TEM form of the momentum equation may be written as

$$\frac{\partial \overline{u}}{\partial t} = \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\overline{v'b'}\right) + F_m, \tag{10.101}$$

where $F_m = \overline{v'\zeta'} + f_0\overline{v}^*$ represents forces from the momentum flux and Coriolis force. The first term on the right-hand side certainly does not look like a force; however, it turns out to be directly proportional to the *form drag* between isentropic layers. Recall from section 3.5 that the form drag, $\tau_d$, at an interface between two layers of shallow water is

$$\tau_d = -\overline{\eta'\frac{\partial p'}{\partial x}}, \tag{10.102}$$

where $\eta$ is the interfacial displacement. But from (10.79) $\eta' = -b'/N^2$ and with this and geostrophic balance we have

$$\tau_d = \frac{\rho_0 f_0}{N^2}\overline{v'b'}. \tag{10.103}$$

Thus, the vertical component of the EP flux (i.e., the meridional buoyancy flux) is in fact a real stress acting on a fluid layer and equal to the momentum flux caused by the wavy interface. The net momentum convergence into an infinitesimal layer of mean thickness $\overline{h}$ is then [cf. (3.63)],

$$F_d = \overline{h}\frac{\partial \tau_d}{\partial z} = \overline{h}\rho_0 f_0 \frac{\partial}{\partial z}\left(\frac{\overline{v'b'}}{N^2}\right), \tag{10.104}$$

and a layer of mean thickness $\overline{h}$ is accelerated according to

$$\frac{\partial \overline{u}}{\partial t} = f_0 \frac{\partial}{\partial z}\left(\frac{\overline{v'b'}}{\partial_z \overline{b}}\right) + F_m. \tag{10.105}$$

## TEM, Residual Velocities, Non-acceleration, and All That

For a Boussinesq quasi-geostrophic system, the TEM form of the unforced momentum equation and the thermodynamic equation are:

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \nabla \cdot \boldsymbol{\mathcal{F}}, \qquad \frac{\partial \overline{b}}{\partial t} + \overline{w}^* \frac{\partial}{\partial z}\overline{b}_0 = \overline{S}, \tag{T.1}$$

where $\partial \overline{b}_0/\partial z = N^2$, $\overline{S}$ represents diabatic effects, $\boldsymbol{\mathcal{F}}$ is the Eliassen–Palm (EP) flux and its divergence is the potential vorticity flux; that is $\nabla_x \cdot \boldsymbol{\mathcal{F}} = \overline{v'q'}$. The residual velocities are

$$\overline{v}^* = \overline{v} - \frac{\partial}{\partial z}\left(\frac{1}{N^2}\overline{v'b'}\right), \qquad \overline{w}^* = \overline{w} + \frac{\partial}{\partial y}\left(\frac{1}{N^2}\overline{v'b'}\right). \tag{T.2}$$

Spherical coordinate and ideal gas versions of these take a similar form. We may define a meridional overturning streamfunction such that $(\overline{v}^*, \overline{w}^*) = \left(-\partial\psi^*/\partial z, \partial\psi^*/\partial y\right)$, and using thermal wind to eliminate time-derivatives in (T.1) we obtain

$$f_0^2 \frac{\partial^2 \psi^*}{\partial z^2} + N^2 \frac{\partial^2 \psi^*}{\partial y^2} = f_0 \frac{\partial}{\partial z}\overline{v'q'} + \frac{\partial \overline{S}}{\partial y}. \tag{T.3}$$

The above manipulations may seem formal, in that they simply transform the momentum and thermodynamic equation from one form to another. However, the resulting equations have two potential advantages over the untransfomed ones.

*(i)* The residual meridional velocity is approximately equal to the average thickness-weighted velocity between two neighbouring isentropic surfaces, and so is a measure of the total (Eulerian mean plus eddy) meridional transport of thickness or buoyancy.

*(ii)* The EP flux is directly related to certain conservation properties of waves. The divergence of the EP flux is the meridional flux of potential vorticity:

$$\boldsymbol{\mathcal{F}} = -(\overline{u'v'})\,\mathbf{j} + \left(\frac{f_0}{N^2}\overline{v'b'}\right)\mathbf{k}, \qquad \nabla \cdot \boldsymbol{\mathcal{F}} = \overline{v'q'}. \tag{T.4}$$

Furthermore, the EP flux satisfies, to second order in wave amplitude,

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \boldsymbol{\mathcal{F}} = \mathcal{D}, \qquad \text{where} \quad \mathcal{A} = \frac{\overline{q'^2}}{2\partial\overline{q}/\partial y}, \quad \mathcal{D} = \frac{\overline{D'q'}}{\partial\overline{q}/\partial y}. \tag{T.5}$$

The quantity $\mathcal{A}$ is a *wave activity density,* and $\mathcal{D}$ is its dissipation. For nearly plane waves, $\mathcal{A}$ and $\boldsymbol{\mathcal{F}}$ are connected by the *group velocity property,*

$$\boldsymbol{\mathcal{F}} = (\mathcal{F}^y, \mathcal{F}^z) = \boldsymbol{c}_g \mathcal{A}, \tag{T.6}$$

where $\boldsymbol{c}_g$ is the group velocity of the waves. If the waves are steady ($\partial\mathcal{A}/\partial t = 0$) and dissipationless ($\mathcal{D} = 0$) then $\nabla \cdot \boldsymbol{\mathcal{F}} = 0$ and using (T.1) and (T.3) there is no wave-induced acceleration of the mean flow; this is the 'non-acceleration' result. Commonly there is enstrophy dissipation, or wave-breaking, and $\nabla \cdot \boldsymbol{\mathcal{F}} < 0$; such *wave drag* leads to flow deceleration and/or a poleward residual meridional velocity.

The appearance of the buoyancy flux is really a consequence of the way we have chosen to average the equations: obtaining (10.105) involved averaging the forces over an isentropic layer, and given this it can only be the residual circulation that contributes to the Coriolis force. One might say that the vertical component of the EP flux is a force in drag, masquerading as a buoyancy flux.

## 10.6 ✦ INFLUENCE OF EDDIES ON THE MEAN FLOW IN THE EADY PROBLEM

We now consider the eddy fluxes in the Eady problem, and, in particular, how these might feed back on to the mean flow. Because of the simplicity of the setting the problem can be fully solved in both the Eulerian or residual frameworks and it is therefore a very instructive, albeit somewhat algebraically complex, example.[10]

### 10.6.1 Formulation

Let us first distinguish between the basic flow, the zonal mean fields, and the perturbation. The basic flow is the flow around which the equations of motion are linearized; this flow is unstable, and the perturbations, assumed to be small, grow exponentially with time. Because the perturbations are formally always small they do not affect the basic flow, but they do produce changes in the zonal mean velocity and buoyancy fields. In Eulerian form this is represented by,

$$\frac{\partial \overline{u}}{\partial t} = f_0 \overline{v} - \frac{\partial \overline{u'v'}}{\partial y}, \qquad \frac{\partial \overline{b}}{\partial t} = -N^2 \overline{w} - \frac{\partial \overline{b'v'}}{\partial y}, \tag{10.106}$$

and the TEM version of these equation is

$$\frac{\partial \overline{u}}{\partial t} = f_0 \overline{v}^* + \overline{v'q'}, \qquad \frac{\partial \overline{b}}{\partial t} = -N^2 \overline{w}^*, \tag{10.107}$$

where in the Eady problem $\partial_y(\overline{u'v'})$ and $\overline{v'q'}$ are both zero. We can calculate the perturbation quantities from the solution to the Eady problem (e.g., calculate $\overline{v'b'}$) and thus infer the structure of the mean flow tendencies $\partial\overline{u}/\partial t$ and $\partial\overline{b}/\partial t$ and the meridional circulation, $(\overline{v}, \overline{w})$ or $(\overline{v}^*, \overline{w}^*)$. All of these fields are perturbation quantities and all are exponentially growing, and so in reality they will eventually have a finite effect on the pre-existing zonal flow, but in the Eady problem, or any similar linear problem, such rectification is assumed to be small and is neglected.

Using the thermal wind relation, $f_0 \partial_z \overline{u} = -\partial_y \overline{b}$ to eliminate time derivatives in (10.106) gives an equation for the meridional streamfunction $\psi_E$, namely,

$$\frac{L^2}{L_d^2} \frac{\partial^2 \psi_E}{\partial z^2} + \frac{\partial^2 \psi_E}{\partial y^2} = -\frac{1}{N^2} \frac{\partial^2 \overline{b'v'}}{\partial y^2}, \tag{10.108}$$

where $(\overline{v}, \overline{w}) = (-\partial\psi_E/\partial z, \partial\psi_E/\partial y)$ and we have non-dimensionalized $z$ with $D$ and $y$ with $L$. The boundary conditions are that $\psi_E = 0$ at $y = 0, L$ and $z = 0, D$. Similarly, and analogously to to (10.63), we obtain an equation for the residual streamfunction, $\psi^*$, namely

$$\frac{L^2}{L_d^2} \frac{\partial^2 \psi^*}{\partial z^2} + \frac{\partial^2 \psi^*}{\partial y^2} = 0, \tag{10.109}$$

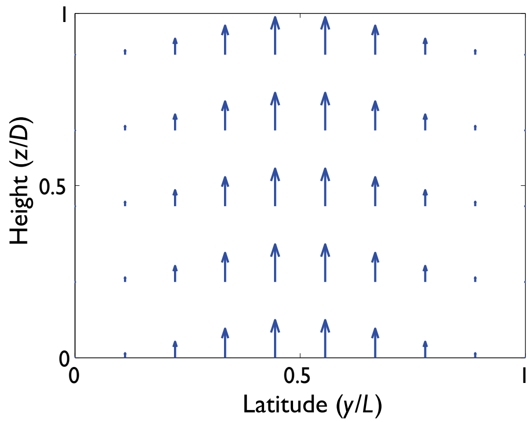


**Fig. 10.2** The Eliassen–Palm vector in the Eady problem.

where now the boundary conditions are that $N^2\overline{w}^* = \partial\overline{v'b'}/\partial y$ at the upper and lower boundaries, and $\overline{v} = 0$ at the lateral boundaries. In terms of the residual streamfunction this is

$$\psi^* = \frac{1}{N^2}\overline{v'b'}, \text{ at } z = 0,\ 1, \qquad \psi^* = 0, \text{ at } y = 0,\ 1. \tag{10.110}$$

The residual and overturning circulations are related by (10.58a), and (10.108) and (10.109) are, at one level, simply different representations of the same problem, connected by a simple mathematical transformation. However, the residual streamfunction better represents the total transport of the fluid. Equation (10.109) is particularly simple, because of the absence of potential vorticity fluxes in the interior, and it is apparent that the residual circulation is driven by boundary sources. We care only about the spatial structure of the right-hand sides of (10.108) and of the boundary conditions of (10.110). The former is given by

$$-\frac{\partial^2\overline{b'v'}}{\partial y^2} \propto -\frac{\partial^2}{\partial y^2}\sin^2 ly = -2l^2\cos 2ly. \tag{10.111}$$

The eddy heat fluxes in the Eady problem are independent of height, as may be calculated explicitly from the solutions of chapter 9. In fact, the result follows without detailed calculation, by first noting that the eddy potential vorticity flux is zero because the basic state has zero QG potential vorticity and therefore none may be generated. Further, because the basic state does not vary in $y$ there can be no momentum flux convergence in the $y$-direction, and so the momentum flux itself is zero if it is zero on the boundary. Thus [using for example (10.27) and (10.28)] the eddy heat flux is independent of height and the EP vectors are directed purely vertically (Fig. 10.2).

The boundary conditions for the residual circulation are

$$\psi^*(y, 0) = \psi^*(y, 1) \propto \sin^2 ly. \tag{10.112}$$

### 10.6.2 Solution

The solutions to (10.108) and (10.109) may be obtained either analytically or numerically. In a domain $0 < y < 1$ and $0 < z < 1$ the residual streamfunction for $l = \pi$ is given by:

$$\begin{aligned} \psi^* &= \sum_{n=1}^{\infty} A_n \sin[(2n-1)ly] \frac{\cosh[L_d\pi(2n-1)(z-0.5)/L]}{\cosh[L_d\pi(2n-1)/2L]}, \\ A_n &= \frac{2}{\pi(2n-1)} - \frac{1}{\pi(2n-1)-2l} - \frac{1}{\pi(2n-1)+2l}. \end{aligned} \tag{10.113}$$

The solution is obtained by first projecting the boundary conditions [proportional to $\sin^2 ly$, or $(1-\cos 2ly)/2$] on to the eigenfunctions of the horizontal part of the Laplacian (i.e., sine functions), and this gives the coefficients of $A_n$. The vertical structure is then obtained by solving $(L/L_d)^2\partial_z^2\psi^* = -\partial_y^2\psi^*$, which gives the cosh functions. The series converges very quickly, and the first term in the series captures the dominant structure of the solution, essentially because, for $l = \pi$, $\sin ly$ is not unlike $\sin^2 ly$ on the interval $[0, 1]$.

The Eulerian circulation is obtained from the residual circulation using (10.58a) and so by the addition of a field independent of $z$ and proportional to $\sin^2 ly$. The resulting structure is dominated by this and the first term of (10.113) (proportional to $\sin ly$) and, noting that the circulation is symmetric about $z = 0.5$, we obtain a circulation dominated by a single cell, with equatorward motion aloft and poleward motion near the surface (Fig. 10.3). The heat flux convergence in high latitudes is leading to mean rising motion, with the precise shape of the streamfunction determined by the boundary conditions. Although this is true, the heat flux arises *because* of the motion of fluid parcels, so it may be a little misleading to infer, as one might from the Eulerian streamfunction, that the heat flux *causes* the individual parcels to rise or sink in this fashion. The residual streamfunction is a better indicator of the total mass transport and, perhaps as one might intuitively expect, these show parcels rising in the low latitudes and sinking in high latitudes, providing a tendency to flatten the isopycnals and to reduce the meridional temperature gradient.

The residual circulation also shows fluid entering or leaving the domain at the boundary — what does this represent? Suppose that instead of solving the continuous problem we had posed the problem in a finite number of layers (and we explicitly consider the two-layer problem below). As the number of layers increases the solutions to the linear baroclinic instability problem approaches that of the Eady problem (e.g., Fig. 9.13); however, as we saw in section 10.3 the residual circulation is closed in the layered model, and the sum over all the layers of the meridional transport vanishes. Now, in the layered model the vertical boundary conditions are built in to the representation by way of a redefinition of the potential vorticity of the top and bottom layers, so that, in the layered version of the Eady problem there appears to be a potential vorticity gradient in these two layers, instead of a buoyancy gradient at the boundary. The residual circulation is then closed by a return flow that occurs only in the top and bottom layers, and as the number of layers increases this flow is confined to a thinner and thinner layer, and to a delta-function in the continuous limit. To indicate this we have placed arrows just above and below the domain in Fig. 10.3. (This equivalence between boundary conditions and delta-function sources is the same as that giving rise to the delta-function boundary layer of section 5.4.3.)

The effect on the mean flow is inferred directly from the residual circulation: the mean flow acceleration is proportional to $\overline{v}^*$ and the buoyancy tendency is proportional to $-\overline{w}^*$, and

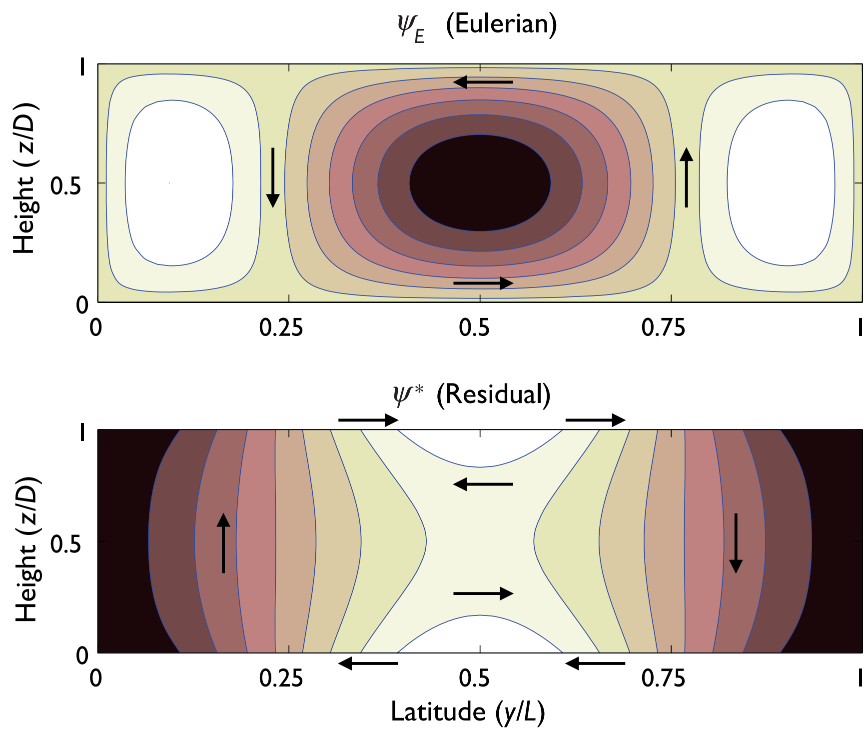


**Fig. 10.3** The Eulerian streamfunction (top) and the residual streamfunction for the Eady problem, calculated using (10.108) and (10.109), with $L^2/L_d^2 = 9$.

these are plotted in Figs. 10.4 and 10.5. Because there is no momentum flux convergence in the problem the zonal flow tendency is entirely baroclinic — its vertical integral is zero — and over most of the domain is such as to reduce the mean shear. Consistently (using thermal wind) the buoyancy tendency is such as to reduce the meridional temperature gradient; that is, the instabilities act to transport heat polewards and so reduce the instability of the mean flow.

### 10.6.3 The two-level problem

The residual circulation and mean-flow tendencies can also be calculated for the two-level (Phillips) problem, with the $\beta$-effect. The potential vorticity fluxes in each layer are non-zero and the mean flow equations are, for $i = 1, 2$,

$$\frac{\partial \overline{u}_i}{\partial t} = f_0 \overline{v}_i^* + \overline{v_i' q_i'}, \qquad \frac{\partial \overline{b}}{\partial t} = -N^2 \overline{w}^*. \tag{10.114}$$

The vertical velocity and buoyancy are evaluated at mid-depth, and the thermal wind equation is $\overline{u}_1 - \overline{u}_2 = -(D/2)\partial_y \overline{b}$ and, by mass conservation, $\overline{v}_1^* = -\overline{v}_2^*$. If we define a residual streamfunction $\psi^*$ such that

$$\overline{v}_1^* = -\overline{v}_2^* = \psi^*, \qquad \overline{w}^* = \frac{\partial \psi^*}{\partial y}, \tag{10.115}$$

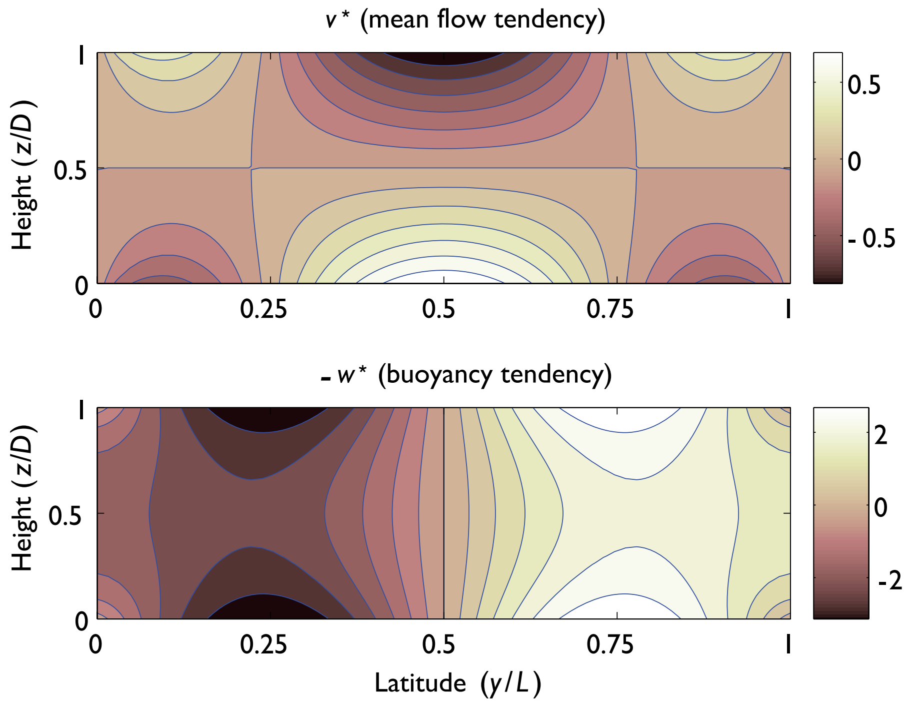


**Fig. 10.4** The tendency of the zonal mean flow ($\partial\overline{u}/\partial t$) and the buoyancy ($\partial\overline{b}/\partial t$) for the Eady problem. Lighter (darker) shading means a positive (negative) tendency, but the units themselves are arbitrary.

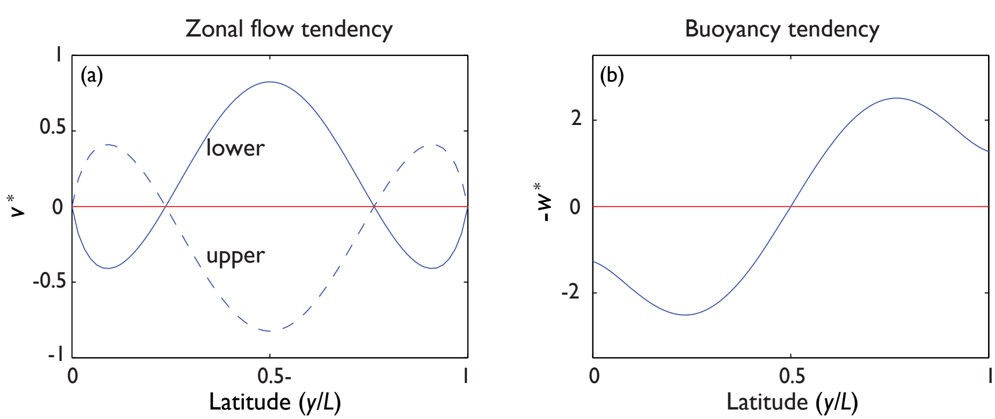


**Fig. 10.5** (a) The tendency of the zonal mean flow ($\partial\overline{u}/\partial t$) just below the upper lid (dashed) and just above the surface (solid) in the Eady problem. The vertically integrated tendency is zero. (b) The vertically averaged buoyancy tendency.

then eliminating time derivatives in (10.114) gives an equation for the residual streamfunction,

$$\frac{\partial^2 \psi^*}{\partial y^2} - \frac{k_d^2}{2}\psi^* = \frac{2 f_0 L^2}{N^2 D}(\overline{v_1' q_1'} - \overline{v_2' q_2'}), \tag{10.116}$$

where $k_d^2/2 = [2f_0/(NH)]^2$ and $D$ is the total depth of the fluid, and we have non-dimensionalized vertical scales by $D$ and horizontal scales by $L$. As in the Eady problem it is only the spatial structure of the terms on the right-hand side that are relevant, and these may be calculated from the solutions to the two-level instability problem. The main difference from the Eady problem is that the potential vorticity fluxes are non-zero, even in the case with $\beta = 0$: effectively, the boundary fluxes of the Eady problem are absorbed into the potential vorticity fluxes of the two layers. Solving for the residual circulation and interpreting the mean-flow tendencies is left as an exercise for the reader (problem 10.8).

## 10.7 ✦ NECESSARY CONDITIONS FOR INSTABILITY

*Let's take a taxi to the finish line.*
Chris Garrett, Ocean Science Meeting, Hawaii 2002.

As we noted in chapter 9, necessary conditions for instability, or sufficient conditions for stability, can be very useful because when satisfied they obviate the need to perform a detailed calculation. In the remainder of this chapter we use the conservation of wave activities — pseudomomentum and pseudoenergy — to derive such conditions. In sections 9.3 and 9.4.3 we derived such conditions assuming the instability to be of normal-mode form. Here we give derivations that are both more general and, in some ways, simpler; they utilize the fact that the potential vorticity flux may be written as a divergence of a vector and therefore vanishes when integrated over a domain, aside from possible boundary contributions.

### 10.7.1 Stability conditions from pseudomomentum conservation

Consider the perturbation enstrophy equation,

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{q'^2} = -\frac{\partial \overline{q}}{\partial y}\nabla_x \cdot \boldsymbol{\mathcal{F}}, \tag{10.117}$$

where $\boldsymbol{\mathcal{F}}$ is the Eliassen–Palm flux given by (10.27), the overbar is a zonal mean and the divergence is in $y$-$z$ plane. Dividing by $\partial\overline{q}/\partial y$ and integrating over a domain $A$ which is such that the Eliassen–Palm flux vanishes at the boundaries gives the pseudomomentum conservation law,

$$\int_A \frac{\partial}{\partial t}\left(\frac{\overline{q'^2}}{\partial_y \overline{q}}\right) \mathrm{d}y\,\mathrm{d}z = 0. \tag{10.118}$$

Equation (10.118) implies that, in the *norm* $\left[q'^2/\partial_y\overline{q}\right]$, the perturbation cannot grow unless $\partial\overline{q}/\partial y$ changes sign somewhere in the domain, or at the boundaries. This result does not depend upon the instability being of normal-mode form. The simplest result of all occurs in a barotropic problem with no vertical variation. Then $\partial\overline{q}/\partial y = \partial/\partial\overline{\zeta}_a y = \beta - \partial^2\overline{u}/\partial y^2$, and demanding that this must change sign for an instability reprises the inflection point (Rayleigh–Kuo) condition. In the more general case, if $\partial\overline{q}/\partial y$ changes sign along a vertical line then the instability is called a baroclinic instability, and if it changes sign along a horizontal line the instability is barotropic — these may be taken as the definitions of those terms. A mixed instability has a change of sign along both horizontal and vertical lines.

### 10.7.2 Inclusion of boundary terms

Suppose now the flow is contained between two flat boundaries, at $z = 0$ and $z = H$. The relevant equations of motion are the potential vorticity evolution in the interior, supplemented by the thermodynamic equation at the boundary. For unforced and inviscid flow these give [cf. (10.11) and (10.14)]

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\frac{\overline{q'^2}}{\partial_y \overline{q}}\right) = -\overline{v'q'}, \qquad 0 < z < H, \tag{10.119}$$

and

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\frac{\overline{b'^2}}{\partial_y \overline{b}}\right) = -\overline{v'b'}, \qquad z = 0, H. \tag{10.120}$$

The poleward flux of potential vorticity is

$$\overline{v'q'} = -\frac{\partial}{\partial y}\overline{u'v'} + \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\overline{v'b'}\right), \tag{10.121}$$

and integrating this expression with respect to both $y$ and $z$ gives

$$\int_A \overline{v'q'}\,\mathrm{d}y\,\mathrm{d}z = \left[\frac{f_0}{N^2}\overline{v'b'}\right]_0^H, \tag{10.122}$$

assuming that the meridional boundaries are at quiescent latitudes. Integrating (10.119) over $y$ and $z$, and using (10.122) gives

$$\frac{\partial}{\partial t}\iint \frac{1}{2}\frac{\overline{q'^2}}{\partial_y \overline{q}}\,\mathrm{d}y\,\mathrm{d}z = -\left[\frac{f_0}{N^2}\overline{v'b'}\right]_0^H. \tag{10.123}$$

Using (10.120) to eliminate $\overline{v'b'}$ finally gives

$$\frac{\partial}{\partial t}\left\{\iint \frac{1}{2}\frac{\overline{q'^2}}{\partial_y \overline{q}}\,\mathrm{d}y\,\mathrm{d}z - \int\left[\frac{1}{2}\frac{f_0}{N^2}\frac{\overline{b'^2}}{\partial_y \overline{b}}\right]_0^H \mathrm{d}y\right\} = 0. \tag{10.124}$$

If this expression is positive or negative definite the perturbation cannot grow and therefore the basic state is stable. Stability thus depends on the meridional gradient of potential vorticity in the interior, and the meridional gradient of buoyancy at the boundary. If $\partial\overline{q}/\partial y$ changes sign in the interior, or $\partial\overline{b}/\partial y$ changes sign at the boundary, we have the potential for instability. If these are both one signed, then various possibilities exist, and using the thermal wind relation ($f_0\partial\overline{u}/\partial z = -\partial\overline{b}/\partial y$) we obtain the following.

I. *A stable case:*

$$\frac{\partial\overline{q}}{\partial y} > 0 \;\text{ and }\; \left.\frac{\partial u}{\partial z}\right|_{z=0} < 0 \;\text{ and }\; \left.\frac{\partial u}{\partial z}\right|_{z=H} > 0 \implies \text{stability.} \tag{10.125}$$

Stability also ensues if all inequalities are switched.

II. *Instability via interior-surface interactions:*

$$\frac{\partial \overline{q}}{\partial y} > 0 \text{ and } \left.\frac{\partial u}{\partial z}\right|_{z=0} > 0 \text{ or } \left.\frac{\partial u}{\partial z}\right|_{z=H} < 0 \implies \text{potential instability.} \tag{10.126}$$

The condition $\partial q/\partial y > 0$ and $(\partial u/\partial z)_{z=0} > 0$ is the most common criterion for instability that is met in the atmosphere. In the troposphere we can sometimes ignore contributions of the buoyancy fluxes at the tropopause ($z = H$), and stability is then determined by the interior potential vorticity gradient and the surface buoyancy gradient. Similarly, in the ocean contributions from the ocean floor are normally very small.

III. *Instability via edge wave interaction:*

$$\left.\frac{\partial u}{\partial z}\right|_{z=0} > 0 \text{ and } \left.\frac{\partial u}{\partial z}\right|_{z=H} > 0 \implies \text{potential instability.} \tag{10.127}$$

(And similarly, with both inequalities switched.) Such an instability may occur where the troposphere acts like a lid, as for example in the Eady problem. If $\partial\overline{q}/\partial y = 0$ and there is no lid at $z = H$ (e.g., the Eady problem with no lid) then the instability disappears.

One consequence of the upper boundary condition is that it provides a condition on the depth of the disturbance. In the Eady problem the evolution of the system is determined by temperature evolution at the surface,

$$\frac{\mathrm{D}b}{\mathrm{D}t} = 0 \qquad \text{at} \quad z = 0,\ H, \tag{10.128}$$

(where $b = f_0\partial\psi/\partial z$) and zero potential vorticity in the interior, which implies that

$$\nabla^2\psi + k_d^2 H^2\frac{\partial^2\psi}{\partial z^2} = 0 \qquad 0 < z < H, \tag{10.129}$$

where $k_d = f_0/(HN)$. Assuming a solution of the form $b \sim \sin kx$ then the Poisson equation (10.129) becomes

$$H^2 k_d^2\frac{\partial^2\psi}{\partial z^2} = k^2\psi, \tag{10.130}$$

with solutions $\psi = A\exp(-\alpha z) + B\exp(\alpha z)$, where $\alpha^2 = k^2N^2/f_0^2$. The scale height of the disturbance is thus

$$h \sim \frac{f_0 L}{2\pi N}. \tag{10.131}$$

where $L \sim 2\pi/k$ is the horizontal scale of the disturbance. If the upper boundary is higher than this, it cannot interact strongly with the surface, because the disturbances at either boundary decay before reaching the other. Put another way, if the structure of the disturbance is such that it is shallower than $H$, the presence of the upper boundary is not felt. In the Eady problem, we know that the upper boundary must be important, because it is only by its presence that the flow can be unstable. Thus, all unstable modes in the Eady problem must be 'deep' in this sense, which can be verified by direct calculation. This condition gives rise to a physical interpretation of the high-wavenumber cut-off: if $L$ is too small, the modes are too shallow to span the full depth of the fluid, and from (10.131) the condition for stability is thus

$$L < L_c = 2\pi \frac{NH}{f_0} \qquad \text{or} \qquad K > K_c = \frac{f_0}{NH} = L_d^{-1}. \tag{10.132}$$

where $L_c$ and $K_c$ are the critical length scales and wavenumbers. Wavenumbers larger than the reciprocal of the deformation radius are stable in the Eady problem. If $\beta$ is non-zero, this condition does not apply, because the necessary condition for instability can be satisfied by a combination of a surface temperature gradient and an interior gradient of potential vorticity provided by $\beta$, as in condition (II) in section 10.7.2. Thus, we may expect that, if $\beta \neq 0$, higher wavenumbers ($k > k_d$) may be unstable but if so they will be shallow, and this may be confirmed by explicit calculation (see Figs. 9.12 and 9.19). In the two-level model shallow modes are, by construction, not allowed so that high wavenumbers will be stable, with or without beta.

## 10.8 ✦ NECESSARY CONDITIONS FOR INSTABILITY: USE OF PSEUDOENERGY

In this section we derive another necessary condition for instability, sometimes called an 'Arnold condition', that is based on the conservation properties of energy and enstrophy. Such conditions can be derived more generally by variational methods, and these lead to somewhat stronger results (in particular, nonlinear results that do not require the perturbation to be small) but our derivations will be elementary and direct.[11]

### 10.8.1 Two-dimensional flow

First consider inviscid, incompressible two-dimensional flow governed by the equation of motion

$$\frac{\partial q}{\partial t} + J(\psi, q) = 0, \tag{10.133}$$

where $q = \zeta + f = \nabla^2\psi + f$ is the absolute vorticity and $\psi$ is the streamfunction. In a steady state, the streamfunction and the potential vorticity are functions of each other so that

$$q = Q(\Psi) \qquad \text{and} \qquad \psi = \Psi(Q), \tag{10.134}$$

where $Q$ is a differentiable but otherwise arbitrary function of its argument, and $\Psi$ its functional inverse. Equation (10.133) is then

$$\frac{\partial q}{\partial t} = -\frac{\mathrm{d}Q}{\mathrm{d}\psi} J(\Psi, \Psi) = 0 \tag{10.135}$$

and all steady solutions are of the form (10.134). We shall prove that if $\mathrm{d}\Psi/\mathrm{d}Q > 0$ then the flow is stable, in a sense to be made explicit below. Consider the evolution of perturbations about such a steady state, so that

$$q = Q + q', \qquad \psi = \Psi + \psi', \tag{10.136}$$

and we suppose that the perturbation vanishes at the domain boundary or that the boundary conditions are periodic. The potential vorticity perturbation satisfies, in the linear approximation,

$$\frac{\partial q'}{\partial t} + J(\psi', Q) + J(\Psi, q') = 0. \tag{10.137}$$

Now, because potential vorticity is conserved on parcels, any function of potential vorticity is also materially conserved, and in particular

$$\frac{\mathrm{D}\Psi(q)}{\mathrm{D}t} = \frac{\partial \Psi}{\partial t} + J(\psi, \Psi) = 0. \tag{10.138}$$

Linearizing this using (10.136) gives

$$\frac{\mathrm{d}\Psi}{\mathrm{d}Q}\frac{\partial q'}{\partial t} + J(\psi', \Psi) + J\left(\Psi, \frac{\mathrm{d}\Psi}{\mathrm{d}Q}q'\right) = 0. \tag{10.139}$$

We now form an energy equation from (10.137) by multiplying by $-\psi'$ and integrating over the domain. Integrating the first term by parts we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \frac{1}{2}(\nabla\psi')^2\,\mathrm{d}A = \int \psi' J(\Psi, q')\,\mathrm{d}A. \tag{10.140}$$

Similarly, from (10.139) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \frac{1}{2}\frac{\mathrm{d}\Psi}{\mathrm{d}Q}q'^2\,\mathrm{d}A = -\int \left[q' J(\psi', \Psi) + q' J\left(\Psi, \frac{\mathrm{d}\Psi}{\mathrm{d}Q}q'\right)\right]\,\mathrm{d}A. \tag{10.141}$$

The second term in square brackets vanishes. This follows using the property of Jacobians, obtained by integrating by parts, that

$$\langle aJ(b,c)\rangle = \langle bJ(c,a)\rangle = \langle cJ(a,b)\rangle = -\langle cJ(b,a)\rangle, \tag{10.142}$$

where the angle brackets denote horizontal integration. Using this we have

$$\begin{aligned}\left\langle q' J\left(\Psi, \frac{\mathrm{d}\Psi}{\mathrm{d}Q}q'\right)\right\rangle = -\left\langle \frac{\mathrm{d}\Psi}{\mathrm{d}Q}q' J(\Psi, q')\right\rangle &= -\frac{1}{2}\left\langle \frac{\mathrm{d}\Psi}{\mathrm{d}Q}J(\Psi, q'^2)\right\rangle \\ &= -\frac{1}{2}\left\langle q'^2 J\left(\frac{\mathrm{d}\Psi}{\mathrm{d}Q}, \Psi\right)\right\rangle = 0.\end{aligned} \tag{10.143}$$

Adding (10.140) and (10.141) the remaining nonlinear terms cancel and we obtain the conservation law,

$$\boxed{\begin{gathered}\widehat{H} = \frac{1}{2}\int\left[(\nabla\psi')^2 + \frac{\mathrm{d}\Psi}{\mathrm{d}Q}q'^2\right]\,\mathrm{d}A \\ \frac{\mathrm{d}\widehat{H}}{\mathrm{d}t} = 0\end{gathered}}\,. \tag{10.144}$$

The quantity $\widehat{H}$ is known as the *pseudoenergy* of the disturbance and because it is a conserved quantity, quadratic in the wave amplitude, it is (like pseudomomentum) a wave activity. Its conservation holds whether the disturbance is growing, decaying or neutral.

If $\mathrm{d}\Psi/\mathrm{d}Q$ is positive everywhere the pseudoenergy is a positive-definite quantity, and the growth of the disturbance is then strictly limited, and the basic state is said to be *stable in the sense of Liapunov*. This means that the magnitude of the perturbation, as measured by some norm, is bounded by its initial magnitude. In this case we define the norm

$$||\psi||^2 \equiv \int\left[(\nabla\psi)^2 + \frac{\mathrm{d}\Psi}{\mathrm{d}Q}(\nabla^2\psi)^2\right]\,\mathrm{d}A, \tag{10.145}$$

so that

$$||\psi'(t)||^2 = ||\psi'(0)||^2. \tag{10.146}$$

If $\mathrm{d}\Psi/\mathrm{d}Q > 0$ then, although the energy of the disturbance can grow, its final amplitude is bounded by the initial value of the pseudoenergy, because if perturbation energy is to grow perturbation enstrophy must shrink but it cannot shrink past zero. Normal-mode instability, in which modes grow exponentially, is completely precluded.

If the pseudoenergy is *negative definite* then stability is also assured, but this is a less common situation for it demands that $\mathrm{d}\Psi/\mathrm{d}Q$ be sufficiently negative so that the (negative of the) enstrophy contribution is always larger than the energy contribution, and this can usually only be satisfied in a sufficiently small domain. To see this, suppose that $q' = \nabla^2\psi'$, and that in the domain under consideration the Laplacian operator has eigenvalues $-k^2$ where

$$\nabla^2\psi' = -k^2\psi' \tag{10.147}$$

and the smallest eigenvalue, by magnitude, is $k_0^2$. Then, using Poincaré's inequality,

$$\int (\nabla^2\psi')^2\,\mathrm{d}A \geq k_0^2 \int (\nabla\psi')^2\,\mathrm{d}A, \tag{10.148}$$

a sufficient condition to make $\widehat{H}$ negative definite is that

$$\frac{\mathrm{d}\Psi}{\mathrm{d}Q} < -\frac{1}{k_0^2}. \tag{10.149}$$

As the domain gets bigger, $k_0$ diminishes and this condition becomes harder to satisfy.[12]

*Parallel shear flow and Fjørtoft's condition*

Consider the stability of a zonal flow (i.e., a flow in the $x$-direction), that varies only with $y$. The flow stability condition is then

$$\frac{\mathrm{d}\Psi}{\mathrm{d}Q} = \frac{\mathrm{d}\Psi/\mathrm{d}y}{\mathrm{d}Q/\mathrm{d}y} = -\frac{U - U_s}{\beta - U_{yy}} > 0, \tag{10.150}$$

where $U_s$ is a constant, representing an arbitrary, constant, zonal flow. The last equality follows because the problem is Galilean invariant, and we are therefore at liberty to choose $U_s$ arbitrarily. To connect this with Fjørtoft's condition (chapter 9) multiply the top and bottom by $(\beta - U_{yy})$, whence we see that a sufficient condition for stability is that $(U - U_s)(\beta - U_{yy})$ is everywhere negative. The derivation here, unlike our earlier one in section 9.3.2, makes it clear that the condition does not only apply to normal-mode instabilities.

### 10.8.2 ✦ Stratified quasi-geostrophic flow

The extension of the pseudoenergy arguments to quasi-geostrophic flow is mostly straightforward, but with a complication from the vertical boundary conditions at the surface and at an upper boundary, and the trusting reader may wish to skip straight to the results, (10.155)–(10.157).[13] For definiteness, we consider Boussinesq, $\beta$-plane quasi-geostrophic flow confined between flat rigid surfaces at $z = 0$ and $z = H$. The interior flow is governed by the familiar potential vorticity equation $\mathrm{D}q/\mathrm{D}t = 0$ and the buoyancy equation $\mathrm{D}b/\mathrm{D}t = 0$ at the two boundaries, where

$$q = \nabla^2\psi + \beta y + \frac{\partial}{\partial z}\left(S(z)\frac{\partial \psi}{\partial z}\right), \qquad b = f_0\frac{\partial \psi}{\partial z}, \tag{10.151}$$

and $S(z) = f_0^2/N^2$ is positive. The basic state ($\psi = \Psi, q = Q, b = B_1, B_2$) satisfies

$$\begin{gathered} \psi = \Psi(Q), \quad 0 < z < H, \\ \psi = \Psi_1(B_1), \quad z = 0 \qquad \text{and} \qquad \psi = \Psi_2(B_2), \quad z = H. \end{gathered} \tag{10.152}$$

Analogous to the barotropic case, we obtain the equations of motion for the interior perturbation

$$\frac{\partial q'}{\partial t} + J(\psi', Q) + J(\Psi, q') = 0, \tag{10.153a}$$

$$\frac{\mathrm{d}\Psi}{\mathrm{d}Q}\frac{\partial q'}{\partial t} + J(\psi', \Psi) + J\left(\Psi, \frac{\mathrm{d}\Psi}{\mathrm{d}Q}q'\right) = 0, \tag{10.153b}$$

and at the two boundaries

$$\frac{\partial b'}{\partial t} + J(\psi', B_i) + J(\Psi_i, b') = 0, \tag{10.154a}$$

$$\frac{\mathrm{d}\Psi_i}{\mathrm{d}B_i}\frac{\partial b'}{\partial t} + J(\psi', \Psi_i) + J\left(\Psi_i, \frac{\mathrm{d}\Psi_i}{\mathrm{d}B_i}b'\right) = 0, \tag{10.154b}$$

for $i = 1, 2$. (By $\mathrm{d}\Psi_i/\mathrm{d}B_i$ we mean the derivative of $\Psi_i$ with respect to its argument, evaluated at $B_i$.) From these equations, we form the pseudoenergy by multiplying (10.153a) by $-\psi'$, (10.153b) by $q'$, and (10.154a) by $\psi'$, (10.154b) by $b'$. After some manipulation we obtain the pseudoenergy conservation law:

$$\boxed{\begin{aligned} \widehat{H} &= \mathcal{E} + \mathcal{Z} + \mathcal{B}_1 + \mathcal{B}_2 \\ \frac{\mathrm{d}\widehat{H}}{\mathrm{d}t} &= 0 \end{aligned}}\,, \tag{10.155}$$

where

$$\begin{aligned} \mathcal{E} &= \frac{1}{2}\left\{(\nabla\psi')^2 + S\left(\frac{\partial \psi'}{\partial z}\right)^2\right\}, \qquad & \mathcal{Z} &= \frac{1}{2}\left\{\frac{\mathrm{d}\Psi}{\mathrm{d}Q}q'^2\right\}, \\ \mathcal{B}_1 &= \frac{1}{2}\left\langle \frac{S(0)}{f_0}\frac{\mathrm{d}\Psi_1}{\mathrm{d}B_1}b'(0)^2\right\rangle, \qquad & \mathcal{B}_2 &= -\frac{1}{2}\left\langle \frac{S(H)}{f_0}\frac{\mathrm{d}\Psi_2}{\mathrm{d}B_2}b'(H)^2\right\rangle. \end{aligned} \tag{10.156}$$

where the curly brackets denote a three-dimensional integration over the fluid interior, and the angle brackets denote a horizontal integration over the boundary surfaces at 0 and $H$. The pseudoenergy $\widehat{H}$ is positive-definite, and therefore stability is assured in that norm, if all of the following conditions are satisfied:

$$\frac{\mathrm{d}\Psi}{\mathrm{d}Q} > 0, \qquad \frac{1}{f_0}\frac{\mathrm{d}\Psi_1}{\mathrm{d}B_1} > 0, \qquad \frac{1}{f_0}\frac{\mathrm{d}\Psi_2}{\mathrm{d}B_2} < 0. \tag{10.157}$$

If the flow is compressible, the potential vorticity is $q = \nabla^2\psi + \beta y + \rho_R^{-1}\partial_z(\rho_R S \partial_z \psi)$, where $\rho_R = \rho_R(z)$, but the final stability conditions are unaltered. If the upper boundary is then removed to infinity where $\rho_R(z) = 0$ then only the lower boundary condition contributes to (10.157). In the layered form of the quasi-geostrophic equations the vertical boundary conditions are built in to the definitions of potential vorticity in the top and bottom layers. In this case, a sufficient condition for stability is that $\mathrm{d}\Psi/\mathrm{d}Q > 0$ in each layer. Indeed, an alternate derivation of (10.155)–(10.157) would be to incorporate the boundary conditions on buoyancy into the definition of potential vorticity by the delta-function construction of section 5.4.3.

*Zonal shear flow*

Consider now zonally uniform zonal flows, such as might give rise to baroclinic instability in a channel. The fields are then functions of $y$ and $z$ only, and the sufficient conditions for stability are:

$$\begin{aligned} \frac{\mathrm{d}\Psi}{\mathrm{d}Q} &= \frac{\partial\Psi/\partial y}{\partial Q/\partial y} = -\frac{U}{\mathrm{d}Q/\mathrm{d}y} > 0, \\ \frac{\mathrm{d}\Psi_1}{\mathrm{d}B_1} &= \frac{\mathrm{d}\Psi_1/\mathrm{d}y}{\mathrm{d}B_1/\mathrm{d}y} = \frac{U(0)}{\mathrm{d}U(0)/\mathrm{d}z} > 0, \\ \frac{\mathrm{d}\Psi_2}{\mathrm{d}B_2} &= \frac{\mathrm{d}\Psi_2/\mathrm{d}y}{\mathrm{d}B_2/\mathrm{d}y} = \frac{U(H)}{\mathrm{d}U(H)/\mathrm{d}z} < 0. \end{aligned} \tag{10.158}$$

using the thermal wind relation, and setting $f_0 = 1$ (its value is irrelevant). These results generalize Fjørtoft's condition to the stratified case,[14] and as in that case we are at liberty to add a uniform zonal flow to all the velocities.

### 10.8.3 ✦ Applications to baroclinic instability

We may use the stability conditions derived above to provide a few more results about baroclinic instability, including an alternate derivation of the minimum shear criterion in two-layer flow, and a derivation of the high-wavenumber cut-off to instability. In what follows we do not derive any new criteria; rather, the derivations make it apparent that the criteria are not restricted to perturbations of normal-mode form.

*Minimum shear in two-layer flow*

We consider two layers of equal depth, on a flat-bottomed $\beta$-plane with basic state

$$\Psi_1 = -U_1 y, \qquad \Psi_2 = -U_2 y \tag{10.159a}$$

$$Q_1 = \beta y - \frac{k_d^2}{2}(U_2 - U_1)y, \qquad Q_2 = \beta y - \frac{k_d^2}{2}(U_1 - U_2)y. \tag{10.159b}$$

This state is characterized by $Q_i = \gamma_i \Psi_i$ where

$$\gamma_1 = -\frac{(\beta + k_d^2\widehat{U})}{(\overline{U} + \widehat{U})}, \qquad \gamma_2 = -\frac{(\beta - k_d^2\widehat{U})}{(\overline{U} - \widehat{U})}, \tag{10.160}$$

with $\overline{U} = (U_1 + U_2)/2$ and $\widehat{U} = (U_1 - U_2)/2$. The barotropic flow does not affect the stability properties, so without loss of generality we may *choose* $\overline{U} < -\widehat{U}$, and this makes $\gamma_1 > 0$. Then

$\gamma_2$ is also positive if $\beta > k_d^2\widehat{U}/2$. Thus, a sufficient condition for stability is that

$$\widehat{U} < \frac{\beta}{k_d^2}, \tag{10.161}$$

as obtained in chapter 9. However, we now see that the stability condition does not apply only to normal-mode instabilities.[15]

Use of pseudomomentum conservation provides an alternative derivation of the same result. The flow will also be stable if in both layers $\partial Q/\partial y > 0$, for then the conserved pseudomomentum will be positive definite. If $U_1 > U_2$ then, from (10.159) $\mathrm{d}Q_1/\mathrm{d}y > 0$. The flow will be stable if $\mathrm{d}Q_2/\mathrm{d}y > 0$, and this gives

$$\widehat{U} = \frac{1}{2}(U_1 - U_2) < \frac{\beta}{k_d^2}, \tag{10.162}$$

as in (10.161).

*The high-wavenumber cut-off in two-layer baroclinic instability*

We can use a pseudoenergy argument to show that there is a high-wavenumber cut-off to two-layer baroclinic instability, with the basic state (10.159). The conserved pseudoenergy analogous to (10.155) and (10.156) is readily found to be

$$\widehat{H} = \left\langle (\nabla\psi_1')^2 + (\nabla\psi_2')^2 + \frac{1}{2}k_d^2(\psi_1' - \psi_2')^2 + \frac{q_1'^2}{\gamma_1} + \frac{q_2'^2}{\gamma_2} \right\rangle = 0. \tag{10.163}$$

Let us choose (without loss of generality) the barotropic flow to be $\overline{U} = \beta/k_d^2$. We then have $\gamma_1 = \gamma_2 = -1/k_d^2$, and the pseudoenergy is then just the actual energy minus $k_d^{-2}$ times the total enstrophy. If we define $\psi = (\psi_1' + \psi_2')/2$ and $\tau = (\psi_1' - \psi_2')/2$ then, using (12.31a) and (12.34), (10.163) may be expressed as

$$\widehat{H} = \left\langle (\nabla\psi)^2 + (\nabla\tau)^2 + k_d^2\tau^2 - k_d^{-2}\left\{(\nabla^2\psi)^2 + \left[(\nabla^2 - k_d^2)\tau\right]^2\right\}\right\rangle. \tag{10.164}$$

Now, let us express the fields as Fourier sums,

$$(\tau, \psi) = \sum_{k,l}(\widetilde{\tau}_{k,l}, \widetilde{\psi}_{k,l})\mathrm{e}^{\mathrm{i}(kx+ly)}. \tag{10.165}$$

(This expression assumes a doubly-periodic domain; essentially the same end-result is obtained in a channel.) The pseudoenergy may then be written as

$$\widehat{H} = \sum_{k,l}\left[K^2\widetilde{\psi}_{k,l}^2(k_d^2 - K^2) + K'^2\widetilde{\tau}_{k,l}^2(k_d^2 - K'^2)\right] \tag{10.166}$$

where $K^2 = k^2 + l^2$ and $K'^2 = K^2 + k_d^2$. If the deformation radius is sufficiently large (or the domain sufficiently small) that $K^2 > k_d^2$, then the pseudoenergy is *negative-definite,* so the flow is stable, no matter what the shear may be. Such a situation might arise on a planet whose circumference was less than the deformation radius, or in a small ocean basin. In the linear problem, in which perturbation modes do not interact, horizontal wavenumbers with $k^2 > k_d^2$ are stable and there is thus a high-wavenumber cut-off to instability, as was found in chapter 9 by direct calculation.

*Notes*

1 After **?**.

2 **??** and **?**. See also problem 10.3.

3 These restrictions on the basic state are not necessary to prove orthogonality, but they make the algebra simpler. Also, we pay no attention here to the nature of the eigenvalues of (10.45), which, in general, consist of both a discrete and a continuous spectrum. See Farrell (1984) and **?**.

4 The TEM was introduced by **??** and **?**. A precursor is the paper of **?**, who noted the shortcomings of zonal averaging in uncovering the meaning of indirect cells in laboratory experiments, and by extension the atmosphere.

5 This problem can be worked around in some cases (**?**Greatbatch 1998).

6 The main result of this subsection was originally obtained by **?**. I thank A. Plumb for a discussion about the derivation given here. See also de Szoeke & Bennett (1993) for related earlier work, and Juckes (2001) and Nurser & Lee (2004) for generalizations.

7 Non-acceleration arguments have a long history, with contributions from Charney & Drazin (1961), **?**, **?** and **?**. **?** put these results in the context of the EP flux and the TEM formalism, and **?** reviews and provides examples.

8 Conservation laws of this ilk, their connection to the underlying symmetries of the basic state and (relatedly) their finite-amplitude extension, are discussed by McIntyre & Shepherd (1987) and **?**. Conservation of momentum is related to the translational invariance of the medium; conservation of $\mathcal{A}$ may be shown to be related to the translational invariance of the basic state, and hence the appellation 'pseudomomentum'.

9 See **?**.

10 Steve Garner and Raffaele Ferrari both provided very helpful input to this section.

11 The original papers are **??**, with a number of results being subsequently developed by **?**. See **?** for a review.

12 The stability criterion is sometimes referred to as 'Arnold's second condition'. More discussion, especially with regard to boundary conditions, is given in McIntyre & Shepherd (1987).

13 **?**, but the method we use is more direct.

14 Pedlosky (1964) derived these conditions by a normal-mode approach.

15 **?** and **?** further consider the finite-amplitude case.

*Further reading*

Andrews, D. G., Holton, J. R. & Leovy, C.B., 1987. *Middle Atmosphere Dynamics.*
Provides a discussion of a number of topics in wave dynamics and wave–mean-flow interaction, including the TEM, mainly in the context of stratospheric dynamics.

Buhler, O. 2009. *Waves and Mean Flows.*
This book provides a comprehensive and readable discussion of waves, mean flows and their interaction, including the Transformed Eulerian Mean, the Generalized Lagrangian Mean, and more.

*Problems*

10.1 Prove that

$$\langle aJ(b,c)\rangle = \langle bJ(c,a)\rangle = \langle cJ(a,b)\rangle = -\langle cJ(b,a)\rangle \qquad \text{(P10.1)}$$

where the angle brackets denote a horizontal integration, and the boundary conditions correspond to no-normal-flow or periodic.

10.2 Consider an axisymmetric barotropic shear flow given by

$$\zeta = \begin{cases} 2\Omega & r \le R \\ 0 & r > R \end{cases} \tag{P10.2}$$

where $\Omega$ and $R$ are constants. Thus, the inner region is in solid body rotation and the outer region is irrotational, and we suppose that the velocity is continuous. [Is this implied by (P10.2) or is it an extra condition?] Suppose that the boundary between the two regions is perturbed. Find the phase speed of this disturbance in terms of its azimuthal wavenumber. Show that for large wavenumbers this reduces to the dispersion relation for a point jet [i.e., $c = U_0 + (\zeta_1 - \zeta_2)/2k$ — see (9.37)], and define what 'large' means in this context.

10.3 ✦ Show that perturbations to a horizontally sheared flow are orthogonal with respect to the wave activity norm. You may restrict attention to two-dimensional (barotropic) flow on the $\beta$-plane.
*Partial solution.* The eigenvalue value equation is

$$(\overline{u}\nabla_k^2 + \partial_y\overline{q})\psi = c\nabla_k^2\psi \tag{P10.3}$$

where $\nabla_k^2 = \partial_{yy} - k^2$. If $q = \nabla_k^2\psi$ then the eigenvalue equation may be written

$$Mq \equiv (\overline{u} + \partial_y\overline{q}\,\nabla_k^{-2})q = cq \tag{P10.4}$$

or

$$Nq \equiv \left(\frac{\overline{u}}{\partial_y\overline{q}} + \nabla_k^{-2}\right)q = c\frac{q}{\partial_y\overline{q}} \tag{P10.5}$$

The operator $N$ on the left-hand side of (P10.5) is self-adjoint (show this) so that the eigenfunctions associated with two different eigenvalues are orthogonal with respect to $\partial_y\overline{q}$; that is, for $n \neq$ m,

$$\int\int \frac{q_n q_m}{\partial_y\overline{q}}\,\mathrm{d}y = 0. \tag{P10.6}$$

If this derivation holds, why is it not simply the case that, from (P10.4), that

$$\int q_n q_m\,\mathrm{d}y = 0, \tag{P10.7}$$

for $n \neq m$? (Is $M$ self-adjoint?)

10.4 Obtain an expression for the EP flux due to equatorial Kelvin waves. What is the sign of the wave drag in the regions of Kelvin wave generation and dissipation?

10.5 ✦ Can the high-wavenumber cut-off to instability in the Eady problem be obtained by wave-activity arguments (e.g., by proving the pseudoenergy is negative definite, as in the two-layer problem). If so, do so.

10.6 ✦ *Stability conditions in the continuously stratified QG model*
Consider the modified Eady problem (Boussinesq, uniform stratification, flow contained between two flat horizontal surfaces at $0$ and $H$), but instead of a uniform shear suppose that the basic state is given by

$$U = -U_0\cos(\pi pz/H) \tag{P10.8}$$

where $U_0 > 0$, and allow $\beta$ to be non-zero.

*(a)* Show there is a critical shear, and that this diminishes as $p$ increases.

*(b)* Show there is a high-wavenumber cut-off to instability.

Compare the results to those of the two-layer model.

*Sketch of solution.*

*(a)* The basic state has no temperature gradient at the boundary, and so stability is assured if $\partial Q/\partial y$ is positive everywhere. The basic state potential vorticity is $Q = \nabla^2\Psi + (f_0^2/N^2)\partial^2\Psi/\partial z^2 + \beta y$, so that

$$\frac{\mathrm{d}Q}{\mathrm{d}y} = \beta - k_d^2\pi^2 p^2 U_0 \cos(\pi p z/H), \tag{P10.9}$$

where $k_d^2 = f_0^2 H^2/N^2$ and $U = -\partial\Psi/\partial y$. Thus, stability is assured if $U_0 < (\beta/k_d^2\pi^2 p^2)$.

*(b)* Choose a barotropic flow equal to $-\beta/(k_d^2\pi^2 p^2)$ so that the mean flow is given by $U = U_0\cos(\pi p z/H) - \beta/(k_d^2 p^2)$. Then $\gamma(z) \equiv Q/\Psi = -k_d^2\pi^2 p^2$. Expand the perturbation streamfunction as $\psi = \sum_{\mathbf{k},\alpha}\psi_{\mathbf{k},\alpha}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{x}}\cos\alpha z$ and obtain an expression for the pseudoenergy analogous to (10.166), and find the conditions under which it is sign-definite.

10.7 Obtain, or at least verify, the solution (10.113). Plot it for various values of deformation radii (using appropriate computer software).

10.8 ✦ Obtain and plot the residual circulation in the linear two-level (Phillips) baroclinic instability problem, both for $\beta = 0$ and $\beta \neq 0$. Also obtain and plot the corresponding tendencies of the zonally averaged zonal wind and buoyancy fields, and interpret your results. A good answer will include a comparison of the solutions with and without beta, and a comparison of the solutions with those of the Eady problem. [Reading **?** may be helpful.]

10.9 ✦ *Balance of terms in the mean flow equations*

*(a)* In the Eady problem the mean flow evolves according to (10.106). To what degree is there an instantaneous balance between the terms on the right-hand side? That is, is the mean-flow evolution a residual between two larger terms? (The answer is trivial for the zonal flow evolution, but less so for buoyancy.)

*(b)* Repeat this problem for the two-layer problem, in both Eulerian and TEM forms. For the latter, calculate the balance between the potential vorticity flux, the residual meridional flow and the zonal flow tendency.

# Part III

# LARGE-SCALE ATMOSPHERIC CIRCULATION

*Catch a wave and you're sitting on top of the world.*
Brian Wilson and Mike Love (The Beach Boys), *Catch a Wave.*

# CHAPTER SIXTEEN

# Planetary Waves and Zonal Asymmetries

PLANETARY WAVES are large-scale Rossby waves in which the potential vorticity gradient is provided by differential rotation (i.e., the beta effect), and they are ubiquitous in Earth's atmosphere. They propagate horizontally over the two Poles, and they propagate vertically into the stratosphere and beyond. In the previous chapter we saw that it is the propagation of Rossby waves away from their mid-latitude source that gives rise to the mean eastward eddy-driven jet. In this chapter we will see that it is the dynamics of such waves that determines the large-scale *zonally asymmetric* circulation of the mid-latitude atmosphere. Our task is to try to understand all this, and to this end the chapter itself has two main topics. In the first few sections we discuss the properties and propagation of planetary waves, in many ways these sections being a continuation of chapter 6. We then look more specifically at stationary planetary waves forced by surface variations in topography and thermal properties, for it is these waves that give rise to the zonally asymmetric circulation. We will find, perhaps not surprisingly, that the stationary wave patterns depend both on the surface boundary conditions and the zonally-averaged flow itself.

In proceeding this way we are dividing our task of constructing a theory general circulation of the extratropical atmosphere into two. The first task (chapters 14 and 15) was to understand the zonally averaged circulation and the transient zonal asymmetries by supposing that, to a first approximation, this circulation is qualitatively the same as it would be if the boundary conditions were zonally symmetric, with no mountains or land–sea contrasts. Given the statistically zonally symmetric circulation, the second task, and the one now confronting us, is to understand the zonally asymmetric circulation. We may do this by supposing that the latter is a perturbation on the former, and using a theory linearized about the zonally symmetric state. It is by no means obvious that such a procedure will be successful, for it depends on the nonlinear interactions among the zonal asymmetries being weak. We might make some a priori estimates that suggest that this might be the case, but the ultimate justification for the approach lies in its a posteriori success.

In our treatment of stationary waves we do not include the effects of transient eddies, that is the effects of equilibrated, finite-amplitude baroclinic systems. This would be the most difficult aspect of calculating the stationary wave response, although their effects may be included diagnostically by evaluating their associated heat and momentum fluxes from observations and adding them to the right-hand sides of the appropriate equations. However, we will find that the calculations are quite revealing even if the effects of transient eddies are omitted entirely. In our discussion of stationary waves we will focus first on the response to orography at the lower boundary, and then consider thermodynamic forcing — arising, for example, from an inhomogeneous surface temperature field.

## 16.1 ROSSBY WAVE PROPAGATION IN A SLOWLY VARYING MEDIUM

In chapters 6 and 7 we looked at wave propagation using linearized equations of motion. We now focus and extend this discussion by looking at Rossby wave propagation in a medium in which the parameters (such as the zonal wind and the stratification) vary spatially. Such a situation is, of course, occurs in the real atmosphere. If the parameters do vary then waves may propagate into a region in which they amplify, perhaps violating the initial assumption of linearity, so let us first look at what the conditions for linearity are.

### 16.1.1 Conditions for linearity

We often linearize the equations of motion in a rather formal way, just by eliminating the nonlinear terms, simply to better understand the behaviour of the system. However, if we are also hoping that the linear equations are an accurate representation of the dynamics we must usually assume that the perturbation quantities are small compared to the background state, or at least that the nonlinear terms are small. This is, of course, not always the case and indeed it may be that in course of propagation the waves amplify and may even *break.* Wave breaking is familiar to anyone who has been to the beach and watched water waves move toward the shore and crash in the 'surf zone' as the mean depth becomes too shallow to support laminar surface waves. Manifestly, the linear approximation breaks down at this point. More generally, wave breaking simply refers to an irreversible deformation of material surfaces, generally leading to dissipation. Since Rossby waves generally grow in amplitude as they propagate up we can expect Rossby wave breaking to occur somewhere in the atmosphere, but waves can also break as they propagate laterally, if and when they grow in size to such an extent that the nonlinear terms in the equations of motion become important.

To examine this consider the quasi-geostrophic potential vorticity equation,

$$\left(\frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla\right) q = 0, \qquad q = \beta y + \nabla^2 \psi' + \frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \psi'}{\partial z}\right). \tag{16.1a,b}$$

The derivation of this equation was given in chapter 5 and all the terms are defined there. In brief, $q$ is the quasi-geostrophic potential vorticity and $\psi$ the streamfunction, $f_0$ is the Coriolis parameter and $\rho_R$ is a density profile, a function of z only. Breaking the above equation up into mean and perturbation quantities in the usual way we obtain

$$\left(\frac{\partial}{\partial t} + \overline{u}(y,z)\frac{\partial}{\partial x}\right) q' + v'\frac{\partial \overline{q}}{\partial y} = -\left(\frac{\partial u'q'}{\partial x} + \frac{\partial v'q'}{\partial y}\right). \tag{16.2}$$

In the linear approximation we neglect the terms on the right-hand side and, seeking wave-like solutions of the form $\psi = F(x - ct)$ we obtain

$$(\overline{u} - c)\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = 0. \tag{16.3}$$

For the linear approximation to be valid the terms in this equation must be larger than the nonlinear terms in (16.2), and this will be the case if

$$|\overline{u} - c| \gg |u'| \qquad \text{and} \qquad \left|\frac{\partial \overline{q}}{\partial y}\right| \gg \left|\frac{\partial q'}{\partial y}\right|. \tag{16.4a,b}$$

Although it is common in elementary treatments of wave dynamics to treat the case in which $\overline{u}$ is a constant, we may also consider the case in which $\overline{u}$ varies slowly, either in latitude or height or both, and (16.3) approximately holds locally. If a wave propagates into a region in which $\overline{u} = c$ then the linear criterion *must* break down. Regions where $\overline{u} = c$ are called *critical surfaces, critical lines, critical layers, critical heights* or *critical latitudes,* depending on context, and we discuss their effects on waves more below. Note that the location of a critical surface does not depend on the frame of reference used to measure the velocities.

Before proceeding further we write down for reference a few results for the simplest case when $\partial\overline{q}/\partial y, \overline{u}, N^2$ and $\rho_R$ are all constant (and refer to section 6.5 as needed). Without undue attention to boundary conditions we seek solutions of the form

$$\psi' = \mathrm{Re}\,\widetilde{\psi}\mathrm{e}^{\mathrm{i}(kx+ly+mz-\omega t)}, \tag{16.5}$$

and obtain the dispersion relation

$$\omega = \overline{u}k - \frac{k\partial\overline{q}/\partial y}{k^2 + l^2 + Pr^2m^2}. \tag{16.6}$$

where $Pr = f_0/N$ is the Prandtl ratio. The components of the group velocity are given by

$$c_g^x = \overline{u} + \frac{(k^2 - l^2 - Pr^2m^2)\partial_y\overline{q}}{(k^2 + l^2 + Pr^2m^2)^2}, \qquad c_g^y = \frac{2kl\partial_y\overline{q}}{(k^2 + l^2 + Pr^2m^2)^2},$$
$$c_g^z = \frac{2kmPr^2\partial_y\overline{q}}{(k^2 + l^2 + Pr^2m^2)^2}. \tag{16.7a,b,c}$$

where $\partial\overline{q}/\partial y = \beta$.

### 16.1.2 Conditions for wave propagation

Let us now consider Rossby wave propagation in a medium in which the zonal wind varies slowly with latitude and height, assuming for simplicity that the density, $\rho_R$, is a constant. The equation of motion is

$$\left(\frac{\partial}{\partial t} + \overline{u}(y,z)\frac{\partial}{\partial x}\right)q' + v'\frac{\partial\overline{q}}{\partial y} = 0. \tag{16.8}$$

Because the coefficients of the equation are not constant we cannot assume harmonic solutions in the $y$ and $z$ directions; rather, we seek solutions of the form

$$\psi' = \widetilde{\psi}(y,z)\mathrm{e}^{\mathrm{i}k(x-ct)}. \tag{16.9}$$

If the parameters in (16.8) are varying slowly compared to the wavelength of the waves then a dispersion relation still exists (as discussed in section 6.3), but the relation will be of the form $\omega = \Omega(\boldsymbol{k}; \boldsymbol{x}, t)$; where the function $\Omega$ varies slowly in space. Now, if the medium is not an explicit function of $x$ or of time the $x$-wavenumber and the frequency will be a constant, and hence $c$ is constant too, and we can use the dispersion relation to find what are effectively the other wavenumbers in the problem. Using (16.9) in (16.8) we find (if, for simplicity, $N^2$ is a constant)

$$\frac{\partial^2 \widetilde{\psi}}{\partial y^2} + \frac{f_0^2}{N^2}\frac{\partial^2 \widetilde{\psi}}{\partial z^2} + n^2(y,z)\widetilde{\psi} = 0 \tag{16.10a}$$

where

$$n^2(y,z) = \frac{\partial \overline{q}/\partial y}{\overline{u} - c} - k^2. \tag{16.10b}$$

Note that in determining $n$ this way we are assuming the frequency is known and using the dispersion relation to determine the quantity $n$. Eq. (16.10) is similar to the Rayleigh equation encountered in chapter 9. The quantity $n$ is the *refractive index* and it greatly affects how the waves propagate: solutions are wavelike when $n^2$ is positive and evanescent when $n^2$ is negative. (To see this in a simple case, suppose there is no $z$-variation so that $\partial^2\widetilde{\psi}/\partial y^2 + n^2\widetilde{\psi} = 0$. If $n$ is constant and real we have harmonic solutions in the $y$-direction of the form $\exp(iny)$. If $n^2 < 0$ the solutions will evanesce.) Indeed, waves tend to propagate toward regions of large $n^2$ and turn away from regions of negative $n^2$, as we will see in the examples to follow.

The value of $n^2$ will become very large if and as $\overline{u}$ approaches $c$ from above and the waves, being very short, will tend to break. If $\overline{u}$ continues to diminish and becomes smaller than $c$ then $n^2$ switches from being large and positive to large and negative. If $n^2$ diminishes because $\partial\overline{q}/\partial y$ diminishes then it will transition smoothly to a negative value. The location where $\overline{u} = c$ is called a critical surface (or line). The location where $n^2$ passes through zero is called a turning surface (or line).

The bounds on $n^2$ can be translated into bounds on the zonal phase speed $c$. Given a zonal wind $\overline{u}$, $c$ is bounded by

$$\overline{u} - \frac{\partial\overline{q}/\partial y}{k^2 + \gamma^2} < c < \overline{u}. \tag{16.11}$$

At the upper bound (a critical surface) the wavelength is small and wave breaking is likely to occur. At the lower bound (a turning surface) the refractive index tends to zero and the wave length tends to infinity. Waves will tend to propagate away from regions with a small $n$ and be refracted toward regions of large $n$. The bounds can also be expressed in terms of the zonal velocity:

$$0 < \overline{u} - c < \frac{\partial\overline{q}/\partial y}{k^2 + \gamma^2}. \tag{16.12}$$

This form is useful when considering a situation in which the wave speed is given, for example by boundary conditions; equation (16.12) then tells us what under what configurations of zonal velocity wave propagation can occur. The lower bound corresponds to a critical surface and the upper bound to a turning surface.

It is algebraically complicated to extend our analysis further in the three-dimensional case, so let us now consider the cases in which the inhomogeneities in the medium occur separately in the horizontal and vertical.

## 16.2 HORIZONTAL PROPAGATION OF ROSSBY WAVES

Consider the purely horizontal problem with linearized equation of motion is

$$\left(\frac{\partial}{\partial t} + \overline{u}(y)\frac{\partial}{\partial x}\right)q' + v'\frac{\partial \overline{q}}{\partial y} = 0, \tag{16.13}$$

where $q' = \nabla^2\psi'$, $v' = \partial\psi'/\partial x$ and $\partial\overline{q}/\partial y = \beta - \overline{u}_{yy}$. If $\overline{u}$ and $\partial\overline{q}/\partial y$ do not vary in space then we may seek wavelike solutions in the usual way and obtain the dispersion relation

$$\omega \equiv ck = \overline{u}k - \frac{\partial\overline{q}/\partial y}{k^2 + l^2} \tag{16.14}$$

where $k$ and $l$ are the $x$- and $y$-wavenumbers.

If the parameters vary in the $y$-direction then we seek a solution of the form $\psi' = \widetilde{\psi}(y)\exp[\mathrm{i}k(x-ct)]$ and obtain, analogously to (16.10)

$$\frac{\partial^2\widetilde{\psi}}{\partial y^2} + l^2(y)\widetilde{\psi} = 0, \quad \text{where} \quad l^2(y) = \frac{\partial\overline{q}/\partial y}{\overline{u} - c} - k^2 \tag{16.15a,b}$$

If the parameter variation is sufficiently small, occurring on a spatial scale longer than the wavelength of the waves, then we may expect that the disturbance will propagate locally as a plane wave. The solution is then of WKB form (see appendix to chapter 7) namely

$$\widetilde{\psi}(y) = A_0 l^{-1/2}\exp(\mathrm{i}\int l\,\mathrm{d}y) \tag{16.16}$$

where $A_0$ is a constant. The phase of the wave in the $y$-direction, $\theta$, is evidently given by $\theta = \int l\,\mathrm{d}y$, so that the local wavenumber is given by $\mathrm{d}\theta/\mathrm{d}y = l$. The group velocity is calculated in the normal way using the dispersion relation (16.14) and we obtain

$$c_g^x = \overline{u} + \frac{(k^2 - l^2)\partial\overline{q}/\partial y}{(k^2 + l^2)^2}, \qquad c_g^y = \frac{2kl\,\partial\overline{q}/\partial y}{(k^2 + l^2)^2}. \tag{16.17a,b}$$

where $\partial\overline{q}/\partial y = \beta - \overline{u}_{yy}$ and $l$ is given by (16.15b), with both quantities varying slowly in the $y$-direction.

### 16.2.1 Wave amplitude

As a Rossby wave propagates its amplitude is not necessarily constant because, in the presence of a shear, the wave may exchange energy with the background state and the WKB solution, (16.15), tells us that the variation goes like $l^{-1/2}(y)$. This variation can be understood from somewhat more general considerations. As discussed in chapter 10 an inviscid, adiabatic wave will conserve its wave activity meaning that

$$\frac{\partial\mathcal{A}}{\partial t} + \nabla\cdot\boldsymbol{\mathcal{F}} = 0, \tag{16.18}$$

where, we recall from section 10.2.1, $\mathcal{A}$ is a quantity quadratic in the wave amplitude and $\boldsymbol{\mathcal{F}}$ is the flux of $\mathcal{A}$, and the two are related by the group velocity property $\boldsymbol{\mathcal{F}} = \boldsymbol{c}_g\mathcal{A}$. In the zonally-averaged case the wave activity and flux for the quasi-geostrophic equations are given by

$$\mathcal{A} = \frac{\overline{q'^2}}{2\partial\overline{q}/\partial y}, \qquad \boldsymbol{\mathcal{F}} = -\overline{u'v'}\,\mathbf{j} + \frac{f_0}{N^2}\overline{v'b'}\,\mathbf{k}, \tag{16.19}$$

with $\boldsymbol{\mathcal{F}}$ is the Eliassen–Palm (EP) flux. If the waves are steady then $\nabla \cdot \boldsymbol{\mathcal{F}} = 0$, and in the two-dimensional case under consideration this means that $\partial \overline{u'v'}/\partial y = 0$. Thus $u'v' = kl|\widetilde{\psi}|^2 =$ constant, and since $k$ is constant along a ray the amplitude of a wave varies like

$$|\widetilde{\psi}| = \frac{A_0}{\sqrt{l(y)}} \tag{16.20}$$

as in the WKB solution. The energy of the wave then varies like

$$\text{Energy} = (k^2 + l^2)\frac{A_0^2}{l}. \tag{16.21}$$

### 16.2.2 Two examples

To illustrate the above ideas in a concrete fashion we consider two simple examples, one with a turning line and one with a critical line.

*Waves with a turning latitude*

A turning line arises where $l = 0$ and it corresponds to the lower bound of $c$ in (16.11). The line arises if the potential vorticity gradient diminishes to such an extent that $l^2 < 0$ and the waves then cease to propagate in the $y$-direction. This may happen even in in unsheared flow as a wave propagates polewards and the magnitude of beta diminishes.

As a wave packet approaches a turning latitude then $n$ goes to zero so the amplitude, and the energy, of the wave approach infinity. However, the wave will never reach the turning latitude because the meridional component of the group velocity is zero, as can be seen from the expressions for the group velocity, (16.17). As a wave approaches the turning latitude $c_g^x \to (\beta - \overline{u}_{yy})/k^2$ and $c_g^y \to 0$, so the group velocity is purely zonal and indeed as $l \to 0$

$$\frac{c_g^x - \overline{u}}{c_g^y} = \frac{k}{2l} \to \infty. \tag{16.22}$$

Because the meridional wavenumber is small the wavelength is large, so we do not expect the waves to break. Rather, we intuitively expect that a wave packet will turn — hence the eponym ‘turning latitude’ — and be reflected.

To illustrate this, consider waves propagating in a background state that has a beta effect that diminishes polewards but no horizontal shear. To be concrete suppose that $\beta = 5$ at $y = 0$, diminishing linearly to $\beta = 0$ at $y = 0$, and that $\overline{u} - c = 1$ everywhere. There is no critical line but depending on the $x$-wavenumber there may be a turning line, and if we choose $k = 1$ then the turning line occurs when $\beta = 1$ and so at $y = 0.8$. Note that the turning latitude depends on the value of the $x$-wavenumber — if the zonal wavenumber is larger then waves will turn further south. The parameters are illustrated in Fig. 16.1.

For a given zonal wavenumber ($k = 1$ in this example) the value of $l^2$ is computed using (16.15b), and the components of the group velocity using (16.17), and these are illustrated in Fig. 16.2. Note that we may choose either a positive or a negative value of $l$, corresponding to northward or southward oriented waves, and we illustrate both in the figure. The value of $l^2$ becomes zero at $y = 0.8$, and this corresponds to a turning latitude. The values of the wave

## Rossby Wave Propagation in a Slowly Varying Medium

The linear equation of motion is, in terms of streamfunction,

$$\left(\frac{\partial}{\partial t}+\overline{u}(y,z)\frac{\partial}{\partial x}\right)\left[\nabla^2\psi'+\frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial\psi'}{\partial z}\right)\right]+\frac{\partial\psi'}{\partial x}\frac{\partial\overline{q}}{\partial y}=0. \tag{RP.1}$$

We suppose that the parameters of the problem vary slowly in $y$ and/or $z$ but are uniform in $x$ and $t$. The frequency and zonal wavenumber are therefore constant. We seek solutions of the form $\psi'=\widetilde{\psi}(y,z)\mathrm{e}^{\mathrm{i}k(x-ct)}$ and find (if, for simplicity, $N^2$ and $\rho_R$ are constant)

$$\frac{\partial^2\widetilde{\psi}}{\partial y^2}+\frac{f_0^2}{N^2}\frac{\partial^2\widetilde{\psi}}{\partial z^2}+n^2(y,z)\widetilde{\psi}=0 \tag{RP.2a}$$

where

$$n^2(y,z)=\frac{\partial\overline{q}/\partial y}{\overline{u}-c}-k^2. \tag{RP.2b}$$

The value of $n^2$ must be positive in order that waves can propagate, and so waves cease to propagate when they encounter either

(i) *A turning line,* where $n^2=0$, or

(ii) *A critical line,* where $\overline{u}=c$ and $n^2$ becomes infinite.

The bounds may usefully be expressed as a condition on the zonal flow:

$$0<\overline{u}-c<\frac{\partial\overline{q}/\partial y}{k^2}. \tag{RP.3}$$

If the length scale over which the parameters of the problem vary is much longer than the wavelengths themselves we can expect the solution to look locally like a plane wave and a WKB analysis can be employed. In the purely horizontal problem we assume a solution of the form $\psi'=\widetilde{\psi}(y)\mathrm{e}^{\mathrm{i}k(x-ct)}$ and find

$$\frac{\partial^2\widetilde{\psi}}{\partial y^2}+l^2(y)\widetilde{\psi}=0,\qquad l^2(y)=\frac{\partial\overline{q}/\partial y}{\overline{u}-c}-k^2. \tag{RP.4}$$

The solution is of the form

$$\widetilde{\psi}(y)=Al^{-1/2}\exp\left(\pm\mathrm{i}\int l\,\mathrm{d}y\right). \tag{RP.5}$$

Thus, $l(y)$ is the local $y$-wavenumber, and the amplitude of the solution varies like $l^{-1/2}$. At a critical line the amplitude of the wave will go to zero although the energy may become very large, and since the wavelength is small the waves may break. At a turning line the amplitude and energy will both be large, but since the wavelength is long the waves will not necessarily break. A similar analysis may be employed for vertically propagating Rossby waves.

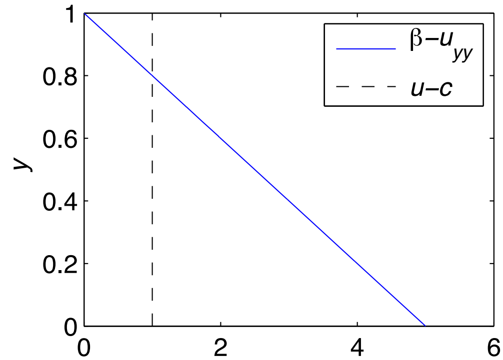


**Fig. 16.1** Parameters for the first example considered in section 16.2.2, with all variables nondimensional. The zonal flow is uniform with $u = 1$ and $c = 0$ (so that $\overline{u}_{yy} = 0$) and $\beta$ diminishes linearly as $y$ increases polewards as shown. With zonal wavenumber $k = 1$ there is a turning latitude at $y = 0.8$, and the wave properties are illustrated in Fig. 16.2.

amplitude and energy are computed using (16.20) and (16.21) (with an arbitrary amplitude at $y = 0$) and these both become infinite at the turning latitude.

What is happening physically? We may suppose that at any given location in the domain there is a source of waves. In the real atmosphere baroclinic disturbances in mid-latitudes are one such source. Waves propagate away from the source (as they must, since the waves must carry energy away), and this determines the sign of $n$ of any particular wave packet. The disturbance may in general consist of many zonal wavenumbers and many frequencies (or phase speeds, $c$), but the dispersion relation must be satisfied for each pair and this determines the meridional wavenumber via an equation such as (16.15). As the wave packet propagates away from the source then, as we noted in section 6.3 on ray theory, if the medium is zonally symmetric the $x$-wavenumber, $k$, is preserved. If the medium is not time-varying then the frequency, and therefore the wave speed $c$, are also preserved. We may approximately construct a ray by following the arrows in Fig. 16.2, and we see that a ray propagating polewards will bend eastward as it approaches the turning latitude Although its amplitude will become large it will not necessarily break because the wavelength is large; in fact, the packet may be reflected southward. We may heuristically construct a ray trajectory by drawing a line that is always parallel to the arrows marking the group velocity. Indeed, the entire procedure might be thought of as an Eulerian analogue of ray theory; rather than following a wave packet we just evaluate the field of group velocity, and if there is no explicit time dependence in the problem a ray follows the arrows.

### *Waves with a critical latitude*

A critical line occurs when $\overline{u} = c$, corresponding to the upper bound of $c$ in (16.11), and from (16.15) we see that at a critical line the meridional wavenumber approaches infinity. From (16.17) we see that both the $x$- and $y$-components of the group velocity are zero — a wave packet approaching a critical line just stops. Specifically, as $l$ becomes large

$$c_g^x - \overline{u} \to 0, \qquad c_g^y \to 0, \qquad \frac{c_g^x - \overline{u}}{c_g^y} \to -\frac{l}{k} \to -\infty. \tag{16.23}$$

From (16.20) the amplitude of the wave packet also approaches zero, but its energy approaches infinity. Since the wavelength is very small we expect the waves to *break* and deposit their momentum, and this situation commonly arises when Rossby waves excited in midlatitudes propagate equatorward and encounter a critical latitude in the subtropics.

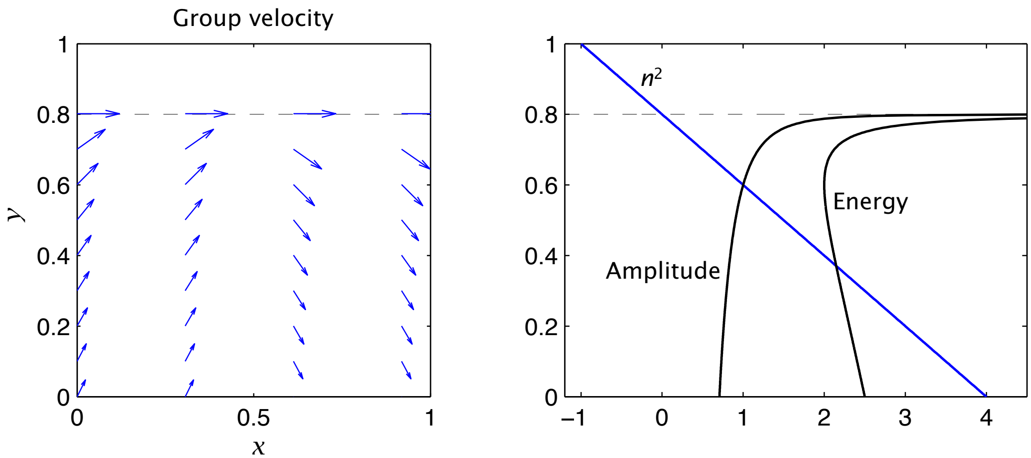


**Fig. 16.2** Left: The group velocity evaluated using (16.17) for the parameters illustrated in Fig. 16.1, which give a turning latitude at $y = 0.8$. For $x < 0.5$ we choose positive values of $n$, and a northward group velocity, whereas for $x > 0.5$ we choose negative values of $n$. Right panel: Values of refractive index squared ($n^2$), the energy and the amplitude of a wave. $n^2$ is negative for $y > 0.8$. See text for more description.

To illustrate this let us construct background state that has an eastward jet in midlatitudes becoming westward at low latitudes, with $\beta$ constant chosen to be large enough so that $\beta - \overline{u}_{yy}$ is positive everywhere. (Specifically, we choose $\beta = 1$ and $\overline{u} = -0.03\sin(8\pi y/5 + \pi/2) - 0.5)$, but the precise form is not important.) If $c = 0$ then there is a critical line when $\overline{u}$ passes through zero, which in this example occurs at $x = 0.2$. (The value of $\overline{u} - c$ is small at $y = 1$, but no critical line is actually reached.) These parameters are illustrated in Fig. 16.3. We also choose $k = 5$, which results in a positive value for $l^2$ everywhere.

As in the previous example we compute the value of $l^2$ using (16.15b) and the components of the group velocity using (16.17), and these are illustrated in Fig. 16.4, with northward propagating waves shown for $x < 0.5$ and southward propagating waves for $x > 0.5$. The value of $n^2$ increases considerably at the northern and southern edges of the domain, and is actually

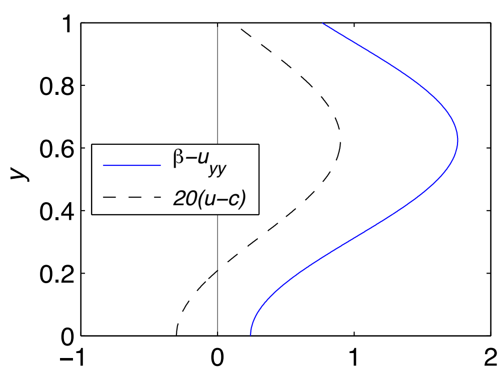


**Fig. 16.3** Parameters for the second example considered in section 16.2.2, with all variables nondimensional. The zonal flow has a broad eastward jet and $\beta$ is constant. There is a critical line at $y = 0.2$, and with zonal wavenumber $k = 5$ the wave properties are illustrated in Fig. 16.4.

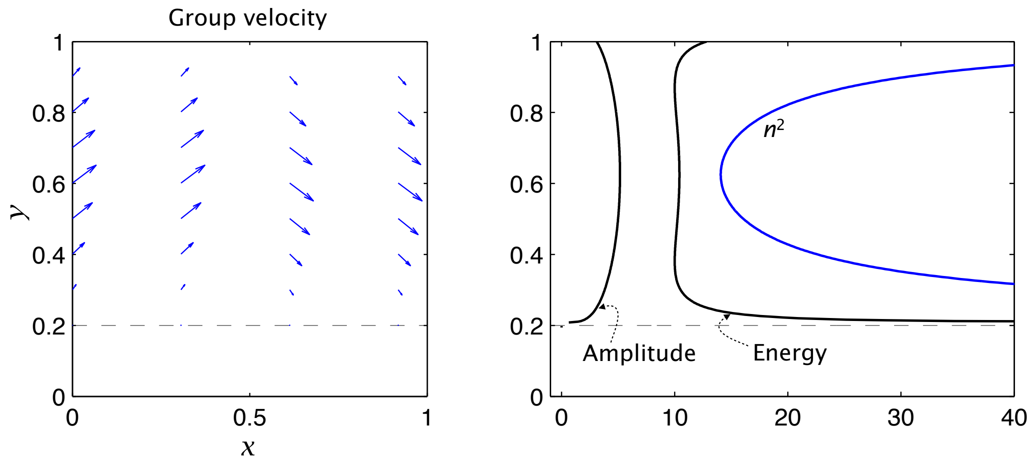


**Fig. 16.4** Left: The group velocity evaluated using (16.17) for the parameters illustrated in Fig. 16.1, which give a critical line at $y = 0.2$. For $x < 0.5$ we choose positive values of $n$, and a northward group velocity, whereas for $x > 0.5$ we choose negative values of $n$. Right panel: Values of refractive index squared, the energy and the amplitude of a wave. The value of $n^2$ becomes infinite at the critical line. See text for more description.

infinite at the critical line at $y = 0.2$. Using (16.20) the amplitude of the wave diminishes as the critical line approaches, but the energy increases rapidly, suggesting that the linear approximation will break down. The waves will actually stall before reaching the critical layer, because both the $x$ and the $y$ components of the group velocity become very small. Also, because the wavelength is so small we may expect the waves to break and deposit their momentum, but a full treatment of waves in the vicinity of a critical layer requires a nonlinear analysis.

The situation illustrated in this example is of particular relevance to the maintenance of the zonal wind structure in the troposphere. Waves are generated in midlatitude and propagate equatorward and on encountering a critical layer in the subtropics they break, deposit westward momentum and retard the flow, as the reader who braves the next section will discover explicitly.

## 16.3 ✦ ROSSBY WAVE ABSORPTION NEAR A CRITICAL LAYER

We noted in the last section that as a wave approaches a critical latitude the meridional wavenumber $l$ becomes very large, but the group velocity itself becomes small. These observations suggest that the effects of friction might become very large and that the wave would deposit its momentum, thereby accelerating or decelerating the mean flow, and if we are willing to make one or two approximations we can construct an explicit analytic model of this phenomena. Specifically, we will need to choose a simple form for the friction and assume that the background properties vary slowly, so that we can use a WKB approximation. Note that we have to include some form of dissipation, otherwise the Eliassen–Palm flux divergence is zero and there is no momentum deposition by the waves.

### 16.3.1 A model problem

Consider horizontally propagating Rossby waves obeying the linear barotropic vorticity equation on the beta-plane (vertically propagating waves may be considered using similar techniques). The equation of motion is

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\nabla^2\psi + \beta^*\frac{\partial \psi}{\partial x} = -r\nabla^2\psi, \tag{16.24}$$

where $\beta^* = \beta - \overline{u}_{yy}$. The parameter $r$ is a drag coefficient that acts directly on the relative vorticity. It is not a particularly realistic form of dissipation but its simplicity will serve our purpose well. We shall assume that $r$ is small compared to the Doppler-shifted frequency of the waves and seek solutions of the form

$$\psi'(x, y, t) = \widetilde{\psi}(y)\mathrm{e}^{\mathrm{i}(k(x-ct))}. \tag{16.25}$$

Substituting into (16.24) we find, after a couple of lines of algebra, that $\widetilde{\psi}$ satisfies, analogously to (16.15),

$$\frac{\partial^2\widetilde{\psi}}{\partial y^2} + l^2(y)\widetilde{\psi} = 0, \quad \text{where} \quad l^2(y) = \frac{\beta^*}{\overline{u} - c - \mathrm{i}r/k} - k^2. \tag{16.26a,b}$$

Evidently, as with the inviscid case, if the zonal wind has a lateral shear then $l$ is a function of $y$. However, $l$ now has an imaginary component so that the wave decays away from its source region. We can already see that if $\overline{u} = c$ the decay will be particularly strong.

### 16.3.2 WKB solution

Let us suppose that the zonal wavenumber is small compared to the meridional wavenumber $l$, which will certainly be the case approaching a critical layer. If $r \ll k(\overline{u}-c)$ then the meridional wavenumber is given by

$$l^2(y) \approx \left[\frac{\beta^*(\overline{u} - c + \mathrm{i}r/k)}{(\overline{u} - c)^2 + r^2/k^2}\right] \approx \frac{\beta^*}{\overline{u} - c}\left[1 + \frac{\mathrm{i}r}{k(\overline{u} - c)}\right] \tag{16.27}$$

whence

$$l(y) \approx \left(\frac{\beta^*}{\overline{u} - c}\right)^{1/2}\left[1 + \frac{\mathrm{i}r}{2k(\overline{u} - c)}\right]. \tag{16.28}$$

The streamfunction itself is then given by, in the WKB approximation,

$$\widetilde{\psi} = Al^{-1/2}\exp\left(\pm\mathrm{i}\int^y l\,\mathrm{d}y'\right). \tag{16.29}$$

just as in (16.16). But now the wave will decay as it moves away from its source and deposit momentum into the mean flow, as we now calculate.

The momentum flux, $F_k$, associated with the wave with $x$-wavenumber of $k$ is given by

$$F_k(y) = \overline{u'v'} = -\mathrm{i}k\left(\psi\frac{\partial\psi^*}{\partial y} - \psi^*\frac{\partial\psi}{\partial y}\right), \tag{16.30}$$

and using (16.28) and (16.29) in (16.30) we obtain

$$F_k(y) = F_0 \exp\left(\pm \mathrm{i} \int_0^y (l - l^*)\, \mathrm{d}y'\right) = F_0 \exp\left(\int_0^y \frac{\pm r \beta^{*1/2}}{k(\overline{u} - c)^{3/2}}\, \mathrm{d}y'\right). \tag{16.31}$$

In deriving this expression we use that fact that the amplitude of $\widetilde{\psi}$ (i.e., $l^{-1/2}$) varies only slowly with $y$ so that when calculating $\partial\widetilde{\psi}/\partial y$ its derivative may be ignored. In (16.31) $F_0$ is the value of the flux at $y = 0$ and the sign of the exponent must be chosen so that the group velocity is directed away from the wave source region. Clearly, if $r = 0$ then the momentum flux is constant.

The integrand in (16.31) is the attenuation rate of the wave and it has a straightforward physical interpretation. Using the real part of (16.28) in (16.17b), and assuming $|l| \gg |k|$, the meridional component of the group velocity is given by

$$c_g^y = \frac{2kl\,\beta^*}{(k^2 + l^2)^2} \approx \frac{2k\,\beta^*}{l^3} = \frac{2k(\overline{u} - c)^{3/2}}{\beta^{*1/2}}. \tag{16.32a,b}$$

Thus we have

$$\text{Wave attenuation rate} = \frac{r\beta^{*1/2}}{k(\overline{u} - c)^{3/2}} = \frac{2 \times \text{Dissipation rate, } 2r}{\text{Meridional group velocity, } c_g^y}. \tag{16.33}$$

As the group velocity diminishes the dissipation has more time to act and so the wave is preferentially attenuated. We give a further interpretation of this result in the next subsection.

How does this attenuation affect the mean flow? The mean flow is subject to many waves and so obeys the equation

$$\frac{\partial \overline{u}}{\partial t} = -\sum_k \frac{\partial F_k}{\partial y} + \text{viscous terms}. \tag{16.34}$$

Because the amplitude varies only slowly compared to the phase, the amplitude of $\partial F_k/\partial y$ varies mainly with the attenuation rate (16.33) and is largest near a critical layer. Consider a Rossby wave propagating away from some source region with a given frequency and $x$-wavenumber. Because $k$ is negative a Rossby wave always carries westward (or negative) momentum with it. That is, $F_k$ is always negative and increases (becomes more positive) as the wave is attenuated; that is to say, if $r \neq 0$ then $\partial F_k/\partial y$ is positive and from (16.34) the mean flow is accelerated *westward* as the wave dissipates. This acceleration will be particularly strong if the wave approaches a critical layer where $\overline{u} = c$. Indeed, such a situation arises when Rossby waves, generated in mid-latitudes, propagate equatorward. As the waves enter the subtropics $\overline{u} - c$ becomes smaller and the waves dissipate, producing a westward force on the mean flow, even though a true critical layer may never be reached. Globally, momentum is conserved because there is an equal and opposite (and therefore eastward) wave force at the wave source producing an eddy-driven jet, as discussed in the previous chapter.

### 16.3.3 Interpretation using wave activity

We can derive and interpret the above results by thinking about the propagation of wave activity. For barotropic Rossby waves, multiply (16.24) by $\zeta/\beta^*$ and zonally average to obtain the wave activity equation,

$$\frac{\partial \mathcal{A}}{\partial t} + \frac{\partial \mathcal{F}}{\partial y} = -\alpha \mathcal{A}, \tag{16.35}$$

where $\mathcal{A} = \overline{\zeta'^2}/2\beta^*$ is the wave activity density, $\partial \mathcal{F}/\partial y = \overline{v'\zeta'}$ is its flux divergence, and $\alpha = 2r$. Referring as needed to the discussion in sections 10.2.1 and 10.2.2, the flux obeys the group velocity property so that

$$\frac{\partial \mathcal{A}}{\partial t} + \frac{\partial}{\partial y}(\boldsymbol{c}_g \mathcal{A}) = -\alpha \mathcal{A}. \tag{16.36}$$

Let us suppose that the wave is in a statistical steady state and that the spatial variation of the group velocity occurs on a longer spatial scale than the variations in wave activity density, consistent with the WKB approximation. We then have

$$c_g^y \frac{\partial \mathcal{A}}{\partial y} = -\alpha \mathcal{A}. \tag{16.37}$$

which integrates to give

$$\mathcal{A}(y) = \mathcal{A}_0 \exp\left( - \int^y \frac{\alpha}{c_g^y} \,\mathrm{d}y' \right). \tag{16.38}$$

That is, the attenuation rate of the wave activity is the dissipation rate of wave activity divided by the group velocity, as in (16.31) and (16.33) (note that $\alpha = 2r$). The wave-activity method of derivation suggests that this result is a general one, not restricted to Rossby waves, and indeed in section 17.4.2 we will find that the attenuation rate of vertically propagating gravity waves is given by essentially the same expression.

The divergence of wave activity will lead to a force on the mean zonal flow, much as discussed in section 15.1. For definiteness, suppose that waves propagate away from a mid-latitude source in the Northern Hemisphere. South of the source $c_g^y$ is negative and north of the source $c_g^y$ is postive. In either case, from (16.38) the wave activity density decreases away from the source and, with reference to (15.35a), the ensuing force on the mean flow is negative, or westward.

## 16.4 VERTICAL PROPAGATION OF ROSSBY WAVES

We now consider in more detail the vertical propagation of Rossby waves, forced by bottom topography, in a stratified atmosphere.[1] The vertical propagation is important not just because it must be taken into account to obtain an accurate picture of the tropospheric response to topographic and thermal forcing, but because it can excite motion in the stratosphere, as considered in chapter 17. We will continue to use the stratified quasi-geostrophic equations, but we now allow the model to be compressible and semi-infinite, extending from $z = 0$ to $z = \infty$.

Let us now consider how Rossby waves propagate in an inhomogeneous, stratified medium. It is simplest to first consider the problem slightly generally, without regard to boundary conditions, for this reveals some of the essential conditions under which Rossby waves propagate. In section 16.5 we will be a little more definite and consider the lower boundary conditions and the requirements for waves propagate vertically, possibly into the stratosphere. Our governing equation is the quasi-geostrophic potential vorticity equation and with applications to the stratosphere in mind we will use log-pressure co-ordinates so that the equation of motion is

$$\frac{\partial q}{\partial t} + J(\psi, q) = 0, \qquad q = \nabla^2\psi + \beta y + \frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \psi}{\partial z}\right), \tag{16.39}$$

where $\rho_R = \rho_0 \mathrm{e}^{-z/H}$ with $H$ being a specified density scale height, typically $RT(0)/g$.

### 16.4.1 Conditions for wave propagation

Let us linearize (16.39) about a zonal wind that depends only on $z$; that is, we let

$$\psi = -\overline{u}(z)y + \psi', \tag{16.40}$$

and obtain

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = 0, \qquad \frac{\partial \overline{q}}{\partial y} = \beta - \frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \overline{u}}{\partial z}\right), \tag{16.41}$$

or equivalently, in terms of streamfunction,

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\left[\nabla^2\psi' + \frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \psi'}{\partial z}\right)\right]& \\ + \frac{\partial \psi'}{\partial x}\left[\beta - \frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \overline{u}}{\partial z}\right)\right] &= 0.\end{aligned} \tag{16.42}$$

The first term in square brackets is the perturbation potential vorticity, $q'$ and the second term equals $\partial\overline{q}/\partial y$.

We may seek solutions to (16.42) of the form

$$\psi' = \mathrm{Re}\,\widetilde{\psi}(z)\mathrm{e}^{\mathrm{i}(kx+ly-kct)}. \tag{16.43}$$

Solutions of (16.42) then satisfy

$$\left[\frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \widetilde{\psi}}{\partial z}\right)\right] = \widetilde{\psi}\left(K^2 - \frac{\partial\overline{q}/\partial y}{\overline{u} - c}\right) \tag{16.44}$$

Let us simplify by assuming that both $\overline{u}$ and $N^2$ are constants, so that $\partial\overline{q}/\partial y = \beta$. Eq. (16.44) further simplifies if we define

$$\Phi(z) = \widetilde{\psi}(z)\left(\frac{\rho_R}{\rho_R(0)}\right)^{1/2} = \widetilde{\psi}(z)\mathrm{e}^{-z/2H} \tag{16.45}$$

whence we obtain

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2} + m^2\Phi = 0, \qquad \text{where} \tag{16.46a,b}$$

where

$$m^2 = \frac{N^2}{f_0^2}\left(\frac{\beta}{\overline{u} - c} - K^2 - \gamma^2\right), \tag{16.47}$$

where $\gamma^2 = f_0^2/(4N^2H^2) = 1/(2L_d)^2$, where $L_d$ is the deformation radius as sometimes defined (i.e., $L_d = NH/f_0$). If the parameters on the right-hand side of (16.47) are constant then so is $m$ and (16.46) has solutions of the form $\Phi(z) = \Phi_0 e^{imz}$ so that the streamfunction itself varies as

$$\psi' = \mathrm{Re}\,\Phi_0 e^{i(kx+ly+mz-kct)+z/2H}. \tag{16.48}$$

In the (more realistic) case in which $m$ varies with height then, if the variation is slow enough, the solution looks locally like a plane wave and WKB techniques may be used to find a solution, as we discuss further in section 16.6. But even then essentially the same conditions for propagation apply, so for now let us suppose $m$ is constant.

For waves to propagate upwards we require that $m^2 > 0$ and, from (16.47), that

$$\boxed{0 < \overline{u} - c < \frac{\beta}{K^2 + \gamma^2} \equiv u_c}, \tag{16.49}$$

where $u_c$ is the Rossby critical velocity. For waves of some given frequency ($\omega = kc$) the above expression provides a condition on $\overline{u}$ for the vertical propagation of planetary waves. For stationary waves $c = 0$ and the criterion is

$$0 < \overline{u} < u_c. \tag{16.50}$$

That is to say, the vertical propagation of stationary Rossby waves occurs only in westerly winds, and winds that are weaker than some critical magnitude that depends on the scale of the wave. We return to this condition in section 16.5. If the waves can take any frequency there is no such condition on $\overline{u}$, for (16.47) is just a form of the dispersion relation and (16.49) is naturally satisfied.

### 16.4.2 Dispersion relation and group velocity

Noting that $\omega = ck$ and rearranging (16.47) we obtain the dispersion relation for three-dimensional Rossby waves

$$\omega = \overline{u}k - \frac{\beta k}{K^2 + \gamma^2 + m^2 f_0^2/N^2}. \tag{16.51}$$

The three components of the group velocity for these waves are then:

$$c_g^x = \overline{u} + \frac{\beta[k^2 - (l^2 + m^2 f_0^2/N^2 + \gamma^2)]}{(K^2 + m^2 f_0^2/N^2 + \gamma^2)^2}, \tag{16.52a}$$

$$c_g^y = \frac{2\beta kl}{(K^2 + m^2 f_0^2/N^2 + \gamma^2)^2}, \qquad c_g^z = \frac{2\beta km f_0^2/N^2}{(K^2 + m^2 f_0^2/N^2 + \gamma^2)^2}. \tag{16.52b,c}$$

The propagation in the horizontal is analogous to the propagation in a shallow water model [c.f. (6.64b)]; note also that higher baroclinic modes (bigger $m$) will have a more westward group velocity. The vertical group velocity is proportional to $m$, and for waves that propagate signals upward we must choose $m$ to have the same sign as $k$ so that $c_g^z$ is positive. If there is no mean flow then the zonal wavenumber $k$ is negative (in order that frequency is positive) and $m$ must then also be negative. Energy then propagates upward but the phase propagates downward.

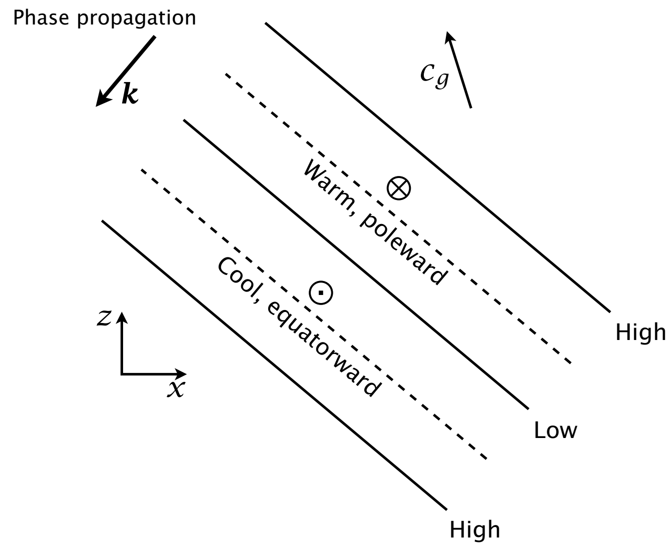


**Fig. 16.5** A schematic east-west section of an upwardly propagating Rossby wave. The slanting lines are lines of constant phase and 'high' and 'low' refer to the pressure or streamfunction values. Both $k$ and $m$ are negative so the phase lines are oriented up and to the west. The phase propagates westward and downward, but the group velocity is upward.

## 16.5 ROSSBY WAVES EXCITED AT THE LOWER BOUNDARY

We now derive some explicit solutions for Rossby waves excited at a lower boundary by topography. Rossby waves may also be excited by thermal anomalies at the lower boundary, although in Earth's atmosphere their amplitude is somewhat smaller, and the treatment of such waves is similar in may ways but left for the reader to explore elsewhere.[2]

### 16.5.1 Lower boundary conditions

The surface boundary condition of vertical velocity is determined by the thermodynamic equation and the upper boundary condition is determined by a radiation condition.

The lower boundary is obtained using the thermodynamic equation,

$$\frac{\partial}{\partial t}\left(\frac{\partial \psi}{\partial z}\right) + J\left(\psi, \frac{\partial \psi}{\partial z}\right) + \frac{N^2}{f_0} w = 0, \tag{16.53}$$

along with an equation for the vertical velocity, $w$, at the lower boundary. This is

$$w = \boldsymbol{u} \cdot \nabla h_b + r\zeta, \tag{16.54}$$

where the two terms respectively represent the kinematic contribution to vertical velocity due to flow over topography and the contribution from Ekman pumping, with $r$ a constant, and the effects are taken to be additive. Linearizing the thermodynamic equation about the zonal flow and using (16.54) gives

$$\frac{\partial}{\partial t}\left(\frac{\partial \psi'}{\partial z}\right) + \overline{u}\frac{\partial}{\partial x}\frac{\partial \psi'}{\partial z} - v'\frac{\partial \overline{u}}{\partial z} = -\frac{N^2}{f_0}\left(\overline{u}\frac{\partial h_b}{\partial x} + r\nabla^2\psi'\right), \qquad \text{at } z = 0. \tag{16.55}$$

### 16.5.2 Model solution

We look for solutions of (16.41) and (16.55) in the form

$$\psi' = \operatorname{Re}\widetilde{\psi}(z)\sin ly\,\mathrm{e}^{\mathrm{i}k(x-ct)} \qquad \text{with} \qquad h_b = \operatorname{Re}\widetilde{h}_b \sin ly\,\mathrm{e}^{\mathrm{i}kx} \tag{16.56}$$

Solutions must then satisfy

$$\left[\frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial\widetilde{\psi}}{\partial z}\right)\right] = \widetilde{\psi}\left(K^2 - \frac{\partial\overline{q}/\partial y}{\overline{u}-c}\right) \tag{16.57}$$

in the interior, and the boundary condition

$$(\overline{u}-c)\frac{\partial\widetilde{\psi}}{\partial z} - \widetilde{\psi}\frac{\partial\overline{u}}{\partial z} + \frac{\mathrm{i}rN^2K^2}{kf_0}\widetilde{\psi} = -\frac{N^2\overline{u}\widetilde{h}_b}{f_0}, \qquad \text{at } z = 0, \tag{16.58}$$

as well as a radiation condition at plus infinity (and we must have that $\rho_0\widetilde{\psi}^2$ be finite). Let us simplify by considering the case of constant $\overline{u}$ and $N^2$ and with $r = 0$. As before we let $\Phi(z) = \widetilde{\psi}(z)\exp(-z/2H)$ and obtain the interior equation

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2} + m^2\Phi = 0, \qquad \text{where} \quad m^2 = \frac{N^2}{f_0^2}\left(\frac{\beta}{\overline{u}-c} - K^2 - \gamma^2\right), \tag{16.59a,b}$$

and $\gamma^2 = f_0^2/(4N^2H^2) = 1/(2L_d)^2$, where $L_d$ is the deformation radius. The surface boundary condition is

$$(\overline{u}-c)\left(\frac{\mathrm{d}\Phi}{\mathrm{d}z} + \frac{\Phi}{2H}\right) = -\frac{N^2\overline{u}\widetilde{h}_b}{f_0}, \qquad \text{at } z = 0. \tag{16.60}$$

*Stationary waves*

Stationary waves have $\omega = ck = 0$. In this case (16.59) has a solution $\Phi = \Phi_0\exp(\mathrm{i}mz)$ provided $m^2$ is positive where

$$m = \pm\frac{N}{f_0}\left(\frac{\beta}{\overline{u}} - K^2 - \gamma^2\right)^{1/2}. \tag{16.61}$$

We must choose the sign of $m$ to ensure that the group velocity, and hence the wave activity, is directed away from the energy source, and if $k < 0$ then $m$ must be negative.

The condition $m^2 > 0$ holds if

$$\boxed{0 < \overline{u} < \frac{\beta}{K^2+\gamma^2}}, \tag{16.62}$$

and this is illustrated in Fig. 16.6. Stationary, vertically oscillatory modes can exist only for zonal flows that are eastwards and that are less than the critical velocity $U_c = \beta/(K^2+\gamma^2)$. One way to interpret this condition is note that in a resting medium the Rossby wave frequency has a minimum value (and maximum absolute value), when $m = 0$, of

$$\omega = -\frac{\beta k}{K^2+\gamma^2}. \tag{16.63}$$

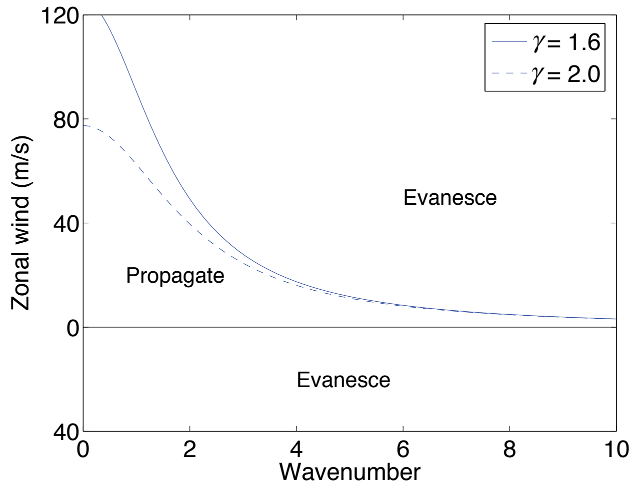


**Fig. 16.6** The boundary between propagating waves and evanescent waves as a function of zonal wind and wavenumber, using (16.62), for a couple of values of $\gamma$. With $N = 2 \times 10^{-2}\,\mathrm{s}^{-1}$, $\gamma = 1.6$ ($\gamma = 2$) corresponds to a scale height of 7.0 km (5.5 km) and a deformation radius $NH/f$ of 1400 km (1100 km).

Note too that in a frame moving with speed $\overline{u}$ our Rossby waves (stationary in the Earth's frame) have frequency $-\overline{u}k$, and this is the forcing frequency arising from the now-moving bottom topography. Thus, (16.62) is equivalent to saying that for oscillatory waves to exist *the forcing frequency must lie within the frequency range of vertically propagating Rossby waves.*

For westward flow, or for sufficiently strong eastward flow, the waves decay exponentially as $\Phi = \Phi_0 \exp(-\alpha z)$ where

$$\alpha = \frac{N}{f_0}\left(K^2 + \gamma^2 - \frac{\beta}{\overline{u}}\right)^{1/2}. \tag{16.64}$$

Note that the critical velocity $u_c = (\beta/K^2 + \gamma^2)$ is a function of wavenumber, and that it increases with horizontal wavelength. Thus, for a given eastward flow long waves may penetrate vertically when short waves are trapped, an effect sometimes referred to as 'Charney–Drazin filtering'. One important consequence of this is that the stratospheric motion is typically of larger scales than that of the troposphere, because waves tend to be excited first in the troposphere (by baroclinic instability and by flow over topography, among other things), but the shorter waves are trapped and only the longer ones reach the stratosphere. In the summer, the stratospheric winds are often westwards and all waves are trapped in the troposphere; the eastward stratospheric winds that favour vertical penetration occur in the other three seasons, although very strong eastward winds can suppress penetration in mid-winter.

Finally, the surface boundary condition, (16.60) gives

$$\Phi_0 = \frac{N^2\widetilde{h}_b/f_0}{(\alpha, -\mathrm{i}m) - (2H)^{-1}}, \tag{16.65}$$

where $(\alpha, -\mathrm{i}m)$ refers to the (trapped, oscillatory) case. Equation (16.65) indicates that resonance is possible when $\alpha = 1/(2H)$, and from (16.64) this occurs when $K^2 = \beta/\overline{u}$, that is when barotropic Rossby waves are stationary. This wave resonates because the wave is a solution of the unforced (and inviscid) equations and, because $\widetilde{\psi} = \Phi \exp[z/(2H)]$, $\widetilde{\psi} = 1$ and has uniform vertical structure. If $K > K_s$ then $\alpha > 1/(2H)$ and the forced wave (i.e., the amplitude of $\psi$) decays with height with no phase variation. If $\alpha < 1/(2H)$ then $\widetilde{\psi}$ increases with height

## Stationary, Topographically Forced Solutions

Collecting the results in section 16.5.2, the stationary solutions of (16.41) and (16.55) are [check equation numbers xxx]:

$$\psi'(x,y,z) = \mathrm{Re}\ \mathrm{e}^{\mathrm{i}mz}\mathrm{e}^{z/2H}\mathrm{e}^{\mathrm{i}kx}\sin ly \frac{f_0\widetilde{h}_b\left[\mathrm{i}m - (2H)^{-1}\right]}{K_s^2 - K^2}, \qquad m^2 > 0 \tag{T.1a}$$

$$\psi'(x,y,z) = \mathrm{Re}\ \mathrm{e}^{[(2H)^{-1}-\alpha]z}\mathrm{e}^{\mathrm{i}kx}\sin ly \frac{N^2\widetilde{h}_b}{f_0\left[\alpha - (2H)^{-1}\right]}, \qquad m^2 < 0 \tag{T.1b}$$

where

$$m = \pm\frac{N}{f_0}\left(\frac{\beta}{\overline{u}} - K^2 - \gamma^2\right)^{1/2} \tag{T.2}$$

and

$$\alpha = +\frac{N}{f_0}\left(K^2 + \gamma^2 - \frac{\beta}{\overline{u}}\right)^{1/2}, \tag{T.3}$$

and $\gamma = f_0/(2NH)$. If $m^2 > 0$ the solutions are propagating, or radiating, waves in the vertical direction. If $m^2 < 0$ the energy of the solution, $|\rho_R\psi'^2|$, is vertically evanescent. The condition $m^2 > 0$ is equivalent to

$$0 < \overline{u} < \frac{\beta}{K^2 + (f_0/2NH)^2}, \tag{T.4}$$

so that vertical penetration is favoured when the winds are weakly eastwards, and the range of $\overline{u}$-values that allows this is larger for longer waves.

In order that the energy propagate upwards the vertical component of the group velocity must be positive, and hence $k$ and $m$ must have the same sign.

(although $\rho_R|\widetilde{\psi}|^2$ decreases with height), and this occurs when $(K_s^2 - \gamma^2)^{1/2} < K < K_s$. If $(K_s^2 - \gamma^2)^{1/2} > K$ then the amplitude of $\phi$, (i.e., $\rho_R|\widetilde{\psi}|^2$) is independent of height; their vertical structure is oscillatory, like $\exp(\mathrm{i}mz)$. The solutions are collected for convenience in the box above.

### 16.5.3 More properties of the solution

The various dynamical fields associated with the solution can all be easily constructed from (T.1), and a few simple properties of the solution are worth noting explicitly. In some cases the explicit calculation is left as a problem to the reader — see problems 16.6 and 16.7.

*Polarization relations.* The polarization relations are the amplitude phase relations between the various fields. These are [xxx]

*Amplitudes and phases.* The decaying solutions have no vertical phase variations (a property known as 'equivalent barotropic') and the streamfunction is exactly in phase or out of phase with the topography according as $K > K_s$ and $\alpha > (2H)^{-1}$, or $K < K_s$ and $\alpha < (2H)^{-1}$. In the latter case the amplitude of the streamfunction actually increases with height, but the energy, proportional to $\rho_R|\psi'^2|$ falls. The oscillatory solutions have (if there is no shear) constant energy with height but a shifting phase. The phase of the streamfunction at the surface may be in or out of phase with the topography, depending on $m$, but the potential temperature, $\partial\psi/\partial z$ is always out of phase with the topography. That is, positive values of $h_b$ are associated with cool fluid parcels.

*Vertical energy propagation.* As noted, the energy propagates upwards for the oscillatory waves. This may be verified by calculating $\overline{p'w'}$ (the vertical component of the energy flux), where $p'$ is the pressure perturbation, proportional to $\psi'$, and $w'$ is the vertical velocity perturbation. To this end, linearize the thermodynamic equation (16.53) to give

$$\frac{\partial}{\partial t}\left(\frac{\partial \psi'}{\partial z}\right) + \overline{u}\frac{\partial}{\partial x}\frac{\partial \psi'}{\partial z} - \frac{\partial \overline{u}}{\partial z}\frac{\partial \psi'}{\partial x} + \frac{N^2}{f_0}w' = 0. \tag{16.66}$$

Then, multiplying by $\psi'$ and integrating by parts gives a balance between the second and fourth terms,

$$N^2\overline{\psi' w'} = \overline{u}\overline{b'v'}, \tag{16.67}$$

where $b' = f_0\partial\psi'/\partial z$ and $v' = \partial\psi'/\partial x$. Thus, the upward transfer of energy is proportional to the poleward heat flux. Evidently, the transfer of energy is upward when $km > 0$, and from (16.52), this corresponds to the condition that the vertical component of group velocity is positive, which has to be the case from general arguments. For Rossby waves $k < 0$ so that upward energy propagation requires $m < 0$ and therefore downward phase propagation.

*Meridional heat transport.* The meridional heat transport associated with a wave is

$$\rho_R\overline{v'b'} = \rho_R f_0 \overline{\frac{\partial \psi'}{\partial x}\frac{\partial \psi'}{\partial z}}. \tag{16.68}$$

For an oscillatory wave this can readily be shown to be positive. In particular, it is proportional to $km/(K_s^2 - K^2)$, and this is positive because $km > 0$ is the condition that energy is directed upwards and $K_s^2 > K^2$ for oscillatory solutions. The meridional transport associated with a trapped solution is identically zero.

*Form drag.* If the waves propagate energy upwards, there must be a surface interaction to supply that energy. There is a force due to *form drag* associated with this interaction, given by

$$\text{form drag} = \overline{p'\frac{\partial h_b}{\partial x}} \tag{16.69}$$

(see chapter 3). In the trapped case, the streamfunction is either exactly in or out of phase with the topography, so this interaction is zero. In the oscillatory case

$$\overline{\psi'\frac{\partial h_b}{\partial x}} = \frac{f_0\widetilde{h}_b^2 km}{4(K_s^2 - K^2)}, \tag{16.70}$$

where the factor of 4 arises from the $x$ and $y$ averages of the squares of sines and cosines. The rate of doing work is $\overline{u}$ times (16.70).

## 16.6 ✦ VERTICAL PROPAGATION OF ROSSBY WAVES IN SHEAR

In the real atmosphere the zonal wind and the stratification change with height and there may be regions in which propagation occurs and regions where it does not, and in this section we illustrate that phenomena with two examples. In one example the zonal wind increases sufficiently with height that wave propagation ceases because the wind is too strong, and in the other the zonal wind decreases aloft and becomes negative (westward), again causing wave propagation to cease. If the zonal wind and the stratification both vary sufficiently slowly with height — meaning that the scale of the variation is much greater than a vertical wavelength — then *locally* the solution will look like a plane wave and the analysis is straightforward, very similar to that performed in sections 7.5 (where we looked at internal waves with varying stratification) and section 16.2.2 (where we looked at horizontally propagating Rossby waves).

For simplicity we consider Rossby waves in a flow with vertical shear but no horizontal shear, with constant stratification and constant density. With reference to section 16.1.2, the equation of motion is

$$\left(\frac{\partial}{\partial t} + \overline{u}(z)\frac{\partial}{\partial x}\right) q' + \beta v' = 0. \tag{16.71}$$

We seek solutions of the form

$$\psi' = \widetilde{\psi}(z)\mathrm{e}^{\mathrm{i}k(x-ct)+ly}, \tag{16.72}$$

obtaining

$$\frac{\partial^2\widetilde{\psi}}{\partial z^2} + m^2(z)\widetilde{\psi} = 0 \tag{16.73a}$$

where

$$m^2(z) = \frac{N^2}{f_0^2}\left[\frac{\beta}{\overline{u}-c} - (k^2+l^2)\right]. \tag{16.73b}$$

The WKB solution to (16.73) (see appendix to chapter 7) is

$$\widetilde{\psi}(z) = Am^{-1/2}\,\mathrm{e}^{\mathrm{i}\int m\,\mathrm{d}z}. \tag{16.74}$$

where $A$ is a constant. The local vertical wavenumber is just $m$ itself (for this is the derivative of the phase), and the amplitude varies like $m^{-1/2}$. This variation of amplitude is consistent with the conservation of wave activity, which in this case means that the Eliassen–Palm flux is constant. As there is no horizontal divergence in this problem, the constancy of $\boldsymbol{\mathcal{F}}$ in (16.19) implies $\partial_z\overline{v'b'} = 0$ and therefore

$$km|\widetilde{\psi}|^2 = \text{constant}. \tag{16.75}$$

Since the horizontal wavenumber is constant the dependence of the amplitude on $m^{-1/2}$ immediately follows. The energy of the wave is not constant unless there is no shear, since it may be extracted or given up to the mean flow.

As discussed in earlier sections wave propagation requires than $m^2$ be positive. For stationary waves ($c = 0$) this gives the condition that $0 < \overline{u} < \beta/(k^2+l^2)$ At the lower bound there is a critical layer and $m^2 \to \infty$. At the upper bound $m^2 = 0$ and this is a turning layer. Let us illustrate the behaviour in these regions with two examples.

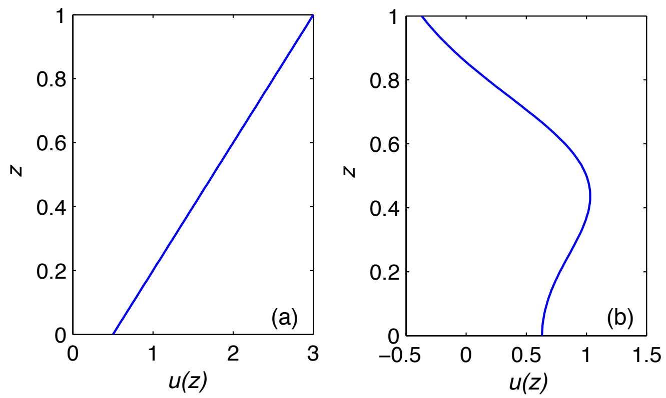


**Fig. 16.7** Two profiles of non-dimensional zonal wind used in the calculations illustrated in Fig. 16.8 and Fig. 16.9. (a) is a uniform shear that gives rise to a turning latitude, and (b) shows a profile in which the zonal wind diminishes to zero aloft, giving rise to a critical layer.

### 16.6.1 Two examples

*Waves with a turning layer*

Consider Rossby waves propagating in a background state in which the zonal wind increases uniformly with height, as in Fig. 16.7a, but in which all other parameters are constant. Specifically, we choose (nondimensional) values of $\beta = 5$, $k = l = 1$ and $c = 0$ (the reader may re-dimensionalize). We also scale the vertical coordinate so that $Pr = 1$. For the profile chosen $m^2$ is positive for $\overline{u} < 2.5$ and so for $0 < z < 0.8$, as shown in Fig. 16.8. For $l$ fixed and $m$ given by (16.73b) we calculate the group velocity using (16.7) and these are displayed in Fig. 16.8. We choose upwardly propagating waves (i.e. $m > 0$),; in any physical situation the group velocity will be directed away from the source, and we are assuming this occurs at the surface. We also show equatorward moving waves for $y < 0.5$ and poleward moving waves for $y > 0.5$, but this is more for illustrative purposes. The right-hand panel of the figure shows the value of $m^2$ diminshing with height, along with the vertical profiles of the amplitude (which goes like $m^{-1/2}$) and the energy (which goes like $(k^2 + l^2 + m^2)m$).

We see from Fig. 16.8 that the group velocity turns away from the turning line, and we can understand this from the ratio of the group velocities given in (16.7), namely

$$\frac{c_g^z}{c_g^y} = \frac{Pr^2 m}{l}. \tag{16.76}$$

The group velocity is purely horizontal at the turning line. The amplitude of the waves is infinite, but the waves do not necessarily break because the vertical wavelength is very large.

*Waves with a critical layer*

Now consider waves in a zonal wind that initially increases with height and then decreases and becomes negative, as illustrated in Fig. 16.7b. There is a critical layer where $\overline{u}$ passes through zero, but all the other parameters are the same as in the previous example. The value of $m^2$ now generally increases with height, as illustrated in the right-hand panel of Fig. 16.9, becoming infinite at the critical layer and negative above it. The amplitude of the wave, being proportional to $m^{-1/2}$ actually goes to zero at the critical layer but the energy increases without bound.

The group velocity, shown in the left panel of Fig. 16.8, turns upward and toward critical layer and, from (16.76), is purely vertical at the critical layer. The trajectory of a wave following

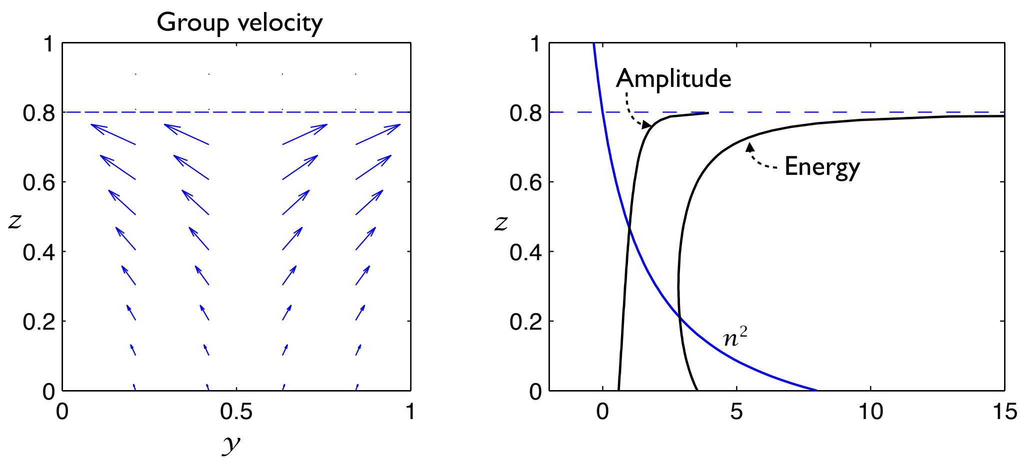


**Fig. 16.8** Left panel: group velocity vectors calculated using (16.7) for the parameters shown in Fig. 16.7a, assuming a source at $y = 0.5$. Right: profiles of $m^2$, wave amplitude and energy. The horizontal line at $z = 0.8$ marks a turning surface, and the group velocity turns away from it.

a ray approaching a critical layer can be calculated in a similar fashion to that of a gravity wave in section 7.5.2, but this is left as an exercise.

## 16.7 FORCED AND STATIONARY ROSSBY WAVES

We now turn our attention to understanding the large-scale zonally asymmetric circulation of the atmosphere, much of which is determining by the presence of stationary Rossby waves forced by topographic and thermal anomalies at the surface.[3]

### 16.7.1 A simple one-layer case

Many of the essential ideas can be illustrated by a one-layer quasi-geostrophic model, with potential vorticity equation

$$\frac{Dq}{Dt} = 0, \qquad q = \zeta + \beta y - \frac{f_0}{H}(\eta - h_b), \tag{16.77}$$

where $H$ is the mean thickness of the layer, $\eta$ is the height of the free surface, $h_b$ is the bottom topography, and the velocity and vorticity are given by $\boldsymbol{u} = (g/f_0)\nabla^\perp\eta \equiv (g/f_0)\mathbf{k}\times\nabla\eta$ and $\zeta = (\partial v/\partial x - \partial u/\partial y) = (g/f_0)\nabla^2\eta$. Linearizing (16.77) about a flat-bottomed state with zonal flow $\overline{u}(y) = -(g/f_0)\partial\overline{\eta}/\partial y$ gives

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = 0, \tag{16.78}$$

where $q' = \zeta' - (f_0/H)(\eta' - h_b)$ and $\partial\overline{q}/\partial y = \beta + \overline{u}/L_d^2$ with $L_d = \sqrt{gH}/f_0$, the radius of deformation. Eq. (16.78) may be written, after the cancellation of a term proportional to

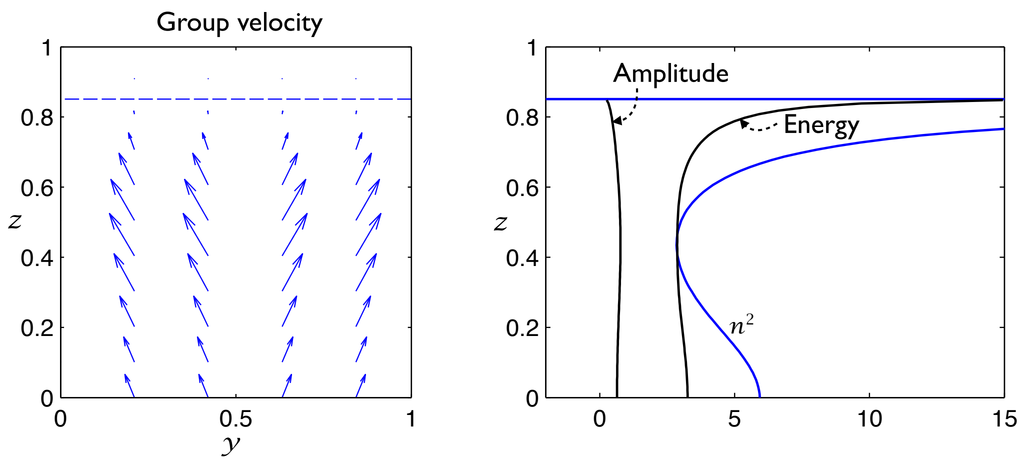


**Fig. 16.9** Left panel: group velocity vectors calculated using (16.7) for the parameters shown in Fig. 16.7b. Right: profiles of $m^2$, wave amplitude and energy. The horizontal line at $z \approx 0.85$ marks a critical surface; the group velocity turns toward it but its amplitude diminishes as the critical surface is approached.

$\overline{u}\partial\eta'/\partial x$, as

$$\frac{\partial}{\partial t}\left(\zeta' - \frac{\psi'}{L_d^2}\right) + \overline{u}\frac{\partial \zeta'}{\partial x} + \beta v' = -\overline{u}\frac{\partial \widehat{h}}{\partial x}, \tag{16.79}$$

where $\psi' = (g/f_0)\eta'$ and $\widehat{h} = h_b f_0/H = h_b g/(L_d^2 f_0)$.

The solution of this equation consists of the solution to the homogeneous problem (with the right-hand side equal to zero, as considered in section 6.4 on Rossby waves) and the particular solution. We proceed by decomposing the variables into their Fourier components

$$(\zeta', \psi', \widehat{h}) = \mathrm{Re}\,(\tilde{\zeta}, \widetilde{\psi}, \widetilde{h}_b) \sin ly\, \mathrm{e}^{\mathrm{i}kx}. \tag{16.80}$$

where such decomposition is appropriate for a channel, periodic in the $x$-direction and with no variation at the meridional boundaries, $y = (0, L)$. The full solution will be a superposition of such Fourier modes and, because the problem is linear, these modes do not interact. The free Rossby waves, the solution to the homogeneous problem, evolve according to

$$\psi = \mathrm{Re}\,\widetilde{\psi} \sin ly\, \mathrm{e}^{\mathrm{i}(kx-\omega t)}, \tag{16.81}$$

where $\omega$ is given by the dispersion relation [cf. (6.63)]

$$\omega = k\overline{u} - \frac{k\partial_y \overline{q}}{K^2 + k_d^2} = \frac{k(\overline{u}K^2 - \beta)}{K^2 + k_d^2}, \tag{16.82a,b}$$

where $K^2 = k^2 + l^2$ and $k_d = 1/L_d$. Stationary waves occur at the wavenumbers for which $K = K_S \equiv \sqrt{\beta/\overline{u}}$. To the free waves we add the solution to the steady problem,

$$\overline{u}\frac{\partial \zeta'}{\partial x} + \beta v' = -\overline{u}\frac{\partial \widehat{h}}{\partial x}, \tag{16.83}$$

which is, using the notation of (16.80)

$$\widetilde{\psi} = \frac{\widetilde{h}_b}{(K^2 - K_s^2)}. \tag{16.84}$$

Now, $\widetilde{h}_b$ is a complex amplitude; thus, for $K > K_s$ the streamfunction response is *in phase* with the topography. For $K^2 \gg K_s^2$ the steady equation of motion is

$$\overline{u}\frac{\partial \zeta'}{\partial x} \approx -\overline{u}\frac{\partial \widehat{h}}{\partial x}, \tag{16.85}$$

and the topographic vorticity source is balanced by zonal advection of relative vorticity. For $K^2 < K_s^2$ the streamfunction response is *out of phase* with the topography, and the dominant balance for very large scales is between the meridional advection of planetary vorticity, $v\partial f/\partial y$ or $\beta v$, and the topographic source. For $K = K_s$ the response is infinite, with the stationary wave resonating with the topography. Now, any realistic topography can be expected to have contributions from *all* Fourier components. Thus, for *any* given zonal wind there will be a resonant wavenumber and an infinite response. This, of course, is not observed, and one reason is that the real system contains friction. The simplest way to include this is by adding a linear damping to the right-hand side of (16.79), giving

$$\frac{\partial}{\partial t}\left(\zeta' - \frac{\psi'}{L_d^2}\right) + \overline{u}\frac{\partial \zeta'}{\partial x} + \beta v' = -r\zeta' - \overline{u}\frac{\partial \widehat{h}}{\partial x}. \tag{16.86}$$

The free Rossby waves all decay monotonically to zero (problem 16.6). However, the steady problem, obtained by omitting the first term on the left-hand side, has solutions

$$\widetilde{\psi} = \frac{\widetilde{h}_b}{(K^2 - K_s^2 - \mathrm{i}R)}, \tag{16.87}$$

where $R = (rK^2/\overline{u}k)$, and the singularity has been removed. The amplitude of the response is still a maximum for the stationary wave, and for this wave the phase of the response is shifted by $\pi/2$ with respect to the topography. The solution is shown in Fig. 16.10.[4] It is typical that for a mountain range whose Fourier composition contains all wavenumbers, there is a minimum in the streamfunction a little downstream of the mountain ridge.

### 16.7.2 Application to Earth's atmosphere

*With three parameters, I can fit an elephant.*
William Thomson, Lord Kelvin (1824–1907).

Rather surprisingly, given the complexity of the real system and the simplicity of the model, when used with realistic topography a one-layer model gives reasonably realistic answers for the Earth's atmosphere. Thus, we calculate the stationary response to the Earth's topography using (16.86), using a reasonably realistic representation of the Earth's topography and, with qualification, the zonal wind. The zonal wind on the left-hand side of (16.86) is interpreted as the wind in the mid-troposphere, whereas the wind on the right-hand side is better interpreted as the surface wind, and so perhaps is about 0.4 times the mid-troposphere wind. Since the

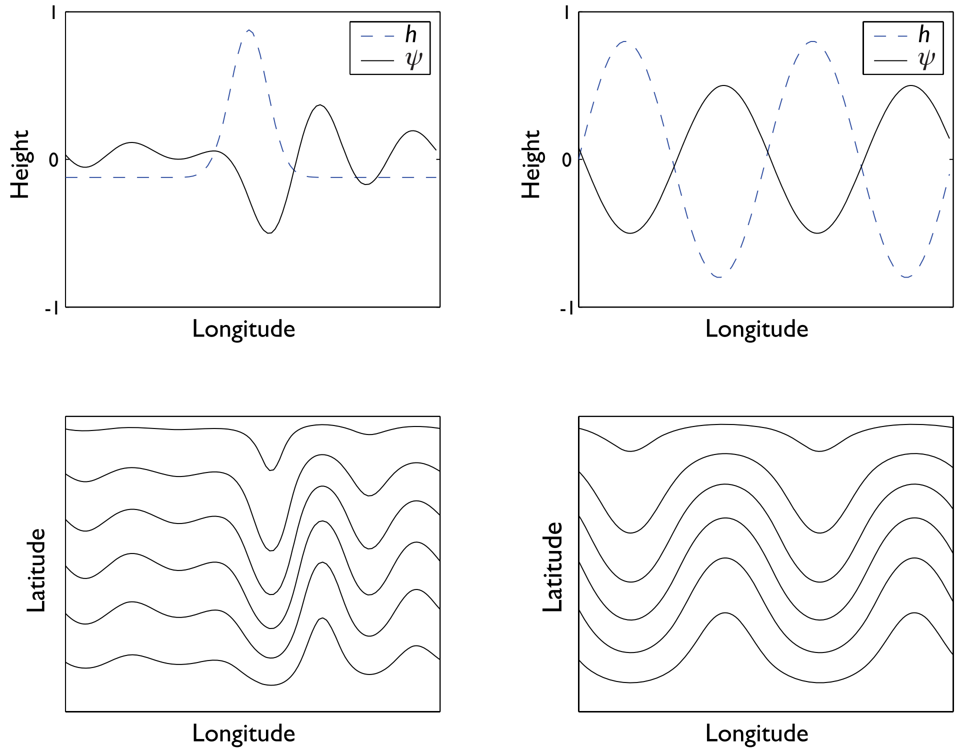


**Fig. 16.10** The response to topographic forcing, i.e., the solution to the steady version of (16.86), for topography consisting of an isolated Gaussian ridge (left panels) and a pure sinusoid (right panels). The wavenumber of the stationary wave is about 4 and $r/(\overline{u}k) = 1$. The upper panels show the amplitude of the topography (dashed curve) and the perturbation streamfunction response (solid curve). The lower panels are contour plots of the streamfunction, including the mean flow. With the ridge, the response is dominated by the resonant wave and there is a streamfunction minimum, a 'trough', just downstream of the ridge. In the case on the right, the flow cannot resonate with the topography, which consists only of wavenumber 2, and the response is exactly out of phase with the topography.

problem is linear, this amounts to tuning the amplitude of the response. The results, obtained using a rather crude representation of the Earth's topography, are plotted in Fig. 16.11. Also plotted is the observed time averaged response of the real atmosphere (the 500 mb height field at 45° N). The agreement between model and observation is quite good, but this must be regarded as somewhat fortuitous if only because the other main source of the stationary wave field — thermal forcing — has been completely omitted from the calculation. Quantitative agreement is thus a consequence of the aforesaid tuning. Nevertheless, the calculation does suggest that the stationary, zonally asymmetric, features of the Earth's atmosphere arise via the interaction of the zonally symmetric wind field and the zonally asymmetric lower boundary, and that these may be calculated to a reasonable approximation with a linear model.

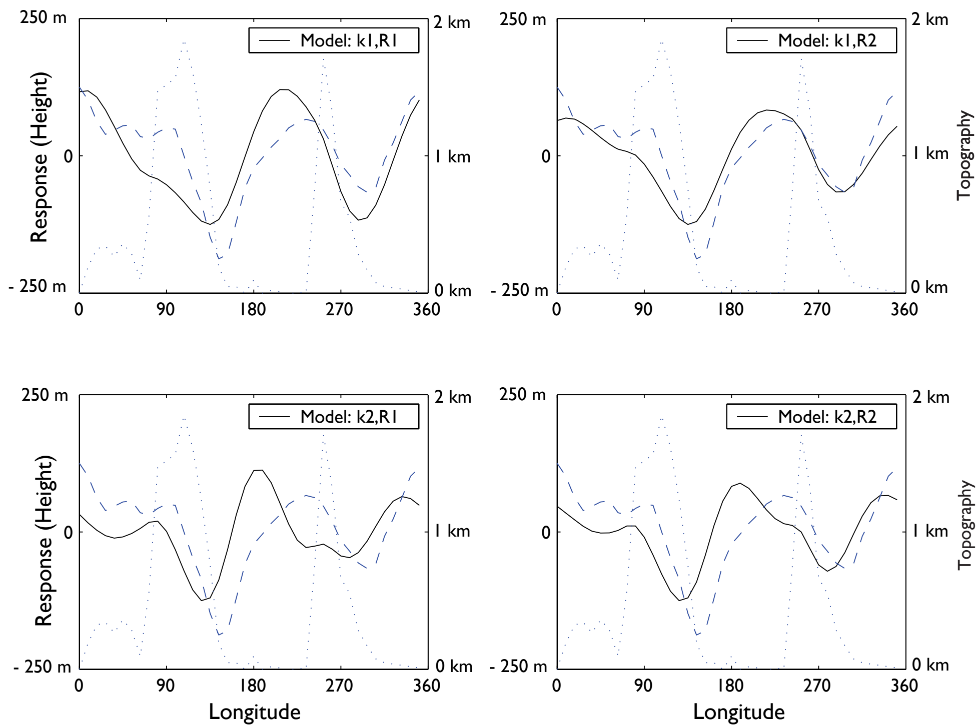


**Fig. 16.11** Solutions of the Charney–Eliassen model. The solid lines are the steady solution of (16.86) using the Earth's topography at 45° N with two values of friction (R1 ≈ 6 days, R2 ≈ 3 days) and two values of resonant zonal wavenumber (2.5 for k1, 3.5 for k2), corresponding to zonal winds of approximately 17 and 13 m s$^{-1}$. The solutions are given in terms of height, $\eta'$, where $\eta' = f_0\psi'/g$, with the scale on the left of each panel. The dashed line in each panel is the observed average height field at 500 mb at 45° N in January. The dotted line is the topography used in the calculations, with the scale on the right of each panel.

### 16.7.3 ✦ One-dimensional Rossby wave trains

Although the Fourier analysis above gives exact (linear) results, it is not particularly revealing of the underlying dynamics. We see from Fig. 16.10 that the response to the Gaussian ridge is largely downstream of the ridge, and this suggests that it will be useful to consider the response as being due to Rossby *wavetrains* being excited by local features. This is also suggested by Fig. 16.12, which shows that the response to realistic topography is relatively local, and may be considered to arise from two relatively well-defined wavetrains each of finite extent one coming from the Rockies and the other from the Himalayas.

One way to analyse these wavetrains, and one which also brings up the concept of group velocity in a natural way, is to exploit (as in section 15.1.2) a connection between changes in wavenumber and changes in frequency. Consider the linear barotropic vorticity equation in the form

$$\frac{\partial}{\partial t}(\zeta - k_d^2\psi) + \overline{u}\frac{\partial \zeta}{\partial x} + \beta\frac{\partial \psi}{\partial x} = -r\zeta, \tag{16.88}$$

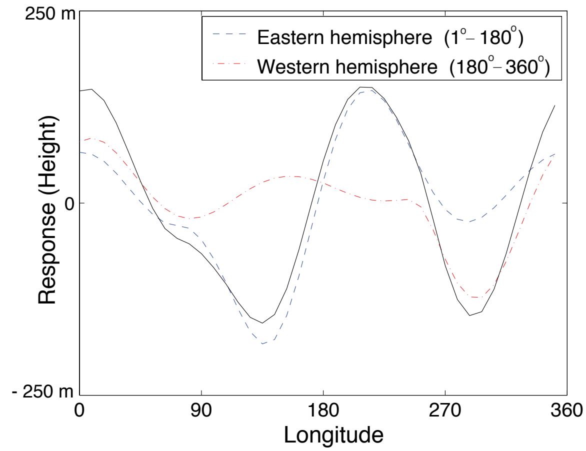


**Fig. 16.12** The solution of the upper left-hand panel Fig. 16.11 (solid line), and the solution divided into two contributions (dashed lines), one due to the topography only of the western hemisphere (i.e., with the topography in the east set to zero) and the other due to the topography only of the eastern hemisphere.

where $r$ is a frictional coefficient, which we presume to be small. Setting $k_d = 0$ for simplicity, the linear dispersion relation is

$$\omega = \overline{u}k - \frac{\beta k}{K^2} - \mathrm{i}r \equiv \omega_R(k,l) - \mathrm{i}r, \tag{16.89}$$

where $K^2 = k^2 + l^2$ and $\omega_R(k,l)$ is the inviscid dispersion relation for Rossby waves. Now, if there is a local source of the waves, for example an isolated mountain, we may expect to see a *spatial* attenuation of the wave as it moves away from the source. We may then regard the system as having a fixed, real frequency, but a changing, possibly complex, wavenumber. To determine this wavenumber for stationary waves (and so with $\omega = 0$), for small friction we expand the dispersion relation in a Taylor series about the inviscid value of $\omega_R$ at the real stationary wavenumber $k_s$, where $k_s = (K_s^2 - l^2)^{1/2}$ and $K_s = \sqrt{\beta/\overline{u}}$. This gives

$$\omega + \mathrm{i}r = \omega_R(k,l) \approx \omega_R(k_s,l) + \left.\frac{\partial \omega_R}{\partial k}\right|_{k=k_s} k' + \cdots. \tag{16.90}$$

Thus, $k' \approx \mathrm{i}r/c_g^x$, where $c_g^x$ is the zonal component of the group velocity evaluated at a fixed position and at the stationary wavenumber; using (6.60b) this is given by

$$c_g^x = \left.\frac{\partial \omega_R}{\partial k}\right|_{k=k_s} = \frac{2\overline{u}k_s^2}{k_s^2 + l^2}. \tag{16.91}$$

The solution therefore decays away from a source at $x = 0$ according to

$$\psi \sim \exp(\mathrm{i}kx) = \exp\left[\mathrm{i}(k_s + k')x\right] \approx \exp(\mathrm{i}k_s x - rx/c_g^x) \tag{16.92}$$

and, because $c_g^x > 0$, the response is east of the source. The approximate solution for the streamfunction (denoted $\psi_\delta$) of (16.86) in an infinite channel, with the topography being a $\delta$-function mountain ridge at $x = x'$, and with all fields varying meridionally like $\sin ly$, is thus

$$\psi_\delta(x - x', y) \sim \begin{cases} 0 & x \le x' \\ -\dfrac{1}{k_s} \sin ly \sin[k_s(x - x')] \exp[-r(x - x')/c_g^x] & x \ge x'. \end{cases} \tag{16.93}$$

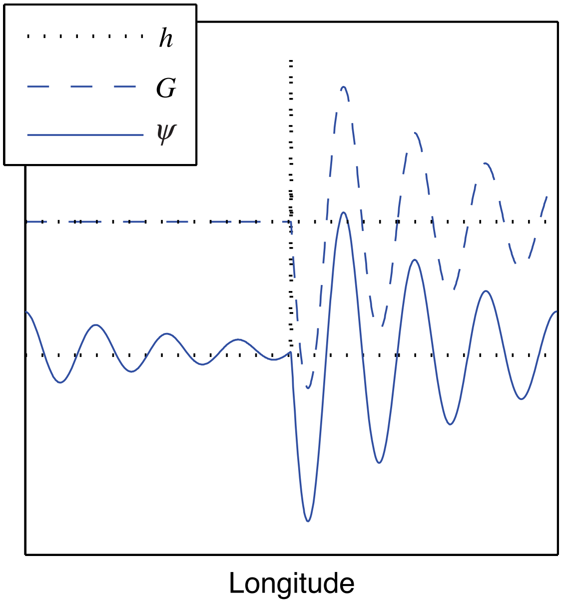


**Fig. 16.13** A one-dimensional Rossby wave train excited by uniform eastward flow over a $\delta$-function mountain ridge ($h$) in the centre of the domain. The upper curve, $G$, shows the Green's function (16.93), whereas the lower curve shows the exact (linear) response, $\psi$, in a re-entrant channel calculated numerically using the Fourier method. The two solutions are both centred around zero and offset for clarity; the only noticeable difference is upstream of the ridge, where there is a finite response in the Fourier case because of the progression of the wavetrain around the channel. The stationary wavenumber is 7.5.

In the more general problem in which the topography is a general function of space, every location constitutes a separate source of wavetrains, and the complete (approximate) solution is given by the integral

$$\psi'(x, y) = \int_{-\infty}^{\infty} \widehat{h}(x)\psi_\delta(x - x', y)\,\mathrm{d}x'. \tag{16.94}$$

The field $\psi_\delta(x - x', y)$, is known as the 'Green function' for the problem, sometimes denoted $G(x - x', y)$.

Example solutions calculated using both the Fourier and Green function methods are illustrated in Fig. 16.13. As in Fig. 16.10 there is a trough immediately downstream of the mountain, a result that holds for a broad range of parameters. In these solutions, the streamfunction decays almost completely in one circumnavigation of the channel, and thus, downstream of the mountain, both methods give virtually identical results. Such a correspondence will not hold if the wave can circumnavigate the globe with little attenuation, for then resonance will occur and the Green function method will be inaccurate; thus, whether the resonant picture or the wavetrain picture is more appropriate depends largely on the frictional parameter. A frictional time scale of about 10 days is often considered to approximately represent the Earth's atmosphere, in which case waves are only slightly damped on a global circumnavigation, and the

Fourier picture is natural with the possibility of resonance. However, the smaller (more frictional) value of 5 days seems to give quantitatively better results in the barotropic problem, and the solution is more evocative of wavetrains. The larger friction may perform better because it is crudely parameterizing the meridional propagation and dispersion of Rossby waves that is neglected in the one-dimensional model.[5]

### 16.7.4 The adequacy of linear theory

Having calculated some solutions, we are in a position to estimate, *post facto,* the adequacy of the linear theory by calculating the magnitude of the omitted nonlinear terms. The linear problem here differs in kind from that which arises when using linear theory to evaluate the stability of a flow, as in chapter 9. In that case, we assume a small initial perturbation and the initial evolution of that perturbation is then accurately described, by construction, using linear equations. However, the amplitude of the perturbation is arbitrary, for it may grow exponentially and its size at any given time is proportional to the magnitude of the initial perturbation, which is assumed small but which is otherwise unconstrained. In contrast, when we are calculating the stationary linear response to flow over topography or to a thermal source, the amplitude of the solution is *not* arbitrary; rather, it is determined by the parameters of the problem, including the size of the topography, and represents a real quantity that might be compared to observations. Of course, because the problem is linear, the amplitude of the solution is directly proportional to the magnitude of the topography or thermal perturbation.

From (16.87), and recalling that the amplitude of $\widetilde{h}$ is scaled relative to the real topography by the factor $(g/L_d^2 f_0)$, we crudely estimate the amplitude of the response to topography to be

$$|\psi'| \sim \frac{\alpha g h_b}{f_0} \approx \alpha \times 10^8\,\mathrm{m^2\,s^{-1}} = 2\times 10^7\,\mathrm{m^2\,s^{-1}}, \qquad |\eta'| = \frac{f_0|\psi'|}{g} \sim \alpha h_b \approx 0.2\,\mathrm{km}. \tag{16.95}$$

where the non-dimensional parameter $\alpha$ accounts for the distance of the response from resonance and the ratio of the length scale to the deformation scale. Choosing $\alpha = 0.2$ and $h_b = 1\,\mathrm{km}$ gives the numerical values above, which are similar to those calculated more carefully, or observed (Fig. 16.11).

If linear theory is to be accurate, we must demand that the self advection of the response is much smaller than the advection by the basic state, and so that

$$|J(\psi', \nabla^2\psi')| \ll |\overline{u}\frac{\partial}{\partial x}\nabla^2\psi'|, \tag{16.96}$$

or, again rather crudely, that $|\psi'/L| \ll \overline{u}$. For $L = 5000\,\mathrm{km}$ we have $\psi'/L = 4\,\mathrm{m\,s^{-1}}$, which is a few times smaller than a typical mid-troposphere zonal flow of $20\,\mathrm{m\,s^{-1}}$, suggesting that the linear approximation may hold water. However, the inequality is by no means so well satisfied that we can state without equivocation that the linear approximation is a good one, especially as a different choice of numerical factors would give a different answer, and the use of a simple barotropic model also implies inaccuracies. Rather, we conclude that we must carefully calculate the linear response, and both compare it with the observations and calculate the implied nonlinear terms, before concluding that linear theory is appropriate. In fact linear theory seems to do quite well, and certainly gives qualitative insight into the nature of the mean mid-latitude zonal asymmetries.

## 16.8 ✦ EFFECTS OF THERMAL FORCING

How does thermal forcing influence the stationary waves? To give an accurate answer for the real atmosphere is a little more difficult than for the orographic case where the forcing can be included reasonably accurately in a quasi-geostrophic model with a term $\overline{u} \cdot \nabla h_b$ at the lower boundary. Anomalous (i.e., variations from a zonal or temporal mean) thermodynamic forcing typically also arises initially at the lower boundary through, for example, variations in the surface temperature. However, such anomalies may be felt throughout the lower troposphere on a relatively short time scale by way of such non-geostrophic phenomena as convection, so that the effective thermodynamic source that should be applied in a quasi-geostrophic calculation has a finite vertical extent. However, an accurate parameterization of this may depend on the structure of the atmospheric boundary layer and this cannot always be represented in a simple way.[6] Because of such uncertainties our treatment concentrates on the fundamental and qualitative aspects of thermal forcing.

The quasi-geostrophic potential vorticity equation, linearized around a uniform zonal flow, is [cf. (16.42)]

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\left[\nabla^2\psi' + \frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \psi'}{\partial z}\right)\right] + \frac{\partial \psi'}{\partial x}\left[\beta - \frac{f_0^2}{\rho_R}\frac{\partial}{\partial z}\left(\frac{\rho_R}{N^2}\frac{\partial \overline{u}}{\partial z}\right)\right] = \frac{f_0}{N^2}\frac{\partial Q}{\partial z} \equiv T, \tag{16.97}$$

where $Q$ is the source term in the (linear) thermodynamic equation,

$$\frac{\partial}{\partial t}\left(\frac{\partial \psi'}{\partial z}\right) + \overline{u}\frac{\partial}{\partial x}\frac{\partial \psi'}{\partial z} - v'\frac{\partial \overline{u}}{\partial z} + \frac{N^2}{f_0}w' = \frac{Q}{f_0}. \tag{16.98}$$

A particular solution to (16.97) may be constructed if $\overline{u}$ and $N^2$ are constant, and if $Q$ has a simple vertical structure. If we again write $\psi' = \mathrm{Re}\,\widetilde{\psi}(z)\sin ly \exp(ikx)$ and let $\Phi(z) = \widetilde{\psi}(z)\exp(-z/2H)$ we obtain

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}z^2} + m^2\Phi = \frac{T}{\mathrm{i}k\overline{u}}\mathrm{e}^{-z/2H}, \qquad \text{where} \quad m^2 = \frac{N^2}{f_0^2}\left(\frac{\beta}{\overline{u}} - K^2 - \gamma^2\right). \tag{16.99}$$

If we let $T = T_0\exp(-z/H_Q)$, so that the heating decays exponentially away from the Earth's surface, then the particular solution to the stationary problem is found to be

$$\widetilde{\psi} = \mathrm{Re}\,\frac{\mathrm{i}\widehat{T}\mathrm{e}^{-z/H_Q}}{k\overline{u}\left[(N/f_0)^2(K_s^2 - K^2) + H_Q^{-2}(1 + H_Q/H)\right]}, \tag{16.100}$$

where $\widehat{T}$ is proportional to $T$. This solution does not satisfy the boundary condition at $z = 0$, which in the absence of topography and friction is

$$\overline{u}\frac{\partial}{\partial x}\frac{\partial \psi'}{\partial z} - v'\frac{\partial \overline{u}}{\partial z} = \frac{Q(0)}{f_0}. \tag{16.101}$$

A homogeneous solution must therefore be added, and just as in the topographic case this leads to a vertically radiating or a surface trapped response, depending on the sign of $m^2$. One way to calculate the homogeneous solution is to first use the linearized thermodynamic equation (16.98), or the linearized vorticity equation (16.103), to calculate the vertical velocity at the surface implied by (16.100), $w_p(0)$ say. We then notice that the homogeneous solution is effectively forced by an equivalent topography given by $h_e = -w_p(0)/(\mathrm{i}k\overline{u(0)})$, and so proceed as in the topographic case. The complete solution is rather hard to interpret, and is in any case available only in special cases, so it is useful to take a more qualitative approach.

### 16.8.1 Thermodynamic balances

It is the properties of the particular solution that distinguish the response to thermodynamic forcing from that due to topography, because the homogeneous solutions of the two cases are similar. And far from the source region, the homogeneous solution will dominate, giving rise to wavetrains as discussed previously.

We can determine many of the properties of the response to thermodynamic forcing by considering the balance of terms in the steady linear thermodynamic equation, which we write as

$$\overline{u}\frac{\partial}{\partial x}\frac{\partial \psi'}{\partial z} - \frac{\partial \psi'}{\partial x}\frac{\partial \overline{u}}{\partial z} + \frac{N^2}{f_0}w' = \frac{Q}{f_0} \equiv R \tag{16.102a}$$

or

$$f_0\overline{u}\frac{\partial v'}{\partial z} - f_0 v'\frac{\partial \overline{u}}{\partial z} + N^2 w' = Q. \tag{16.102b}$$

The vorticity equation is

$$\overline{u}\frac{\partial \zeta'}{\partial x} + \beta v' = \frac{f_0}{\rho_R}\frac{\partial \rho_R w'}{\partial z}. \tag{16.103}$$

Assuming that the diabatic forcing is significant, we may imagine three possible simple balances in the thermodynamic equation:

*(i)* zonal advection dominates, and $v' = \partial\psi'/\partial x \sim QH_Q/(f_0\overline{u})$;

*(ii)* meridional advection dominates, and $v' \sim QH_u/(f_0\overline{u})$;

*(iii)* vertical advection dominates, and $w' \sim Q/N^2$. Then, for large enough horizontal scales the balance in the vorticity equation is $\beta v' \sim f_0 w'_z$ and $v' \sim f_0 Q/(\beta N^2 H_Q)$. For smaller horizontal scales advection of relative vorticity may dominate that of planetary vorticity, and $\beta$ is replaced by $\overline{u}K^2$.

Here, $H_Q$ is the vertical scale of the source (so that $\partial Q/\partial z \sim Q/H_Q$) and $H_u$ is the vertical scale of the zonal flow (so that $\partial\overline{u}/\partial z \sim \overline{u}/H_u$). We also assume that the vertical scale of the solution is $H_Q$, so that $\partial v'/\partial z \sim v'/H_Q$. Which of the above three balances is likely to hold? Heuristically, we might suppose that the balance with the smallest $v'$ will dominate, if only because meridional motion is suppressed on the $\beta$-plane. Then, zonal advection dominates meridional advection if $H_u > H_Q$, and vice versa. Defining $\widehat{H} = \min(H_u, H_Q)$ then horizontal advection will dominate vertical advection if

$$\mu_1 = \frac{\beta N^2 H_Q \widehat{H}}{\overline{u} f_0^2} \ll 1. \tag{16.104}$$

More systematically, we can proceed in *reductio ad absurdum* fashion by first neglecting the vertical advection term in (16.102), and seeing if we can construct a self-consistent solution. If

$\psi' = \mathrm{Re}\,\widetilde{\psi}_p(z)\mathrm{e}^{\mathrm{i}kx}$, and noting that $\overline{u}\partial\widetilde{\psi}_p/\partial z - \widetilde{\psi}_p\partial\overline{u}/\partial z = \overline{u}^2(\partial/\partial z)(\widetilde{\psi}_p/\overline{u})$ we obtain

$$\widetilde{\psi}_p = \frac{\mathrm{i}\overline{u}}{kf_0}\int_z^\infty \frac{\widetilde{Q}}{\overline{u}^2}\,\mathrm{d}z, \tag{16.105}$$

where $\widetilde{Q}$ denotes the Fourier amplitude of $Q$. Then, from the vorticity equation (16.103), we obtain the (Fourier amplitude of the) vertical velocity

$$\widetilde{w}_p = \frac{-\mathrm{i}k}{f_0\rho_R}\int_z^\infty \rho_R\overline{u}(K_s^2 - K^2)\widetilde{\psi}_p\,\mathrm{d}z. \tag{16.106}$$

Using this one may, at least in principle, check whether the vertical advection in (16.102) is indeed negligible. If $\overline{u}$ is uniform (and so $H_u \gg H_Q$) then we find

$$\widetilde{\psi}_p \propto \frac{\mathrm{i}QH_Q}{kf_0\overline{u}} \qquad \text{and} \qquad \widetilde{w}_p \propto \frac{QH_Q^2(K_s^2 - K^2)}{f_0^2}. \tag{16.107a,b}$$

Using this, vertical advection indeed makes a small contribution to the thermodynamic equation provided that

$$\mu_2 = \frac{N^2H_Q^2|K_s^2 - K^2|}{f_0^2} \ll 1 \tag{16.108}$$

If $K_s^2 \gg K^2$ and $\widehat{H} = H_Q$ then (16.108) is equivalent to (16.104). If $\overline{u}$ is not constant and if $H_u \ll H_Q$ then $H_u$ replaces $H_Q$ and the criterion for the dominance of horizontal advection becomes

$$\mu = \frac{N^2\widehat{H}H_Q|K_s^2 - K^2|}{f_0^2} \ll 1. \tag{16.109}$$

This is the condition that the first term in the denominator of (16.100) is negligible compared with the second. For a typical tropospheric value of $N^2 = 10^{-4}\,\mathrm{s}^{-1}$ and for $K > K_S$ we find that $\mu \approx (H_Q/7\,\mathrm{km})^2$, and so we can expect $\mu < 1$ in extra-equatorial regions where the heating is shallow. At low latitudes $f_0$ is smaller, $\beta$ is bigger and $\mu \approx (H_Q/1\,\mathrm{km})^2$, and we can expect $\mu > 1$. However, there is both uncertainty and variation in these values.

*Equivalent topography*

In the case in which zonal advection dominates, the equivalent topography is given by

$$h_e = \frac{-\widetilde{w}_p(0)}{\mathrm{i}k\overline{u(0)}} = \frac{1}{\overline{u(0)}f_0\rho_R(0)}\int_0^\infty \rho_R\overline{u}(K_s^2 - K^2)\widetilde{\psi}_p\,\mathrm{d}z, \tag{16.110}$$

where $\widetilde{\psi}_p$ is given by (16.105). The point to notice here is that if $K < K_s$ the equivalent topography is in phase with $\psi_p$.

### 16.8.2 Properties of the solution

In the tropics $\mu$ may be large for $H_Q$ greater than a kilometre or so. Heating close to the surface cannot produce a large vertical velocity and will therefore produce a meridional velocity. However, away from the surface the heat source will be balanced by vertical advection. For scales such that $K < K_S$, a criterion that might apply at low latitudes for wavelengths longer than a few thousand kilometres, the associated vortex stretching $f\partial w/\partial z > 0$ is balanced by $\beta v$ and a poleward meridional motion occurs. This implies a trough west of the heating and/or a ridge east of the heating, although the use of quasi-geostrophic theory to draw tropical inferences may be a little suspect.

In mid-latitudes $\mu$ is typically small and horizontal advection locally balances diabatic heating. In this case there is a trough a quarter-wavelength downstream from the heating, and equatorward motion at the longitude of the source. [To see this, note that if the heating has a structure like $\cos kx$ then from either (16.100) or (16.105) the solution goes like $\psi_p \propto -\sin kx$.] The trough may be warm or cold, but is often warm. If $H_Q \ll H_u$, as is assumed in obtaining (16.100), then $\theta$ is positive and warm. This is because zonal advection dominates and so the effect of the heating is advected downstream. If $H_Q \gg H_u$ and meridional advection is dominant, then the trough is still warm provided $Q$ decreases with height. The vertical velocity can be inferred from the vorticity balance. If $f_0\partial w/\partial z \approx \beta v$ and if $w = 0$ at the surface (in the absence of Ekman pumping and any topographic effects) there is *descent* in the neighbourhood of a heat source. This counter-intuitive result arises because it is the horizontal advection that is balancing the diabatic heating. (This result cannot be inferred from the particular solution alone.) If the advection of relative vorticity balances vortex stretching, the opposite may hold.

The homogeneous solution can be inferred from (16.110) and (T.1). Consider, for example, waves that are trapped ($m^2 < 0$) but still have $K < K_S$; that is $K^2 < K_S^2 < K^2 + \gamma^2$. The homogeneous solution forced by the equivalent topography is out of phase with that topography, and so out of phase with $\psi_p$, using (16.110). For still shorter waves, $K > K_s$, the homogeneous solution is in phase with the equivalent topography, and so again out of phase with $\psi_p$. Thermal sources produced by large-scale continental land masses may have $K^2 < K_s^2$ and, if $K^2 + \gamma^2 < K_S^2$ they will produce waves that penetrate up into the stratosphere and typically these solutions will dominate far from the source. Evidently though, the precise relationship between the particular and homogeneous solution is best dealt with on a case-by-case basis. A few more general points are summarized in the box on page 713.

### 16.8.3 Numerical solutions

The numerically calculated response to an isolated heat source is illustrated in Figs. 16.14 and 16.15. The first figure shows the response to a 'deep' heating at 15° N. As the reasoning above would suggest, the vertical velocity field (not shown) is upwards in the vicinity of the source. Away from the source, the solution is dominated by the homogeneous solutions in the form of wavetrains, as described in section 16.7.3, with a simple vertical structure. (In fact, the pattern is quite similar to that obtained with a suitably forced barotropic model, as was found earlier in the topographically forced case.)

Figure 16.15 shows the response to a perturbation at 45° N, and again the solutions are qualitatively in agreement with the reasoning above. The local heating is balanced by an equatorward wind, and there is a surface trough about 20° east of the source, and an upper-level pressure maximum, or ridge, about 60° east. The scale height of the wind field, $H_u$ is about 8 km, greater than that of the source, and the balance in the thermodynamic equation is between the zonal advection of the temperature anomaly $\overline{u}\theta'_x$ and the heat source, so producing a temperature maximum downstream. Again, the far field is dominated by the wavetrain of the homogeneous solution.

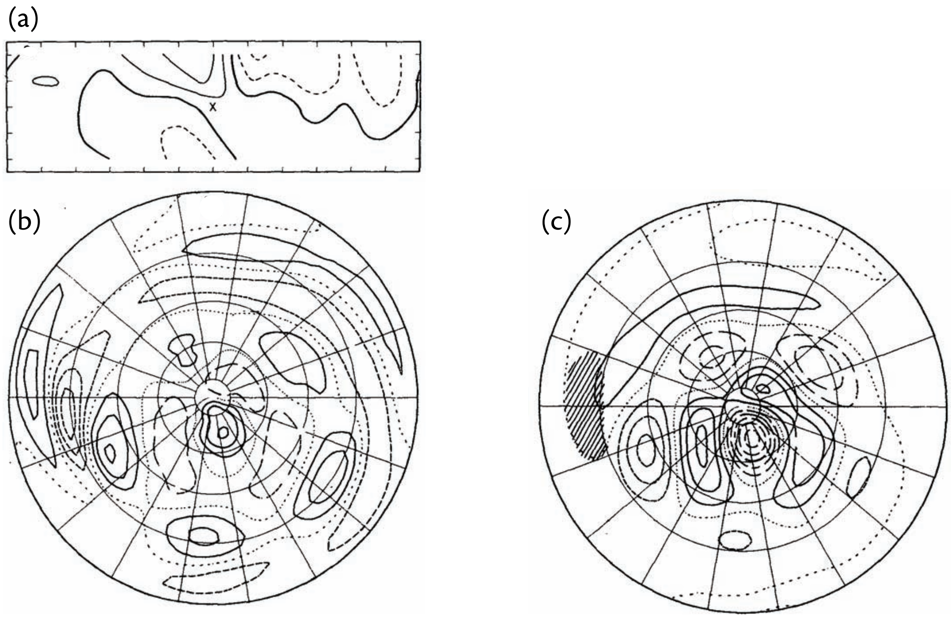


**Fig. 16.14** Numerical solution of a baroclinic primitive equation model with a deep heat source at 15° N and a zonal flow similar to that of northern hemisphere winter. (a) Height field in a longitude height at 18° N (vertical tick marks at 100, 300, 500, 700 and 900 mb); (b) 300 mb vorticity field; (c) 300 mb height field. The cross in (a) and the hatched region in (c) indicate the location of the heating.[8]

Finally, we show a calculation (Fig. 16.16) that, although linear, includes realistic forcing from topography, heat sources and observed transient eddy flux convergences, and uses a realistic zonally averaged zonal flow, although some physical parameters (e.g., representing frictional and diffusion) in the calculation must be changed in order that a steady solution can be achieved. Such a calculation is likely to be the most accurate achievable by a linear model, and discrepancies from observations indicate the presence of nonlinearities that are neglected in the calculation. In fact, a generally good agreement with the observed fields is found, and provides some *post facto* justification for the use of linear, stationary wave models.[13]

In such realistic calculations it is virtually impossible to see the wavetrains emerging from isolated features like the Rockies or Himalayas, because they are combined with the responses from all the other sources included in the calculation. Breaking up the forcing into separate contributions from orographic forcing, heating, and the time averaged momentum and heat fluxes from transient eddies reveals that all of these separate contributions have a non-neglible influence. We should also remember that the effects of the fluxes from the transient eddies are not explained by such a calculation, merely included in a diagnostic sense. Nevertheless, the agreement does reveal the extent to which we might understand the steady zonally asymmetric circulation of the real atmosphere as the response due to the interaction of a zonally uniform zonal wind with the asymmetric features of the Earth's geography and transient eddy field. The quasi-stationary response of the planetary waves to surface anomalies, and the interaction of transient eddies with the large-scale planetary wave field, are important factors in the natu-

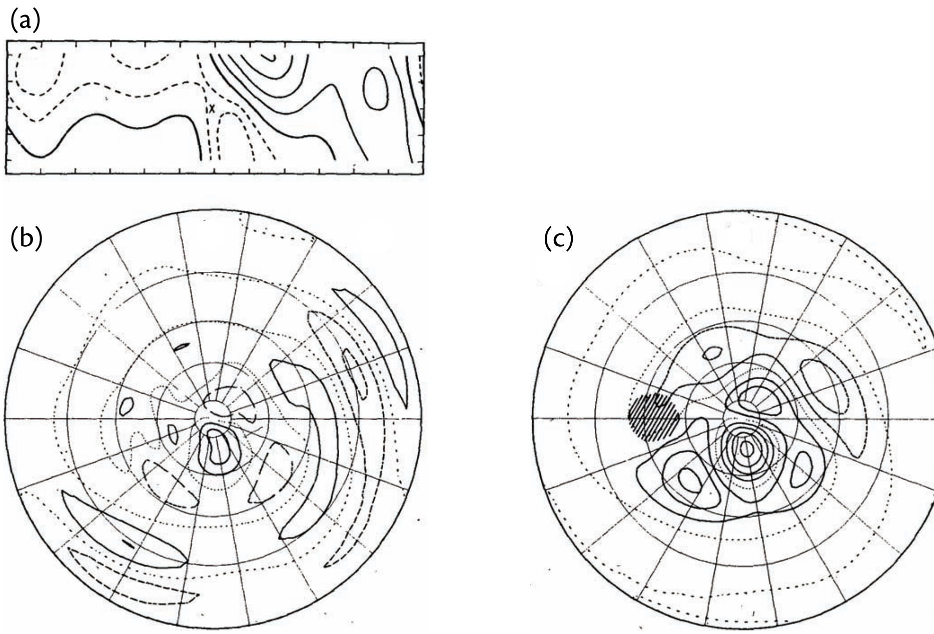


**Fig. 16.15** As for Fig. 16.14, but now the solution of a baroclinic primitive equation model with a deep heat source at 45° N. (a) Height field in a longitude height at 18° N; (b) 300 mb vorticity field; (c) 300 mb height field. The cross in (a) and the hatched region in (c) indicate the location of the heating.[10]

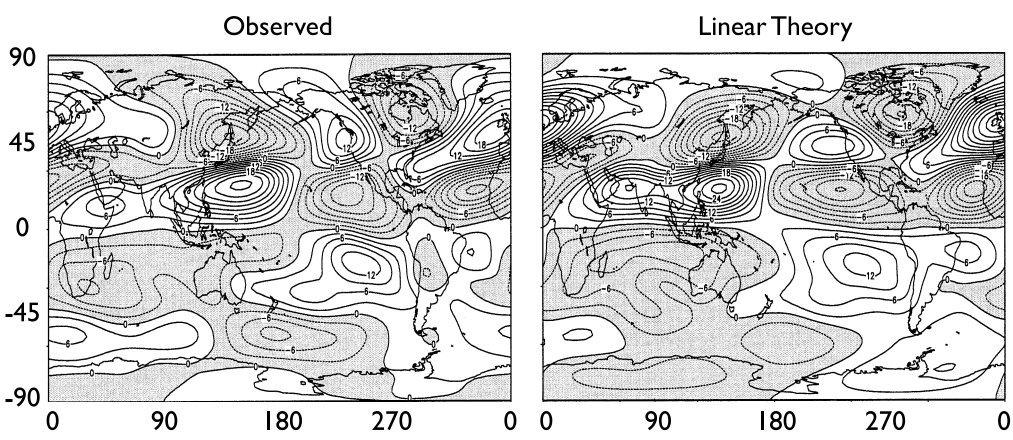


**Fig. 16.16** Left: the observed stationary (i.e., time-averaged) streamfunction at 300 mb (about 7 km altitude) in northern hemisphere winter. Right: the steady, linear response to forcing by orography, heat sources and transient eddy flux convergences, calculated using a linear model with the observed height-varying zonally averaged zonal wind. Contour interval is $3 \times 10^6\ \mathrm{m^2\,s^{-1}}$, and negative values are shaded. Note the generally good agreement, and also the much weaker zonal asymmetries in the southern hemisphere.[12]

### Thermal Forcing of Stationary Waves: Salient Points

*(i)* The solution is composed of a particular solution and a homogeneous solution.

*(ii)* The homogeneous solution may be thought of as being forced by an 'equivalent topography', chosen so that the complete solution satisfies the boundary condition on vertical velocity at the surface.

*(iii)* For a localized source, the far field is dominated by the homogeneous solution. This solution has the same properties as a solution forced by real topography. Thus, it may include waves that penetrate vertically into the stratosphere as well as wavetrains propagating around the globe with an equivalent barotropic structure.

*(iv)* In the extratropics, a diabatic heating is typically balanced by horizontal advection, producing a trough a quarter wavelength east (downstream) of a localized heat source. The heat source is balanced by advection of cooler air from higher latitudes, and there may be sinking air over the heat source. This can occur when $\mu \ll 1$ [see (16.109)].

*(v)* In the tropics, a heat source may be locally balanced by vertical advection, that is adiabatic cooling as air ascends. This can occur when $\mu \gg 1$.

*(vi)* In the real atmosphere, the stationary solutions must coexist with the chaos of time-dependent, nonlinear flows. Thus, they are likely to manifest themselves only in time averaged fields and in a modified form.

ral variability of climate, and their understanding remains a difficult challenge and a topic of research for dynamical meteorologists.

## 16.9 ✦ WAVE PROPAGATION USING RAY THEORY

Rossby waves, of course, propagate meridionally as well as zonally. Furthermore, one of the major mountain ranges on the Earth — the Himalayas — is fairly localized in the meridional direction, and even though the Rockies and Andes do form a convenient meridional ridge, the Rossby waves they generate will still propagate both zonally and meridionally. Furthermore, the coefficients of the linear equations of motion vary with space: on the sphere $\beta$ is a function of latitude and in general topography is a function of both latitude and longitude. Given this complexity, we cannot solve the full problem except numerically, but a few ideas from wave tracing illustrate many of the features of the response, and indeed of the stationary wave pattern in the Earth's atmosphere.[14]

### 16.9.1 Ray tracing

[This section may no longer be needed. xxx]

Let us first recall a few results about rays and ray tracing that we encountered in chapter 6. Most of the important properties of a wave, such as the energy (if conserved) and the wave activity, propagate along rays at the group velocity. Rays themselves are lines that are parallel to the group velocity, generally emanating from some wave source. A ray is perpendicular to the local wave front, and in a homogeneous medium a wave propagates in a straight line. In non-homogeneous media the group velocity varies with position; however, if the medium varies only slowly, on a scale much larger than that of the wavelength of the waves, the wave activity still propagates along rays at the group velocity.

Let us represent a wave by

$$\psi(x_i, t) = \Psi(x_i, t)\mathrm{e}^{\mathrm{i}\theta(x_i,t)} \tag{16.111}$$

where the amplitude, $\Psi$, varies more slowly than the phase, $\theta$. [We use subscripts ($i$, $j$, etc.) to denote Cartesian axes, repeated subscripts are to be summed over, and $(x_i, t)$ means $(x, y, z, t)$.] Locally, the frequency, $\omega$, and wavenumber, $k_i$ satisfy

$$k_i = \frac{\partial\theta}{\partial x_i}, \qquad \omega = -\frac{\partial\theta}{\partial t}, \tag{16.112a,b}$$

and these imply

$$\frac{\partial k_i}{\partial t} = -\frac{\partial\omega}{\partial x_i}. \tag{16.113}$$

The frequency is, in general, a function of wavenumber, position and time, with the time dependence arising if the medium varies in time. The dispersion relation is an equation of the form

$$\omega = \Omega(k_i; x_i, t) \tag{16.114}$$

again with $x_i$ and $t$ varying only slowly, and the local group velocity is given by $c_{gi} = \partial\omega/\partial k_i$.

As derived in section 6.3 the wavevector and the frequency may both vary with position and time. The wavenumber varies according to

$$\frac{\partial k_i}{\partial t} + c_{gj}\frac{\partial k_i}{\partial x_j} = -\left(\frac{\partial\omega}{\partial x_i}\right)_{k_i}. \tag{16.115}$$

The left-hand side is the change in wavenumber along a ray. If the frequency is constant in space the right-hand side vanishes and the wavenumber is simply propagated at the group velocity. If the frequency is independent of a particular coordinate then the corresponding wavenumber is constant along the ray. If the frequency changes with position (as in general it will), then the wavenumber will change along a ray, and thus so will the direction of propagation — the wave is *refracted*. Note that we can write (16.115), and the definition of group velocity, in the compact forms

$$\frac{\mathrm{D}_{c_g} k_i}{\mathrm{D}t} = -\frac{\partial\Omega}{\partial x_i}, \qquad \frac{\mathrm{D}_{c_g} x_i}{\mathrm{D}t} = \frac{\partial\Omega}{\partial k_i}, \tag{16.116a,b}$$

where $\mathrm{D}_{c_g}/\mathrm{D}t \equiv \partial/\partial t + (\boldsymbol{c}_g\cdot\nabla)$.

The variation of the frequency is given by

$$\frac{\partial\omega}{\partial t} = \frac{\partial\Omega}{\partial t} + \frac{\partial\Omega}{\partial k_i}\frac{\partial k_i}{\partial t} = \frac{\partial\Omega}{\partial t} - \frac{\partial\Omega}{\partial k_i}\frac{\partial\omega}{\partial x_i} \tag{16.117}$$

or

$$\frac{\mathrm{D}_{c_g}\omega}{\mathrm{D}t} = \frac{\partial \omega}{\partial t} + \boldsymbol{c}_g \cdot \nabla\omega = \frac{\partial \Omega}{\partial t}. \tag{16.118}$$

If the frequency is not an explicit function of time then the frequency is constant along a ray.

One practical result of all this that in problems of the form

$$\frac{\partial}{\partial t}\nabla^2\psi + \beta(y)\frac{\partial \psi}{\partial x} = 0, \tag{16.119}$$

both the frequency and the $x$-wavenumber are constant along a ray. The wavenumber is not constant in the $y$-direction because the frequency is a function of $y$.

### 16.9.2 Rossby waves and Rossby rays

If the topography is localized, then ray theory provides a useful way of calculating and interpreting the response to a flow over that topography. On the $\beta$-plane and away from the orographic source the steady linear response to a zonally uniform but meridionally varying zonal wind will obey

$$\overline{u}(y)\frac{\partial}{\partial x}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\psi' + \beta\frac{\partial \psi'}{\partial x} = 0. \tag{16.120}$$

In fact, an equation of this form applies on the sphere. To see this, we transform the spherical coordinates $(\lambda, \vartheta)$ into Mercator coordinates with the mapping[15]

$$x = a\lambda, \quad \frac{1}{a}\frac{\partial}{\partial \lambda} = \frac{\partial}{\partial x}, \qquad y = \frac{a}{2}\ln\left(\frac{1+\sin\vartheta}{1-\sin\vartheta}\right), \quad \frac{1}{a}\frac{\partial}{\partial \vartheta} = \frac{1}{\cos\vartheta}\frac{\partial}{\partial y}. \tag{16.121}$$

The spherical-coordinate vorticity equation then becomes

$$\overline{u}_M\frac{\partial}{\partial x}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\psi' + \beta_M\frac{\partial \psi'}{\partial x} = 0, \tag{16.122}$$

where $\overline{u}_M = \overline{u}/\cos\vartheta$ and

$$\beta_M = \frac{2\Omega}{a}\cos^2\vartheta - \frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{1}{\cos^2\vartheta}\frac{\mathrm{d}}{\mathrm{d}y}(\overline{u}_M\cos^2\vartheta)\right] = \cos\vartheta\left(\beta_s + \frac{1}{a}\frac{\partial\overline{\zeta}}{\partial\vartheta}\right), \tag{16.123}$$

where $\beta_s = 2a^{-1}\Omega\cos\vartheta$. Thus, $\beta_M$ is the meridional gradient of the absolute vorticity, multiplied by the cosine of latitude. An advantage of Mercator coordinates over their spherical counterparts is that (16.122) has a Cartesian flavour to it, with the metric coefficients being absorbed into the parameters $\overline{u}_M$ and $\beta_M$. Of course, unlike the case on the true $\beta$-plane, the parameter $\beta_M$ is not a constant, but this is not a particular disadvantage if $\overline{u}_{yy}$ is also varying with $y$.

Having noted the spherical relevance we revert to the $\beta$-plane and seek solutions of (16.120) with the form $\psi' = \widetilde{\psi}(y)\exp(\mathrm{i}kx)$, whence

$$\frac{\mathrm{d}^2\widetilde{\psi}}{\mathrm{d}y^2} = \left(k^2 - \frac{\beta}{\overline{u}}\right)\widetilde{\psi} = \left(k^2 - K_s^2\right)\widetilde{\psi}, \tag{16.124}$$

where $K_s = (\beta/\overline{u})^{1/2}$. From this equation it is apparent that if $k < K_s$ the solution is harmonic in $y$ and Rossby waves may propagate away from their source. On the other hand, wavenumbers $k > K_s$ are trapped near their source; that is, short waves are meridionally trapped by eastward flow.

Without solving (16.124), we can expect an isolated mountain to produce two wavetrains, one for each meridional wavenumber $l = \pm(K_s^2 - k^2)^{1/2}$. These wavetrains will then propagate along a ray, and given the dispersion relation this trajectory can be calculated (usually numerically) using the expressions of the previous section. The local dispersion relation of Rossby waves is

$$\omega = \overline{u}k - \frac{\beta k}{k^2 + l^2}, \tag{16.125}$$

so that their group velocity is

$$c_g^x = \frac{\partial \omega}{\partial k} = \overline{u} - \frac{\beta(l^2 - k^2)}{(k^2 + l^2)^2} = \frac{\omega}{k} + \frac{2\beta k^2}{(k^2 + l^2)^2}, \tag{16.126a}$$

$$c_g^y = \frac{\partial \omega}{\partial l} = \frac{2\beta k l}{(k^2 + l^2)^2}. \tag{16.126b}$$

The sign of the meridional wavenumber thus determines whether the waves propagate polewards (positive $l$) or equatorwards (negative $l$). Also, because the dispersion relation (16.126) is independent of $x$ and $t$, the zonal wavenumber and frequency in the wave group are constant along the ray, and the meridional wavenumber then adjusts to satisfy the local dispersion relation (16.125). Thus, from (16.124), the meridional scale becomes larger as $K_s$ approaches $k$ from above and an incident wavetrain is reflected, its meridional wavenumber changes sign, and it continues to propagate eastwards.

Stationary waves have $\omega = 0$, and the trajectory of a ray is parameterized by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{c_g^y}{c_g^x} = \frac{l}{k}. \tag{16.127}$$

For a given zonal wavenumber the trajectory is then fully determined by this condition and that for the local meridional wavenumber which from (16.125) is

$$l^2 = K_s^2 - k^2. \tag{16.128}$$

Finally, from (16.126) the magnitude of the group velocity is

$$|c_g| = [(c_g^x)^2 + (c_g^y)^2]^{1/2} = 2\frac{k}{K_s}\overline{u}, \tag{16.129}$$

which is double the speed of the projection of the basic flow, $\overline{u}$, onto the wave direction. Given the above relations, and the zonal wind field, we can compute rays emanating from a given source, although the calculation must still be done numerically. One example is given in Fig. 16.17.

### ✦ *A WKB solution*

[Part of this section is redundant with the WKB appendix to chapter 7. xxx]

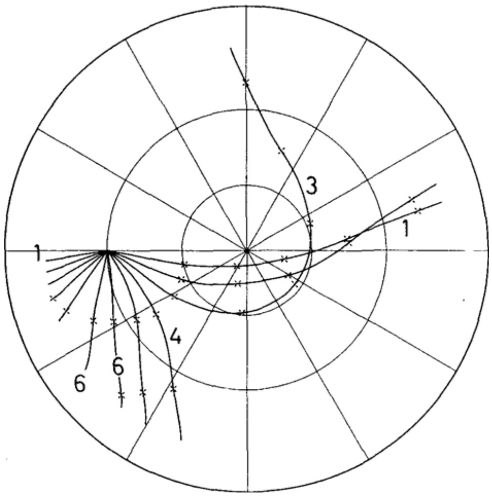


**Fig. 16.17** The rays emanating from a point source at 30° N and 180° (nine o'clock), calculated using the observed value of the wind at 300 mb.[17] The crosses mark every 180° of phase, and mark the positions of successive positive and negative extrema. The numbers indicate the zonal wavenumber of the ray. The ray paths may be compared with the full linear calculation shown in Fig. 16.18.

Some information about the wave amplitudes along a ray can be obtained using a WKB approach, as described in the appendix to chapter **??**. Let us write (16.124) as

$$\frac{\mathrm{d}^2\widetilde{\psi}}{\mathrm{d}y^2} + l^2\widetilde{\psi} = 0 \tag{16.130}$$

where

$$l^2(y) = K_s^2 - k^2. \tag{16.131}$$

If $l(y)$ is sufficiently slowly varying in $y$ (i.e., if $|\mathrm{d}l^{-1}/\mathrm{d}y| \ll 1$) then we may seek a solution of the form

$$\widetilde{\psi} = A\mathrm{e}^{\mathrm{i}g(y)}. \tag{16.132}$$

This leads to an approximate solution for $g(y)$, namely

$$g(y) = \int^y l(y)\,\mathrm{d}y + \frac{\mathrm{i}}{2}\ln l(y), \tag{16.133}$$

and the approximate solution for the stationary streamfunction is then

$$\psi(x, y) = Al^{-1/2}\exp\left[i\left(kx + \int^y l(y)\,\mathrm{d}y\right)\right], \tag{16.134}$$

where $A$ is a constant. Consider, for example, the disturbance excited by an isolated low-latitude peak, with $\overline{u}$ increasing, and so $K_s$ decreasing, polewards of the source. Assuming that initially there exists a zonal wavenumber $k$ less than $K_s$ then two eastward propagating wavetrains are excited. The meridional wavenumber of the poleward wavetrain diminishes according to (16.131), so that, using (16.127), the ray becomes more zonal. At the latitude where $k = K_s$, the 'turning latitude' the wave is reflected but continues propagating eastwards. The southward propagating wavetrain is propagating into a medium with smaller $\overline{u}$ and larger $K_s$. At the critical latitude, where $\overline{u} = 0, l \to \infty$ but $c_g^x$ and $c_g^y$ both tend to zero, but [using (16.126)] in such a way that $c_g^x/c_g^y \to 0$. That is, the rays become meridionally oriented and their speed tends to zero. At this latitude the waves may be absorbed, but the analysis is specialized and beyond our scope.[18] Finally, we mention without derivation that for zonal flows with constant angular velocity the trajectories are great circles.

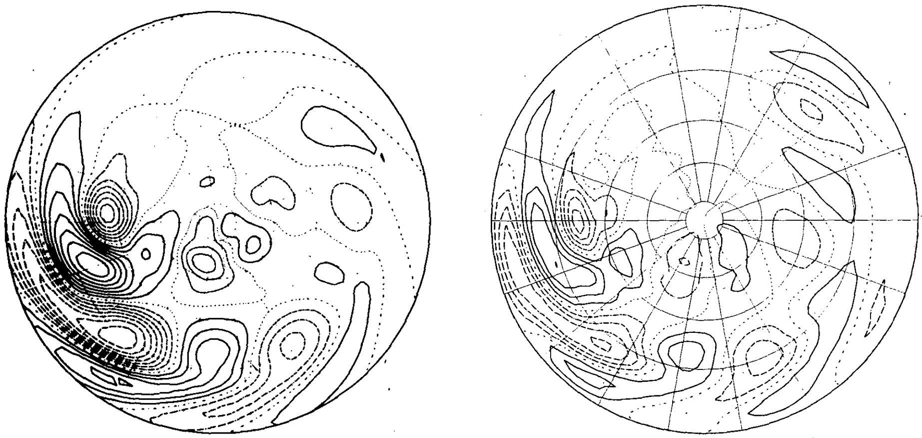

**Fig. 16.18** The linear stationary response induced by circular mountain at 30° N and at 180° longitude (nine o'clock). The figure on the left uses a barotropic model, whereas the figure on the right uses a multi-layer baroclinic model.[20] In both cases the mountain excites a low-wavenumber polar wavetrain and a higher-wavenumber subtropical train.

### 16.9.3 Application to an idealized atmosphere

Given the complexity of the real atmosphere, and the availability of computers, it is probably best to think of the remarks above as helping us interpret more complete numerical, but still linear, calculations of stationary Rossby waves — for example, numerical solutions of the stationary barotropic vorticity equation in spherical coordinates,

$$\frac{\overline{u}}{a\cos\vartheta}\frac{\partial\zeta'}{\partial\lambda} + v'\left(\frac{1}{a}\frac{\partial\overline{\zeta}}{\partial\vartheta} + \beta\right) = -\frac{\overline{u} f_0}{aH\cos\vartheta}\frac{\partial h_b}{\partial\lambda} - r\zeta', \tag{16.135}$$

where $[u, v] = a^{-1}[-\partial\psi/\partial\vartheta, (\partial\psi/\partial\lambda)/\cos\vartheta]$, $\beta = 2\Omega a^{-1}\cos\vartheta$ and $\zeta = \nabla^2\psi$. The last term in (16.135) crudely represents the effects of friction and generally reduces the sensitivity of the solutions to resonances. Solutions to (16.135) may be obtained first by discretizing and then numerically inverting a matrix, and although the actual procedure is quite involved it is analogous to the Fourier methods used earlier for the simpler one-dimensional problem. Such linear calculations, in turn, help us interpret the stationary wave pattern from still more complete models and in the Earth's atmosphere.

Figure 16.18 shows the stationary solution to the problem with a realistic northern hemisphere zonal flow and an isolated, circular mountain at 30° N. The topography excites two wavetrains, both of which slowly decay downstream because of frictional effects, rather like the one-dimensional wavetrain in Fig. 16.13. The polewards propagating wavetrain develops a more meridional orientation, corresponding to a smaller meridional wavenumber $l$, before moving southwards again, developing a much more zonal orientation eventually to decay completely as it meets the equatorial westward flow. The equatorially propagating train decays a little more rapidly than its polewards moving counterpart because of its proximity to the critical latitude. More complicated patterns naturally result if a realistic distribution of topography is used, as we see later in Fig. 16.16. We can see wavetrains emanating from both the Rockies and the Himalayas, but distinct poleward and equatorward wavetrains are hard to discern.

*Notes*

1 Much of this development stems from Charney & Drazin (1961).

2 A quite extensive discussion is given by Pedlosky (1987a).

3 Much of our basic understanding in this area stems from conceptual and numerical work on forced Rossby waves by Charney & Eliassen (1949), who looked at the response to orography using a barotropic model. This was followed by a study by Smagorinsky (1953) on the response to thermodynamic forcing using a baroclinic, quasi-geostrophic model. Seeking more realism later studies have employed the primitive equations and spherical coordinates in studies that are at least partly numerical (e.g., Egger 1976, and a host of others), although most theoretical studies perforce still use the quasi-geostrophic equations. We also draw from various review articles, among them Smith (1979), Dickinson (1980), Held (1983) (particularly for sections 16.7.3 and 16.8) and Wallace (1983). See also the collection in the *Journal of Climate*, vol. 15, no. 16, 2002.

4 To obtain the solutions shown in Fig. 16.10 and Fig. 16.11, the topography is first specified in physical space. Its Fourier transform is taken and the streamfunction in wavenumber space is calculated using (16.87). The inverse Fourier transform of this gives the streamfunction in physical space.

5 The difference between wavetrains emanating from an isolated topographic feature and a global resonant response is relevant for intra-seasonal variability, which might be considered a quasi-stationary response to slowly changing boundary conditions like the sea-surface temperature. If resonance is important, we might expect to see global-scale anomalies, whereas the viewpoint of damped wave-trains is more local. This whole area is one of continuing, active, research with deep roots going back to Namias (1959) and Bjerknes (1959) and beyond.

A different point of view, one that we do not explore in this book, is that the zonally asymmetric features of the Earth's atmosphere are predominantly due to *nonlinear* effects. One possibility is that eddies might significantly modify (and perhaps amplify and sustain) stationary patterns through their large-scale turbulent transfers; see, for example, Green (1977) and Shutts (1983). We could incorporate such effects into a linear model by including the eddy effects as forcing term on the right-hand side of a linear equation such as (16.79), or its two- or three-dimensional analogue, although the forcing term would have to be calculated using a nonlinear theory or taken from observations. Different again is the notion, inspired by models of low-order dynamical systems, that the atmosphere might have *regimes* of behaviour, and that the zonally asymmetric patterns are manifestations of the time spent in a particular regime before transiting to another. See for example Kimoto & Ghil (1993) and Palmer (1997).

6 Because of such difficulties, understanding the effects of sea-surface temperature anomalies on the atmosphere has become largely the subject of GCM experiments, and one plagued with ambiguous results that depend in part on the particular configuration of the GCM. Some of the modelling issues are reviewed by Kushnir *et al.* (2002).

8 From Hoskins & Karoly (1981).

10 From Hoskins & Karoly (1981).

12 Adapted from Held *et al.* (2002).

13 Such solutions are nearly always most easily obtained numerically. One way is to use a Fourier method described earlier. A related method is to write the equations in finite difference form, schematically as $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{F}$, where $\boldsymbol{X}$ is the vector of all the model fields, $\boldsymbol{F}$ represents the known forcing and $\boldsymbol{A}$ is a matrix obtained from the equations of motion and boundary conditions, and solve for $\boldsymbol{X}$. A quite different method is to use a nonlinear time-dependent model, such as a GCM: prescribe or hold steady the zonally averaged zonal flow as well as all the zonally

asymmetric forcing terms, but multiply the asymmetric terms by a small number (e.g., 0.01) to ensure the response is linear; then calculate the steady response by forward time integration, and then divide that solution by the small number to obtain the final solution.

14 The description of the stationary waves in terms of wavetrains comes from Hoskins & Karoly (1981), with some earlier theoretical results having been derived by Longuet-Higgins (1964).

15 Steers (1962) and Phillips (1973).

17 From Hoskins & Karoly (1981).

18 At the critical latitude the WKB analysis fails and both dissipative and nonlinear effects are likely to play a role. See Dickinson (1968) and Tung (1979).

20 From Grose & Hoskins (1979) and Hoskins & Karoly (1981).

## *Problems*

16.1 Consider the barotropic vorticity equation on the $\beta$-plane, with an uneven lower surface, satisfying the equation of motion

$$\frac{\mathrm{D}q}{\mathrm{D}t} = 0, \qquad q = \nabla^2\psi + \beta y + h_b(x, y), \tag{P16.1}$$

where $h_b$ is proportional to the bottom topography, assumed to be small. Linearize this about a constant zonal flow, $U$, and seek *steady-state* solutions of the form $\psi = \mathrm{Re}\,\widetilde{\psi}\mathrm{e}^{\mathrm{i}(kx+ly)}$, with the topography similarly represented. Show that the response is infinite (a resonance) if its wavenumber is equal to that of stationary, free, barotropic Rossby waves.

Suppose that friction is introduced, so that the equation of motion becomes $\mathrm{D}q/\mathrm{D}t = -r\zeta$. Show that the response is now always finite. If the mountain is a single sinusoid, $h_b = H \sin kx$, sketch the response (i.e., the streamfunction field).

16.2 *(a)* Explore the response of the single-layer quasi-geostrophic system to flow over topography. Using Matlab, or otherwise, first obtain a response similar to that in Fig. 16.10. Then vary the frictional time scale, the wavenumber of the stationary Rossby wave and the structure of the topography. Show that when the topography contains a resonant wavenumber a trough in the streamfunction often occurs just downstream of the mountain peak, and that this is to be expected from the analytic solution.

*(b)* Suppose that the basic flow is uniform and *westwards*. Obtain and discuss the form of any stationary solutions on the $f$-plane and $\beta$-plane.

16.3 ✦ Using an atlas, or obtaining the information from the literature or on-line, obtain a rough representation of the Earth's topography at 45° S and express it as a Fourier series. Then obtain (e.g., using Matlab) the barotropic stationary response to this topography — that is, the solution to (16.86). Explore the sensitivity of the solution to variations in $\overline{u}$, to using a different $\overline{u}$ on the left- and right-hand sides of (16.86), to the frictional parameter $r$, and to the deformation radius $L_d$. Artificially flatten the topography over South America and comment on how the results vary. Finally, discuss whether your calculations are qualitatively and quantitatively in accord with observations. [This problem develops a calculation similar to that of the well-known paper of Charney & Eliassen (1949).]

16.4 Obtain an expression analogous to (16.94) for the case with a finite deformation radius ($k_d \neq 0$). Compare the two results and explain the differences, if any.

16.5 Using log-pressure coordinates, show that the surface boundary condition analogous to (16.55) is

$$\frac{\partial}{\partial t}\left(\frac{\partial\psi'}{\partial Z} - \frac{N^2}{g}\psi'\right) + \overline{u}\frac{\partial}{\partial x}\frac{\partial\psi'}{\partial Z} - v'\frac{\partial\overline{u}}{\partial Z} = -\frac{N^2}{f_0}\left(\overline{u}\frac{\partial h_b}{\partial x} + \alpha\nabla^2\psi'\right) \quad \text{at } Z = 0, \tag{P16.2}$$

where here $Z$ is proportional to log-pressure.
*Hint:* Note that the relation between $W = \mathrm{D}Z/\mathrm{D}t$ and the real vertical velocity is $w = (f_0/g)\partial\psi/\partial t + RT/(gH)W$, and choose $H = RT(0)/g$.

16.6 The vertical energy flux in a radiating wave is proportional to $\rho_R \overline{p'w'}$ where the overline denotes a horizontal average and $p'$ and $w'$ are the pressure and the vertical velocity, respectively.

*(a)* For the oscillatory solution, with $m^2 > 0$, and without explicitly invoking group velocity, show that if the energy flux is to be directed upwards then the product $km$ must be positive, where $k$ and $m$ are the zonal and vertical wavenumbers.

*(b)* For the trapped solution with $m^2 < 0$, explicitly show that the vertical propagation of energy is zero.

N. B. In this problem and the next it is important to take the real part of each field properly before evaluating the averages. Relatedly, if $h = \mathrm{Re}\, h_b \mathrm{e}^{\mathrm{i}kx}$ where $h_b = h_{br} + \mathrm{i}h_{bi}$ then $h_b = h_{br}\cos kx - h_{bi}\sin kx$, but with little loss of generality one may choose either $h_{br}$ or $h_{bi}$ equal to zero.

16.7 Obtain an expression for the meridional heat flux associated with the solutions (T.1). In particular, show that for $m^2 > 0$ it is proportional to $|h_b|^2 km/(K_s^2 - K^2)$ and therefore deduce that it is positive for an upwardly propagating wave. Show that for the trapped solutions the meridional heat flux is zero.

16.8 Evaluate the wave activity density (pseudomomentum) associated with the solutions (T.1), and the associated EP flux. Show that the group velocity property is satisfied, and that the transport of wave activity is directed upwards for oscillatory solutions.

16.9 ✦ Obtain the homogeneous solution to (16.97) that, when added to the particular solution, properly satisfies the boundary condition (16.101). Discuss the solution, and in particular show that the total response remains bounded even as the denominator in (16.100) goes to zero.

16.10 *Overturning circulation and downward control*

*(a)* Beginning with momentum and thermodynamic equations in the residual quasi-geostrophic forms

$$\frac{\partial \overline{u}}{\partial t} + \frac{f_0}{\rho}\frac{\partial \psi^*}{\partial z} = \mathcal{F} + \mathcal{D}, \qquad \frac{\partial \overline{b}}{\partial t} + \frac{N^2}{\rho}\frac{\partial \psi^*}{\partial y} = H, \tag{16.3a,b}$$

and using thermal wind relation to eliminate time derivatives, obtain an elliptic equation for $\psi^*$. (You may use Cartesian geometry throughout this problem.)

*(b)* Derive a diagnostic relation that must be satisfied between $H$ (the heating) and $\mathcal{F} + \mathcal{D}$ that must be satisfied in a steady state.

*(c)* ✦ In a steady state, suppose that $\mathcal{F}$ is one-signed and non-zero only in the middle of the domain, and that $\mathcal{D}$ appropriately balances it in an integral sense. Obtain approximate or numerical solutions of your elliptic equation with $\mathcal{D}$ non-zero at the top or with $\mathcal{D}$ non-zero at the bottom of the domain, and $\psi^* = 0$ at the domain boundaries. You may assume the domain has finite vertical extent and that $\rho = 1$. Make and state appropriate assumptions as to the nature of $H$, or anything else, as needed.

*(d)* ✦ Suppose that wave activity is generated near the Earth's surface and propagates to the stratosphere where the waves break. Assume that $\mathcal{D}$ is non-zero only in an isolated region above the region of wave breaking. Discuss whether angular momentum conservation can be satisfied. Compare this case with the one in which $\mathcal{D}$ is non-zero in a surface boundary layer. As examples, suppose that $\mathcal{D}$ is a drag or a viscous force on the zonal wind.

*And beyond it, the deep blue air, that shows*
*Nothing, and is nowhere, and is endless.*
Philip Larkin, High Windows.

# CHAPTER SEVENTEEN

# The Stratosphere

THE *STRATOSPHERE* IS THE REGION OF THE ATMOSPHERE above the troposphere and below the mesosphere; thus, it extends from the tropopause at a height (depending on latitude) of 9–16 km, or a pressure of around 200–300 hPa, to the stratopause at about 50 km or about 1 hPa (see Fig. 15.22 on page 660). In this chapter our goal is, simply put, to provide an introductory explanation of the dynamics governing the structure and variability of the stratosphere. The *middle atmosphere* is the somewhat larger region that also includes the mesosphere, and so that extends up to the mesopause at about 90 km or $2 \times 10^{-3}$ hPa, but we won't consider the mesosphere here.[1]

The outline of this chapter is roughly as follows. We begin with a rather descriptive overview of the stratosphere as a whole. Then, starting in section 17.2, we discuss the Rossby and gravity waves that in many ways serve to drive the circulation. We come back to the circulation itself in section 17.5, focusing mainly on the generation of zonal flows and the meridional residual overturning circulation. We round out the chapter with discussions of two striking examples of stratospheric variability, namely the quasi-biennial oscillation in section 17.7, and extratropical variability and sudden warmings in section 17.8, with these possibly unfamiliar terms to be defined in the sections ahead.

## 17.1 A DESCRIPTIVE OVERVIEW

In the troposphere the stratification is determined by dynamical processes — largely by convection at low latitudes and additionally by baroclinic instability at high latitudes — and the tropopause is the height to which the dynamical activity reaches, as discussed in chapter 15. In contrast, in the stratosphere the temperature is determined to a much greater degree by radiative processes and the dynamics are, compared to those in the tropopause, slow. Over much of the stratosphere the temperature actually increases with height, and this is due to a layer of

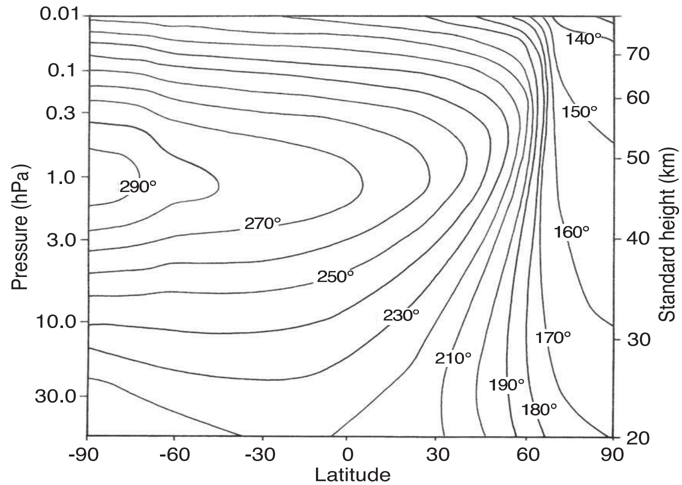


**Fig. 17.1** The zonally averaged radiative-equilibrium temperature in January; that is, the temperature that would arise in the absence of fluid motion in the stratosphere.[3]

ozone that absorbs solar radiation in the mid-stratosphere between about 20 and 30 km. If there were no ozone we would certainly have a tropopause and a stratosphere, but the temperature in the stratosphere would increase much less with height than it in fact does.

The resulting radiative-equilibrium temperature for the month of January is illustrated in Fig. 17.1. The calculation to produce this takes as given the downward solar radiation at the top of the atmosphere and the upward solar and infra-red radiative flux from the troposphere, as well as the physical properties of the fluid, and then calculates the temperature that would ensue without any fluid motion. There is quite a strong lateral gradient in the winter hemisphere and a weaker reversed temperature in the summer hemisphere, and in fact the part of the stratosphere with the highest radiative equilibrium temperature is the upper-stratosphere summer pole, at around 1 hPa. The observed zonally averaged temperature and zonal-wind structure are plotted in Fig. 17.2. From these we infer the following.

- The stratosphere is very stably stratified, with a typical lapse rate corresponding to $N \approx 2 \times 10^{-2}$ s, about twice that of the troposphere on average. This is in part due to the absorption of solar radiation by ozone between 20 and 50 km.
- In the summer the solar absorption at high latitudes leads to a reversed temperature gradient (warmer pole than equator) and, by thermal wind balance, a negative vertical shear of the zonal wind. The temperature distribution is, in fact, not far from the radiative equilibrium distribution, and over much of the summer stratosphere the mean winds are indeed negative (westwards).
- In winter high latitudes receive very little solar radiation and there is a strong meridional temperature gradient and consequently a strong vertical shear in the zonal wind. Nevertheless, this temperature gradient is significantly weaker than the radiative equilibrium temperature gradient, implying a poleward heat transfer by the fluid motions.

The observed fields are a consequence of the dynamics as well as the radiative forcing and we may ask how and why does the dynamics of the stratosphere differ from the tropopause? One main reason is that the higher stratification of the stratosphere tends to weaken the baroclinic instability, and the instability that does occur is at a much larger scale. A typical value

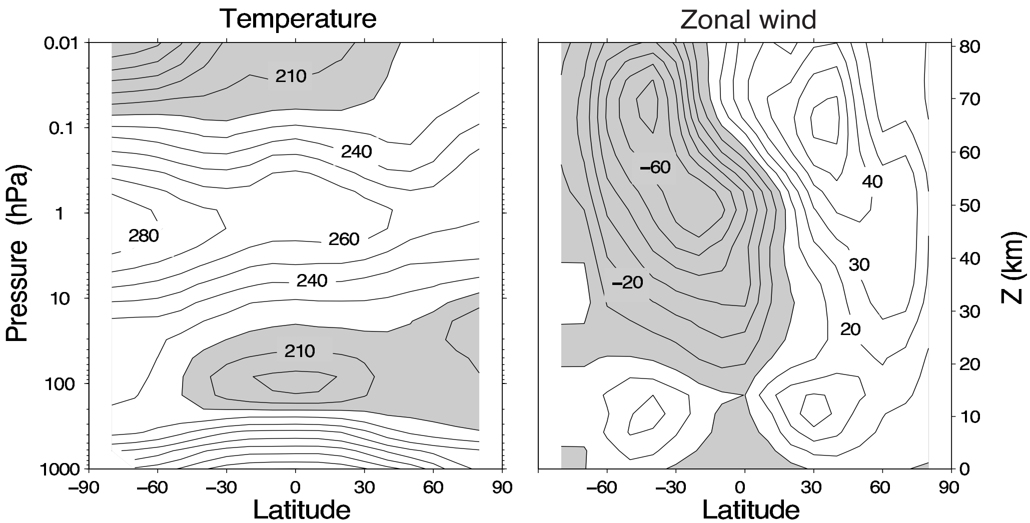


**Fig. 17.2** The zonally averaged temperature and zonal wind in January. The temperature contour interval is 10 K, and values less than 220 K are shaded. Zonal wind contours are 10 m s$^{-1}$ and negative (westward) values are shaded.[5]

of the static stability in the stratosphere is $N \approx 2 \times 10^{-2}\,\mathrm{s}^{-1}$, and using a height scale of 20 km gives a value of the deformation radius $NH/f$ of about 4000 km, as opposed to a canonical 1000 km in the troposphere. Thus, even with the same horizontal temperature gradient as the troposphere, a typical instability scale (of the stratosphere in isolation) is large, perhaps at wavenumber 2 rather than wavenumber 8. The growth rate is also much less than in the troposphere: the Eady growth rate is given by $\sigma_E \equiv 0.31\Lambda H/L_d = 0.31U/L_d$, where $\Lambda$ is the shear, so even when a horizontal temperature gradient is present the growth rate will typically be several times smaller than its tropospheric value. Of course, if baroclinic instability has a modal form then the instability grows at the same rate in the stratosphere as the tropospheric one (it is the same mode), but in this case the higher lapse rate suppresses the amplitude of the stratospheric instability, as shown in Fig. 9.21.

Weak baroclinic instability is not the only process that leads to a circulation in the stratosphere. A second is the propagation of gravity and Rossby waves from the troposphere to the stratosphere. The waves themselves are obviously variable and if and when they break they will drive a mean flow, and a good fraction of this chapter will be devoted to describing that process. But before heading into that topic we'll say a few more words about the circulation itself, and it is convenient to divide it into two parts:

*(i)* A quasi-horizontal circulation.

*(ii)* A meridional overturning circulation that is most usefully described as residual circulation using the TEM formalism.

We now discuss these separately.

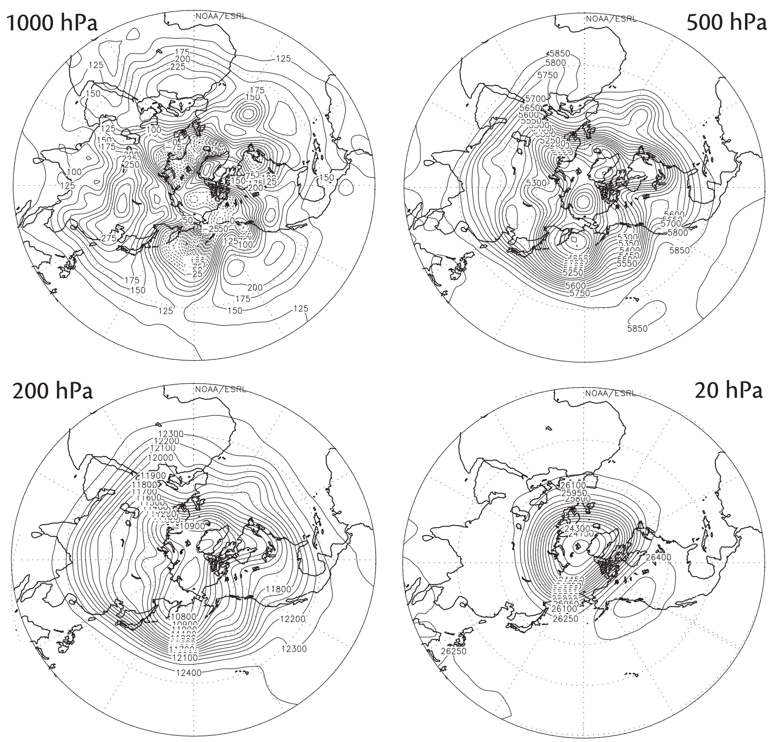


**Fig. 17.3** The geopotential height on February 1, 2000, at various levels in the atmosphere — 1000 and 500 hPa are in the troposphere, 200 hPa is around the tropopause and 20 hPa is in the mid stratosphere, at about 30 km. Note the increase in scale of the synoptic features with height.

### 17.1.1 The quasi-horizontal circulation

In the extra-tropics the stratification is high and the Rossby number small and, at least to the extent that the scales of motion are not truly hemispheric the circulation is well described by the quasi-geostrophic equations. Now, not only does any stratospheric baroclinic instability tend to occur on a large scale, but so does any wave activity that arises from the propagation of Rossby waves up from the troposphere. This is is because of Charney–Drazin filtering, discussed in sections 16.4 and 16.5, summarized in Fig. 16.6: the smaller the wavelength the smaller is the range of zonal winds through which the waves can propagate. If the wind is too high the waves encounter a turning surface, whereas if the wind is too low they encounter a critical layer. Thus, we would expect that the general horizontal scale of motion is larger in the stratosphere that in the troposphere, and this is borne out by inspection of Fig. 17.3, which shows geopotential height at various levels. The complex patterns of the lower and mid troposphere are well filtered, and in mid-stratosphere the pattern is dominated by wavenumbers 1 and 2. Indeed it seems from the figure (which is typical) that much of the motion is concentrated around a *polar vortex*.

Looking at geopotential (which roughly corresponds to a streamfunction) gives a somewhat misleading impression of the lack of activity away from the poles. Here, because diabatic effects occur on a somewhat longer time scale than advective processes, the flow may be characterized

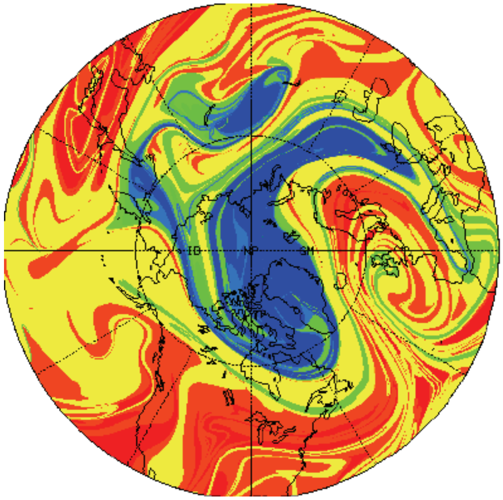

**Fig. 17.4** The tracer distribution in the northern hemisphere lower stratosphere on 28 January 1992. The tracer was initialized on 16 January by setting it equal to the potential vorticity field calculated from an observational analysis, and then advected for 12 days by the observed wind fields.[7]

by the advection of potential vorticity on more slowly evolving isentropic surfaces, as illustrated in Figs. 17.4 and 17.5. Both the potential vorticity and the tracer are evocative of the flows in chapters 11 and 12. We see Rossby waves breaking and vortices stretched into filaments and tendrils, the features of an enstrophy cascade. We also perceive some idea of the spectral non-locality of the enstrophy transfer — a single large vortex overturns and breaks and there is little sense of an orderly cascade of enstrophy to dissipative scales. For this reason, the mid-latitude region is known as the *surf zone*. It is precisely this wave breaking that gives rise to the enstrophy flux to small scales and its dissipation, and which in turn gives rise to the overturning circulation that we discuss below.

As we noted, the surf-zone does not usually extend to the pole, and in winter dense cold air over the pole forms itself into a cyclonic vortex, apparent in both Fig. 17.3 and 17.5. Although the vortex is ultimately the result of diabatic forcing, and has a preferred location, the tendency of quasi-two-dimensional flow to organize itself into vortices (as we see in Figs. 9.6 and 11.9) undoubtedly contributes to its coherence and isolation from the rest of the hemisphere. The boundary of the vortex, as measured by the value of the potential vorticity or of the tracer, is quite sharp with the value of PV often jumping by a factor of 2 or so, and the vortex is quite persistent — it is a near-permanent feature of the winter hemisphere. Within the vortex potential vorticity tends to homogenize, and once formed the main communication that the vortex has with the surf zone is via occasional wave breaking at its boundary. It is interesting that, although the potential vorticity gradient is strong at the edge of the vortex, the exchange of properties is weak, implying a failure of notions of diffusion.

Stable as it is, the polar vortex is nevertheless sometimes disrupted by wave activity from below; this tends to occur when the wave activity itself is quite strong, and when the mean conditions are such as to steer that wave activity polewards. Occasionally, this activity is sufficiently strong so as to cause the vortex to break down, or to split into two smaller vortices, and so allow warm mid-latitude air to reach polar latitudes — an event known as a *stratospheric sudden warming*, and one such is illustrated in Fig. 17.6. We come back to the mechanism of such warmings later in the chapter.

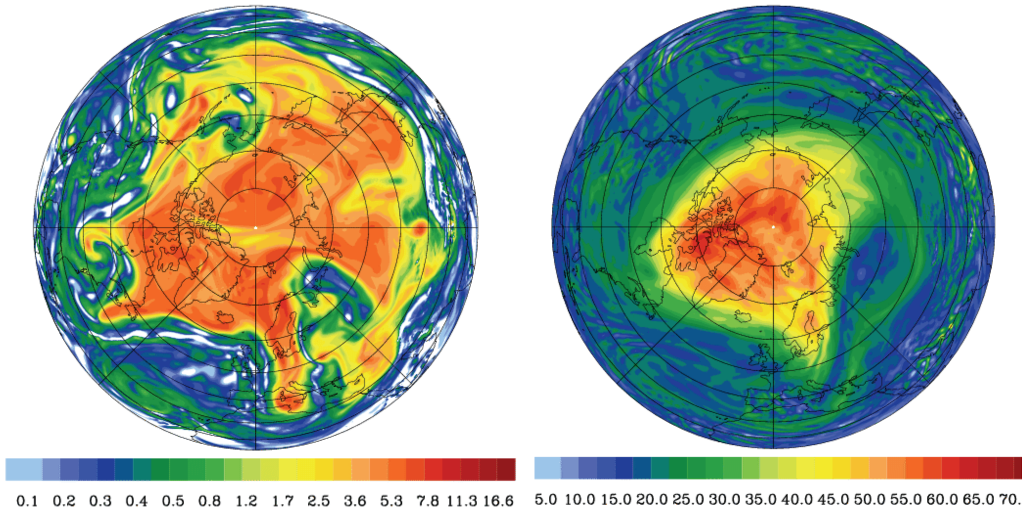


**Fig. 17.5** The potential vorticity on two isentropic surfaces, the 310 K surface (left) and the 475 K surface (right), on January 19, 2005. The shaded bar is in PV units. The 310 K surface is mainly in the troposphere (see Fig. 15.16) where baroclinic instability is abundant. The 475 K surface is at about 20 km altitude, and on it we see a polar stratospheric vortex with a fairly sharp boundary where the PV gradient is high, and a mid-latitude region of smaller-scale features and wave breaking.[9]

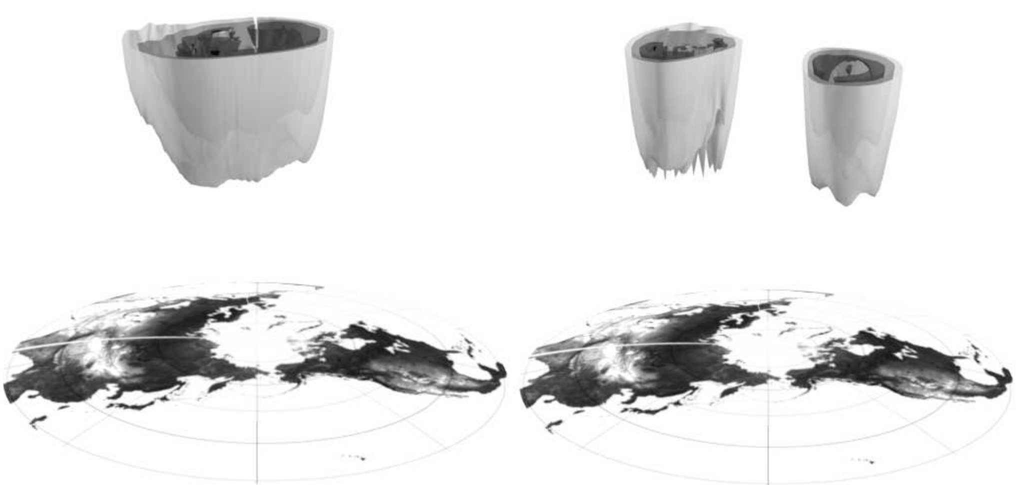

**Fig. 17.6** The edge of the stratospheric polar vortex in 1984. Plotted is the 35 PVU isosurface of $Q^* = Q(\theta/\theta_0)^{4.1}$, where $Q$ is Ertel PV and $\theta_0 = 475\,\mathrm{K}$. The vertical co-ordinate is potential temperature. Like $Q$, $Q^*$ is materially conserved in adiabatic flow. It is approximately constant at the vortex edge, roughly compensating for the change in density with height that affects the Ertel PV. The left panel shows the vortex in a fairly usual state, and the right panel shows a split vortex following a stratospheric sudden warming.[11]

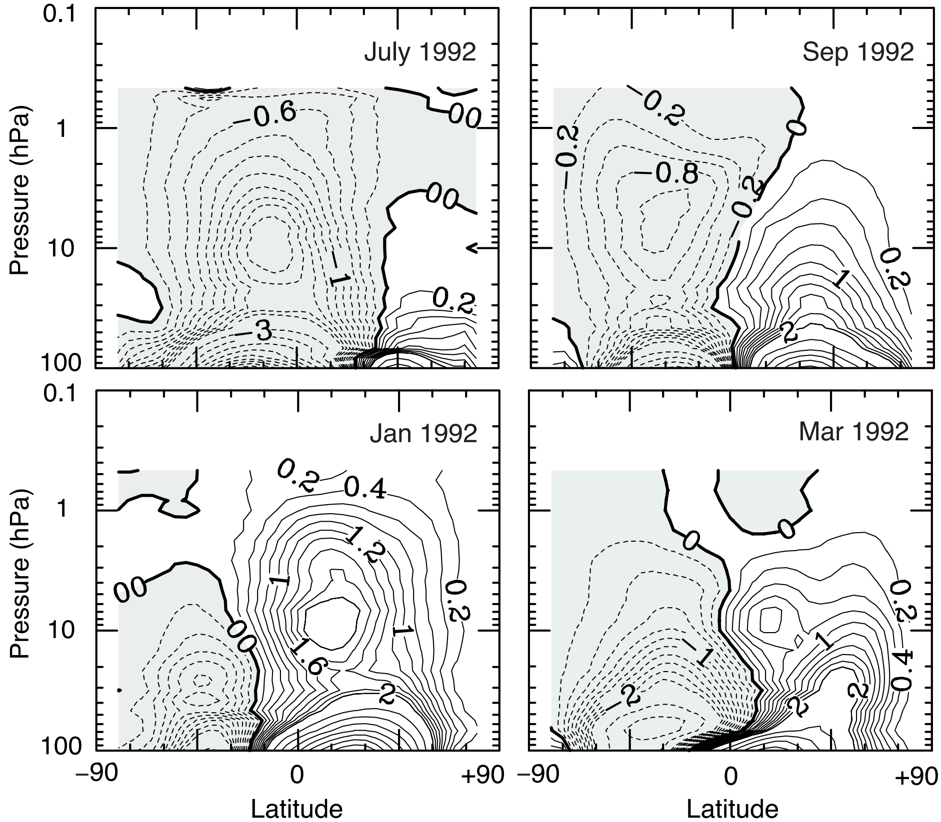


**Fig. 17.7** The observed thickness-weighted (residual) streamfunction in the stratosphere, in Sverdrups ($10^9\,\mathrm{kg\,s^{-1}}$). The circulation is clockwise where the contours are solid. Note the stronger circulation in the winter hemispheres, whereas the equinoctial circulations (September, March) are more inter-hemispherically symmetric.[13]

### 17.1.2 The overturning circulation

That there is a meridional overturning circulation in the stratosphere was inferred by A. Brewer and G. Dobson based on observations of water vapour and chemical transport, and is often called the *Brewer–Dobson circulation*.[14] Brewer and Dobson both inferred the circulation on the basis of tracer transports, rather than performing an Eulerian average of velocity measurements (which would have been impossible then, and is still difficult now). Thus, the circulation they inferred was, in modern parlance, a residual circulation and some modern observations of this circulation are shown in Fig. 17.7. The figure actually shows the observed thickness-weighted circulation, which is almost equivalent to the residual circulation (section 10.3.3), and which so represents both the Eulerian mean and eddy-contributed components. We see a single, equator-to-pole cell in each hemisphere, stronger in the winter hemisphere where it goes high into the stratosphere. There is also a distinct lower branch to the circulation, present in all seasons although strongest in winter, that is confined to the lower stratosphere and is in some ways a vertical extension of (the residual circulation of) the tropospheric Ferrel Cell.

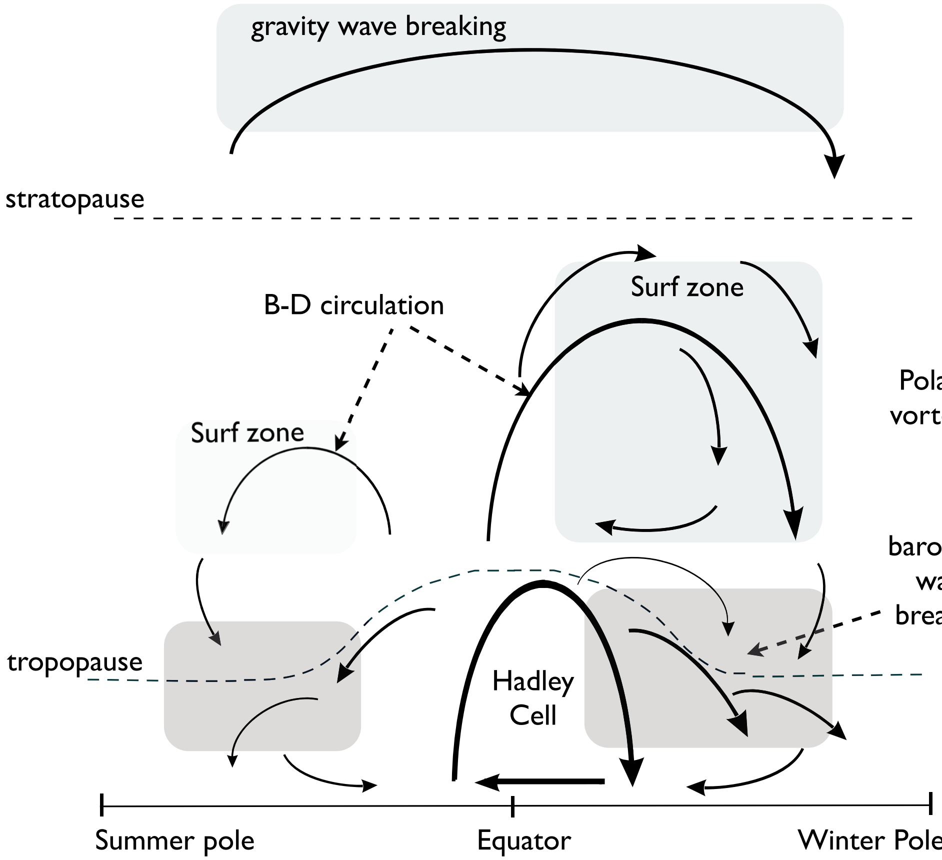


**Fig. 17.8** A schema of the residual mean meridional circulation of the atmosphere. The solid arrows indicate the residual circulation (B-D for Brewer–Dobson) and the shaded areas the main regions of wave breaking (i.e., enstrophy dissipation) associated with the circulation. In the surf zone the breaking is mainly that of planetary Rossby waves, and in the troposphere and lower stratosphere the breaking is that of baroclinic eddies. The surf zone and residual flow are much weaker in the summer hemisphere. Only in the Hadley Cell is the residual circulation comprised mainly of the Eulerian mean; elsewhere the eddy component dominates.

Not all the upper circulation is ventilated by the troposphere — some of it recirculates within the stratosphere. This circulation and some of the associated dynamics is illustrated schematically in Fig. 17.8,[15] and three regions may usefully be delineated: (i) A tropical region; (ii) a mid-latitude region; (iii) the polar vortex. The tropical region is relatively quiescent, an area of upward motion where air is drawn up from the troposphere. In mid-latitudes the residual flow is generally polewards before sinking at high latitudes. In winter the extreme cold leads to the formation of the polar vortex, a strong cyclonic vortex that appears quite isolated from mid-latitudes although, especially in the Northern Hemisphere, it is not always centred over the pole. Obtaining a better understanding this figure is one of the goals of this chapter, and for this we turn to dynamics. Stratospheric dynamics are in large measure dependent on the waves that exist in the stratosphere so first we must discuss them.

## 17.2 WAVES IN THE STRATOSPHERE

Both gravity waves and Rossby waves are important in the stratosphere and we have already discussed aspects of both. In chapters 6 and 16 we looked at Rossby waves, in chapter 7 we looked rather generally at gravity waves, and in chapter 8 we looked at equatorial waves and mixed Rossby-gravity waves. Our goal in this section and the one following is to bring these various strands together and see how they apply to and affect the stratosphere.

We will first write down the appropriate equations of motion and discuss the upward propagation of gravity waves and Rossby waves in mid-latitudes. It is the upward propagation and breaking of Rossby waves in mid-latitudes that is primarily responsible for maintaining the meridional overturning circulation (MOC) in the stratosphere, commonly called the Brewer–Dobson circulation, and we discuss that later in the chapter. In section 17.3 we focus attention on the equatorial region, where Rossby waves and gravity waves are intertwined. It turns out to be the upward propagation of such waves that is responsible for maintaining one of the most dramatic and manifest examples of wave-induced variability in the Earth's atmosphere, namely the quasi-biennial oscillation in the equatorial stratosphere, which we discuss in section 17.7.

### 17.2.1 Linear equations of motion

Because we are dealing explicitly with compressible atmosphere we will use, at least initially, the ideal gas equations rather than the Boussinesq equations, and use log-pressure coordinates, as discussed in section 2.6.3. (Nevertheless we will often find that, to a decent approximation, the equations reduce to the Boussinesq form, with compressibility having only a small effect.) In contrast to our previous treatment of gravity waves we will restrict attention to the hydrostatic case, thereby limiting ourselves to relatively large scales.

On a $\beta$-plane the equations of motion in log-pressure coordinates, linearized about a resting state, may be written as

$$\frac{\partial u}{\partial t} - fv = -\frac{\partial \Phi}{\partial x}, \qquad \frac{\partial v}{\partial t} + fu = -\frac{\partial \Phi}{\partial y}, \tag{17.1a}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{\rho_R}\frac{\partial(\rho_R w)}{\partial z} = 0, \tag{17.1b}$$

$$\frac{\partial}{\partial t}\frac{\partial \Phi}{\partial z} + wN_*^2 = 0. \tag{17.1c}$$

The equations are, respectively, the two horizontal momentum equations, the mass continuity equation and the thermodynamic equation. The notation is as in section 2.6.3 except we use lower case $z$ and $w$; thus, $z = -H\ln(p/p_0)$ where $p_0$ is a constant reference pressure and $H$ is a reference height — for example, $H = RT_0/g$ where $T_0$ is a reference temperature. The density profile $\rho_R$ is an exponential, $\rho_R(z) = \rho_0 \exp(-z/H)$, where we may take $\rho_0 = 1$, and $N_*^2$ is a reference stratification similar to but not exactly the same as the buoyancy frequency; we will take it to be constant and drop the subscript $*$. The Coriolis parameter $f$ varies as $f = f_0 + \beta y$; when we consider equatorial waves we will take $f_0 = 0$, and when we consider gravity waves in mid-latitudes we will take $\beta = 0$.

As in section 16.4 it is convenient to extract that part of the solution that grows exponentially with height, and so seek wave solutions of the form

$$[u, v, w, \Phi] = [\widetilde{u}(y), \widetilde{v}(y), \widetilde{w}(y), \widetilde{\Phi}(y)]\,\mathrm{e}^{z/2H}\,\mathrm{e}^{\mathrm{i}(kx+mz-\omega t)}. \tag{17.2}$$

We cannot assume a simple harmonic form in the $y$-direction because the equations of motion have coefficients that depend on $y$. Substituting (17.2) into (17.1) yields

$$-\mathrm{i}\omega\widetilde{u} - f\widetilde{v} = -\mathrm{i}k\widetilde{\Phi}, \qquad -\mathrm{i}\omega\widetilde{v} + f\widetilde{u} = -\frac{\partial\widetilde{\Phi}}{\partial y}, \tag{17.3a}$$

$$\mathrm{i}k\widetilde{u} + \frac{\partial\widetilde{v}}{\partial y} + \mathrm{i}\left(m + \frac{\mathrm{i}}{2H}\right)\widetilde{w} = 0, \tag{17.3b}$$

$$-\mathrm{i}\omega\left(\frac{1}{2H} + \mathrm{i}m\right)\widetilde{\Phi} + \widetilde{w}N^2 = 0. \tag{17.3c}$$

Perhaps surprisingly, in many situations we can ignore the factor $1/2H$ in this system. Many observed stratospheric waves have a vertical wavelength, $\lambda$, is of order a few kilometers and usually less than 10 km. Also, $T_0 = 240\,\mathrm{K}$ then $H \approx 7\,\mathrm{km}$. The $1/2H$ factor is small when $m \gg 1/2H$ or $4\pi H/\lambda \gg 1$. In fact, it is the square of this ratio that needs to be large, and this is true for all but the deepest stratospheric waves. Compressibility remains in the system because, using (17.2), all the perturbation variables grow exponentially with height, albeit slowly.

In the sections that follow we look some of the waves supported by this system. The analysis is more complicated for equatorial region, where gravity and Rossby waves are intertwined, so we begin with the mid-latitudes.

### 17.2.2 Waves in mid-latitudes

In mid-latitudes there is a good frequency separation between Rossby waves and gravity waves so they can be treated separately; also, since we have already treated both in chapters 7 and 16 our discussion is brief and focuses on their stratospheric relevance.

*Rossby waves*

If we neglect the factor of $1/H^2$ the $x$- and $z$-components of the group velocity are

$$c_g^x = \frac{(k^2 - l^2 - f_0^2 m^2/N^2)\beta}{(k^2 + l^2 + f_0^2 m^2/N^2)^2}, \qquad c_g^z = \frac{2km\beta f_0^2/N^2}{(k^2 + l^2 + f_0^2 m^2/N^2)^2}. \tag{17.4a,b}$$

Since $k < 0$ for Rossby waves then, in order for the waves to be upwardly propagating, (17.4b) requires that $m < 0$. Thus, the lines of constant phase tilt westward with height, as illustrated in Fig. **??**. The ratio of the vertical to the horizontal components of group velocity is not, unlike the case with gravity waves, a simple function of the wavenumbers and it is not possible to determine wither $c_g^x$ is positive or negative without knowing the value of $l$, the meridional wavenumber. To obtain a typical value of the vertical group velocity in the atmosphere we may take $k^{-1} = 1000\,\mathrm{km}$, $m^{-1} = 10\,\mathrm{km}$, $f_0/N = 10^{-2}$, $\beta = 10^{-11}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$, giving $c_g^z \sim 0.1\,\mathrm{m\,s}^{-1} \approx 10\,\mathrm{km/day}$.

Because Rossby waves grow in amplitude as they ascend the linearity assumption will fail and the waves may break, and in doing so they will deposit momentum. It is this wave breaking that is responsible for the production of the stratospheric meridional overturning circulation that we discuss in section 17.5 and 17.6.

*Gravity waves*

Gravity waves may also propagate up into the stratosphere from the troposphere. If we take $f$ to be a constant, $f_0$, then, except for the factor of $\rho_R$ the set (17.1) is the same as the $f$-plane hydrostatic Boussinesq equations, namely (7.155) on page 320. It is a straightforward matter to show that the dispersion relation is

$$\omega^2 = \frac{f^2 m'^2 + (k^2 + l^2)N^2}{m'^2} = f^2 + \frac{N^2(k^2 + l^2)}{m'^2}, \tag{17.5}$$

where $m'^2 = m^2 + 1/4H^2$. As noted above the factor of $1/4H^2$ (which also appeared previously in (7.231)) is often small and we shall ignore it. If we suppose that the horizontal component of the wave vector is aligned with the $x$-axis (i.e. $l = 0$) then the group velocity components are

$$c_g^x = \frac{N^2 k}{\omega m^2} = \frac{N^2}{\omega m}\cos\vartheta, \qquad c_g^z = -\frac{N^2(k^2 + l^2)m}{\omega m^4} = \frac{-N^2}{\omega m}\cos^2\vartheta. \tag{17.6a,b}$$

where $\cos^2\vartheta = k^2/(k^2 + m^2) \approx k^2/m^2 \ll 1$. The above expressions are most easily derived directly from (17.5) but are also the hydrostatic limit of (7.151). The directional aspects of these expressions are the same as those given for the non-rotating case in (7.75) with $\sin\vartheta = 1$, consistent with the hydrostatic limit — indeed we can obtain (17.6) from (7.75) by setting $\cos\lambda = 1, \sin\vartheta = 1$, $\omega = N\cos\vartheta$ and $\kappa = m$. Thus, the relation of the group velocity to the phase speed is much the same as for the gravity waves considered in chapter 7, and in particular we have $c_g^z/c_g^x = -k/m$. If the waves are generated in the troposphere then $c_g^z$ must be positive and so $m$ must be negative. In mid-latitudes there is however no requirement that the horizontal propagation be in any particular direction. The distribution of velocity, pressure and temperature are illustrated in the two panels of Fig. 17.9 for waves with a positive and negative horizontal wavenumber and a negative vertical wavenumber.

## 17.3 WAVES IN THE EQUATORIAL STRATOSPHERE

We now focus attention on the problem of equatorial waves in the stratosphere — both Kelvin waves and Rossby-gravity waves — for these turn out to be particularly important in generating stratospheric variability. One of our goals is to show that there can be vertically propagating waves with both eastward and westward phase speeds, even at relatively low frequencies. We first look at Kelvin waves as these provide a gentle introduction via a simple special treatment, and follow this by a more general treatment that includes Rossby-gravity waves.

### 17.3.1 Kelvin waves

We obtain the Kelvin wave solution by setting $\widetilde{v} = 0$ everywhere in (17.3), whence, after eliminating $\widetilde{w}$, (17.3) straightforwardly becomes

$$\omega\widetilde{u} = k\widetilde{\Phi}, \qquad f\widetilde{u} = -\frac{\partial\widetilde{\Phi}}{\partial y}, \qquad \omega\left(m^2 + \frac{1}{4H^2}\right)\widetilde{\Phi} - N^2 k\widetilde{u} = 0 \tag{17.7a,b,c}$$

Eqs. (17.7a,b) give $\omega\partial\widetilde{u}/\partial y + kf\widetilde{u} = 0$, which upon integration and with $f = \beta y$ yields

$$\widetilde{u}(y) = \widetilde{u}_0 \mathrm{e}^{-\beta y^2/2c_p}, \tag{17.8}$$

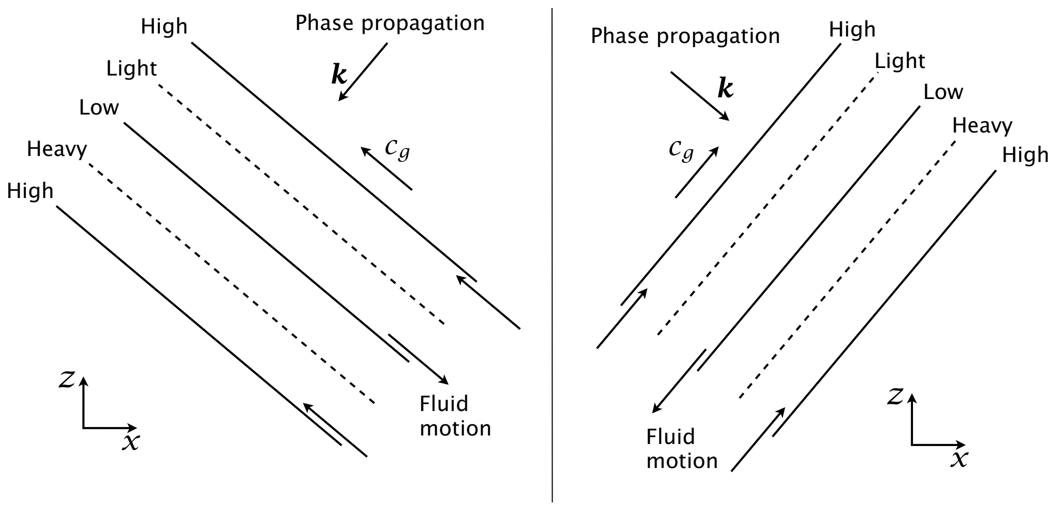


**Fig. 17.9** Phase relationships for two examples of upwardly propagating gravity waves. The sketch on the left shows waves propagating to the left, with $k < 0$, and the one on the right sketch shows waves with $k > 0$. Lines of constant phase are solid and dotted lines, with 'high' and 'low' referring to pressure and 'light' and 'heavy' referring to density and corresponding to warm and cold, respectively. In both sketches $m$ is negative and the group velocity is directed upwards and phase propagates downwards. For hydrostatic flow the phase lines would be nearly horizontal. The figure may be compared with Fig. 7.4.

where $c_p = \omega/k$ and $\tilde{u}_0$ is the value of $\tilde{u}$ at the equator. The exponential fall-off is familiar from our earlier studies of Kelvin waves (e.g., section 3.7.3 and many parts of chapter 8) and requires that $c_p > 0$, meaning that the phase speed of the waves is eastward. Also, since $\omega > 0$ by convention, the $x$-wavenumber is positive (i.e, $k > 0$). The dispersion relation for Kelvin waves follows easily from (17.7a,c) and is

$$\omega^2 = \frac{N^2 k^2}{m^2 + 1/4H^2}. \tag{17.9}$$

Aside from the factor of $1/4H^2$, which in any case is often small compared to $m^2$, (17.9) is essentially the same as the dispersion relation for hydrostatic gravity waves, namely (7.61) on page 297. The zonal and vertical components of the group velocity are

$$c_g^x = \frac{N}{(m^2 + 1/4H^2)^{1/2}}, \qquad c_g^z = \frac{\partial \omega}{\partial m} = \frac{-Nkm}{(m^2 + 1/4H^2)^{3/2}}. \tag{17.10}$$

Now, for upwardly propagating waves (and so for waves that emanate from the troposphere) we require $c_g^z > 0$ and therefore (because $k > 0$) $m < 0$. The combined conditions of $k > 0$ and $m < 0$ means that the phase lines tilt eastward with height, as in the right panel of Fig. 17.9. Finally, note that the frequency of Kelvin waves, unlike inertia-gravity waves, is uninfluenced by rotation and thus, as seen in Fig. 8.2, can extend over a broad range.

### 17.3.2 A more general treatment of equatorial waves

For simplicity let us assume that the scale height $H$ is very large compared to the vertical wavelengths of interest; that is, $m^2 \gg 1/H^2$ and $\exp(-z/H) = 1$. Eqs. (17.1b) and (17.1c) then combine to give

$$\frac{\partial}{\partial t}\frac{\partial^2 \Phi}{\partial z^2} - N^2\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0. \tag{17.11}$$

If we assume a vertical structure of the form $\Phi(x, y, z, t) = \widetilde{\Phi}(x, y, t)\exp(imz)$, and similarly for $u$ and $v$, then we obtain

$$\frac{\partial \widetilde{\Phi}}{\partial t} + \frac{N^2}{m^2}\left(\frac{\partial \widetilde{u}}{\partial x} + \frac{\partial \widetilde{v}}{\partial y}\right) = 0, \tag{17.12a}$$

with corresponding momentum equations

$$\frac{\partial \widetilde{u}}{\partial t} - f\widetilde{v} = -\frac{\partial \widetilde{\Phi}}{\partial x}, \qquad \frac{\partial \widetilde{v}}{\partial t} + f\widetilde{u} = -\frac{\partial \widetilde{\Phi}}{\partial y}. \tag{17.12b,c}$$

Evidently, (17.12) are isomorphic to the linear shallow water equations (8.20) on page 348 with the replacement $c^2 = N^2/m^2$, where, in (8.20), $c^2 \equiv gH_e$ where $H_e$ is the equivalent depth. (Note that $c$ is not necessarily the phase speed in this problem; we will denote that as $c_p$.)

All of the machinery following (8.20) can then be applied to (17.12), with Rossby-gravity waves and Kelvin waves emerging in the same way (indeed the Kelvin waves identified above emerge as a special case); thus, in what follows we draw directly from section 8.2. To avoid conflicts in notation, we use $m$ to denote the vertical wavenumber and $n$ to denote the order of the Hermite function, akin to a meridional wavenumber.

#### *Rossby-gravity waves*

The dispersion relation that emerges from (17.12) is, by analogy with (8.39b), (8.56) and (8.66),

$$\omega^2 - c^2k^2 - \beta\frac{kc^2}{\omega} = (2n+1)\beta c, \qquad n > 0, \tag{17.13a}$$

$$\omega^2 - \omega kc - \beta c = 0, \qquad n = 0, \tag{17.13b}$$

$$\omega = ck. \qquad n = -1. \tag{17.13c}$$

where $c = N/m$. The case with $n = 0$ is the Yanai wave (a Rossby-gravity wave) and the cases with $n \geq 1$ are planetary waves or gravity waves. The '$n = -1$' case is the Kelvin wave, and all cases are illustrated in Fig. 8.2.

From (17.13) we may in principle obtain the vertical group velocity $\partial\omega/\partial m$. However, it is more informative to proceed by noting that for waves whose origin is in the troposphere we may think of the frequency as being given and (17.13) as providing a condition on the vertical wavenumber. This in turn suggests that we nondimensionalize the wavenumbers by defining

$$\widehat{m} = \frac{\omega^2}{\beta N}m = \frac{\omega^2}{\beta c}, \qquad \widehat{k} = \frac{\omega}{\beta}k, \tag{17.14}$$

with the hats denoting nondimensional variables. Eqs. (17.13) become

$$\widehat{m}^2 - (2n+1)\widehat{m} - \widehat{k}^2 - \widehat{k} = 0, \qquad n > 0, \tag{17.15a}$$

$$\widehat{m} - \widehat{k} - 1 = 0, \qquad n = 0 \tag{17.15b}$$

$$\widehat{m} = \widehat{k}, \qquad n = -1. \tag{17.15c}$$

These equations define a set of curves in $\widehat{m}$–$\widehat{k}$ space that are similar to the curves in $\omega$–$k$ space defined by (17.13), although (17.15) are lower order. Eq. (17.15a) has the solution

$$\widehat{m} = \left(n + \tfrac{1}{2}\right) \pm \left[\left(\widehat{k} + \tfrac{1}{2}\right)^2 + n(n+1)\right]^{1/2}, \tag{17.16}$$

and the complete set of curves is plotted in Fig. 17.10. The curves at the top and bottom are gravity waves, corresponding to the positive sign in (17.16), and the planetary waves are the curves just to the left of the origin, corresponding to the negative sign. The $n = 0$ curve (the Yanai wave) and the $n = -1$ curve (the Kelvin wave) are labelled.

We can infer the group velocity from the figure by noting that, since $\beta$ and $N$ are constant, the curves are contours of constant frequency. The group velocity is by definition the gradient of frequency in wavenumber space and so is at right angles to these curves, and is marked by short arrows. How can we determine the direction of the arrows from the curves? It is because the group velocity is directed toward a higher frequency. [Is there a better explanation here, dear reader? xxx] For waves propagating up from the troposphere the group velocity must be upward, and therefore have negative $m$, as we found earlier for the Kelvin wave.

### 17.3.3 Observational evidence

[Dear Reader: Do you have a nice reference or a figure you would like to contribute? Perhaps showing that there really are gravity and Rossby waves in the stratosphere? xxx]

## 17.4 WAVE MOMENTUM TRANSPORT AND DEPOSITION

We now consider the effects momentum transport by waves on the zonal mean flow in the stratosphere. Such transport is responsible both the the quasi-biennial oscillation (QBO) in the equatorial stratosphere and the stratospheric meridional overturning circulation, two phenomena that we consider at length later on. We begin with Rossby waves.

### 17.4.1 Rossby waves

We have seen that Rossby wave arise both in mid-latitudes and in the tropics. Frequency separation is greater in mid-latitudes but dynamics the same etc etc....

The vertical momentum transport of zonal momentum by eddies is given in part by $\overline{w'u'}$, where an overbar denotes a zonal average. However, directly evaluating this expression from the quasi-geostrophic equations is not particularly simple because $w$ is not a first order variable — it results from the divergence of the ageostrophic horizontal velocities. Furthermore, under quasi-geostrophic scaling we actually neglect the vertical eddy flux divergences, but nevertheless the eddy fluxes may certainly make themselves felt aloft, among other things by generating a form stress that acts to transfer momentum vertically and (relatedly) by generating a meridional overturning circulation, as summarized in the shaded box on page 454.[16]

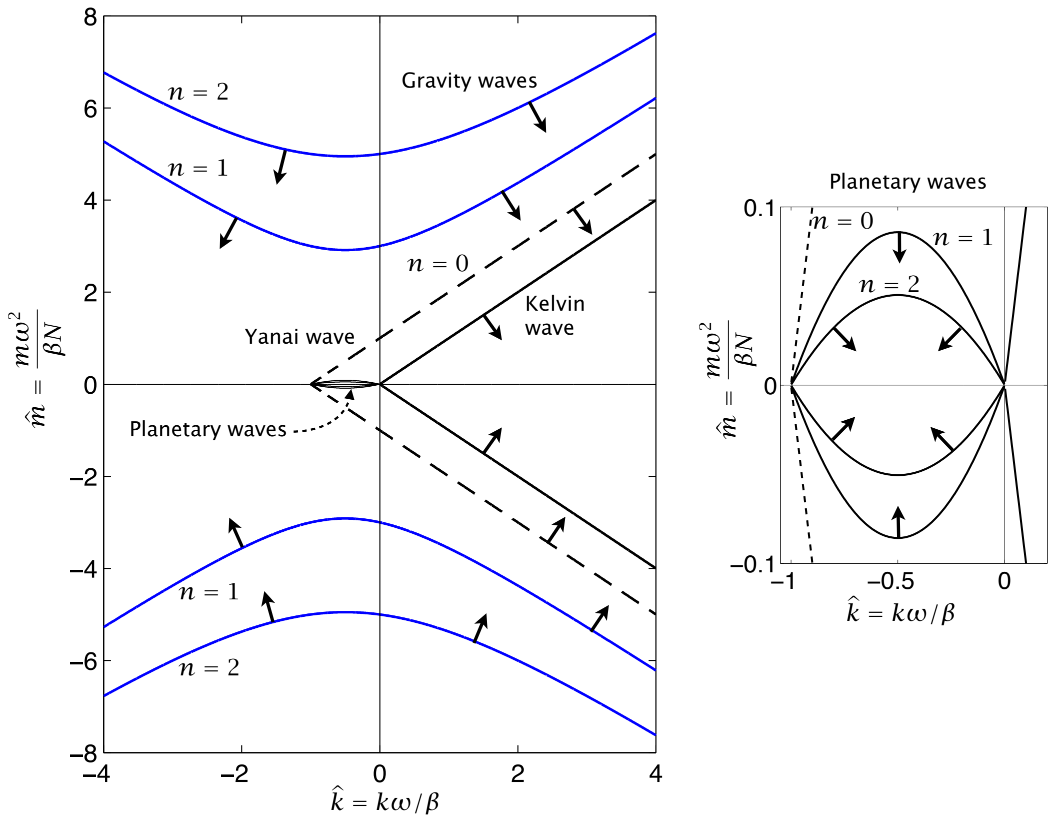


**Fig. 17.10** The dispersion relation plotted in $\widehat{m}$–$\widehat{k}$ space, using (17.15). Shown are gravity waves, the Yanai wave and the Kelvin wave for positive and negative $\widehat{m}$, with the plot at the right showing a magnification of the region near the origin. The arrows in the figure indicate the group velocity, which, being the gradient of the frequency, is perpendicular to the curves. Upward propagating waves occur for negative $m$. Compare with Fig. 8.2.

The better way to proceed is to use the transformed Eulerian mean (TEM) framework of section 10.3, for which the inviscid and adiabatic zonally-averaged momentum and thermodynamic equations are

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \overline{v'q'}, \qquad \frac{\partial \overline{b}}{\partial t} + N^2 \overline{w}^* = 0, \tag{17.17a,b}$$

where potential vorticity flux is related to the EP flux, $\boldsymbol{\mathcal{F}}$, by

$$\overline{v'q'} = \nabla \cdot \boldsymbol{\mathcal{F}}, \qquad \boldsymbol{\mathcal{F}} = -\overline{u'v'}\,\mathbf{j} + \frac{f_0}{N^2}\overline{v'b'}\,\mathbf{k}, \tag{17.18}$$

and for simplicity we use the Boussinesq equations with $b$ as buoyancy. Now, if the eddy fluxes are due to the presence of waves that satisfy a dispersion relation then, as shown in section 10.2.2, the EP flux is related to the group velocity by

$$\boldsymbol{\mathcal{F}} = (\mathcal{F}^y, \mathcal{F}^z) = \mathbf{c}_g \mathcal{A}. \tag{17.19}$$

Combining the above equations we obtain

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \frac{\partial}{\partial y}(c_g^y \mathcal{A}) + \frac{\partial}{\partial z}(c_g^z \mathcal{A}) = \nabla \cdot \boldsymbol{\mathcal{F}}. \tag{17.20}$$

Now, reprising (10.29a), the wave activity itself satisfies a conservation law of the form

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \boldsymbol{\mathcal{F}} = \mathcal{D}, \tag{17.21}$$

where $\mathcal{A} = \overline{q'^2}/2\beta$, which is a positive quantity and $\mathcal{D}$ represents dissipation. Evidently if $\mathcal{D} = 0$ and the waves are steady then $\nabla \cdot \boldsymbol{\mathcal{F}} = 0$ so that the left hand side of (17.20) is zero. (This is close to the non-acceleration result of section 10.5.2.) Evidently in order to get a zonal flow acceleration or an MOC we need to invoke some dissipative or time-dependent processes.

Consider the case in which Rossby waves propagate up from the troposphere, with $c_g^z > 0$. Suppose that there is some dissipation in the system (and/or that Rossby waves break as they ascend) and that wave activity $\mathcal{A}$ falls with height, and suppose further that we are in a steady state. In this case $\nabla \cdot \boldsymbol{\mathcal{F}} < 0$ and from (17.17a) this will produce a mean flow deceleration (i.e., a westward acceleration) and/or a polewards residual flow. The balance between these two possibilities is discussed in section 17.6, but suffice it to say for now that near the equator a zonal acceleration is the more likely outcome than a meridional circulation. Why is there a preferred sense of acceleration when the waves break? Ultimately it is because of the beta effect which distinguishes east from west, and the wave activity (or pseudomomentum) $\mathcal{A}$ is proportional to beta. The beta effect leads to a particular orientation of the phase of Rossby waves and this carries westward momentum away from the source region. When the waves break, be it in the tropospheric subtropics or in the stratosphere, a westward momentum is deposited.

### 17.4.2 Gravity and Kelvin waves

Now consider the vertical momentum transport in gravity waves. If these are uninfluenced by rotation, or if they reside on the $f$-plane, then there is no preferred horizontal direction of propagation. Kelvin waves, on the other hand, propagate their phase eastwards only. The upwards transport of momentum from waves originating in the troposphere can occur only for waves with a positive group velocity, and thus for either of the examples illustrated in Fig. 17.9. Those waves that propagate phase westward (i.e., have $k < 0$) have $\overline{u'w'} < 0$, and those that propagate phase eastward, such as equatorial Kelvin waves, have $\overline{u'w'} > 0$. The contribution to the zonal flow acceleration by the wave transport is given by

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial}{\partial z}\overline{u'w'} + \text{other terms}. \tag{17.22}$$

and if the amplitude of the waves stays constant with height then no mean flow acceleration is induced. However, if the amplitude diminishes with height, because of dissipative processes, then the waves that have a westward (eastward) phase propagation will cause the zonal flow to accelerate westward (eastward). Thus *the dissipation of Kelvin waves as they propagate vertically will cause an eastward acceleration of the zonal flow.*

### 17.4.3 ✦ The processes of wave attenuation

We now explicitly consider the dissipation of waves and the associated momentum deposition as gravity waves propagate vertically. (This section is marked with a black diamond because it is algebraically involved, but the results are needed to understand later sections on the quasi-biennial oscillation. The reader may wish to skim section 16.3 before proceeding, for there we consider the similar but algebraically simpler problem of Rossby wave absorption near a critical layer.) To keep the algebra manageable we will consider the propagation of two-dimensional ($x$–$z$) gravity waves in a Boussinesq fluid uninfluenced by rotation; the processes whereby Kelvin waves and Rossby waves deposit momentum are largely similar but the derivation is more involved. The momentum and buoyancy equations, linearized about a zonal flow $U(z)$ and constant stratification $N^2$, are

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} = -\frac{\partial \phi}{\partial x}, \qquad \frac{\partial w}{\partial t} + U\frac{\partial w}{\partial x} = -\frac{\partial \phi}{\partial z} + b, \tag{17.23a}$$

$$\frac{\partial b}{\partial t} + U\frac{\partial b}{\partial x} + wN^2 = -\alpha b. \tag{17.23b}$$

We include a damping term, $-\alpha b$, where $\alpha$ is a constant, in the buoyancy equation but neglect viscous effects in the momentum equation. We will generally suppose that the damping is small, with $r \ll k(U - c)$. If we cross-differentiate the momentum equation and use the mass continuity equation ($\partial u/\partial x + \partial w/\partial z = 0$) we obtain the linear vorticity equation

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)\nabla^2\psi + \frac{\partial \psi}{\partial x}\frac{\mathrm{d}^2 U}{\mathrm{d}z^2} = \frac{\partial b}{\partial x}, \tag{17.23c}$$

where $\psi$ is such that $w = \partial\psi/\partial x$ and $u = -\partial\psi/\partial z$. We seek solutions of (17.23c,b)in the form

$$[\psi, b] = [\widetilde{\psi}(z), \tilde{b}(z)]\,\mathrm{e}^{\mathrm{i}k(x-ct)}, \tag{17.24}$$

whence

$$\mathrm{i}(-kc + Uk)\left(-k^2\psi + \frac{\mathrm{d}^2\widetilde{\psi}}{\mathrm{d}z^2}\right) + ik\psi\frac{\mathrm{d}^2 U}{\mathrm{d}z^2} = ik\tilde{b} \tag{17.25a}$$

$$\mathrm{i}(-kc + Uk)\tilde{b} + \mathrm{i}kN^2\widetilde{\psi} = -\alpha\tilde{b}. \tag{17.25b}$$

These two equations combine to give

$$\frac{\mathrm{d}^2\widetilde{\psi}}{\mathrm{d}z^2} + m^2(z)\widetilde{\psi} = 0, \tag{17.26a}$$

$$m^2(z) = \left[\frac{N^2[1 + \mathrm{i}r/k(U - c)]}{(U - c)^2 + r^2/k^2} - k^2 - \frac{\mathrm{d}^2 U/\mathrm{d}z^2}{(U - c)}\right]. \tag{17.26b}$$

There are a few notable aspects to this equation.

*(i)* It is an equation for the vertical structure of the streamfunction and has the same form as (7.116) on page 311, where we discussed internal waves in non-constant stratification. If $U$ and $N^2$ were constant in (17.26) then $m$ would be constant and its real part would be the vertical wavenumber.

*(ii)* There is an imaginary component to $m$ which can be expected to cause the solution to decay in the vertical. In most circumstances the decay is slow but in the neighbourhood of a critical layer where $c = U$ then the decay will be rapid. The wave can be expected to deposit momentum and accelerate or decelerate the mean flow, depending on the direction of the phase propagation of the wave. When $r = 0$ there is no dissipation and no deposition.

*(iii)* If $m$ varies sufficiently slowly we can obtain an approximate solution by WKB methods. The methodology will break down at the critical layer but nevertheless it is insightful, so let us proceed.

*WKB solution and momentum flux*

The WKB solution (see the appendix to chapter 7) to (17.26) is

$$\widetilde{\psi}(z) = Am^{-1/2} \exp\left(\pm \mathrm{i} \int^z m \, \mathrm{d}z'\right), \tag{17.27}$$

where $A$ is a constant. The wave momentum flux, $F$, associated with the wave is

$$F_k(z) = \overline{u'w'} = -\mathrm{i}k\left(\psi \frac{\partial \psi^*}{\partial z} - \psi^* \frac{\partial \psi}{\partial z}\right), \tag{17.28}$$

where the overbar denotes a zonal average and the right-hand side is always real, and the subscript on $F$ indicates we are considering the effects of a single wave of wavenumber $k$.

There are two other other simplifying approximations to make. First, we are concerned with low aspect ratio flows ($L_z/L_x \ll 1$) so that the factor of $k^2$ is small compared to $m^2$ and may be neglected. Second, we assume that the variation of the mean flow occurs on a long vertical scale compared to $m$, and so neglect the term in $\mathrm{d}^2U/\mathrm{d}z^2$. With these approximations (17.26b) becomes, if $\alpha/k(U-c) \ll 1$,

$$m(z) \approx \left[\frac{N^2[1 + \mathrm{i}r/k(U-c)]}{(U-c)^2 + \alpha^2/k^2}\right]^{1/2} \approx \frac{N}{U-c}\left[1 + \frac{\mathrm{i}\alpha}{2k(U-c)}\right] \tag{17.29}$$

Now, the fact that $m$ varies only slowly with $z$ (specifically $m^2 \gg |\mathrm{d}m/\mathrm{d}z|$) means that when we take the vertical derivative of $\widetilde{\psi}$ we can ignore the derivative of the vertical derivative of the amplitude, $Am^{-1/2}$. Given this, and using (17.27) and (17.29) in (17.28) we obtain

$$F_k(z) = F_0 \exp\left(\mathrm{i} \int_0^z (m - m^*) \, \mathrm{d}z'\right) = F_0 \exp\left(\int_0^z \frac{-N\alpha}{k(U-c)^2} \, \mathrm{d}z'\right). \tag{17.30}$$

where $F_0$ is the value of the flux at $z = 0$ and we have chosen the sign in the exponent to be appropriate for upwardly propagating gravity waves. The integrand in the right-most expression is the attenuation rate of the wave and, referring to (7.81) on page 303, it can be written as

$$\text{Attenuation rate} = \frac{\alpha}{k(U-c)^2/N} = \frac{\text{Dissipation rate}}{\text{Vertical group velocity}}. \tag{17.31}$$

Essentially, as the wave approaches a critical layer it stalls, giving the dissipative processes more time to act. The result of (17.31) is a general one; we found an almost identical result when looking at the absorption of Rossby waves in section 16.3 — see (16.33). (The dissipation rate in expressions like (16.33) and (17.31) that of wave activity, which in the gravity wave case equals to the thermal dissipation rate.)

*Effect on the mean flow*

If a wave propagating upwards is attenuated there will be a divergence in the eddy momentum flux associated with that wave. In particular, if $r$ in (17.31) is nonzero then momentum flux deposition will increase rapidly as a critical layer is approached (although we cannot expect the derivation to hold at the critical layer itself), $\partial F_k/\partial z$ will be nonzero and the zonal mean flow will be accelerated or decelerated. For definiteness, consider a wave propagating upwards with a positive phase speed (so $m < 0$ and $k > 0$). From (17.30) $F_k$ diminishes with height and the mean flow is accelerated eastward. Similarly, absorption of a wave of negative phase speed leads to a negative, or westward mean-flow acceleration. It is not inconceivable to imagine that the wave deposition will affect the mean flow to an extent that the position of the deposition is significantly altered, leading to interesting dynamical behaviour. Indeed this is precisely what happens in the quasi-biennial oscillation of the equatorial stratosphere. But before we discuss variability let us discuss the maintenance of the mean state.

## 17.5 PHENOMENOLOGY OF THE RESIDUAL OVERTURNING CIRCULATION

We now return to a discussion of the general circulation of the stratosphere and in particular the maintenance of the residual meridional overturning circulation (RMOC), that is the Brewer–Dobson circulation. We expect the circulation to be a consequence of waves coming up from the troposphere and breaking, with both tropospheric baroclinic instability and flow over zonal asymmetries being sources of wave activity. We may naturally ask such questions as whether wavebreaking can give rise to a circulation of the right strength and the right sense, whether it will give the correct seasonal variability, and what determines vertical extent of the circulation. We begin with some elementary theory and phenomenology, for we will find that it will explain a number of the main features of the RMOC. The advanced or confident reader may skip ahead to section 17.6.

### 17.5.1 Wave breaking and residual flow

The equations of motion governing the mean fields are the zonally averaged momentum and thermodynamic equations, which with quasi-geostrophic scaling and in residual form (see section 10.3.1) may be written as

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \nabla \cdot \boldsymbol{\mathcal{F}} + \overline{F}, \tag{17.32a}$$

$$\frac{\partial \overline{\theta}}{\partial t} + \frac{\partial \overline{\theta}}{\partial z} \overline{w}^* = \overline{J}, \tag{17.32b}$$

where $\overline{F}$ represents frictional effects (for example, due to small-scale turbulence) and $\overline{J}$ represents heating, and on the $\beta$-plane the residual velocities are related to the Eulerian means by

$$\overline{v}^* = \overline{v} - \frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R \frac{\overline{v'\theta'}}{\partial_z \overline{\theta}}\right), \qquad \overline{w}^* = \overline{w} + \frac{\partial}{\partial y}\left(\frac{\overline{v'\theta'}}{\partial_z \overline{\theta}}\right). \tag{17.33}$$

The vector $\boldsymbol{\mathcal{F}}$ is the Eliassen–Palm flux, and this is related to the meridional flux of potential vorticity by $\nabla \cdot \boldsymbol{\mathcal{F}} = \overline{v'q'}$. The wave activity itself obeys the Eliassen–Palm relation

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \boldsymbol{\mathcal{F}} = \mathcal{D}, \tag{17.34}$$

where $\mathcal{A}$ is the wave activity, $\boldsymbol{\mathcal{F}}$ is its flux and $\mathcal{D}$ is its dissipation.

From the autumn to the spring, the zonal wind in the stratosphere is generally receptive to planetary-scale Rossby waves propagating up from the troposphere (Fig. 16.6), although at high latitudes in winter there may be a period when the eastward zonal winds are too strong for waves to propagate. If these waves break in the stratosphere then there will be an enstrophy flux to small scales and dissipation. In a statistically-steady state and with small frictional effects the dominant balance in the zonal momentum equation (17.32a) is

$$-f_0 \overline{v}^* \approx \overline{v'q'}, \tag{17.35}$$

where $\overline{v}^*$ is the residual velocity and the potential vorticity flux on the right-hand side is induced by the Rossby wave breaking. In dissipative regions the zonally averaged potential vorticity flux will tend to be down its mean gradient and, if the potential vorticity gradient is polewards (largely because of the $\beta$-effect), the residual velocity will be positive if $f_0$ is positive. That is, the residual flow will be *polewards,* in both hemispheres, and the mechanism giving rise to this is called the 'Rossby wave pump'. Put another way, Rossby waves propagating up from the troposphere break and deposit westward momentum in the stratosphere, and in the mean this wave drag is largely balanced by the Coriolis force on the polewards residual meridional circulation.

This meridional circulation is weakest in summer mainly because linear Rossby waves cannot propagate upwards through the westward mean winds, as illustrated in Fig. 17.11. It is quite striking how the EP vectors avoid the region of westward winds in the summer hemisphere, even though the level of wave activity at low elevations is relatively similar in the summer and winter hemispheres (look between 10 and 15 km in the figure). We can interpret this by noting that for nearly plane waves the EP flux obeys the group velocity property, meaning that $\boldsymbol{\mathcal{F}} = \boldsymbol{c}_g \mathcal{A}$; however, as discussed in section 16.4, if the mean winds are westward the waves evanesce instead of propagate, and thus almost the entire summer hemisphere is shielded from upwardly propagating waves, leaving it in a near-radiative equilibrium state. In the other seasons, the EP flux is able to propagate into the stratosphere and a circulation is generated. This acts to weaken the pole–equator temperature gradient, as we see by inspection of the thermodynamic equation: if the heating is represented by a simple relaxation to a radiative equilibrium state, $\theta_E$, then in a steady state we have

$$N^2 \overline{w}^* = \frac{\theta_E - \theta}{\tau}. \tag{17.36}$$

Poleward flow in mid-latitudes must be supplied by rising air at low latitudes, and sinking air at high latitudes. Thus, from autumn to spring, at low latitudes we have $\theta < \theta_E$ and at high latitudes $\theta > \theta_E$.

Although cause and effect can be very difficult to disentangle in fluid dynamical problems, and the ultimate cause of nearly all fluid motions in the atmosphere is the differential heating from the Sun, it is important to realize that the meridional overturning in the stratosphere

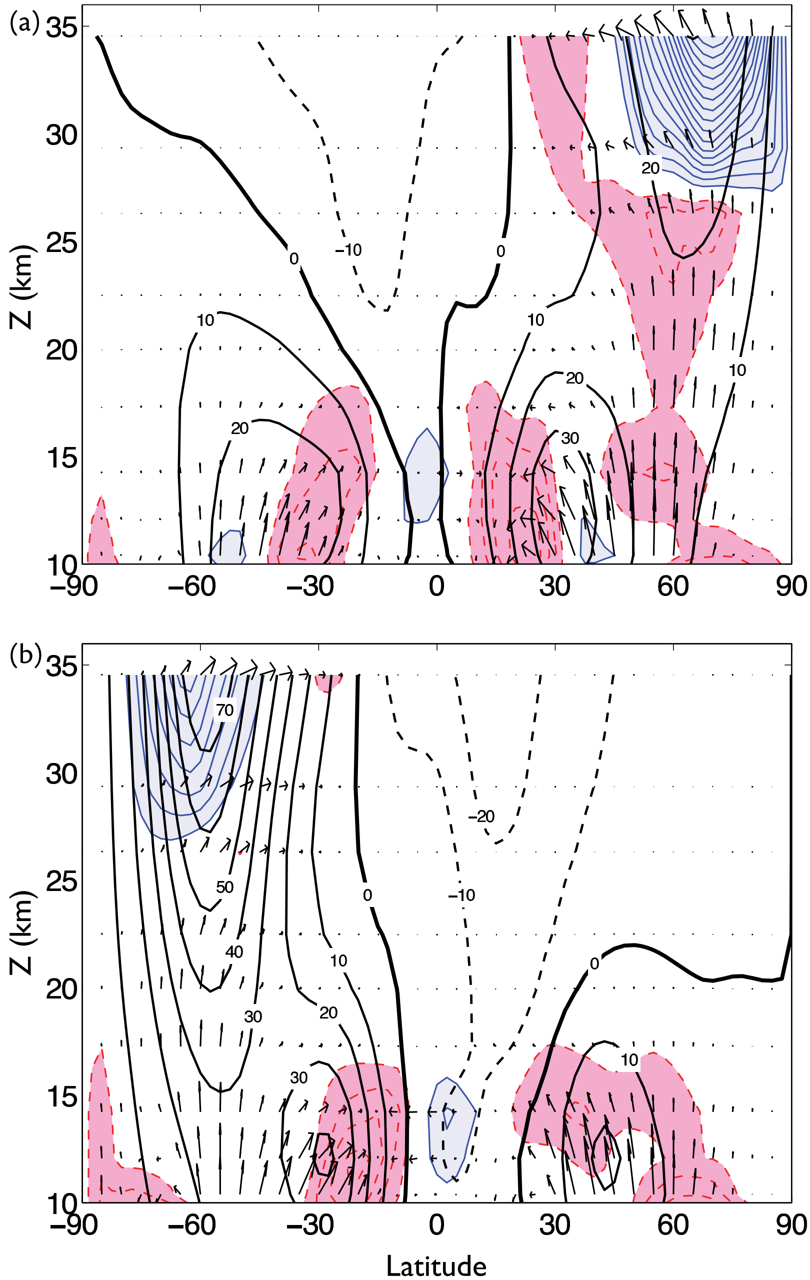


**Fig. 17.11** The EP flux vectors (arrows), the EP flux divergence (shaded and light contours) and the zonally averaged zonal wind (heavy contours) for (a) northern hemisphere winter; (b) southern hemisphere winter. Note the almost zero EP values in the summer hemispheres, and strong convergence at high latitudes in the winter hemispheres, leading to poleward residual flow and/or zonal flow acceleration. The EP divergence is shaded for values greater than $+1\ \mathrm{m\,s^{-1}}$/day, light solid contours) and for values less than $-1\ \mathrm{m\,s^{-1}}$/day (light dashed contours). The vertical coordinate is log pressure, extending between about 260 and 10 mb.

is not a direct response to differential solar heating: note that the most intense solar heating is over the summer pole, yet here there is little or no ascent. Rather, the circulation is more usefully thought of as a response to potential vorticity fluxes which in turn are determined by the upward propagation of Rossby waves from the troposphere combined with the poleward gradient of potential vorticity in the stratosphere. It is salutary to note that without motion we have $\theta = \theta_E$, so there is no net heating at all. So we can actually regard the heating as a consequence of the wave forcing.

## 17.6 ✦ DYNAMICS OF THE RESIDUAL OVERTURNING CIRCULATION

We now discuss the dynamics of the residual meridional overturning circulation in rather more detail than in the previous section, without shying away from the repetition of important matters.[17] The dynamics of the RMOC can be usefully couched as a quasi-linear problem. That is, the governing equations can be written with the linear terms on the left-hand side and the nonlinear terms as forcing terms on the right-hand side. Of course, this cannot be regarded as a full solution, but if the right-hand sides can be determined, if approximately, by independent means then the equations can be solved fairly straightforwardly and the structure of the RMOC so determined. Such a procedure is actually likely to be more successful in the stratosphere than in the troposphere; in the latter, the nonlinear terms are a truly essential part of the solution and cannot properly be separated from the linear dynamics of the RMOC (although we might choose to separate the terms as a exercise, to *diagnose* what forces the RMOC). In the stratosphere the nonlinear terms represent the effects of waves and wave breaking on the mean flow. These waves — both gravity waves and Rossby waves — often have their origins in the troposphere, and although the propagation and breaking of the waves does depend strongly on the background flow the basic features of the forcing of the RMOC can still usefully be considered independently of the RMOC itself.

### 17.6.1 Equations of motion

Away from the equator the Rossby number is small and the equations governing the large scale flow are in good geostrophic balance. The equations of motion governing the mean fields are the zonally averaged momentum and thermodynamic equations, along with the thermal wind equation and the mass continuity equations. In what follows we will write the equations in their full form in spherical co-ordinates using the ideal gas equations in log-pressure co-ordinates, since both sphericity and compressibility are important effects, using the transformed Eulerian mean formalism.

The full equations of motion may be written as

$$\frac{\partial \overline{u}}{\partial t} - f\overline{v}^* = \mathcal{G} + \mathcal{D} = \nabla \cdot \boldsymbol{\mathcal{F}} - \gamma \overline{u}, \tag{17.37a}$$

$$f\frac{\partial \overline{u}}{\partial z} + \frac{R}{aH}\frac{\partial \overline{T}}{\partial \vartheta} = 0, \tag{17.37b}$$

$$\frac{\partial \overline{T}}{\partial t} + \overline{w}^* S = Q_s + Q_l = \mu(T_R - \overline{T}), \tag{17.37c}$$

$$\frac{1}{a\cos\vartheta}\frac{\partial}{\partial \vartheta}(\overline{v}^*\cos\vartheta) + \frac{1}{\rho_R}\frac{\partial}{\partial z}(\rho_R \overline{w}^*) = 0. \tag{17.37d}$$

These equations are, respectively, the zonal momentum equation, the thermal wind equation, the thermodynamic equation and the mass continuity equation, with an overbar denoting a zonal average. The vertical co-ordinate, $z$ is log pressure and $S = H_\rho N^2/R$ where $R$ is the gas constant and $H_\rho$ is the scale height used to define $z$; thus, $z$ has dimensions of height and $\rho_R = \exp(-z/H_\rho)$. (We denote our vertical coordinate as $z$ rather than $Z$ for aesthetic reasons.) The equations have a very similar form if written in height co-ordinates using the anelastic approximation (section 2.5); in that case, $\rho_R$ is a reference profile of density and the thermodynamic equation is written using potential temperature or buoyancy as the thermodynamic variable, the factor $H_\rho/R$ no longer appears, and $z$ really is physical height.

The other variables in (17.37) use our standard notation, with $\overline{v}^*$ and $\overline{w}^*$ being the residual, or transformed Eulerian mean, meridional and vertical velocities. The right-hand side of (17.37a) represents wave forcing and friction: $\mathcal{G} = \nabla\cdot\boldsymbol{\mathcal{F}}$ is the divergence of the Eliassen–Palm flux and $\mathcal{D}$ is the frictional force, which we take to be a simple linear drag. Such a drag is a little arbitrary but its form greatly simplifies the ensuing analysis. On the right-hand side of the thermodynamic equation $Q_s$ and $Q_l$ represent the forcing due to solar and long wave radiation; we take $Q_l = -\mu\overline{T}$ where $\mu$ is a constant thermal damping rate, and we may write $Q_s = \mu T_r(\vartheta, z)$ where $T_R$ is a radiative equilibrium temperature, assumed known. There are no fluid-dynamic wave-forcing terms in the TEM form of the thermodynamic equation. Typically, the momentum dissipation is small and $\gamma \ll \mu$, and indeed we may take $\gamma = 0$ without much loss of realism, except close to the ground.

### 17.6.2 An equation for the MOC

If the right-hand sides are known, (17.37) constitutes a closed set of equations for the response of the temperature and the three components of the velocity to an applied wave force $\mathcal{G}$ and solar heating $Q_s$. Although nominally the equations have two time derivatives, the zonal wind and temperature are related through the thermal wind relation and the equations are indeed balanced and no gravity waves are present. Our interest here is in the meridional overturning circulation and hence $\overline{v}^*$ and $\overline{w}^*$, and it is possible to derive a single equation for either of these variables that provides considerable insight into that circulation; we will focus on $\overline{w}^*$.

The procedure is similar to that used in section 14.5.2 and the reader may wish to review that section before proceeding. Essentially, we differentiate (17.37a) with respect to $z$ and (17.37c) with respect to $\vartheta$ and then use (17.37c) to eliminate the time derivates. We then use the mass continuity equation to obtain a single equation in $\overline{w}^*$. We are particularly interested in the dependence of the MOC on the spatial structure and time dependence of $\mathcal{G}$ and $Q_s$, and to this end it is instructive to consider the case in which the time dependence is harmonic; that is, $\mathcal{G} = \widetilde{\mathcal{G}}\mathrm{e}^{\mathrm{i}\omega t}$, $Q_s = \widetilde{Q}_S\mathrm{e}^{\mathrm{i}\omega t}$ and $\widetilde{w} = \overline{w}^*\,\mathrm{e}^{\mathrm{i}\omega t}$. After a little algebra, we obtain

$$\begin{aligned}\frac{\partial}{\partial z}\left[\frac{1}{\rho_0}\frac{\partial(\rho_0\widetilde{w})}{\partial z}\right] + \left(\frac{\mathrm{i}\omega+\gamma}{\mathrm{i}\omega+\mu}\right)\frac{N^2}{4\Omega^2 a^2\cos\vartheta}\frac{\partial}{\partial\vartheta}\left[\frac{\cos\vartheta}{\sin^2\vartheta}\frac{\partial\widetilde{w}}{\partial\vartheta}\right] \\ = \frac{1}{2\Omega a\cos\vartheta}\frac{\partial}{\partial\vartheta}\left[\frac{\cos\vartheta}{\sin\vartheta}\frac{\partial\widetilde{\mathcal{G}}}{\partial z}\right] + \left(\frac{\mathrm{i}\omega+\gamma}{\mathrm{i}\omega+\mu}\right)\frac{R}{4H\Omega^2 a^2\cos\vartheta}\frac{\partial}{\partial\vartheta}\left[\frac{\cos\vartheta}{\sin^2\vartheta}\frac{\partial\widetilde{Q}_s}{\partial\vartheta}\right]\end{aligned} \tag{17.38a}$$

The equation is quite a TEXful, but it is useful to realize that, schematically and without all the metric factors, (17.38a) is of the form

$$\frac{\partial^2\widetilde{w}}{\partial z^2} + A\frac{N^2}{f^2}\frac{\partial^2\widetilde{w}}{\partial y^2} \sim \frac{1}{f}\frac{\partial}{\partial y}\frac{\partial\widetilde{\mathcal{G}}}{\partial z} + \frac{A}{f^2}\frac{\partial^2\widetilde{Q}_s}{\partial y^2}. \tag{17.38b}$$

where

$$A = \frac{\mathrm{i}\omega + \gamma}{\mathrm{i}\omega + \mu}.$$

Eq. (17.38b) is similar to (10.63); with the addition of diabatic terms and a slight change in notation the latter equation is

$$f_0^2 \frac{\partial^2 \widetilde{\psi}}{\partial z^2} + AN^2 \frac{\partial^2 \widetilde{\psi}}{\partial y^2} = f_0 \frac{\partial \widetilde{\mathcal{G}}}{\partial z} + A \frac{\partial \widetilde{Q}_s}{\partial y}, \tag{17.38c}$$

where $\widetilde{\psi} = \psi^* \mathrm{e}^{\mathrm{i}\omega t}$ is the amplitude of residual streamfunction of the overturning circulation. Since $\overline{w}^* = \partial\psi^*/\partial y$, (17.38b) is almost the $y$-derivative of (17.38c). (We note that in (10.63) we took $\gamma = \mu$ so that $A = 1$.) Much of the physical interpretation in what follows comes from (17.38c), although we will allow $f$ to vary spatially — that is, we use $f$ and not $f_0$.

With quasi-geostrophic scaling the wave forcing term in (17.38) is

$$\mathcal{G} = \overline{v'q'} \tag{17.39}$$

where

$$\overline{v'q'} = -\frac{\partial}{\partial y}\overline{u'v'} + \frac{\partial}{\partial z}\left(\frac{f_0}{N^2}\overline{v'b'}\right) = \nabla \cdot \boldsymbol{\mathcal{F}} \tag{17.40}$$

and

$$\boldsymbol{\mathcal{F}} = -\overline{u'v'}\,\mathbf{j} + \frac{f_0}{N^2}\overline{v'b'}\,\mathbf{k} \tag{17.41}$$

is the Eliassen–Palm flux.

### 17.6.3 The nature of the response

The operator on the left-hand side of (17.38) is elliptic, similar to a that of a Poisson equation. Thus, the response will be much less localized than the forcing itself, a concept that is familiar from potential vorticity inversion. To understand the equation better it is useful to take a heuristic look at some special cases, as follows. The cases are not all 'orthogonal' to each other — thus, for example, the low-latitude limit could be either low frequency or high frequency.

*(i) The aspect ratio of the response*
From (17.38c) the natural aspect ratio of the response, $\alpha_r$ say, is given by

$$\alpha_r = \frac{H_r}{L_r} = \frac{1}{A^{1/2}}\frac{f}{N} = \left(\frac{\mathrm{i}\omega + \mu}{\mathrm{i}\omega + \gamma}\right)^{1/2} \frac{f}{N}. \tag{17.42}$$

where $H_R$ and $L_R$ are the vertical and horizontal scales of the response. If the thermal and mechanical dissipation are zero, or have the same time scale, then $A = 1$, but more generally the presence of dissipation can alter the aspect ratio considerably. Also, the thermal dissipation can be expected to be much stronger than the mechanical dissipation, meaning that $\mu \gg \gamma$. The high- and low-frequency limits then have somewhat behaviour, as we see.

*(ii) The high-frequency limit*

In this case the thermal and mechanical damping are negligible and $A = (\mathrm{i}\omega+\gamma)/(\mathrm{i}\omega+\mu) \approx 1$. Since $\mu$ typically varies between 1/(20 days) in the lower stratosphere and 1/(5 days) in the upper stratosphere, and $1/\gamma$ is an even longer time, phenomena of order a few days fall into this category. Sudden stratospheric warmings are one example, although since the timescale of warmings is of order days thermal effects are not wholly negligible.

Using (17.42) we see that the aspect ratio of the response is simply of the order of Prandtl's ratio. That is, $\alpha_r = H/L \sim f/N$ and since $f \sim 10^{-4}\,\mathrm{s}^{-1}$ and $N \sim 10^{-2}\,\mathrm{s}^{-1}$ or larger, the response to rapid forcing is typically quite shallow. Of course, although the Prandtl ratio is small it is a natural scaling of vertical to horizontal scales in atmospheric dynamics and shallowness should be interpreted in that context. Still, as we approach the equator the response shallows still further, although at the equator itself (17.38) ceases to to valid. A quantitative analysis of the right-hand side of (17.38a) further suggests that both waves (the $\mathcal{G}$ term) and solar forcing act to drive the overturning circulation.

*(iii) The low-frequency limit*

In this case the frequency is less than the thermal damping rate; that is, $\omega \ll \mu$, with one obvious example being the annual cycle. If as is realistic, $\gamma \ll \mu$ then $|A|$ becomes small. The effect of the solar heating thus also becomes small in (17.38c). The response generally deepens with the ratio of the vertical to the horizontal scales being given by

$$\alpha_r = \frac{H_r}{L_r} \approx \left(\frac{1}{A^{1/2}}\right)\frac{f}{N} = \left(\frac{\mu}{\mathrm{i}\omega+\gamma}\right)^{1/2}\frac{f}{N} \gg \frac{f}{N}, \tag{17.43}$$

if $\mu \gg \gamma$. In the steady-state limit the solar heating is balanced by the thermal relaxation and the right-hand side of (17.37c) nearly vanishes. We discuss this limit in more detail below, in the section on downward control.

*(iv) Deep and shallow forces*

A force may be regarded as deep or shallow depending on whether its aspect ratio (vertical to horizontal, $\alpha_F$ say) is greater or less than $f/N$, and it turns out that deep force tends to gives rise to an acceleration of the zonal wind (a non-zero $\partial\overline{u}/\partial t$) whereas a shallow force tends to give rise to a meridional circulation. To see this we will consider the simplified forms of the momentum equation

$$\frac{\partial \overline{u}}{\partial t} - f_0\overline{v}^* = \mathcal{G}, \tag{17.44}$$

along with the MOC equation, (17.38c) which, if $f_0$ and $N$ are both constant, is a Poisson equation with a right-hand side equal to $\partial\mathcal{G}/\partial z$. Solutions can be obtained by Fourier series methods with terms having the form

$$\widetilde{\psi} = \Psi\cosh k_z z \sin k_y y. \tag{17.45}$$

Let us suppose the forcing is also of this form, so that by a deep forcing we mean that $k_y/k_z \gg f_0/N$ and shallow means $k_y/k_z \gg f_0/N$. We'll also suppose that $A = \mathcal{O}(1)$. From (17.38c) we see that

$$(f_0^2k_z^2 - N^2k_y^2)\widetilde{\psi} \sim f_0 k_z G, \tag{17.46}$$

where $G$ is the forcing amplitude. From this we can informally infer the form of the solution for deep and shallow forcing, as follows.

- *Deep forcing:* The dominant balance in (17.46) is between the second term on the left-hand side and the right-hand side and therefore $\widetilde{\psi} \sim f_0 k_z G/(N^2 k_y^2)$ and

$$\widetilde{w} \sim \frac{f_0 k_z G}{N^2 k_y}, \qquad \widetilde{v} \sim \frac{f_0 k_z^2 G}{N^2 k_y^2}. \tag{17.47}$$

  Now look at the momentum equation (17.44). The ratio of the Coriolis force to the forcing on the right-hand side is given by

$$\frac{|f_0 v|}{|\mathcal{G}|} \sim \frac{f_0^2 k_z^2}{N^2 k_y^2} \ll 1, \tag{17.48}$$

  where the inequality follows by definition of what is deep. The dominant balance in the momentum equation must then be between the wave forcing and the acceleration.

- *Shallow forcing:* The dominant balance in (17.46) is between the first term on the left-hand side and the right-hand side and therefore $\widetilde{\psi} \sim G/(f_0 k_z)$ and

$$\widetilde{w} \sim \frac{k_y G}{f_0 k_z}, \qquad \widetilde{v} \sim \frac{G}{f_0}. \tag{17.49}$$

  The ration of the Coriolis term to the forcing term in the momentum equation is now $\mathcal{O}(1)$, and therefore the response to a shallow forcing appears in the meridional circulation rather than as an acceleration.

The underlying reason for the two different responses arises from the need to satisfy the thermal wind equation. If the response to a shallow force were to be in the momentum equation then there would be a tendency to produce a shear, and hence a meridional temperature gradient, but it then becomes difficult to satisfy the thermodynamic equation; a response in the MOC is thus preferred. But if the force is deep then the response can and will be in the form of an acceleration.

*(v) Deep and shallow heating*

A similar analysis may be applied to a heating field, but given the intuition we have just developed about the response to a mechanical forcing there is no need to go through the mathematical details (although these are straightforward). A shallow heating (perhaps more usefully thought of as a broad heating) can and will produce a direct response in the temperature field itself. However, if the heating is deep (or, more usefully, latitudinally confined) then any direct response in the temperature field would have to be associated with a response in the zonal wind. Instead of this, a latitudinally confined heating will tend to produce a response in the meridional circulation.

*(vi) The low-latitude limit*

At low-latitudes most forces become deep because $f$ is small. More precisely, the criterion for deepness, that $H_f/L_f \gg N/f$ where $H_f$ and $L_f$ are the vertical and horizontal scales of the forcing, becomes easier to satisfy. Thus, wavebreaking at low-latitudes is more likely to induce an acceleration than a similar wavebreaking in midlatitudes, which will tend to induce an overturning circulation. By the same token, a heating source at low-latitudes

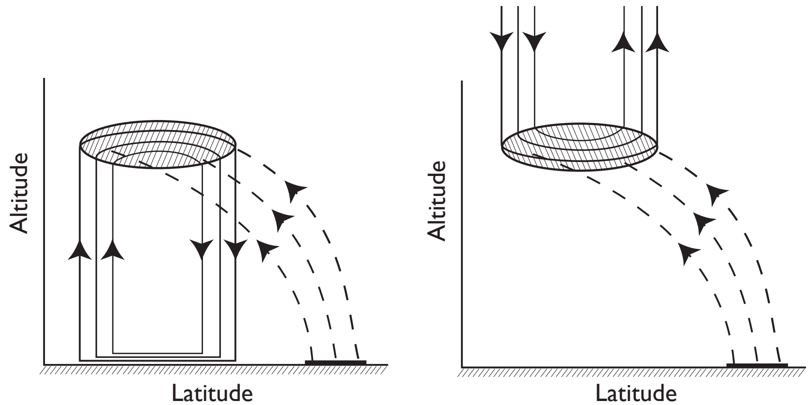


**Fig. 17.12** Idealized example of downward control. Left panel: wave activity propagates upwards (dashed lines) from a tropospheric source, breaking and depositing zonal momentum in the shaded region. This induces an overturning circulation (solid lines) below the region of momentum deposition, connecting the region of wave breaking with a frictional boundary layer. Right panel: putative 'upward control', which would require a frictional sink above the wave breaking region for the response to be steady.

has a greater tendency to induce an overturning circulation than a similar heat source in midlatitudes. Of course, the above statements are rules of thumb, and in any given case one should perform a more quantitative calculation to properly understand the response to any given heating or mechanical forcing.

If $\gamma = 0$ there is no spreading in $y$ in the steady state.

### 17.6.4 The steady-state limit and downward control

Let us now consider in a little more detail the steady-state response in which $\omega/\mu \to 0$, and we also take $\gamma = 0$, so that there is no momentum forcing. Although (17.38) of course still holds, but it is also useful to look directly at the momentum equation and thermodynamic equations. The momentum equation, (17.37a) reduces to a balance between the Coriolis force and the wave driving, namely

$$-f\overline{v}^* = \mathcal{G}, \tag{17.50}$$

and the thermodynamic equation becomes

$$\overline{w}^* S = Q_s + Q_l = \mu(T_R - \overline{T}). \tag{17.51}$$

The thermodynamic equation gives us very little information about the vertical velocity because the right-hand side contains the unknown temperature, $\overline{T}$; rather, we can glean nearly all the information we want from (17.50).

If we differentiate (17.50) with respect to $y$ (or $\vartheta$) and use the mass continuity equation we obtain

$$\frac{1}{\rho_0}\frac{\partial \rho_0 w}{\partial z} = \frac{1}{a\cos\vartheta}\frac{\partial}{\partial\vartheta}\left(\frac{\mathcal{G}\cos\vartheta}{f}\right). \tag{17.52}$$

This is a first-order partial differential equation for the vertical velocity, and we can obtain the vertical velocity itself by a vertical integration, using a single boundary condition. If we require that vertical velocity stays finite at $z \to \infty$, and so that $\rho_0 w = 0$, then we obtain

$$\boxed{\overline{w}^* = \frac{1}{a\rho_0(z)\cos\vartheta}\frac{\partial}{\partial\vartheta}\int_z^\infty \left(\frac{\rho_0(z')\mathcal{G}(\vartheta, z')\cos\vartheta}{f}\right) \mathrm{d}z'}. \tag{17.53a}$$

The quasi-geostrophic version of this equation is just

$$\overline{w}^*(z) = -\frac{1}{\rho_R}\frac{\partial}{\partial y}\int_z^\infty \rho_R(z')\frac{\mathcal{G}(\vartheta, z')}{f_0}\,\mathrm{d}z'. \tag{17.53b}$$

Eq. (17.53) implies that, in the steady-state limit, the vertical velocity at a given height is determined by the wave forcing *above* that height. The physical situation is illustrated in the left panel of Fig. 17.12. Here, a wave source in the troposphere propagates upwards and breaks in the middle atmosphere depositing momentum. This induces a meridional circulation as illustrated in the left panel, with a response *below* the momentum source. The numerically-computed response to an imposed force is illustrated in Fig. 17.13, showing how the response changes depending on the time-scales of the forcing and damping, with the steady-state response illustrated in the bottom panel.

The above derivation may seem a little disingenuous, for surely we might just as well have assumed $w = 0$ at $z = 0$, leading (in the quasi-geostrophic case) to

$$\overline{w}^*(z) = \frac{1}{\rho_R}\frac{\partial}{\partial y}\int_0^z \rho_R\frac{\mathcal{G}+\mathcal{D}}{f_0}\,\mathrm{d}z'. \tag{17.54}$$

This might appear to give 'upward control', as illustrated in the right panel of Fig. 17.12. However, if we are integrating from the ground up then frictional effects are important near the surface and must be included, as represented by the frictional term $\mathcal{D}$ in (17.54). Furthermore, mass conservation demands that

$$\int_0^\infty \rho_R\overline{v}^*\,\mathrm{d}z = 0, \qquad \text{implying} \qquad \int_0^\infty \rho_R(\mathcal{G}+\mathcal{D})\,\mathrm{d}z = 0, \tag{17.55a,b}$$

using (17.37a) for steady state conditions. Thus, above the level of the momentum source $\mathcal{F}$, (17.54) also in fact implies that the vertical velocity is zero, because $\mathcal{F}$ and $\mathcal{D}$ have cancelling effects. Thus, the location of the frictional boundary layer where the momentum is removed distinguishes up from down. Equation (17.55b) tells us that the frictional boundary layer at the bottom must adjust itself to remove the same amount of momentum that is deposited by wave breaking higher up, if there is to be a steady state. If there were a momentum sink above the momentum deposition region there would be no justification for downward control, for we would have to include that frictional term in (17.53). However, it is hard to envision how such a sink could exist without violating angular momentum conservation (see also problem 17.10).

From the point of view of the diagnostic equation for the meridional overturning circulation, in the steady state limit $A = 0$ and the solar forcing on the right-hand side and the $y$-derivative on the left-hand side of (17.38) both vanish, and the equation for the MOC becomes

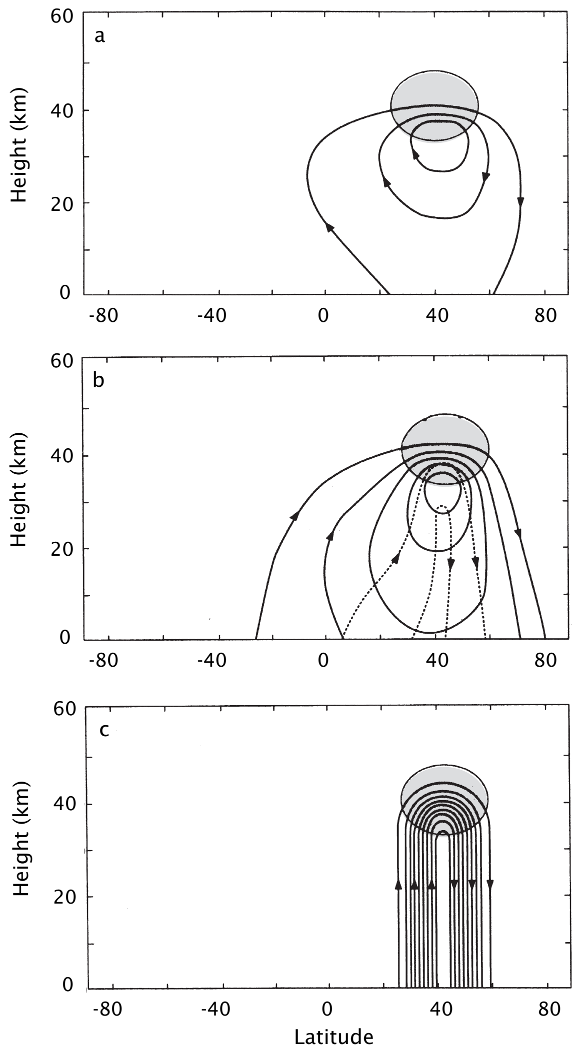


**Fig. 17.13** The numerically-computed response of the meridional overturning circulation to a longitudinally symmetric westward force, with the frequency of the forcing decreasing from top to bottom.[19] Contours are streamlines of the residual circulation, with the same uniform interval in all panels, and the shading denotes the forcing region.

(a) Response to high-frequency forcing, $\omega/\mu \gg 1$, $\omega \gg \gamma$. The response is adiabatic and weakly spreads into the opposite hemisphere.
(b) A lower frequency case with $\omega/\mu = 0.34$, corresponding to an annual cycle and a 20-day thermal relaxation timescale. The solid and dashed lines show the response that is in phase and out of phase with the forcing, respectively.
(c) Steady state response, $\omega/\mu \ll 1$. The circulation increases in magnitude and narrows as the frequency decreases, and in panel (c) it is given using the downward control expression (17.53a).

$$\frac{\partial}{\partial z}\left[\frac{1}{\rho_0}\frac{\partial(\rho_0\widetilde{w})}{\partial z}\right] = \frac{1}{2\Omega a\cos\vartheta}\frac{\partial}{\partial\vartheta}\left[\frac{\cos\vartheta}{\sin\vartheta}\frac{\partial\widetilde{\mathcal{G}}}{\partial z}\right] \tag{17.56a}$$

or, in the quasi-geostrophic limit,

$$f_0\frac{\partial^2\widetilde{\psi}}{\partial z^2} = \frac{\partial\widetilde{\mathcal{G}}}{\partial z}. \tag{17.56b}$$

That is to say, in a steady state the solar forcing provides no input to the meridional circulation! This may seem a little counter-intuitive, but if there is no wave forcing and $\mathcal{G} = 0$ then the vertical velocity is zero and the temperature adjusts to the radiative equilibrium temperature, so that the diabatic forcing is zero, as we now discuss in just a little more detail.

*The temperature field*

Given that the in the steady state vertical velocity field is determined by the wave forcing, the temperature field can be determined diagnostically from the thermodynamic equation. Thus, using (17.37c) with no time-dependence and a vertical velocity given by (17.53), we obtain in the quasi-geostrophic case

$$\mu(\overline{T} - T_r) = \frac{S}{\rho_R}\frac{\partial}{\partial y}\int_z^\infty \rho_R \frac{\mathcal{G}(\vartheta, z')}{f_0}\,\mathrm{d}z', \tag{17.57}$$

with a similar but more complicated expression in the full case. The temperature field at a given height is determined purely by the momentum forcing, being given by the meridional gradient of the zonal force *above* that height.

*An oceanic comparison*

It is instructive to compare downward control with the Stommel problem in oceanography (section 19.1.1). The steady version of (**??**) and the equation for the streamfunction for the *horizontal* flow, $\psi$, in the ocean, (19.6), may respectively be written

$$f_0\frac{\partial \psi^*}{\partial z} = \mathcal{F} + \mathcal{D}, \qquad \beta\frac{\partial \psi}{\partial x} = \mathcal{F}_w + r\nabla^2\psi, \tag{17.58a,b}$$

where $\mathcal{F}_w$ represents the wind forcing at the ocean surface, and the second term on the right-hand side of (17.58b) represents friction. In (17.58a), $\overline{v}^* = -\partial\psi^*/\partial z$ and in (17.58b), $v = \partial\psi/\partial x$. In the ocean interior, the frictional term is negligible, and in solving the resulting first-order equation ($\beta\partial\psi/\partial x = \mathcal{F}_w$) we may apply the boundary condition of $\psi = 0$ only at one meridional boundary. The natural choice is to choose the eastern boundary for this, and then invoke frictional processes to bring $\psi$ to zero on the west. It is a natural choice because Rossby waves propagate westwards ('westward control', as in Fig. 19.12); thus, the boundary current (e.g., the Gulf Stream) is on the west *because* the wind's influence is carried westwards by Rossby waves, not vice versa. Westward control is an enormously robust effect that pervades almost every aspect of large-scale physical oceanography. In the atmospheric case there is no similar mechanism that demands that the influence of the momentum source be propagated downwards. Rather, downward control results *because* the frictional boundary layer is at the bottom, not vice versa.

As we mentioned, the mechanism of downward control is related to that which gives rise to the Ferrel Cell in the troposphere, and that is certainly a strong and robust effect. Whether the downward control effect following wavebreaking *in the stratosphere* is strong enough to influence circulation in the troposphere, or the structure of the tropopause, remains an open question.

## 17.7 THE QUASI-BIENNIAL OSCILLATION

### 17.7.1 A brief review of the observations

The *quasi-biennial oscillation,* or 'QBO' as it commonly called, is a quasi-periodic reversal of the zonal wind in the equatorial stratosphere, as illustrated in Fig. 17.14. It is the most dominant variability of that region, and the following lists some of the main features of the phenomenon.[24] See also Fig. 17.15.

**Table 17.1** Typical, approximate, values of parameters appropriate for waves and background flow in the equatorial lower stratosphere.[21]

| Parameter | Background | Rossby-gravity waves | Kelvin waves |
|---|---|---|---|
| Static stability, $N$ | $2.2 \times 10^{-2}\,\mathrm{s}^{-1}$ | | |
| Beta at equator, $\beta$ | $2.3 \times 10^{-11}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$ | | |
| Coriolis parameter, $f$, at 5° | $1.27 \times 10^{-5}\,\mathrm{s}^{-1}$ | | |
| Wave period | | 4–5 days | 10–20 days |
| Zonal wavelength | | 10,000 km | 20,000–40,000 km |
| Zonal wavenumber | | 4 | 1–2 |
| dimensionally | | $6.3 \times 10^{-7}\,\mathrm{m}^{-1}$ | $1.6\text{–}3.2 \times 10^{-7}\,\mathrm{m}^{-1}$ |
| Meridional scale | | 1,200 km | 1,500 km |
| Vertical wavelength | | 4–8 km | 6–10 km |
| Phase speed, relative to ground | | $20\text{–}25\,\mathrm{m\,s}^{-1}$ (westward) | $25\,\mathrm{m\,s}^{-1}$ (eastward) |
| Amplitudes: | | | |
| zonal velocity | | $2\text{–}3\,\mathrm{m\,s}^{-1}$ | $4\text{–}8\,\mathrm{m\,s}^{-1}$ |
| meridional velocity | | $2\text{–}3\,\mathrm{m\,s}^{-1}$ | 0 |
| vertical velocity | | $1\text{–}2\,\mathrm{mm\,s}^{-1}$ | $1\text{–}2\,\mathrm{mm\,s}^{-1}$ |
| temperature | | 1 K | 2–3 K |
| geopotential height | | 4 m | 30 m |
| $F_0$, wave forcing at 17 km, see (17.60). | | $3\text{–}6 \times 10^{-3}\,\mathrm{m}^2\,\mathrm{s}^{-2}$ | $4\text{–}10 \times 10^{-3}\,\mathrm{m}^2\,\mathrm{s}^{-2}$ |
| Thermal damping rate, $\alpha$ | | $0.5\text{–}1.5 \times 10^{-6}\,\mathrm{s}^{-1}$ | $0.5\text{–}1.5 \times 10^{-6}\,\mathrm{s}^{-1}$ |

- The zonal winds in the equatorial region between about 5 and 100 hPa (about 40 and 18 km) alternate between being eastward and westward with an average period of about 28 months, with the period varying between 22 and 34 months.

- The QBO is almost latitudinally symmetric about the equator. The amplitude is approximately Gaussian with a half width of about 12°.

- The phenomenon is approximately zonal symmetric; that is, the longitudinal variation is small.

- The maximum amplitude of the oscillation is about $30\,\mathrm{m\ s}^{-1}$ ($\pm 15\,\mathrm{m\,s}^{-1}$) at about 20 hPa on the equator. The westward winds are slightly stronger than the eastward winds, after removing the annual cycle.

- The wind pattern descends at about 1 km per month with little loss of amplitude until it reaches 100 hPa, and the cycle begins again. (This does not mean that information propagates downward, as we discuss later.)

- The QBO is mildly synchronized to the annual cycle, with transitions between eastward and westward flow having a tendency to occur more commonly in March–June than in the other months.

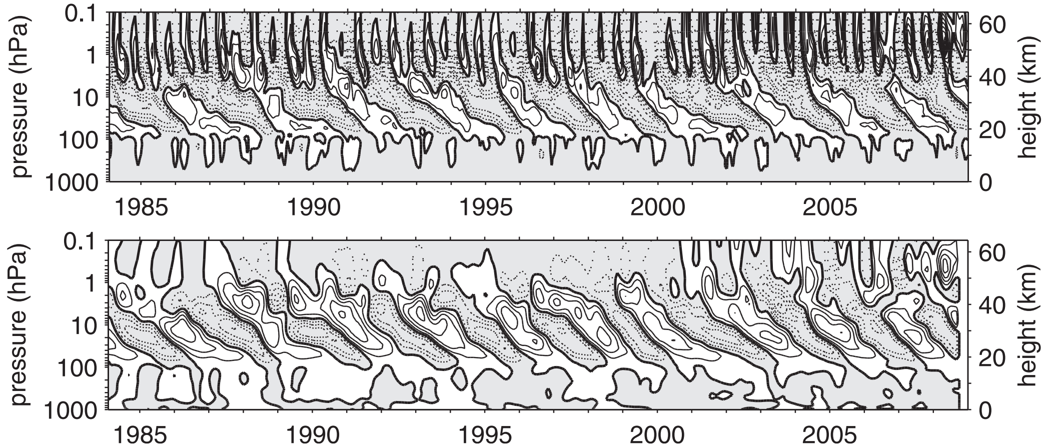


**Fig. 17.14** Time height sections of the observed monthly mean equatorial zonal wind for 1984–2010, with the seasonal cycle removed. Contour interval is $5\,\mathrm{m\,s^{-1}}$ and grey shading (with dotted contours) denote negative, or westward, winds. In the bottom panel the data are band-passed to retain periods between 9 months and 4 years only. An oscillation, the 'quasi-biennial oscillation' or QBO, is quite visible between 20 and 40 km, or about 60 and 3 hPa.[23]

Although we will not discuss it here, another oscillatory phenomena occurs above the QBO known as the semi-annual oscillation, or SAO. The SAO is an oscillation in the zonal wind with an approximate period of six months (it should perhaps be called the quasi semi-annual oscillation) and it occurs between 1 hPa and 0.1 hPa and extends from about 30° N to 30° S (Fig. 17.15).

### 17.7.2 A qualitative discussion of mechanisms

*Candidate mechanisms*

We first note that the QBO *must* involve zonally asymmetric motions. Without such asymmetric or eddying motions there can be no maximum of angular momentum within the fluid interior, and therefore no eastward winds at the equator, as explained in sections 13.5.1 and 14.2.8. Given this, let us ponder for a moment what *might* be the mechanism of the QBO. One might suppose that horizontally propagating planetary waves would be a likely mechanism, for the transport of momentum by Rossby waves is known to be an important mechanism for the maintenance of jets in mid-latitudes. However, the descent of the wind pattern with no loss of amplitude cannot be easily explained by such a mechanism.[27] Other proposed mechanisms involved interactions with the annual cycle or its harmonics (natural enough given the period of the QBO) or invoked external forcing or some nonlinear feedback. However, no candidate mechanism was able to explain all the features noted above, until a mechanism involving the vertical propagation and absorption of gravity waves was proposed, as we now describe.[28]

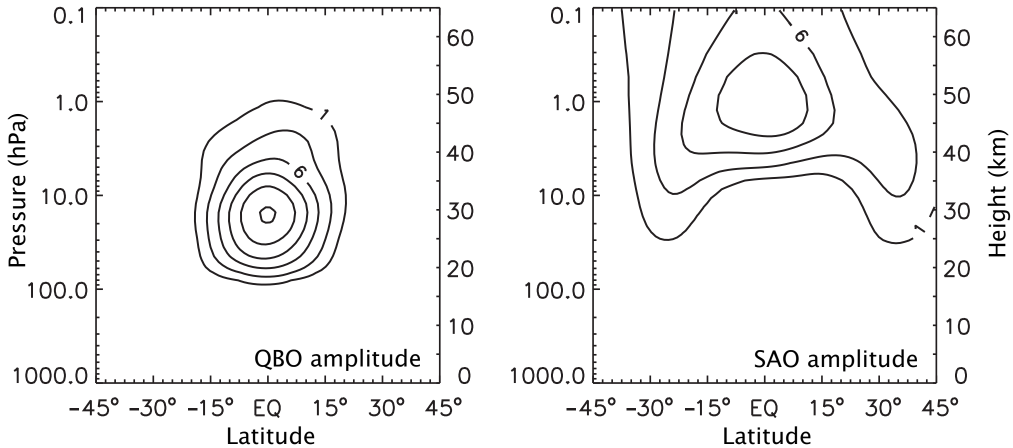


**Fig. 17.15** The amplitudes (essentially the root-mean square of the zonally-averaged wind of the oscillations after temporal filtering, with contours at 1, 3, 6, 9, 12 and 15 m s$^{-1}$) of the QBO and the SAO. The QBO evidently extends roughly from 15° S to 15° N and from 100 hPa to 1 hPa (18 km to 50 km). The SAO is broader, higher, weaker and faster.[26]

### *A gravity wave mechanism*

The mechanism for the QBO that is now generally accepted involves the upward propagation and absorption of gravity waves and the effect of this on the zonal flow. We first describe the mechanism rather roughly and qualitatively. A broad spectrum of gravity waves, with phase speeds in both eastward and westward directions, is generated in the upper equatorial troposphere by deep convection and various other instabilities. The waves will in general have a component of the group velocity that is directed upward, and if these waves are dissipated via mechanical or thermal damping then they will force mean flow accelerations (steady, non-dissipative waves cannot force a mean flow acceleration). A critical level, where the phase speed of the waves equals the speed of the mean flow (i.e., where $c = \overline{u}$) is one place where wave absorption and mean-flow acceleration will be particularly effective, because as a wave approaches a critical level it slows, giving more time for dissipation to act. However, it is not necessary for there to be an actual critical level; indeed, waves approaching a critical level will often be largely dissipated before reaching it.

Let us suppose that initially there is a westward shear (that is $\partial\overline{u}/\partial z > 0$) and that there are upward propagating gravity waves with positive phase speed $c$. These will be very efficiently absorbed as they approach the critical level, depositing momentum and causing the mean flow to accelerate. As pictured in the left panel of Fig. 17.16, this causes the critical level to descend and hence the subsequent absorption of gravity waves and acceleration of the zonal flow will be at a lower level. The wind anomaly thus descends, and so on.

A similar effect will still occur even if there is no critical level, provided there is enough dissipation for the gravity wave to be absorbed. In the right-hand panel of Fig. 17.16 the initial profile is uniform and there is no critical level. If upward propagating gravity waves are nevertheless absorbed somewhere the mean flow will accelerate. Even if this is insufficient to induce a critical level, the difference between the wave speed and the zonal fluid speed will be

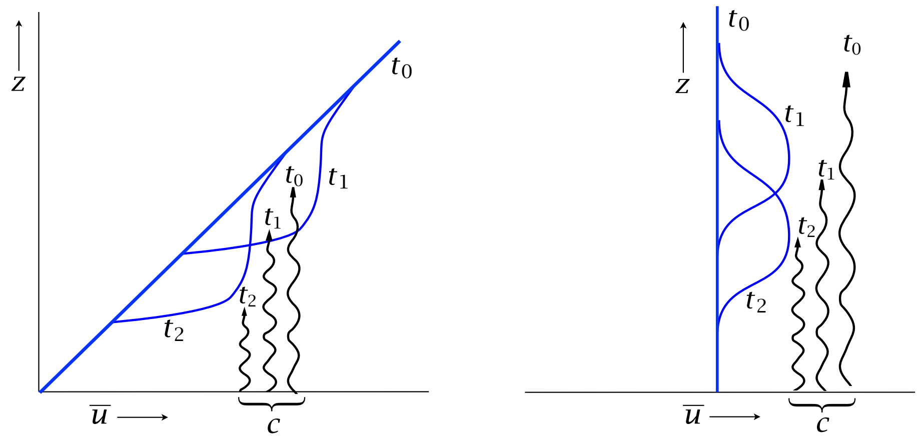


**Fig. 17.16** Schema of the descent of the zonal wind under the action of upward propagating gravity waves and relaxation back to the initial profile. The solid lines are zonal wind profiles at times $t_0$, $t_1$ and $t_2$ and the wavy lines indicate the penetration of gravity waves, all of the same speed $c$, at those times. In the left panel there is initially a westward shear. The gravity waves are absorbed, especially in the vicinity of $\overline{u} = c$, accelerating the mean flow. Subsequent gravity waves are thus absorbed at successively lower levels, causing the wind anomaly to descend. In the right panel the initial wind is uniform but again the wind anomaly descends.

reduced (i.e., $c - \overline{u}$ diminishes) and gravity wave absorption is enhanced. Gravity waves are thus absorbed at a lower level than previously and the anomaly in the zonal wind descends, as before.

Eventually, in the above models, the wind anomaly descends to the level of the gravity wave source. Depending on the strength of the dissipation one might imagine that dissipative processes could then wipe out the wind anomaly completely and the whole process would start all over again, or perhaps a low level westward wind anomaly would persist, so redefining the mean flow. However, in either case the zonal wind anomaly would not change sign (i.e., become westward), as is observed in the real QBO. For that to occur we may invoke a second wave in conjunction with an instability as we now explain.

### *A two-wave model*

Suppose now there are two upward propagating gravity waves with speeds $+c$ and $-c$ (where $c$ itself is positive), each of which will slowly be dissipated as it propagates, with the dissipation enhanced for smaller values of $|\overline{u} - c|$ or $|\overline{u} + c|$, respectively. (There will of course be very large dissipation if there is a true critical level.) Suppose that the mean flow has no shear, then simply by symmetry that state can persist, with the eastward and westward waves being dissipated equally as they ascend with no zonal flow generation. However, that symmetric state is unstable; to see this suppose that there is a small eastward perturbation to the zonal wind, as illustrated in the left panel Fig. 17.17. The eastward propagating wave will then be preferentially dissipated, because $\overline{u} - c$ is smaller for it than for the westward wave. The eastward anomaly in the zonal wind will therefore grow and descend, just as described above. The upward propagation of the eastward wave is then limited, but the westward wave is unconstrained at so it reaches higher levels before eventually being absorbed, providing a westward acceleration to the zonal flow, as

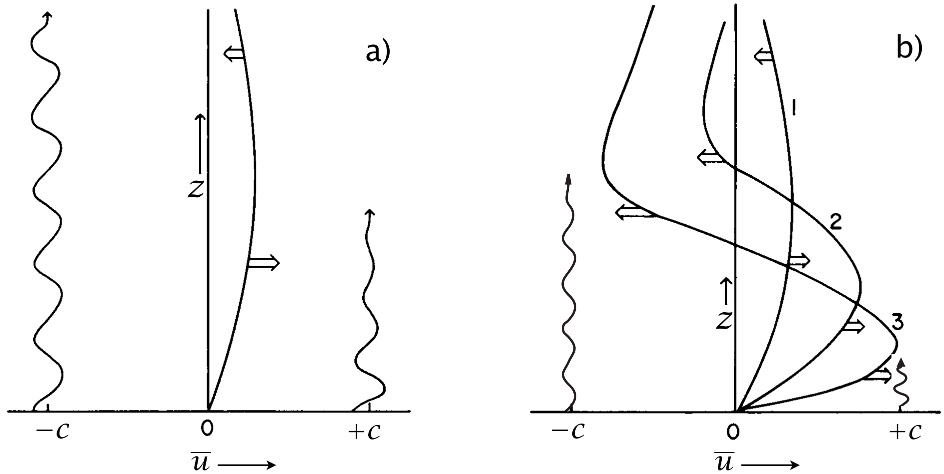


**Fig. 17.17** Schema of the initial instability leading to the QBO.[30] The solid lines show the zonal flow, the wavy arrows indicate the gravity-wave penetration from below and the double arrows indicate wave-induced acceleration. Initially, as in the left panel, a small eastward perturbation is added to a stationary mean flow. The eastward moving wave is preferentially absorbed and the perturbation is amplified and then descends, with the right panel showing the zonal flow at the successive times indicated, with the gravity wave penetration illustrated at $t = 3$. After the flow develops an eastward component the westward wave penetrates higher before being absorbed, inducing a westward flow aloft that then itself descends, making the eastward anomaly thinner. Subsequent stages and the development of a periodic oscillation are illustrated in Fig. 17.18.

illustrated in the right panel of Fig. 17.17.

As the the westward anomaly descends it squeezes the eastward anomaly which becomes thinner and thiner. Dissipative processes then become more efficient and can erode the eastward anomaly completely, with the flow becomes entirely westward, as illustrated in panel (b) of Fig. 17.18. A high level eastward anomaly is then created (panel (c) of Fig. 17.18), descending and squeezing the westward anomaly, and a mirror image of the first stage takes place. The entire cycle repeats itself and an oscillation is born, with the period of the oscillation being determined by the strength of the gravity waves and the rate of dissipation: stronger gravity waves lead to a faster acceleration of the mean flow and so a greater rate of descent and so a shorter period. Finally, note that the waves need not have speeds symmetric on either side of zero, $+c$ and $-c$. Suppose, for example, the wave speeds were both positive, $a$ and $b$ say. The mean flow could accelerate to an average value of $(a + b)/2$, with the flow then oscillating between $a$ and $b$ in a fashion similar to the symmetric case.

### 17.7.3 A quantitative model of the QBO

We now consider the above wave–mean-flow interaction model a little more quantitatively, and our first goal will be to obtain equations of motion for the interaction. To this end we will parameterize the vertical propagation and absorption of gravity waves by simple expressions

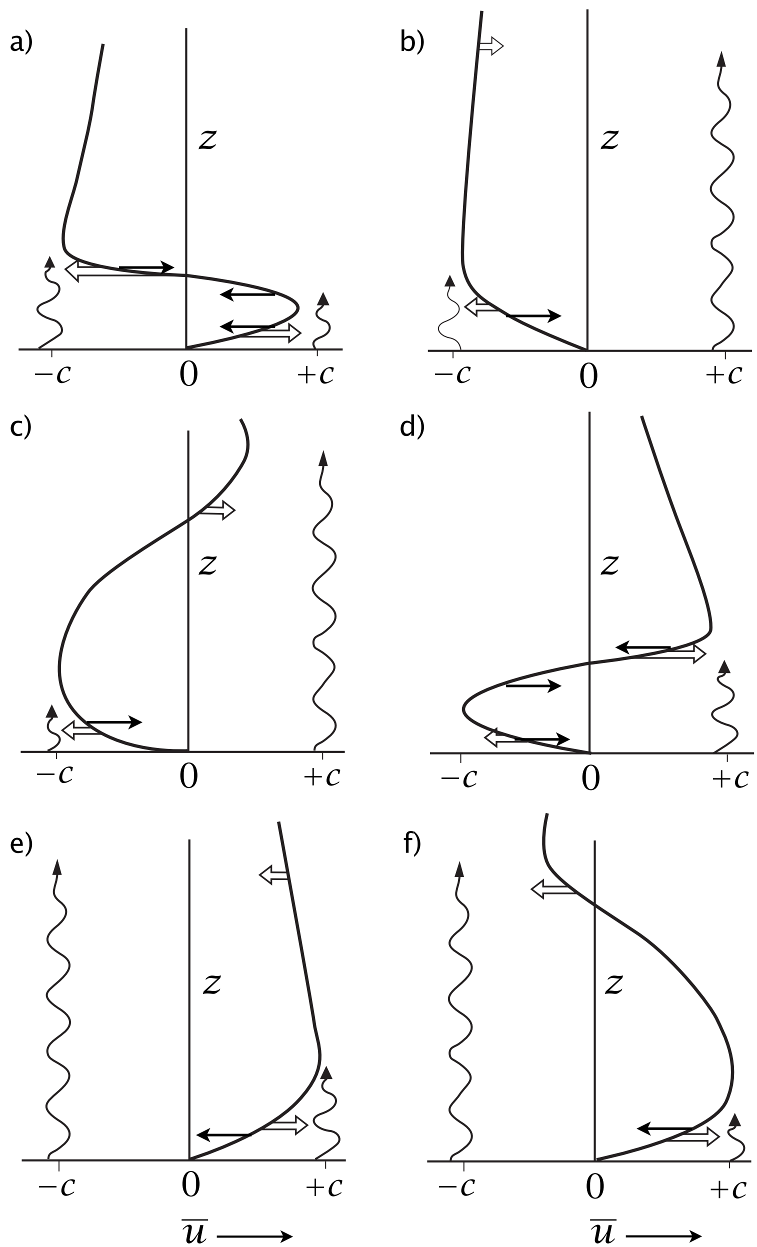


**Fig. 17.18** Schema of the evolution of the QBO with gravity-wave forcing from below, following an initial perturbation illustrated in Fig. 17.17.[32] The solid lines show the mean flow and the wavy lines indicate the propagation of gravity waves. Horizontal double arrows indicate wave forcing and single arrows indicate viscous relaxation. The panels are at successive times, with the top four panels showing a half cycle, and panels (d), (e) and (f) are mirror images of (a), (b) and (c). Wave-induced acceleration of the mean flow occurs preferentially near critical levels where $\overline{u} = c$.

## Essentials of the QBO

### What is the Quasi-Biennial Oscillation?

- The QBO is a quasi-periodic reversal of the zonal-mean zonal winds between about 20 km and 45 km and -15° S to 15° N, with a period of about 28 months. It is the dominant pattern of variability in the equatorial stratosphere and it is the clearest manifestation of a non-directly forced nearly periodic phenomenon in the atmosphere.
- The eastward and westward zonal winds appear to propagate downward at about 1 km per month, reversing at the end of each half cycle.
- The half-amplitude of the zonal wind cycle is about 15 m s$^{-1}$, with the westward winds being slightly stronger.

### What is the mechanism?

- The oscillation is caused by the upward propagation and absorption of Kelvin waves and Rossby-gravity waves at the equator. If a wave has an eastward phase speed then, on absorption, it will cause the mean zonal flow to accelerate eastwards. Furthermore, the absorption is strongest near a critical layer, where the mean zonal wind speed equals the phase speed. An upward propagating Kelvin wave thus causes the mean flow aloft to accelerate eastwards, and then the maximum in eastward winds to move downwards and eventually to be dissipated. An upward propagating Rossby wave then generate a westward zonal wind anomaly aloft, which similarly propagates down. In this way the zonal wind oscillates between positive and negative values, as illustrated in Fig. 17.18.
- The waves are primarily excited by moist convection in the upper tropical troposphere.
- The period is determined by a combination of parameters involving the wave and mean flow. In the simplest model of two upwardly propagating gravity waves the period is given by

$$P = \frac{Akc^3}{\alpha N_0 F_0} \tag{QBO.1}$$

  where $A$ is a nondimensional number weakly dependent on viscosity, and the other parameters, properties of the waves and mean flow, are defined in the text. The period is not proportional to the period of the waves; rather it is inversely proportional to their strength, $F_0$, because stronger waves cause more mean flow acceleration.

### Why is the phenomenon equatorially confined?

- The mechanism requires there to be upwardly propagating waves with very different phase speeds in order that the mean flow can oscillate between the two values. In equatorial regions such a forcing is provided by Rossby waves and Kelvin waves.
- In midlatitudes the tropospheric flow is largely balanced and it is primarily long Rossby waves that reach the stratosphere with a spectrum of phase speeds. If and when they break they would provide a westward acceleration. Furthermore, in mid- and high latitudes an imposed force tends to induce a mean meridional circulation, not a mean flow acceleration (section **??**).

resulting from gravity wave theory described in section 17.3. The absorption leads to a zonal flow acceleration, which in turn affects the wave absorption, and so on.

Let us consider a semi-infinite (no top), non-rotating, stratified fluid subject to a standing wave forcing at the lower boundary. Specifically, the waves are of the form

$$w = \mathrm{Re}\, \widetilde{w}_1(z)\mathrm{e}^{ik(x-ct)} + \widetilde{w}_2(z)\mathrm{e}^{ik(x+ct)}. \tag{17.59}$$

The waves have a dispersion relation as discussed in section 17.3, a positive (upward) group velocity, and we will take $\widetilde{w}_1 = \widetilde{w}_2$. If there is a source of gravity waves such as convection there is no difficulty in exciting waves with either an eastward of westward phase speed: a Kelvin wave has a purely eastward phase speed ($c_p > 0$), a Rossby-gravity wave has a westward phase speed, and gravity waves completely uninfluenced by rotation can have a phase speed in either direction. The Kelvin and Rossby-gravity waves, probably the most important waves for the QBO, typically have zonal wavenumbers 1–4, and so zonal wavelengths greater than 10,000 km, and periods of 3 days or longer.

As the waves propagate up they are dissipated, primarily by thermal rather than viscous dissipation, and their amplitude diminishes in the vertical and consequently they deposit momentum into the mean flow. From the WKB calculation of section 17.4.3 the wave momentum flux, $F_k(z) = \overline{u'w'}$, of a given upwards-propagating wave is of the form

$$\overline{F}_k(z) = \overline{F}_k(0) \exp\left[-\int_0^z g_k(z')\,\mathrm{d}z'\right] \tag{17.60}$$

where the subscript $k$ indicates the zonal wave number and the attenuation rate, $g_k(z)$, for a given upward-propagating internal wave is given by

$$g_k(z) = \frac{\text{damping rate}}{\text{vertical group velocity}} = \frac{\alpha}{k(\overline{u}-c)^2/N}. \tag{17.61}$$

where $c$ is the phase speed of the waves. The mean flow, $\overline{u}(z,t)$, is influenced by many such waves and so evolves according to

$$\frac{\partial \overline{u}}{\partial t} = -\sum_k \frac{\partial \overline{F}_k}{\partial z} + \nu \frac{\partial^2 \overline{u}}{\partial z^2}. \tag{17.62}$$

In writing (17.62) we include a dissipative term but neglect terms representing advection by the mean flow (such as $w\partial\overline{u}/\partial z$) and the Coriolis force $fv$. Eq. (17.62) is a closed partial differential equation in a single unknown for the mean flow. If the forcing consists of two waves, one with a phase speed $c$ that is positive and one with a negative phase speed, the the model produces behaviour that is quite similar to that of the QBO, as we see shortly.

*Direction of influence*

Although both the observations and the schematic solutions illustrated in Fig. 17.18 suggest that influence is somehow propagating downwards, this is in fact not the case when the gravity waves are propagating upward. From (17.60) and (17.62) the wave-driven acceleration of the mean flow, $A_w$ say, is given by

$$A_w = -\frac{\partial \overline{F}}{\partial z} = +\overline{F}(0)g(z)\exp\left[\int_0^z g(z')\,\mathrm{d}z'\right] = +g(z)\overline{F}(z). \tag{17.63}$$

That is, the acceleration is a function only of the profile of $g$ in the region from 0 to $z$, that is the region through which the wave has propagated. Furthermore, attenuation rate $g(z)$ is itself, from (17.61), a function only of the local value of $\overline{u}(z)$ and not of the derivatives of $\overline{u}$. Thus, the wave forcing at some level $z$ is a function only of the profile of $\overline{u}(z')$ for $z' < z$ and independent of the profile at higher altitudes. In other words, and in so far as the diffusivity term in (17.62) is negligible, there is no downward propagation of influence of the mean flow and the mean flow evolution is independent of what takes place above. The physical origin of this result is simply that waves are propagating upward and are absorbed by the mean profile as they ascend. If there were a *source* of waves at very high altitude, or if waves were reflected within the fluid (in which case the first-order WKB approximation is incomplete) then there could be a downward propagation of influence.

### 17.7.4 Scaling and numerical solutions

Scaling the equations — that is, nondimensionalizing in an intelligent way — not only makes numerical integration easier but also indicates what the natural height and time scales are for the problem. Important external parameters that determine the problem are the stratification $N$ (which has units of inverse time, $T^{-1}$), the damping rate $\alpha$ (also units of inverse time) and the strength of the wave forcing, $\overline{F}$ (units of $(L/T)^2$). A natural horizontal scale is the inverse of the wavenumber $k$.

Denoting nondimensional quantities with a hat, let

$$\widehat{F} = \overline{F}/F_0, \qquad \widehat{N} = N/N_0, \tag{17.64}$$

where $F_0 = \overline{F}(0)$ and $N_0$ is a typical value of $N$. If $N$ were uniform then we would simply choose $N_0 = N$ whence $\widehat{N} = 1$. To obtain sensible nondimensional quantities we note that the attenuation rate, $g$, has dimensions of inverse height so using (17.61) we choose a scaling height $H$ as

$$H = \frac{kc^2}{\alpha N_0}. \tag{17.65}$$

This might suggest a time scaling of $T = kc/\alpha N$. However, because we have a forced problem, the form of (17.62) suggests that we choose

$$T = \frac{cH}{F_0} = \frac{kc^3}{\alpha N_0 F_0} \tag{17.66}$$

with a velocity scaling of $U = F_0 T/H = c$ (note that this is not an advective scaling). The nondimensional coefficient of viscosity and thermal damping coefficients are then

$$\widehat{\nu} = \nu \frac{T}{H^2} = \nu \frac{\alpha N_0}{F_0 kc}, \qquad \widehat{\alpha} = \alpha T = \frac{kc^3}{N_0 F_0}. \tag{17.67}$$

The nondimensional equation for the mean flow evolution is then, for a single wave,

$$\frac{\partial \widehat{u}}{\partial \widehat{t}} = -\frac{\partial \widehat{F}}{\partial \widehat{z}} + \widehat{\nu}\frac{\partial^2 \widehat{u}}{\partial \widehat{z}^2}, \tag{17.68a}$$

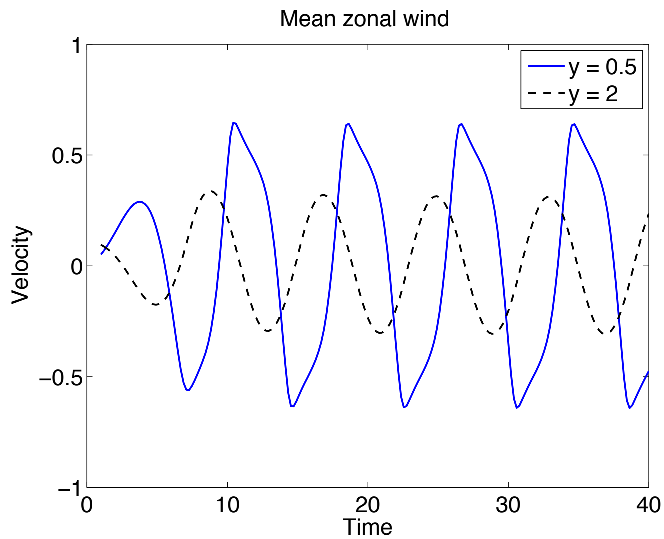


**Fig. 17.19** The evolution of the mean zonal wind at two different levels in a numerical solution of (17.68). All the variables are nondimensional, with the only parameter in the problem being viscosity, and here $\widehat{\nu} = 0.15$. A small perturbation is added to $\widehat{u}$, and relatively quickly the solution becomes periodic.

where

$$\widehat{F}(z) = \exp\left[-\int_0^{\widehat{z}} \frac{1}{(\widehat{u}-1)^2}\,\mathrm{d}\widehat{z}'\right] \tag{17.68b}$$

and the hats indicate nondimensional quantities. The great simplification that (17.68) offers over (17.60)–(17.62) is that in the former there are no parameters, save for the viscosity, and so the time and vertical scales of the problem are laid bare. In particular, if viscosity is small the only significant timescale in the system is (17.69) and the period of the oscillation must be proportional to that, and the vertical scale of the oscillation must be given by (17.65). Evidently the period of the oscillation is inversely proportion to the strength of the waves, but the vertical extent and the amplitude of the oscillation are both independent of the wave strength.

### *A numerical solution*

Eq. (17.68) may readily be numerically integrated and solutions are illustrated in Fig. 17.19 and Fig. 17.20. The simulations clearly show many of the qualitative features of the observed QBO, including the decay of the pattern with height and its downward propagation. The simulations are dependent on the viscosity $\widehat{\nu}$ in that, if $\widehat{\nu} = 0$, the jet at the bottom of the domain cannot be dissipated and the system in fact evolves to a steady state. On the other hand, if the viscosity is large then the jets become too broad, and the boundary layer near $z = 0$ that is evident in Fig. 17.20 is thicker. Still, if the viscosity is small but nonzero then over the bulk of the cycle it plays little role and the oscillation period is only weakly dependent on its value. Thus, for example, in Fig. 17.18 viscosity is needed to wipe out the low level westward jet between panels (d) and (e), but has little role in the rest of the half cycle and so only a small effect on the period. If viscosity is unimportant then the only timescale in the problem is that given by (17.69) and the period is proportional to it. and from numerical integrations we find that it is is given by

$$P = A\frac{kc^3}{\alpha N_0 F_0} \tag{17.69}$$

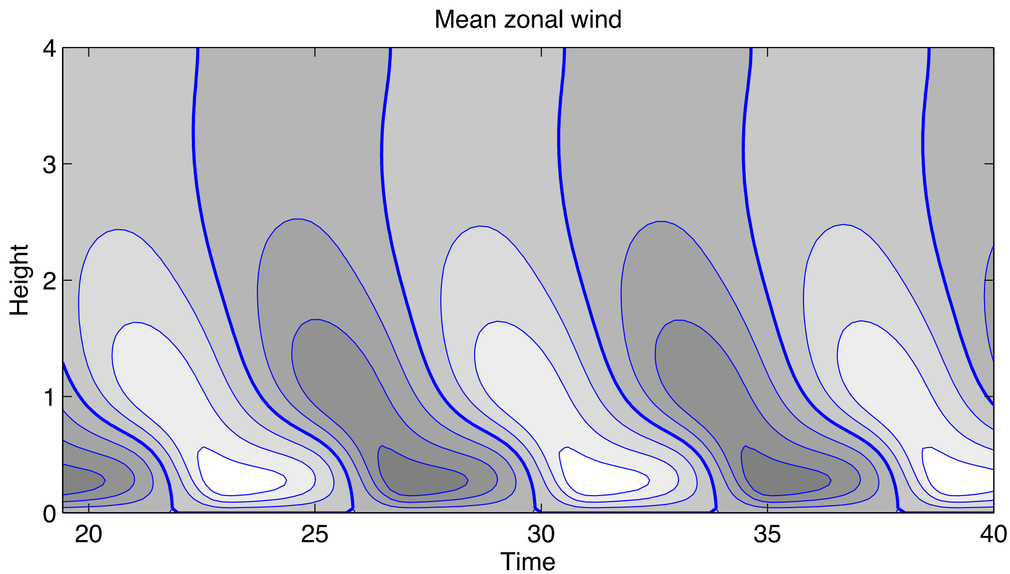


**Fig. 17.20** Time height section of nondimensional zonal wind in a numerical solution of (17.68), showing the last 20 time units of the same integration as Fig. 17.19. Contours are plotted every 0.2 units, positive winds are shaded darker and the zero contour is thicker.

where $A \approx 8$. Note that the period of the QBO is not directly dependent on the period of oscillation of the waves themselves.

### 17.7.5 The role of Rossby wave and Kelvin waves

Thus far our discussion of the mechanism of the QBO has focussed on the upward propagation of somewhat generic gravity waves in which rotation played no role. In fact the waves that propagate into the stratosphere in equatorial regions are of two main types, Kelvin waves and Rossby waves. Kelvin waves are a form of gravity wave but have only an eastward phase propagation, whereas Rossby waves are balanced waves with a westward propagation.[33] The theoretical development paralleling section 17.7.3 is naturally more complex, in part because the problem is now, in principle, a three-dimensional one. However, it is much simplified if we consider motions at the equator and if we take note that, in general, the attenuation rate of a wave is equal to its damping rate divided by its group velocity, as in (17.61). The corresponding attenuation rates for Kelvin and Rossby waves are then given by

$$\text{Kelvin wave:} \qquad g_K(z) = \frac{\alpha}{k_K(\overline{u} - c_K)^2/N}, \tag{17.70a}$$

$$\text{Rossby wave:} \qquad g_R(z) = \frac{\alpha}{k_R(\overline{u} - c_R)^2/N}\left(\frac{\beta}{k_R^2(\overline{u} - c_R)} - 1\right). \tag{17.70b}$$

The Kelvin wave attenuation rate is just the same as that for a non-rotating gravity wave, although the wave speed, $c_K$, is strictly positive. The Rossby wave attenuation rate (whose deriva-

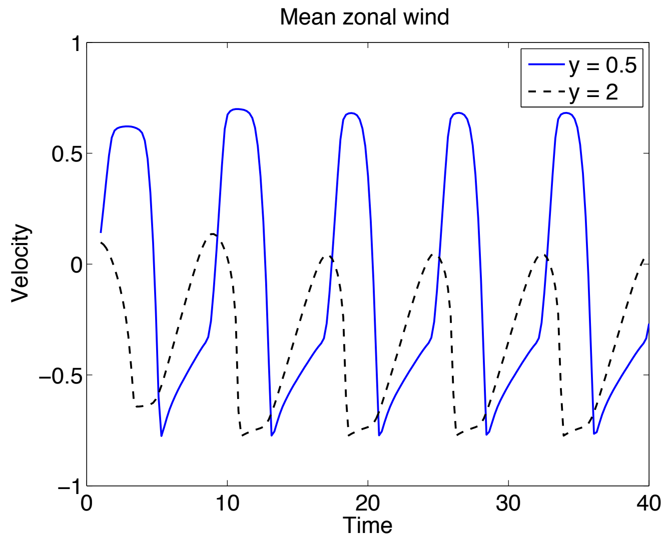


**Fig. 17.21** The evolution of the mean zonal wind at two different levels in a numerical solution using (17.70). All the variables are nondimensional, with the only parameter in the problem being viscosity. A small perturbation is added to $\widehat{u}$, and relatively quickly the solution becomes periodic.

tion requires a little algebra) involves the equatorial beta parameter and a negative phase speed, $c_R$. The full problem is now defined by (17.62) and (17.60), now with $g(z)$ given by (17.70). It is evident that the problem is no longer east-west symmetric; however, the essential structure of the problem remains. Rossby wave absorption is enhanced near a critical layer where $\overline{u} = c_R$, and Kelvin wave absorption still occurs near $\overline{u} = c_K$, so we expect to see an oscillation contained between these two values. Further, just as in the gravity wave problem, influence propagates upward with the waves.

The equation set (17.62), (17.60) and (17.70) may readily be numerically integrated and solutions are illustrated in Fig. 17.21 and Fig. 17.22. The east–west asymmetry arises from the factor

$$\frac{\beta}{k_R^2(\overline{u} - c_R)} - 1. \tag{17.71}$$

It is the factor $g_R$ that is responsible for the westward acceleration of the mean flow, for 'dragging' $\overline{u}$ toward the value $c_R$, which is negative. But $g_R$ is zero when $\overline{u} - c_R = \beta/k_R^2$ which, for our numerical simulation, occurs when $\overline{u} = 1$. Thus, when $\overline{u}$ is close to its eastward (Kelvin-wave induced) peak the westward acceleration is small. Another source of east-west asymmetry, likely to be important in reality, is that likelihood that Rossby waves and Kelvin waves may have different amplitudes. If Kelvin waves were stronger, for example, then the eastward acceleration would be stronger than the westward and that part of the cycle would be faster.

### 17.7.6 General discussion

The above sections have described a couple of relatively simple models that seem to capture the essence of the QBO.[34] The model using a Rossby waves and Kelvin waves is not noticeably more realistic in its predictions; rather, it attractive because Rossby waves and kelvin waves are observed in the equatorial stratosphere so it is more realistic in its assumptions. The observed east–west asymmetry in the observed QBO is not obviously caused by the differences between Rossby waves and Kelvin waves; other possibilities include the effects of a mean circulation

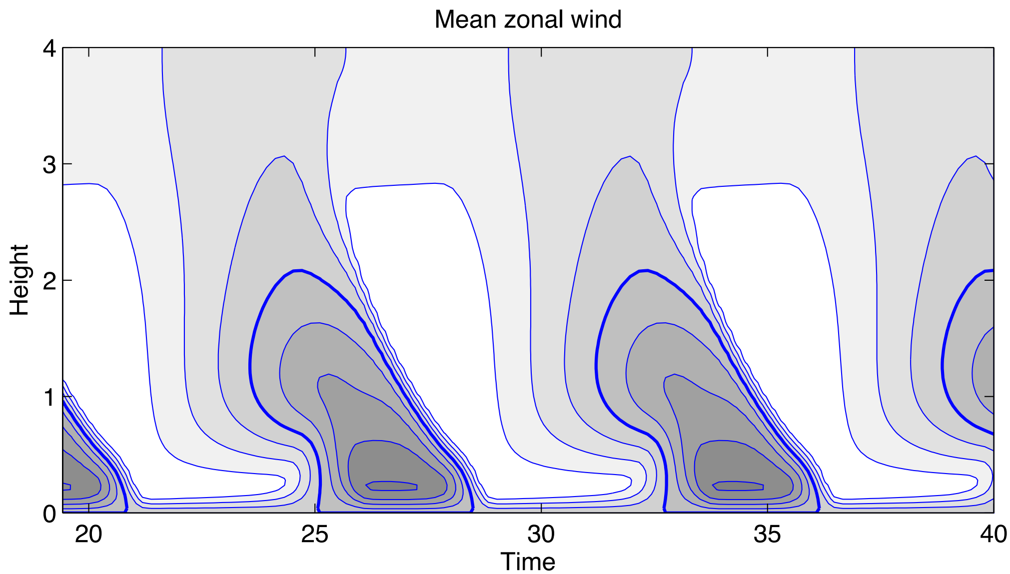


**Fig. 17.22** Time height section of nondimensional zonal wind in a numerical solution using (17.70), showing the last 20 time units of the same integration as Fig. **??**. Contours are plotted every 0.2 units, positive winds are shaded darker and the zero contour is thicker.

and the three-dimensional nature of the problem, and possible differences in the strength of the eastward and westward forcing.

The model with two non-rotating gravity waves is attractive because it allows a more complete analysis of its properties. In particular, the upward propagation of waves leading to a downward propagation of the zonal wind pattern, and the factors determining the period of the oscillation, are made transparent. The period of the problem is given by (17.69). Using table 17.1 as a guide, let us take the following dimensional values of the parameters: $k = 2\times10^{-7}\,\mathrm{m}^{-1}$, $\alpha = 1\times10^{-6}\,\mathrm{s}^{-1}$, $c = 25\,\mathrm{m\,s}^{-1}$, $N_0 = 2.2\times10^{-2}\,\mathrm{s}^{-1}$, $F_0 = \times10^{-2}\,\mathrm{m}^2\,\mathrm{s}^{-2}$. We obtain a timescale of $T = (kc^3/\alpha N_0 F_0) \approx 160$ days or about 5 months and so, using (17.69), a period of 40 months. Obviously there is considerable uncertainty in the parameters chosen and it would not be difficult to choose a set of parameters giving the observed value of about 26 months — or for that matter, to choose a set that gave a still longer period.

The vertical scale of the oscillation is given by (17.65) and with the above set of parameters we obtain $H = (kc^2/\alpha N_0) \approx 6\,\mathrm{km}$. From the numerical simulations we see that the vertical penetration of the phenomena is 2 or 3 times this, so about 15 km. This value again is reasonably close to the observed value of, from Fig. 17.14, about 20 km, but again we should be wary of too close an agreement, especially using a one-dimensional quasi-Boussinesq model. It is interesting that the period and vertical extent of the observed oscillation vary only a little, implying only a little interannual variability in the forcing strength and other parameters of the problem.

The actual waves themselves are primarily generated by convection in the tropical troposphere, then propagating up into the stratosphere. It is difficult to numerically simulate a QBO with an explicit representation of gravity waves because of the large range of scales involved in the problem, although three-dimensional simulations with parameterized gravity waves have been quite successful (see references in endnote 36). However, striking support for the mechanism described above has come from laboratory experiments, using an annulus of stratified water subject to a standing wave forced by a flexible lower boundary.[35] Given a strong enough forcing an oscillating mean flow was generated whose structure was found to be in very good agreement with the two-wave theory. Thus, at the very least, the mechanism does describe a real physical phenomenon.

There are a great many aspects of the QBO that we have not discussed; among the most egregious of our omissions is any serious discussion of latitudinal structure, of the effects of a mean circulation, and of three dimensional numerical simulations. Readers wishing to learn more about these should consult the original literature, and a few references that may serve as an introduction are given here.[36] Finally, to make a personal remark, the QBO is both a curiosity and a triumph. The former because its relationship to and influence on the rest of the circulation, or certainly on human society, is by no means demonstrable to the casual observer. It clearly does not, for example, have the impact of El Niño or of tropospheric weather, and its effect on the rest of the stratosphere is arguably not particularly remarkable.[37] Yet, excepting directly forced oscillations like the diurnal cycle, it is the clearest example of a nearly periodic phenomenon in the atmosphere and its beautiful explanation must rank as major achievement in geophysical fluid dynamics.

## 17.8 VARIABILITY AND EXTRA-TROPICAL WAVE–MEAN-FLOW INTERACTION

Compared to the troposphere the stratosphere has only mild baroclinic instability, both because of the lack of shear because of the larger value of the deformation radius. If we take $N = 2.2 \times 10^{-2}\,\mathrm{s}^{-1}$, $H = 20\,\mathrm{km}$ and $f = 10^{-4}\,\mathrm{s}^{-1}$ then $L_d = NH/f = 4,400\,\mathrm{km}$, and given that for many problems in baroclinic instability the instability scale is a few times that, sometimes there is simply no room for instability on a planet with a radius of 6,000 km. When there is an instability it will be at large scales, perhaps at wavenumbers 1, 2 or 3, as opposed to wavenumber 8 or so in the troposphere. This is not to say there is no variability in the stratosphere, with the variability arising in two main ways.

*(i)* From waves propagating up from the troposphere, with the stratospheric variability reflecting that of troposphere.

*(ii)* From oscillatory or even chaotic flow arising from the interaction, within the stratosphere, of large-scale planetary waves with themselves and with the mean flow. The forcing may still come from the troposphere but, even when this is steady, intra-stratospheric interactions give rise to unsteadiness.

In either case the variability tends to be relatively slow (compared to the troposphere) and at a large scale — the orographic forcing from the troposphere undergoes Charney–Drazin filtering and tends to occurs at wavenumbers 1 and 2, and as noted any baroclinic instability is also at large scale. It is therefore useful to think of the variability as a wave–mean-flow problem rather than as a problem in fully-developed geostrophic turbulence.

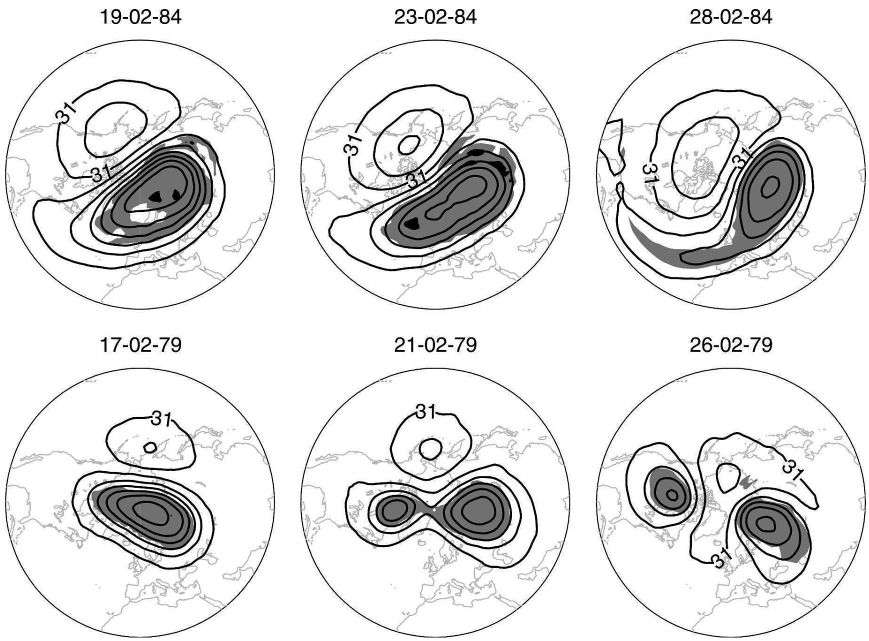


**Fig. 17.23** Time sequence of two stratospheric warmings, with the top row showing the displacement of the polar vortex and the bottom showing a split, with the dates marked. Contours are geopotential height on the 10 hPa surface and shading shows potential vorticity greater than 4 PV units.[39]

### 17.8.1 Upward propagating disturbances and sudden warmings

Consider planetary waves that are excited in the troposphere and propagate upward, as described in chapter 16, with this occurring predominantly in winter when the tropospheric forcing is strongest. The wave activity obeys

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \boldsymbol{\mathcal{F}} = D \tag{17.72}$$

where $\mathcal{A} = \overline{q'^2}/2\overline{q}_y$ and $\nabla \cdot \boldsymbol{\mathcal{F}} = \overline{v'q'}$. Thus, if dissipation is small and $\partial\mathcal{A}/\partial t$ is positive, $\overline{v'q'}$ is negative. Now consider the zonal momentum equation in quasi-geostrophic TEM form, namely

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \overline{v'q'}. \tag{17.73}$$

Rossby waves propagating into the stratosphere or, by the same token, dissipating Rossby waves in a statistically steady state, thus induce a deceleration (i.e., a westward tendency) of the zonal mean flow and/or a poleward meridional flow. (The east-west directionality arises because of the presence of $\overline{q}_y$ (and often $\overline{q}_y \approx \beta$) in the expression for $\mathcal{A}$; on an $f$-plane there is no difference between east and west.) The partitioning between the zonal and meridional flow depends on the time- and space-scale of the forcing, but as noted in section 17.6.3 only for a very deep forcing is the residual circulation response negligible. One way to see this is is to note

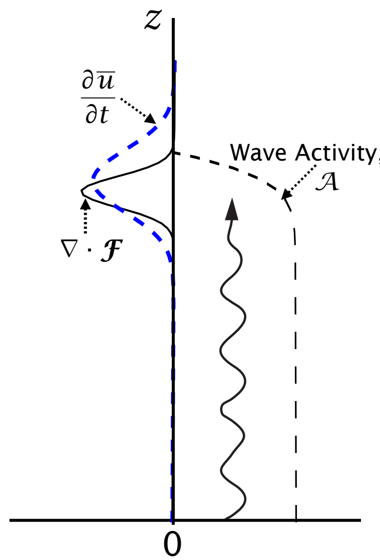


**Fig. 17.24** Schematic of a model of a stratospheric warming. Upward propagating Rossby waves (wavy line) reach the stratosphere and break, and the wave activity (dashed line) diminishes. In the breaking region the EP flux divergence is negative ($\nabla \cdot \mathcal{F} < 0$), inducing a westward acceleration ($\partial\overline{u}/\partial t < 0$) over a somewhat broader region. See also Fig. 17.25.

that the adiabatic buoyancy equation, $\partial b/\partial t + N^2\overline{w}^* = 0$, contains no eddy terms. But if the zonal wind shear changes, the buoyancy must change to maintain thermal wind balance and hence $\overline{w}^* \neq 0$. For a deeper forcing the induced shear will be smaller and hence the response of the temperature residual ciculation are also smaller. A similar diagnosis can be made upon inspection of (10.87), namely

$$\left[\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\right)\right]\frac{\partial \overline{u}}{\partial t} = \frac{\partial^2}{\partial y^2}\overline{v'q'}. \tag{17.74}$$

This equation tells us that the zonal-wind response to an wave forcing will tend to be of larger scale that the forcing itself, because of the elliptic nature of the operator acting on $\partial\overline{u}/\partial t$.

Suppose, then, that Rossby waves propagate upwards from the troposphere and break in the stratosphere. The mean eastward flow (sometimes called the polar night jet) will be weakened, so allowing more waves to propagate up, since strong eastward flow inhibits propagation (Fig. 16.6). If the process continues the winds will eventually reverse, forming a critical layer (as described in section ) where $\overline{u} = 0$. This completely inhibits further upward propagation and wave breaking is intensified, inducing a westward flow at the level to which the propagation reaches. There is a rapid changeover to westward flow and the critical layer descends. (This sequence has a similarity with the westward acceleration phase of the QBO, but in the extra-tropics there is no eastward counterpart as there are no Kelvin wave, and thus not is no oscillation. Rather, the eastward winds of the polar night jet are gradually restored by radiative effects.) By thermal wind, a reduced (or reversed) vertical shear is associated with a reduced (or reversed) meridional temperature gradient, and finally the polar night jet is replaced by a warmer westward flow. Put another way, the deposition of westward wave momentum leads to a warming of the high-latitude stratosphere. Such an event can at times be strong enough to split asunder the cold polar vortex and when it does the event is known as a *sudden stratospheric warming.*[40]

The interaction of the waves and mean flow is schematically illustrated in Fig. 17.24 and Fig. 17.25. In Fig. 17.24 we see a wave propagating up from the troposphere and breaking, with

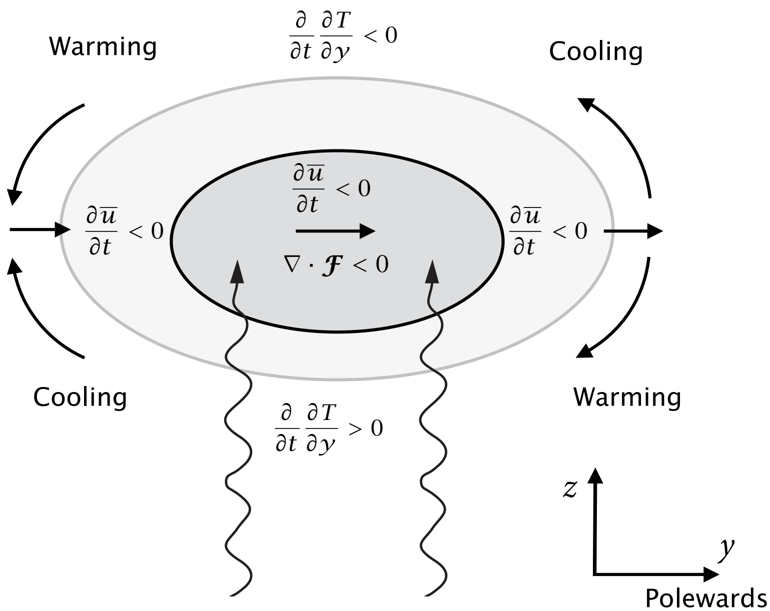


**Fig. 17.25** The circulation induced by a patch of negative EP flux divergence ($\nabla \cdot \boldsymbol{\mathcal{F}} < 0$, dark shading). The westward acceleration ($\partial\overline{u}/\partial t < 0$) occurs there, spreading out over a broader region (light shading). Assuming there is no acceleration far away from the wavebreaking, the temperature response can be inferred from thermal wind, with a warming at the lower poleward end of the breaking, as indicated, and an induced residual circulation as shown by the arrows.

wave activity then falling. The EP flux is negative in the breaking region causing a deceleration of the zonal flow over a somewhat broader region because of the elliptic operator in (17.74). The temperature response, shown in Fig. 17.25, can be inferred from thermal wind balance, noting that above the breaking region $\partial_t(\partial u/\partial z) > 0$ to that $\partial_t(\partial T/\partial y) < 0$, and oppositely for below, and the direction residual circulation follows by noting that adiabatic warming (cooling) results from descent. (The residual circulation may also be inferred from (10.63) or xxx.)

*Observations and numerical simulations*

[This section is not yet completed and the figures may be placeholders.]

To illustrate the above mechanism in a more realistic setting we shown some results from a primitive equation simulation and some observations of a sudden warming in the polar stratosphere.[41]

Figure 17.26 shows the anomalous zonal wind, the temperature and the EP fluxes during the growth and maturation of a composite warming even — that is, the fields are averaged over many warming events. Referenced relative to the peak of the warming, the *onset* refers to days −37 to −23, *growth* to days −22 to −8 and *mature* to days −7 to +7. During the onset and growth phase the upwards EP flux is particularly strong, causing a warming at high latitudes, with winds becoming more westward and descending, in broad consistency with the theoretical model described above.

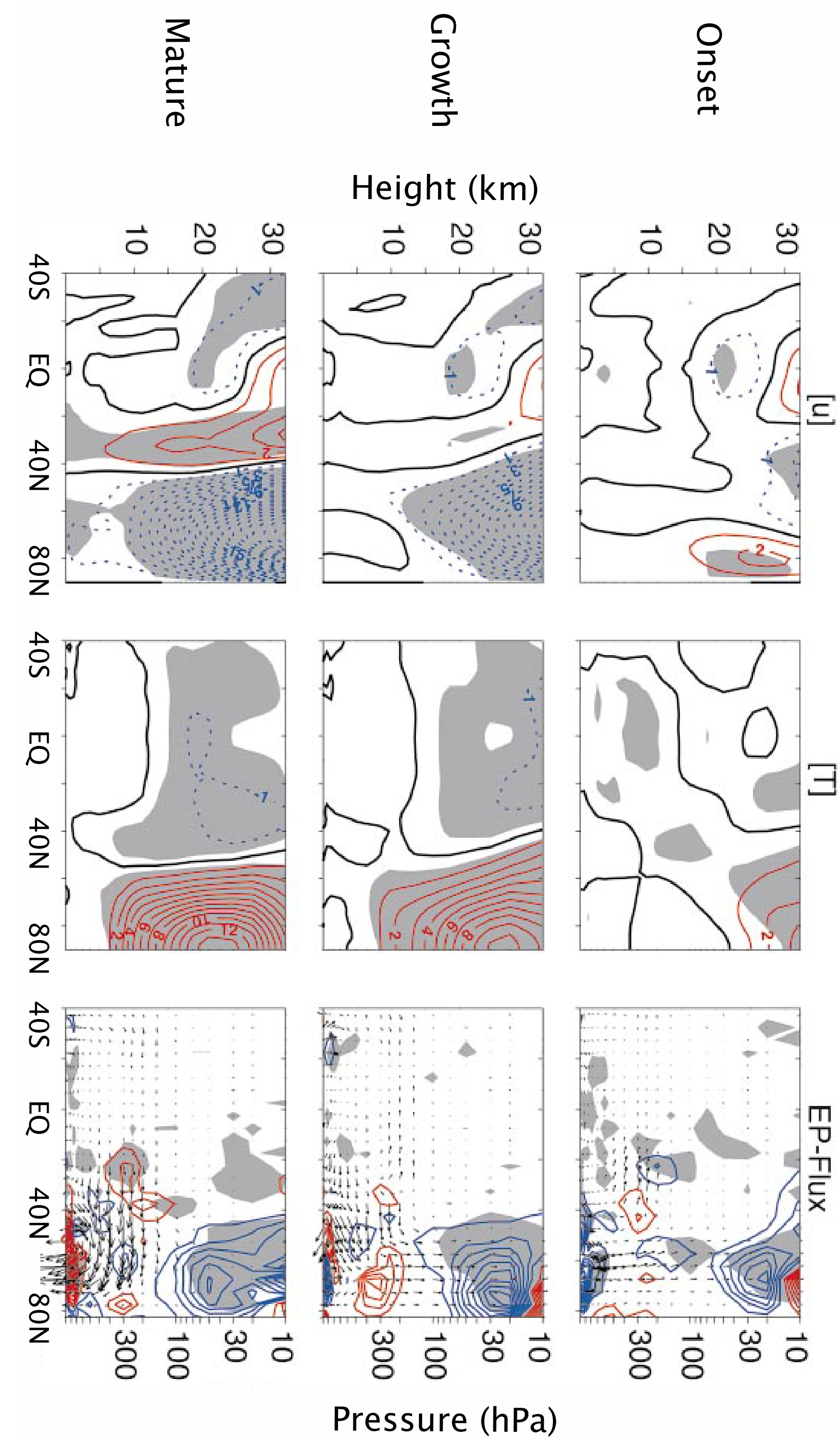


**Fig. 17.26** Anomalous zonal-mean zonal wind, (middle) zonal-mean temperature, and (right) EP flux with its divergence during the growth and maturation on a composite SSW.

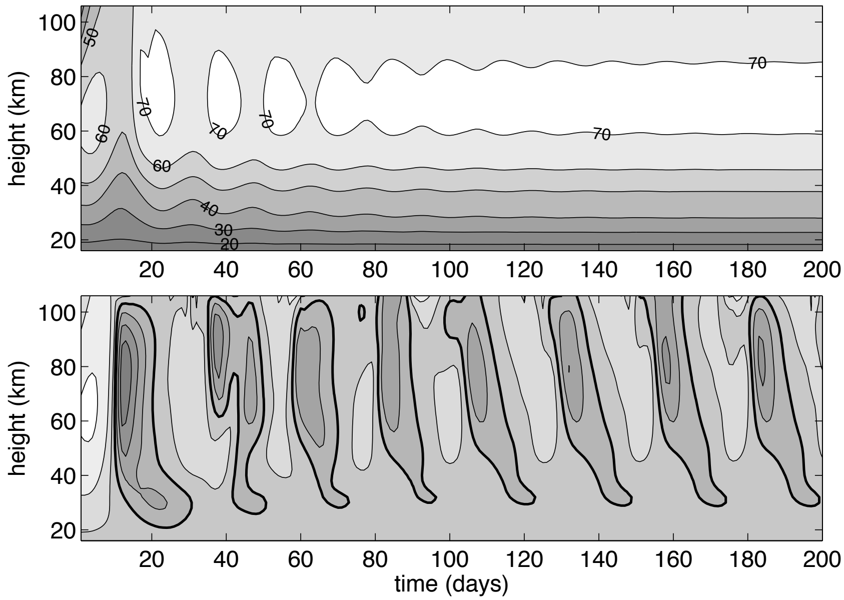


**Fig. 17.27** Evolution of the zonal mean zonal wind in a case with steady wave forcing with a value of 200 in the top panel and 300 in the lower panel. In the bottom panel the contours are every 20 m s$^{-1}$, positive values have lighter shades and the zero contour is heavy.

### 17.8.2 Wave–mean-flow interaction and internal stratospheric variability

As noted above, stratospheric variability need not arise solely from waves propagating up from the tropopause and we can illustrate this with a simple, albeit numerical, model of wave–mean-flow interaction, similar to those discussed in section 10.1.3. Specifically the model consists of the following quasi-geostrophic ideal-gas equations.[42]

For the mean flow we have

$$\frac{\partial \overline{q}}{\partial t} = \overline{F} - \frac{\partial}{\partial y}\overline{v'q'}, \tag{17.75a}$$

where

$$\overline{q}(y,z,t) - \beta y = \left[\frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R \frac{f_0^2}{N^2}\frac{\partial \Psi}{\partial z}\right) + \frac{\partial^2 \Psi}{\partial y^2}\right], \qquad \left(\overline{u}, \frac{R}{Hf_0}\overline{T}\right) = \left(-\frac{\partial \Psi}{\partial y}, \frac{\partial \Psi}{\partial z}\right) \tag{17.75b,c}$$

For the eddies we have

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\frac{\partial \overline{q}}{\partial y} = F', \tag{17.76a}$$

where

$$q'(x,y,z,t) = \nabla^2 \psi' + \frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R \frac{f_0^2}{N^2}\frac{\partial \psi'}{\partial x}\right), \qquad (u', v') = \left(-\frac{\partial \psi'}{\partial y}, \frac{\partial \psi'}{\partial x}\right) \tag{17.76b}$$

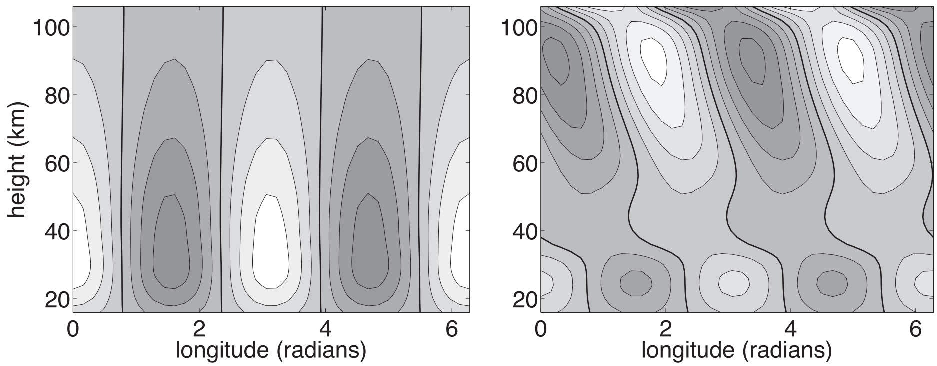


**Fig. 17.28** Snapshots of the wave streamfunction in the steady case (forcing of 200) in the left panel and in the oscillatory case (forcing of 300) in the right panel. Zero contour is heavy.

and

$$\frac{\partial \overline{q}}{\partial y} = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} - \frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R \frac{f_0^2}{N^2}\frac{\partial \overline{u}}{\partial z}\right). \tag{17.76c}$$

The notation follows our usual conventions, with $\rho_R(z)$ being a reference density profile and $\overline{F}$ and $F'$ the forcing/dissipation terms for the mean flow and eddies, respectively. It is the domain, the boundary conditions and forcing that distinguish the model and make it representative of the stratosphere, as we now discuss.

The model domain is a channel nominally centred at 60° and of width 60°, extending upwards to about 100 km. The forcing on the zonal flow is a relaxation back to a specified radiative equilibrium temperature field (or equivalently a thermal wind field). In the results shown below this is independent of time and corresponds to a constant shear of $1\,\mathrm{m\,s^{-1}}$ per kilometre, or a temperature difference of about 15 K across the domain, with a relaxation timescale that varies from 20 days at 20 km to 4 days at 50 km. There is also a weak linear drag on the mean flow. The eddies are forced by imposing a constant perturbation at the lower boundary, with wavenumber 2 in the simulations shown. There is a radiative damping on the eddies ensuring that the eddies are mostly damped before reaching the top of the domain. The vertical variations are represented using finite differencing, whereas in the horizontal both the mean flow and the eddies are expanded in a Fourier series with only a very small number of terms retained. Thus, we write

$$[\overline{q}, \Psi] = [Q_0(z,t), \Psi_0(z,t)]\cos ly, \quad [q', \psi'] = \mathrm{Re}\left[q_0(z,t), \psi_0(z,t)\right]\sin ly \exp(\mathrm{i}kx). \tag{17.77}$$

and after some manipulation we can obtain evolution equations for $Q_0$ and $q_0$ with diagnostic equations for $\Psi_0$ and $\psi_0$. The quadratic terms in the equations of motion create higher order terms that are projected back onto the retained terms. (Aside from the severe horizontal truncation the numerical method used to find results is not a key aspect of the model.)

### *Some numerical results and interpretation*

Results of two numerical integrations are shown in Fig. 17.27 and Fig. 17.28. In one integration the geopotential forcing at the lower boundary has an amplitude of 200 m, whereas in the other it has an amplitude of 300 m. In the first case the flow evolves into an absolute steady state, whereas in the second the mean flow and the waves oscillate with a period of about 25 days, with the mean flow actually becoming negative over half the cycle. The streamfunction in the unsteady case is tilting into the mean shear (the right panel of Fig. 17.28), evocative of baroclinic instability.

The oscillations are in some way evocative of stratospheric warmings. The climatological eastward winds transition rather quickly to westward winds (darker shading in Fig. 17.27), with the westward winds then descending and with a slower recovery back to climatology. There are two points to be made:

*(i)* Interactions that are internal to the stratosphere can give rise to oscillatory motion.

*(ii)* The source of energy for the waves may in part arise from a baroclinic instability and in part from tropospheric forcing.

Having said this, most stratospheric sudden warming can be traced back to a tropospheric disturbance.

*Notes*

1 To read more about the middle atmosphere as a whole see, for example, the review by Hamilton (1998), the collection of articles in *Journal of the Meteorological Society of Japan*, vol. 80, no. 4B, 2002, as well as Andrews *et al.* (1987). (Haynes 2005) provides a nice review of the large-scale dynamics of the extratropical stratosphere.

3 Adapted from Fels (1985) with the help of K. Hamilton.

5 Figure courtesy of J. Wilson of GFDL, using data from Fleming *et al.* (1988).

7 Courtesy of D. Waugh. Colour plots of figures may be downloaded from the web site of this book.

9 Courtesy of A. Dörnbrack.

11 Figure kindly prepared by M. Jucker using ERA-interim data. See also Lait (1994) for discussion of the alternative PV.

13 Adapted from Eluszkiewicz *et al.* (1997).

14 Brewer (1949) and Dobson (1956). Brewer deduced upward motion into the stratosphere at low latitudes based on the water vapour distribution, while Dobson deduced a poleward transport within the stratosphere based on the ozone distribution — the circulation takes ozone from the low latitudes toward the poles. Although originally the Brewer–Dobson circulation was taken to mean the chemical transport circulation, it is now usually taken to mean the residual (thickness or mass) circulation. The two may differ if there is mixing of chemical without mixing of mass, and the chemical transport may differ among chemicals.

15 See also Plumb (2002), which motivated this figure.

16 It turns out that $\overline{u'w'} > 0$ for upward propagating Rossby waves and thus, if the waves were to be dissipated, an eastward acceleration would seemingly be implied. In fact it is the form stress that is the most important aspect of vertical momentum transport in such waves, and when the waves are dissipated a westward acceleration ensues.

17 We particularly draw from Garcia (1987) and Haynes (2005).

19 Adapted from Holton *et al.* (1995).

21 Values are taken from Wallace (1973), Plumb (1984) and Andrews *et al.* (1987).

23 Adapted from Gray (2010). The observations used come mainly from satellites above 10 hPa and from satellites and radiosondes below.

24 A broad overview of QBO is provided by Baldwin *et al.* (2001), with updates and additions by Gray (2010), and a more theoretical review is given by Plumb (1984). The term quasi-biennial oscillation seems to have been coined by Angell & Korshover (1964), although the discovery of the QBO is generally credited to R. J. Reed and R. A. Ebdon, independently and at about the same time (Ebdon 1960, Veryard & Ebdon 1961, Reed 1960, Reed *et al.* 1961).

26 Figure adapted from Gray (2010), who used the method of Pascoe *et al.* (2005).

27 Wallace & Holton (1968). See Baldwin *et al.* (2001) for more discussion and references.

28 The first theory of the QBO along these lines was put forward by Lindzen & Holton (1968), with important clarifications and simplifications provided by Plumb (1977), and it is these models that we draw from. Plumb (1984) provides a useful review of the dynamics. A host of papers elaborating on the basic mechanism have since appeared, discussing such thing as the particular type of gravity waves involved, the role of the Coriolis force and the meridional confinement of the QBO, the possible influence of the solar cycle, the impact on tracer transport and so on.

30 Adapted from Plumb (1984).

32 Adapted from Plumb (1984).

33 Historically, a model with both Kelvin and Rossby waves was the first model to be proposed for the QBO, by Lindzen & Holton (1968) and Holton & Lindzen (1972); the simpler and clarifying model of section 17.7.3 comes from (Plumb 1977).

34 Here, as with some other scientific theories of complex phenomena, it is hard to be absolutely certain that the theory is capturing the phenomenon correctly, for the skeptic can always point to observational disagreements. Yet the analog of deciphering a complex communication seems apt: if after some effort an encrypted signal is deciphered to reveal a meaningful message, it seems perverse to ask whether the deciphering is unique, and whether some other message might have emerged from a different decryption.

35 Plumb (1977).

36 A comprehensive bibliography is provided by Baldwin *et al.* (2001), with some more recent references given by Gray (2010). Examples of additional theoretical development are Dunkerton (1982, 1997), Boyd (1976) and Plumb & Bell (1982), but there are several more. Reasonably realistic simulations of a QBO in an atmospheric GCM have been achieved by Hamilton *et al.* (2001), Scaife *et al.* (2002), Giorgetta *et al.* (2002) and others since. The possible effects of the QBO on the extra-tropical circulation are discussed by Holton & Tan (1980, 1982), Jones *et al.* (1998), Kushner (2010), Labitzke *et al.* (2006), Randel *et al.* (1999), Scott & Haynes (1998a), Dunkerton *et al.* (1988) and others. The list goes on.

37 See references in the previous endnote.

39 From Charlton & Polvani (2007).

40 The model described here was proposed by Matsuno (1971) and although nonlinear and, to a lessor degree, non-geostrophic, effects play a quantitative role Matsuno's model remains the foundation of our understanding. An early review is that of Schoeberl (1978) and there have been numerous modelling and observational studies since then. Thus, to name but a few of many, Dunkerton *et al.* (1981) and Palmer (1981) explored the phenomenon from a TEM perspective, Limpasuvan *et al.* (2000) and Charlton & Polvani (2007) provide a comprehensive view of the observations of warmings using reanalysis datasets, Charlton *et al.* (2007) look at various simulations with GCMs, and Gray *et al.* (2001) look at external influences on the timing of warmings.

41 These simulations were kindly performed by Martin Jucker. See also Jucker *et al.* (2013).

42 This model was introduced by Holton & Mass (1976) and the numerical results we show use a code originally written by J. Holton. Yoden (1987, 1990) explored the bifurcation properties of the model, and the effects of a seasonal cycle, with a view to better understanding the parameters for which steady, oscillatory or chaotic motion was present. Christiansen (1999, 2000), Scaife & James (2000), Scott & Haynes (1998b) and others have explored related behaviour using more complete models, for example with the primitive equations and/or including eddy-eddy interactions and with different boundary conditions. Haynes (2005) provides a review and literature survey.

### *Further reading*

Andrews, D. G., Holton, J. R. & Leovy, C.B., 1987. *Middle Atmosphere Dynamics.*
A textbook discussing both the theory and observations of the middle atmosphere in some detail.

### *Problems*

17.1 Please let me know if you have any.

# Part IV

# LARGE-SCALE OCEANIC CIRCULATION

*With the possible exception of the equator, everything begins somewhere.*
C. S. Lewis

# CHAPTER TWENTYTWO

# Equatorial Circulation of the Ocean

In this chapter we discuss the circulation of the equatorial ocean, with particular emphasis on the equatorial undercurrent.

### A cursory comment on customs

> *I'm getting youngerly every day.*
> Paul Kushner, University of Toronto.

Meteorologists tend to talk about westerly winds — the winds that come from the west — because it is where the winds come from that determines what the weather will be. Oceanographers tend to talk about eastward currents, because this is where the currents will take things (or perhaps oceanographers are just a more forward looking crowd). It can be confusing to mix the conventions, and so to keep thing as clear as possible we will follow the lead of the oceanographers and talk about eastward and westward flow, for both currents and winds. Also, some sources suggest that westward is an adjective and westwards is an adverb, but there is little consistency in usage, especially for the adverb, with a slight tendency to use westwards in British English and westward in American English. We'll use westward (and eastward, etc.) as both adjective and adverb.

## 22.1 THE OBSERVED CURRENTS

[Needs a bit more here, but not a whole book — maybe a few more plots from ECCO or World Ocean Atlas or similar. Let me know if you have suggestions or, better still, nice plots.]

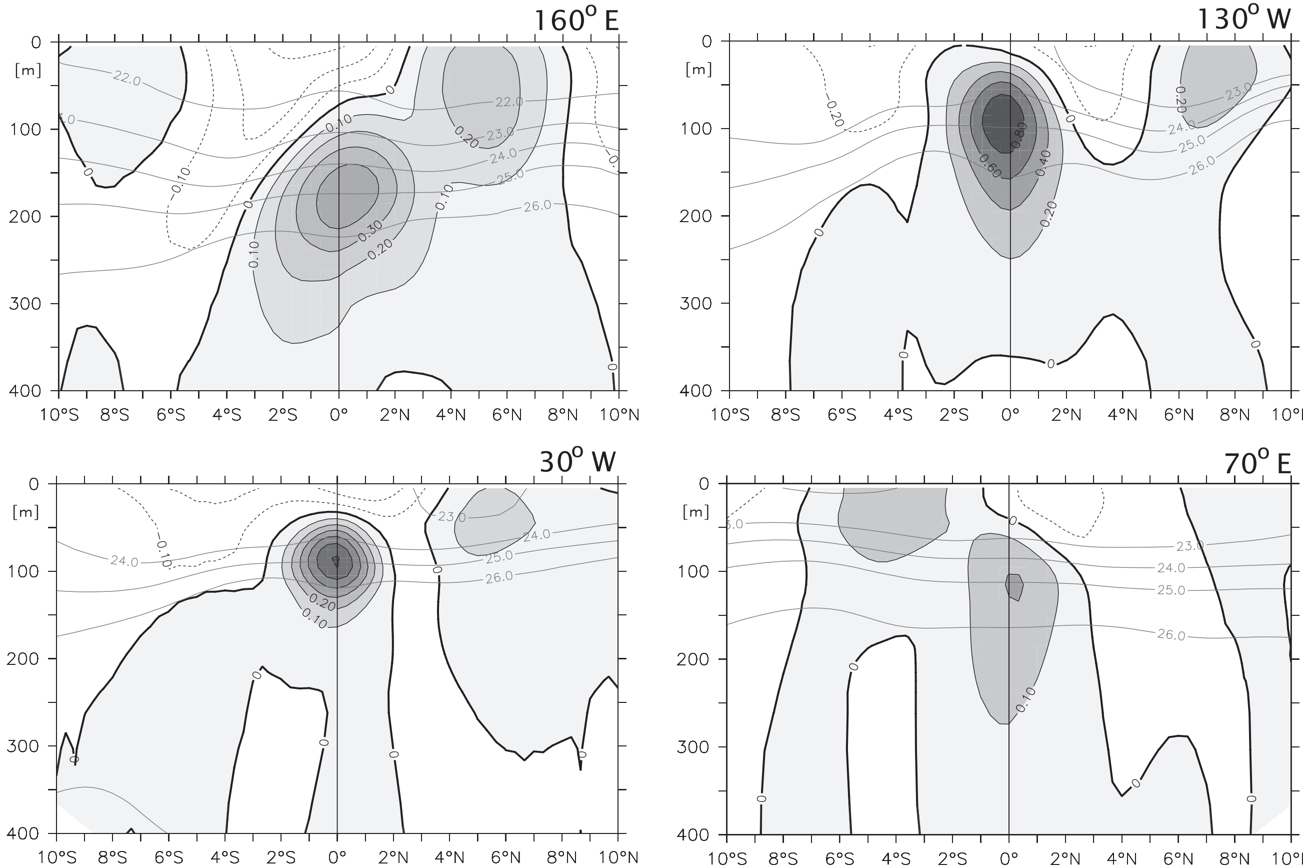


**Fig. 22.1** Sections of the observed mean zonal current (thicker contours and shading) at two longitudes in the Pacific (upper panels), in the Atlantic (lower left) and in the Indian Ocean (lower right). The contours are every 20 cm $s^{-1}$ in the upper two panels and every 10 cm $s^{-1}$ in the lower panels. Note the well-defined eastward undercurrent at the equator in all panels, and a weaker eastward countercurrent at about 6° N and/or 6° S. The thinner, more horizontal lines are isolines of potential density.[2]

In mid- and high-latitudes the large-scale upper-ocean currents are those of the great gyres and, in the Southern Hemisphere, the Antarctic Circumpolar Current. The gyres are very robust features, existing in all the basins, and may be understood as the direct response to the winds, and in particular the curl of the wind stress. In the equatorial regions the currents also display some very robust features, illustrated in Fig. 22.1 and the top panel of Fig. 22.2. The observations shown were mainly made with acoustic Doppler current profilers (ADCP), which measure the currents by measuring the Doppler shift from a sonar. [xxxx] The main features are as follows.

*(i)* A shallow westward flowing surface current, typically confined to the upper 50 m or less, strongest within a few degrees of the equator, although not always symmetric about the equator. Its speed is typically a few tens of centimetres per second.

*(ii)* A strong coherent eastward undercurrent extending to about 200 m depth, confined to within a few degrees of the equator. Its speed may be up to a meter per second or a little more, and it is this current that dominates the vertically integrated transport at the equator. Beneath the undercurrent the flow is relatively weak.

*(iii)* Westward flow on either side of the undercurrent, with eastward countercurrents poleward of this. In the Pacific the countercurrent is strongest in the Northern Hemisphere where it reaches the surface.

## 22.2 DYNAMICAL PRELIMINARIESS

In mid-latitudes the large scale currents system may be understood using the planetary geostrophic equations of motion. Applying these allows us to understand formation of the great wind-driven gyres, with Sverdrup balance providing a solid foundation on which to build. As we approach lower latitudes the Coriolis parameter, $f$ decreases and the Rossby number increases and one might expect that dynamics based on geostrophic balance will ultimately fail. Perhaps surprisingly, it is only very close to the equator that the Rossby number exceeds unity. If we take a velocity of $0.5\,\mathrm{m\ s^{-1}}$ and a length scale of 500 km then the Rossby number at 5° latitude is 0.08, at 2°, 0.2 and at 1°, 0.4. These numbers suggest that until we virtually at the equator (where the Rossby number is infinite) we can use some of the familiar tools from the midlatitude dynamics. Of course at the equator the Coriolis parameter switches sign and this leads to some interesting features. The vertical structure is also a little complex so let us first see the extent to which the familiar Sverdrup balance can explain the vertically integrated flow.

### 22.2.1 The vertically integrated flow and Sverdrup balance

The horizontal momentum may be written

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \boldsymbol{f}\times\boldsymbol{u} = -\nabla\phi + \frac{1}{\rho_0}\frac{\partial \boldsymbol{\tau}}{\partial z} \tag{22.1}$$

where $\boldsymbol{\tau}$ is the stress on the fluid. The mass conservation equation is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{22.2}$$

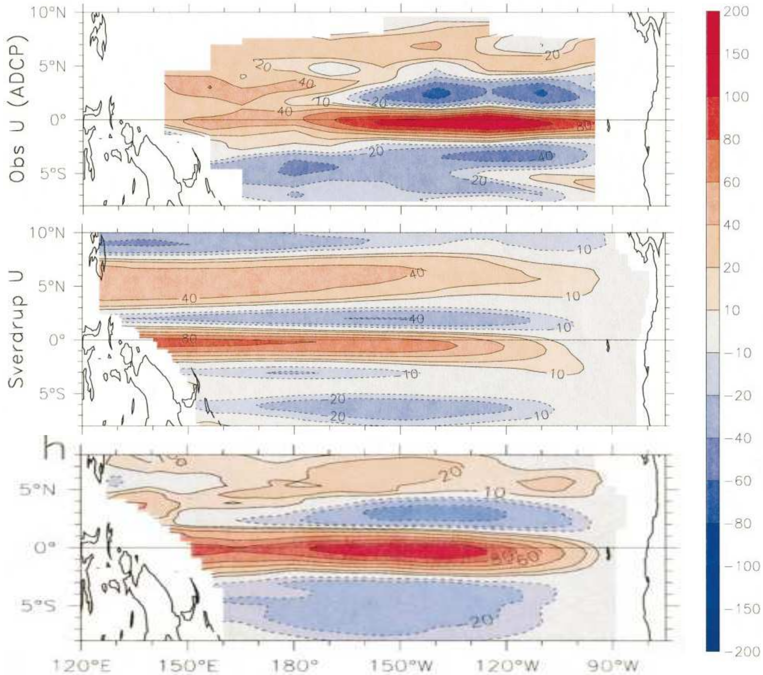


**Fig. 22.2** Vertically integrated zonal transport in the Pacific. Red colours indicate eastward flow, blue colours westward. The top panel shows the observed flow, the middle panel shows the flow calculated using Sverdrup balance with the observed wind, and the bottom panel shows the flow calculated with a 'generalized' Sverdrup balance that includes the nonlinear terms in a diagnostic way.[4]

which on vertical integration over the depth of the ocean

$$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = 0, \tag{22.3}$$

where $U$ and $V$ are the vertically integrated zonal and meridional velocities (e.g., $U = \int u \, \mathrm{d}z$) and we assume the ocean has a flat bottom and a rigid lid at the top. If we assume the flow is steady and integrate (22.1) vertically, then take the curl and use (22.3) we obtain

$$\beta V = \mathrm{curl}_z(\widetilde{\boldsymbol{\tau}}_T - \widetilde{\boldsymbol{\tau}}_B) + \mathrm{curl}_z \boldsymbol{N}, \tag{22.4}$$

where $\widetilde{\boldsymbol{\tau}}$ is the kinematic stress ($\widetilde{\boldsymbol{\tau}} = \boldsymbol{\tau}/\rho_0$ where $\rho_0$ is the reference density of seawater) with the subscripts $T$ and $B$ denoting top and bottom, $\boldsymbol{N}$ represents all the nonlinear terms and $\mathrm{curl}_z$ is defined by $\mathrm{curl}_z \boldsymbol{A} \equiv \partial A^y/\partial x - \partial A^x/\partial y = \mathbf{k} \cdot \nabla \times \boldsymbol{A}$. Equations (22.4) and (22.3) are closed equations for the vertically averaged flow. In oceanography we nearly always deal with the kinematic stress rather than the stress itself, so henceforth we will drop the tilde over the $\tau$ symbol. In the few cases that we need to refer to the actual stress we will denote this by $\tau^*$; thus, $\tau = \tau^*/\rho_0$. We will also drop the adjective 'kinematic' unless there is a specific need to be explicit or avoid ambiguity.

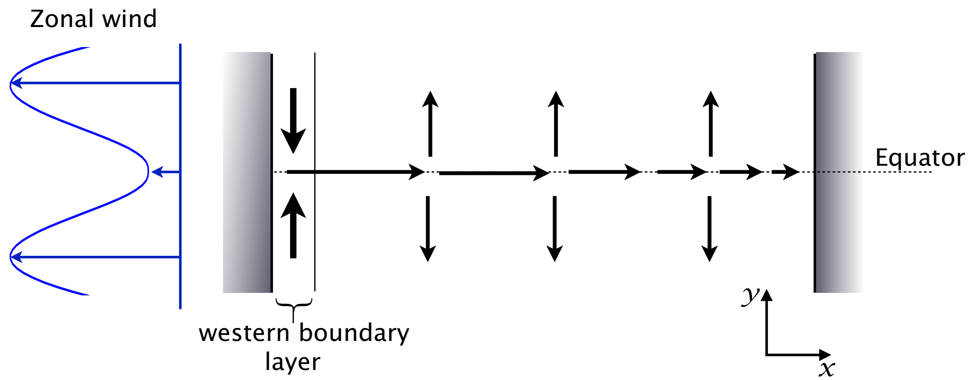


**Fig. 22.3** Schema of Sverdrup flow at the equator between two meridional boundaries. The mean winds are all westward, but with a minimum in magnitude at the equator. By Sverdrup balance, (22.5), the wind stress produces the divergent meridional flow shown, which in turn induces an eastward equatorial zonal flow, strongest in the western part of the basin.

If we neglect the nonlinear terms and the stress at the bottom (we'll come back to these terms later) then (22.4) becomes

$$\beta V = \mathrm{curl}_z \boldsymbol{\tau}_T. \tag{22.5}$$

This is just Sverdrup balance, which first appeared in chapter 14, eq. (19.20). The zonal transport is obtained by differentiating (22.5) with respect to $y$, using (22.3) to replace $\partial_y V$ with $\partial_x U$, and then integrating from the eastern boundary ($x_E$). This procedure gives

$$U = -\frac{1}{\beta}\int_{x_E}^{x} \frac{\partial}{\partial y}\mathrm{curl}_z \boldsymbol{\tau}_T \, \mathrm{d}x' + U(x_E, y). \tag{22.6}$$

We don't integrate from the western boundary because a boundary layer can be expected there, whereas the value of $U$ at the eastern boundary will be small.

If $U(x_E, y) = 0$ and the wind is zonally uniform then (22.6) becomes

$$U(x, y) = \frac{1}{\beta}(x - x_E)\frac{\partial^2 \boldsymbol{\tau}_T}{\partial y^2}. \tag{22.7}$$

That is, the depth integrated flow is proportional to the second derivative of the zonal wind stress, and because $x < x_E$ we have $U \propto -\partial^2 \boldsymbol{\tau}_T/\partial y^2$. Evidently, the result will depend rather sensitively on the wind pattern. Although the zonal wind is generally westward in the tropics there is a minimum in the magnitude of that wind near the equator (that is, a local maximum as schematized in Fig. 22.3) so that $\partial^2 \boldsymbol{\tau}/\partial y^2$ is negative. Thus, using (22.7), $U$ will generally be positive at the equator. Using the observed wind field the Sverdrup flow — that is, the solution of (22.6) with $U(x_E, y) = 0$ — can be calculated and this is plotted in the middle panel of Fig. 22.2. There is a good but not perfect agreement with the observations: the observed flow has its maximum further east. Further, in the western equatorial Pacific the observed eastward flow is quite broad whereas the eastward Sverdrup flow is narrow, flanked on either side by westward flow. Some of the discrepancy can be attributed to the role of the nonlinear and

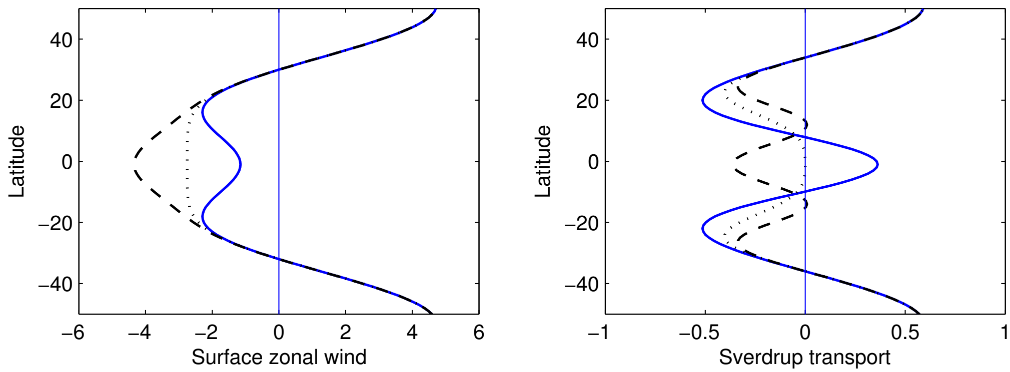


**Fig. 22.4** The left panel shows three putative surface zonal (atmospheric) winds, $u$, all with westward winds in the tropics and with the solid line being the most realistic. The right panel shows the corresponding negative of the second derivative, $-\partial^2 u/\partial y^2$, proportional to the (oceanic) Sverdrup transport, in arbitrary units. The wind represented by solid line gives an eastward transport at the equator, as is observed, with the others differing markedly.

frictional terms, as illustrated in the bottom panel of Fig. 22.2. To obtain the flow illustrated, the calculation proceeds from (22.4) in the same way as before, but now includes the nonlinear terms and a representation of frictional effects in a diagnostic fashion. Thus, for example, the nonlinear terms of the form $\mathrm{curl}_z(\int \boldsymbol{u}\cdot\nabla\boldsymbol{u}\,\mathrm{d}z)$ are evaluated and used to calculate a generalized Sverdrup flow, where the velocities are taken from a nonlinear model forced by the observed winds. Of the nonlinear terms, the largest ones involve the meridional derivatives of the zonal flow, for example $\partial_y(uu_x)$. The effect of the nonlinear terms is to decelerate the eastward flow in the eastern Pacific, with friction tending to damp the flow especially in the central Pacific, and the resulting flow is evidently closer to the observations than is linear Sverdrup balance. Of course the full solution (22.4) must give a vertically integrated flow that closely resembles the observations, because there are only very weak approximations made in deriving it. The success of the Sverdrup theory lies in the extent to which the vertically integrated flow can be satisfied by the simple linear balance (22.5), and then improved by adding nonlinear and dissipative terms in a diagnostic fashion.

### 22.2.2 Delicacy of the Sverdrup flow

Although the calculations of Sverdrup flow do show good agreement with observations the calculation — and, most likely, the observed flow — is rather sensitive to the precise form of the winds, as illustrated in Fig. 22.4. The figure shows three surface zonal wind distributions, with the 'w' shaped solid line having a minimum in the westward flow (i.e., a minimum in the trade winds) at the equator and so being the most realistic. The right-hand panel shows the negative of the second derivative of the winds which is proportional to the zonal Sverdrup flow. Only in the one case does the wind produce an eastward Sverdrup flow. In fact, in the case illustrated with the dashed lines, the small changes in the meridional gradient of the wind between 15° and 20° produce quite large variations in the Sverdrup transport. Given this delicacy, the small difference in the latitudinal variation of the Sverdrup flow and the observed flow, illustrated in the top and middle panels of Fig. 22.2, is not surprising and cannot be considered a major failure of the theory. However, the difference in the longitudinal structure of the two fields is indicative of the importance of other terms in the vorticity balance.

## 22.3 A LOCAL MODEL OF THE EQUATORIAL UNDERCURRENT

The most conspicuous feature of the ocean current system at low latitudes is the equatorial undercurrent, and we now consider its dynamics.[5] The physical picture that we first discuss is a 'local' one, and is essentially the following. The mean winds are westward and provide a stress on the upper ocean, pushing the near-surface waters westward. Given that there is a boundary in the west, the water piles up there so creating a pressure-gradient force that pushes fluid eastward. To some degree the pressure gradient and the wind stress compensate each other leading to a state of no motion. However, the compensation is not perfect. Close to the surface the stress is dominant and a westwards surface current results. Below the surface the pressure gradient dominates, resulting in an eastward flowing undercurrent, as seen in the observations shown in Fig. 22.1.

The above description makes no mention of the Coriolis parameter or Sverdrup balance or the wind-stress curl. On the one hand that suggests that the dynamics are likely to be robust and will not depend in a delicate way on the wind pattern in the same way that the Sverdrup flow does. On the other hand, given the usefulness of the Sverdrupian concept, such a description is also likely to be incomplete. To proceed further we'll construct a small hierarchy of mathematical models of the equatorial current system, beginning with the very simplest model of a homogeneous fluid subject to a uniform westward stress at the surface.

Following this we discuss a more inertial and non-local physical picture, in which the undercurrent may be thought of as being pushed by a pressure head that begins in extra-equatorial regions. In the extreme limiting case of this picture, the winds at the equator have no effect on the undercurrent. The real equatorial undercurrent likely involves a combination of local and inertial dynamics, and is still a topic of research.

### 22.3.1 Response of a homogeneous layer to a uniform zonal wind

Let us first consider the simple case of the response of a layer of homogeneous fluid to a steady zonal wind that is uniform in the $y$-direction. With our usual notation the equations of motion in the presence of momentum and mass forcing are

$$\frac{\mathrm{D}u}{\mathrm{D}t} - fv = -g'\frac{\partial \eta}{\partial x} + \frac{\tau^x}{H} \tag{22.8a}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t} + fu = -g'\frac{\partial \eta}{\partial y} + \frac{\tau^y}{H} \tag{22.8b}$$

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = M. \tag{22.8c}$$

where $(\tau^x, \tau^y)$ are the zonal and meridional kinematic stresses on the fluid, $H$ is the depth of the fluid and $M$ is a mass source, which for now we take to be zero. For steady flow and neglecting the nonlinear terms the equations become

$$-fv = -g'\frac{\partial \eta}{\partial x} + \frac{\tau^x}{H} \tag{22.9a}$$

$$+fu = -g' \frac{\partial \eta}{\partial y} \tag{22.9b}$$

$$H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0. \tag{22.9c}$$

If we take the $y$-derivative of (22.9a) and subtract it from the $x$-derivative of (22.9b), and noting that $\partial \tau^x / \partial y = 0$, we obtain

$$\beta v = 0. \tag{22.10}$$

Thus, using the continuity equation (22.9c), we have $\partial u/\partial x = 0$. That is, the zonal velocity is uniform. If there is a zonal boundary at which $u = 0$ then the zonal flow is zero everywhere and the complete solution is

$$u = 0, \quad v = 0, \quad g\frac{\partial \eta}{\partial x} = \frac{\tau^x}{H}, \quad \frac{\partial \eta}{\partial y} = 0. \tag{22.11}$$

That is to say, the ocean is motionless and the wind stress is balanced by a pressure gradient. If the wind is westward, as it is on the equator, then $\partial \eta / \partial x < 0$ and the thermocline slopes down and deepens toward the west. The fact that there is no flow could of course have been anticipated from Sverdrup balance in the absence of a wind-stress curl. Although the real ocean is not as simple as our model of it, the analysis exposes a truth with some generality: the wind stress is largely opposed by a pressure gradient rather than inducing a large westward acceleration that is eventually halted by friction.

### 22.3.2 An unstratified local model of the equatorial undercurrent

Let us now consider a model with some vertical structure, thereby allowing the wind stress to be taken up in the upper ocean. The wind will still push near-surface water westwards and create a zonal pressure gradient. The deeper water will feel the pressure-gradient force — because the pressure is hydrostatic — but not the wind stress, and so flows eastwards. A simple model that can capture these effects begins with the three-dimensional momentum equations, namely

$$-fv = -\frac{\partial \phi}{\partial x} + \nu_z \frac{\partial^2 u}{\partial z^2} + \nu_h \nabla^2 u, \tag{22.12a}$$

$$fu = -\frac{\partial \phi}{\partial y} + \nu_z \frac{\partial^2 v}{\partial z^2} + \nu_h \nabla^2 v, \tag{22.12b}$$

In these equations the zonal and meridional velocities are, as usual, $u$ and $v$ and $\phi$ is the kinematic pressure. The parameters $\nu_z$ and $\nu_h$ are eddy viscosities acting on vertical and horizontal shear, respectively, and the $\nabla$ operator is purely horizontal (so that $\nabla^2 u = \partial^2 u/\partial x^2 + \partial^2 u/\partial y^2$). Dealing with a horizontal viscosity requires a more mathematically cumbersome treatment that we defer that to section 22.3.3; rather, in its place we will invoke a linear drag whence the momentum equations, along with the mass continuity equation, become

$$-fv = -\frac{\partial \phi}{\partial x} + \nu_z \frac{\partial^2 u}{\partial z^2} - ru, \tag{22.13a}$$

$$fu = -\frac{\partial \phi}{\partial y} + \nu_z \frac{\partial^2 v}{\partial z^2} - rv, \tag{22.13b}$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0. \tag{22.13c}$$

The drag terms are presumed to act throughout the depth of the fluid and so are a little ad hoc, but their presence enables us to construct a simple and very illuminating model. We should also remember that almost *any* frictional terms in a model of the large-scale circulation are to some degree ad hoc: the viscosities $(\nu_z, \nu_z)$ are certainly not molecular viscosities and there is no proper justification for the use of eddy viscosities on momentum.

The vertical friction terms $(\partial^2 u/\partial z^2, \partial^2 v/\partial z^2)$ enable the wind's influence to be felt in the upper ocean via the boundary conditions, namely

$$\nu\frac{\partial u}{\partial z}=\tau^x, \qquad \nu\frac{\partial v}{\partial z}=\tau^y \qquad\qquad \text{at } z=0, \tag{22.14a}$$

$$\nu\frac{\partial u}{\partial z}=0, \qquad \nu\frac{\partial v}{\partial z}=0 \qquad\qquad \text{at } z=-H, \tag{22.14b}$$

where $(\tau^x, \tau^y)$ is the kinematic wind stress. With boundary conditions of $w = 0$ at top and bottom the vertical integral of (22.13c) is

$$\frac{\partial U}{\partial x}+\frac{\partial V}{\partial y}=0, \tag{22.14c}$$

where $(U, V) = \int (u, v)\,\mathrm{d}z$ is the vertically integrated flow. Equation (22.14c) allows for the introduction of a streamfunction $\psi$ such that $U = -\partial\psi/\partial y$ and $V = \partial\psi/\partial x$. Cross-differentiating (22.13a,b) and vertically integrating then gives

$$r\nabla^2\psi+\beta\frac{\partial\psi}{\partial x}=\mathrm{curl}_z\tau. \tag{22.15}$$

This is the equation of Stommel's model, essentially the same as (19.6) [reference to *AOFD* ], and in the absence of the frictional term the vertically integrated flow is given by Sverdrup balance. If the wind has no curl the vertically integrated flow is zero, as before. However, the flow is not zero at each vertical level as we now see.

Let us now assume the flow is unstratified, meaning that the buoyancy $b$ is a constant, which we take to be zero. The hydrostatic relation is $\partial\phi/\partial z = b = 0$ so that $\phi$ is uniform with height. From (22.13a,b) the vertically integrated momentum equations are then

$$H\frac{\partial\phi}{\partial x}=\tau^x-rU+fV, \qquad H\frac{\partial\phi}{\partial y}=\tau^y-rV-fU. \tag{22.16a,b}$$

Let us further suppose that the stress (i.e., $\boldsymbol{\tau} = \nu\partial\boldsymbol{u}/\partial z$) is non-zero only in a shallow layer — an Ekman layer — in the upper ocean. Below this layer we have, from (22.13a,b),

$$-fv=-\frac{\partial\phi}{\partial x}-ru, \qquad fu=-\frac{\partial\phi}{\partial y}-rv. \tag{22.17a,b}$$

and using (22.16) we obtain

$$-fv'=-\frac{\tau^x}{H}-ru', \qquad fu'=-\frac{\tau^y}{H}-rv'. \tag{22.18a,b}$$

where $u' \equiv u - U/H$ and $v' \equiv v - V/H$ is the deviation of the flow from the vertical average (i.e., the deviation from Sverdrup balance). That is to say, we may solve the equations assuming the Sverdrup flow is zero, and add it back in at the end of the day, noting also that the presence of a Sverdrup flow makes no difference to the vertical velocity. Given this, we'll drop the prime on $u'$ and $v'$ unless ambiguity would arise. Solving for $u$ and $v$ gives the expressions for the deep flow, namely

$$u = \frac{-\tau^x r - \tau^y f}{H(r^2 + f^2)}, \qquad v = \frac{\tau^x f - \tau^y r}{H(r^2 + f^2)}. \tag{22.19a,b}$$

The transport in the Ekman layer at the surface is in the opposite direction to the deep flow, in order to satisfy the integral constraints that $\int u \, dz = \int v \, dz = 0$. To complete the solution we use the mass continuity equation, (22.13c), to obtain $w$, giving

$$w = -\frac{(z + H)}{H} \frac{\beta(r^2 - \beta^2 y^2)\tau^x + 2r\beta^2 y \tau^y}{(r^2 + \beta^2 y^2)^2}. \tag{22.20}$$

To better understand these solutions it is useful to look at the nondimensional form and we obtain that by setting

$$(u, v) = (\widehat{u}, \widehat{v})\frac{\tau}{2\Omega H}, \quad y = \widehat{y}a, \quad (\tau^x, \tau^y) = (\widehat{\tau}^x, \widehat{\tau}^y)\tau, \quad \beta = \frac{2\Omega}{a} \tag{22.21}$$

where a hat denotes a nondimensional quantity and $a$ is the radius of Earth. The nondimensional versions of (22.18) are then

$$-\widehat{y}\widehat{v} = -E_r \widehat{u} - \widehat{\tau}^x, \qquad \widehat{y}\widehat{u} = -E_r \widehat{v} - \widehat{\tau}^y, \tag{22.22}$$

where $E_r = r/(2\Omega)$ is a horizontal Ekman number and if, for example, the wind is zonal and westward then $\widehat{\tau}^x = -1$ and $\tau^y = 0$. The nondimensional versions of (22.19) are

$$\widehat{u} = \frac{-E_r \widehat{\tau}^x - \widehat{\tau}^y \widehat{y}}{E_r^2 + \widehat{y}^2}, \qquad \widehat{v} = \frac{\widehat{\tau}^x \widehat{y} - \widehat{\tau}^y E_r}{E_r^2 + \widehat{y}^2}. \tag{22.23a,b}$$

The overall strength of the undercurrent scales, unsurprisingly given the nature of the model, with the wind stress but the Ekman number determines the width and height of the profile. A typical solution is plotted in Fig. 22.6 (along with a solution using harmonic friction that we discuss later). The parameters are $\widehat{\tau}^x = -1, \widehat{\tau}^y = 0$ and $E_r = 8 \times 10^{-3}$, which corresponds to a purely westward wind and a frictional decay timescale of about 10 days. If we further suppose that the dimensional value of the stress is about $4 \times 10^{-2}\,\mathrm{N/m^2}$ and take $H = 100\,\mathrm{m}$ we obtain the dimensional values shown in the plot. The zonal flow as given by (22.19a) is then *eastward*, for the reason we have mentioned before, namely that, overall, the wind is balanced by an opposing pressure gradient and the deep ocean feels the pressure gradient but not the wind stress; thus, the deep zonal flow is in the opposite direction to the surface wind. The deep meridional flow is zero at the equator, where $f = 0$, but is toward the equator in both hemispheres and so induces equatorial upwelling.

The shallow Ekman-layer flow is away from the equator, in order that the vertically integrated flow is zero. A consequence of this is that the vertical velocity is positive — that is, there is *upwelling* at the equator, as can be seen directly from (22.20) when $\tau^y = 0$, $y = 0$ and $\tau^x < 0$.

The zonal undercurrent falls with latitude with a width proportional to $E_r$. The peak value at the equator is proportional to $E_r^{-1}$, so that by reducing the drag we make the equatorial peak sharper. However, and as that scaling suggests, the overall transport is independent of $E_r$ (at least for a constant, zonal stress). To see this we integrate (22.23) with $\widehat{\tau}^x = -1$ and $\widehat{\tau}^y = 0$:

$$\widehat{U}_T = \int_{-\infty}^{\infty} \widehat{u}\,\mathrm{d}\widehat{y} = \int_{-\infty}^{\infty} \frac{E_r}{E_r^2 + \widehat{y}^2}\,\mathrm{d}\widehat{y} = \left[\tan^{-1}\frac{\widehat{y}}{E_r}\right]_{-\infty}^{\infty} = \pi. \tag{22.24}$$

Dimensionally, this translates to

$$U_T = H\int_{-\infty}^{\infty} u\,\mathrm{d}y = \frac{-\pi a \tau^x}{2\Omega}. \tag{22.25}$$

It is pleasing that the total transport of the undercurrent does not depend on the rather poorly-constrained frictional coefficient, although the transport as given by (22.25) is somewhat smaller than observed. This can be guessed from Fig. 22.6 where the parameters are such that the width of the undercurrent is similar to that observed but its magnitude is too low (compare with Fig. 22.1). If we take $\tau^x = 4\times10^{-2}\,\mathrm{N/m^2}$ then using (22.25) we obtain a transport of about $5\times10^6\,\mathrm{m^3\,s^{-1}}$ or 5 Sv whereas the observed transport, with the vertical average (i.e, the Sverdrup flow) removed is 10–15 Sverdrups. Part of the discrepancy may come from the neglect of nonlinearity and stratification, and part of it from there being an inertial component to the equatorial undercurrent that is not a local response to the wind field, as we discuss in section 22.4.

The expressions are also useful when the wind is not purely zonal. In the somewhat less realistic situation in which the wind is northward ($\tau^y > 0, \tau^x = 0$), the deep flow is southward. If the wind blows toward the northwest the undercurrent flows down the pressure gradient to the southeast.

*Vertical structure at the equator*

Because there is no lateral friction the solution at the equator is independent of the solution elsewhere and an analytic form for the vertical profile may easily be obtained. The Coriolis parameter is zero and so, from (22.13), the equations of motion become

$$0 = -\frac{\partial\phi}{\partial x} + \nu\frac{\partial^2 u}{\partial z^2} - ru, \tag{22.26a}$$

$$0 = -\frac{\partial\phi}{\partial y} + \nu\frac{\partial^2 v}{\partial z^2} - rv. \tag{22.26b}$$

If the meridional wind stress at the surface is zero (i.e., $\nu_z\partial v/\partial z = 0$ at $z = 0$) then $v = 0$ everywhere. The zonal pressure gradient is given by (22.16a) and the zonal flow is then given by the solution of

$$\nu_z\frac{\partial^2 u}{\partial z^2} - ru = \frac{\tau^x}{H}, \tag{22.27}$$

with boundary conditions

$$\nu_z\frac{\partial u}{\partial z} = \begin{cases}\tau^x & \text{at } z = 0\\ 0 & \text{at } z = -H.\end{cases} \tag{22.28}$$

The solution is easily found to be

$$u = A\mathrm{e}^{\alpha z} + B\mathrm{e}^{-\alpha z} - \frac{\tau^x}{Hr}, \tag{22.29}$$

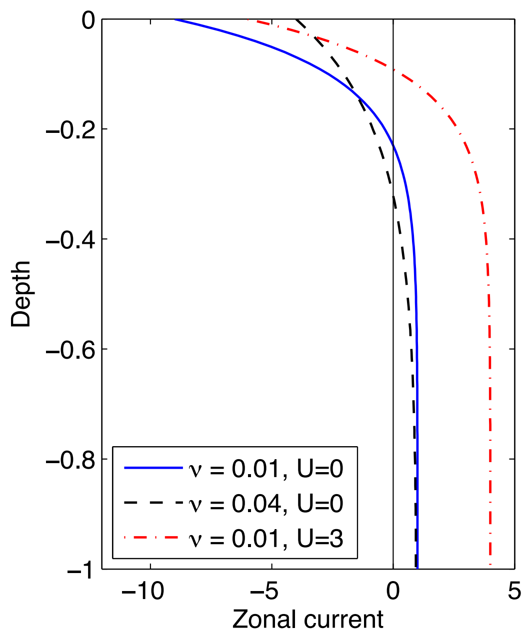


**Fig. 22.5** Vertical profile of the zonal current at the equator, obtained using the analytic solutions (22.29) and (22.30), with $H = 1$, $r = 1$, $\tau^x = -1$ and the values of $\nu$ and $U$ indicated in the legend.

where $\alpha = \sqrt{r/\nu_z}$ and $A$ and $B$ are obtained from the boundary conditions. We find

$$A = \frac{\tau^x}{\sqrt{\nu_z r}} \left( \frac{e^{\alpha H}}{e^{\alpha H} - e^{-\alpha H}} \right), \qquad B = \frac{\tau^x}{\sqrt{\nu_z r}} \left( \frac{e^{-\alpha H}}{e^{\alpha H} - e^{-\alpha H}} \right). \tag{22.30a,b}$$

A key parameter is the depth scale $d = \alpha^{-1} = \sqrt{\nu_z/r}$ that determines the depth to which the surface flow extends: if $\nu_z/r$ is small the flow in the direction of the wind is confined to a shallow layer near the surface with the undercurrent beneath. A couple of example solutions are illustrated in Fig. 22.5.

These solutions indicate one failing of this simple model: the undercurrent is too deep and extends all the way to the bottom of the ocean; evidently the model fails to reproduce a coherent, focussed eastward flowing jet of finite vertical extent such as is seen in Fig. 22.1. The main effect that makes a difference to this is stratification (with nonlinearity an important secondary effect) and we'll consider how this constrains the vertical extent later on.

### *A note on the undercurrent in the presence of a Sverdrup flow*

The zonal winds in the tropics have a minimum in the westward flow, that is a local maximum in $u$, at the equator and produce an eastward vertically integrated (Sverdrup) flow, as sketched by the solid line in Fig. 22.4. It seems natural to associate this flow with the eastwards undercurrent, but in and of itself this be misleading. The Sverdrup flow is produced by the wind-stress curl whereas the undercurrent is a consequence of the wind itself, and the two are not necessarily in the same direction. If, for example, the meridional variation of the wind differed, and were more akin to the dashed line in Fig. 22.4, then the Sverdrup flow would be westward. Whether the undercurrent would be eastward or westward now depends on the relative strength of the Sverdrup flow as well as other parameters, as a very simple argument shows.

## The Mean Equatorial Currents

The main observed features of tropical currents are as follows.

- Vertically integrated flow that is in approximate Sverdrup balance, but with non-negligible contributions from nonlinearity and friction. At the equator this flow is eastward, flanked by narrow westward and eastward moving strips, transitioning to broader westward flow polewards of about 10° that is part of the main subtropical gyres.
- A shallow westward flow at the equator, no more than a few tens of meters deep and a few degrees wide, with speeds of a few tens of centimetres per second.
- A strong eastward flowing undercurrent, typically from about 50 m to 200 m depth and a few degrees wide, with velocities up to a metre per second.

The leading order dynamics of this flow is roughly as follows.

- The zonal Sverdrup flow is proportional to the meridional derivative of the wind-stress curl, and so roughly to $-\partial^2 u_s/\partial y^2$ where $u_s$ is the surface zonal wind. The vertically integrated eastward flow at the equator is thus a response to the *minimum* in the westward trade winds at the equator.
- The shallow surface westward flow, and the strong eastward undercurrent, are primarily a response to the westward winds themselves, rather than the curl of the winds, and so are very robust features. (See the shaded box on page 942 for more about the equatorial undercurrent.)
- If the surface zonal winds were uniformly westward in the tropics, or had a westward maximum at the equator, the vertically integrated flow would be quite different and might be westward because of the dependence of the zonal flow on the second derivative of the wind stress in Sverdrup theory. However, there might well still be an eastward equatorial undercurrent, depending on the strength of the Sverdrup flow.

The deep flow is the superposition of the Sverdrup flow, $U$, and the vertically varying flow, so that at the equator and with $\tau^y = 0$ the deep flow is given by

$$u = \frac{-\tau^x + rU}{Hr}, \qquad v = V/H. \tag{22.31}$$

Plainly, if the magnitude of $U$ is sufficiently large then the zonal undercurrent $u$ will take the sign of $U$, rather than automatically opposing the direction of the wind stress. In this simple linear model, the deep flow is just the sum of two components, one proportional to and opposing the surface wind stress, and one in the direction of the Sverdrup flow. With the wind as it is today, the two effects reinforce each other and for that reason the undercurrent is significantly stronger than the surface flow, but this is not a general rule.

### 22.3.3 ✦ Effect of horizontal viscosity

In this section we will revert to the use of horizontal viscosity in place of a linear drag. As we noted, both horizontal viscosity and linear drag are somewhat ad hoc, so that one purpose of this exercise is to see what aspects of the solution are robust to choices of frictional parameterization.

*Formulating the problem*

As we see from Fig. 22.2, meridional variations tend to occur on a smaller scale than zonal variations so we'll neglect the zonal derivatives in the lateral friction. Our equations of motion then become

$$-fv = -\frac{\partial \phi}{\partial x} + \nu_z \frac{\partial^2 u}{\partial z^2} + \nu_h \frac{\partial^2 u}{\partial y^2}, \tag{22.32a}$$

$$fu = -\frac{\partial \phi}{\partial y} + \nu_z \frac{\partial^2 v}{\partial z^2} + \nu_h \frac{\partial^2 v}{\partial y^2}, \tag{22.32b}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{22.32c}$$

where $f = \beta y$ and with boundary conditions given by (22.14), as before. The vertically integrated horizontal flow, $(U, V)$, satisfies

$$-fV = -H\frac{\partial \phi}{\partial x} + \tau^x + \nu_h \frac{\partial^2 U}{\partial y^2}, \tag{22.33a}$$

$$fU = -H\frac{\partial \phi}{\partial y} + \tau^y + \nu_h \frac{\partial^2 V}{\partial y^2}, \tag{22.33b}$$

$$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = 0, \tag{22.33c}$$

and cross-differentiating leads to an equation similar to (22.15), namely

$$\nu_z \nabla^2 \frac{\partial^2 \psi}{\partial y^2} + \beta \frac{\partial \psi}{\partial x} = \mathrm{curl}_z \boldsymbol{\tau}. \tag{22.34}$$

Once again, in the absence of a wind-stress curl, the vertically integrated flow is zero and the wind stress is balanced by a pressure gradient. The flow relative to the vertical average, $(u', v')$, is given by subtracting (22.33) (divided by $H$) from (22.32) giving

$$-fv' = \nu_z \frac{\partial^2 u'}{\partial z^2} + \nu_h \frac{\partial^2 u'}{\partial y^2} - \frac{\tau^x}{H}, \tag{22.35a}$$

$$fu' = \nu_z \frac{\partial^2 v'}{\partial z^2} + \nu_h \frac{\partial^2 v'}{\partial y^2} - \frac{\tau^y}{H}, \tag{22.35b}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w}{\partial z} = 0. \tag{22.35c}$$

This is independent of the vertical average itself (as was the case with the linear drag) and henceforth, we'll take the vertical averaged flow to be zero and drop the prime on the velocity, with the understanding that it may be added back as needed. A full solution of (22.35) is both difficult to obtain and uninformative, so we will concentrate on various special cases, as follows.

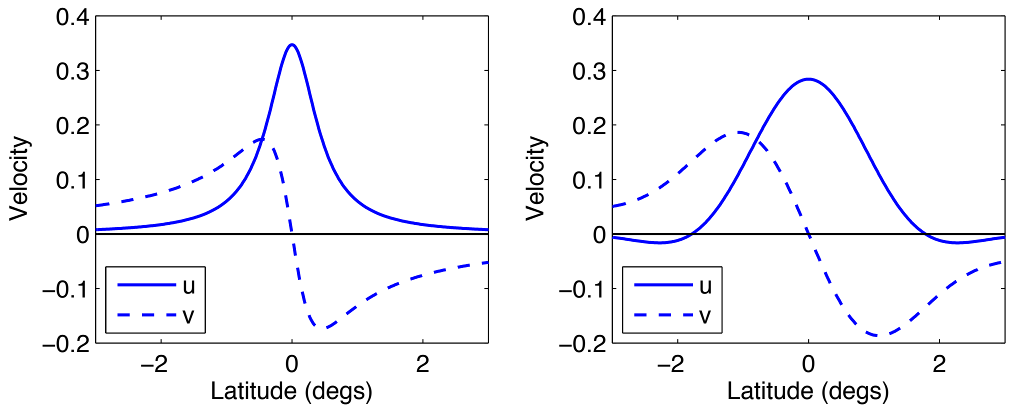


**Fig. 22.6** Horizontal profiles of the undercurrent with friction represented by a linear drag (left) and by a harmonic viscosity (right), nominally in dimensional units (metres/second and degrees).

*Solution away from the equator*

Away from the equator we neglect the horizontal friction terms and (22.35) becomes

$$-fv = \nu_z \frac{\partial^2 u}{\partial z^2} - \frac{\tau^x}{H}, \qquad fu = \nu_z \frac{\partial^2 v}{\partial z^2} - \frac{\tau^y}{H}, \tag{22.36a,b}$$

The particular solution to this is the depth independent flow,

$$v_p = \frac{\tau^x}{fH}, \qquad u_p = \frac{-\tau^y}{fH}. \tag{22.37a,b}$$

To this we must add the solution of the homogenous equation

$$\nu_z \frac{\partial^2 u}{\partial z^2} + fv = 0, \qquad \nu_z \frac{\partial^2 v}{\partial z^2} - fu = 0. \tag{22.38a,b}$$

These are the equations for an Ekman layer, first encountered in chapter 2 [c.f., (2.281)]. As there, the solution spirals down from the surface while decaying exponentially with an e-folding depth of $\sqrt{2\nu_z/f}$. The transport in the Ekman layer, $(\tau^y/f, -\tau^x/f$, is equal and opposite to the transport of the particular solution so that the total transport is zero.

[xx Check on Ekman transport. Also, possibly use $A$ instead of $\nu$ for consistency with chapter 2.]

*Solution below the Ekman layer*

When $f$ is small the lateral friction terms cannot be ignored and we are left with the full problem again. However, below the surface layer (which is the Ekman layer itself except very close to the equator) the vertical friction may be neglected and we can obtain a solution analogous to (22.19). The flow in this deep layer satisfies

$$-fv = \nu_h \frac{\partial^2 u}{\partial y^2} - \frac{\tau^x}{H}, \qquad fu = \nu_h \frac{\partial^2 v}{\partial y^2} - \frac{\tau^y}{H}, \tag{22.39}$$

where $f = \beta y$, and the reader will see that these equations are very similar to (22.18). It is now convenient to nondimensionalize and we do that by setting

$$(u, v) = (\widehat{u}, \widehat{v})\frac{\tau}{2\Omega H}, \quad y = \widehat{y}a, \quad (\tau^x, \tau^y) = (\widehat{\tau}^x, \widehat{\tau}^y)\tau, \quad \beta = \frac{2\Omega}{a} \tag{22.40}$$

where a hat denotes a nondimensional quantity and $a$ is the radius of Earth. The nondimensional versions of (22.39) are then

$$-\widehat{y}\widehat{v} = E_h\frac{\partial^2 \widehat{u}}{\partial \widehat{y}^2} - \widehat{\tau}^x, \qquad \widehat{y}\widehat{u} = E_h\frac{\partial^2 \widehat{v}}{\partial \widehat{y}^2} - \widehat{\tau}^y, \tag{22.41}$$

where $E_h = \nu_h/(2\Omega a^2)$ is a horizontal Ekman number.

The easiest way to obtain a solution is to multiply the second equation by i (i.e., $\sqrt{-1}$) and add to the first, to give

$$E_h\frac{\partial^2 Z}{\partial \widehat{y}^2} - \mathrm{i}\,\widehat{y}Z = T \tag{22.42}$$

where $Z \equiv \widehat{u} + \mathrm{i}\widehat{v}$ and $T = \widehat{\tau}^x + \mathrm{i}\widehat{\tau}^y$, which we henceforth take to be equal to $-1$ (i.e., a purely westward stress). Eq. (22.42) is a particular form of Airy's equation and its solution is given by[6]

$$Z(\widehat{y}) = \int_0^\infty \exp\left[-E_h\alpha^3/3 - \mathrm{i}\,y\alpha\right]\,\mathrm{d}\alpha \tag{22.43}$$

This solution asymptotes to the geostrophic balance $Z = 1/(\mathrm{i}\widehat{y})$ (i.e., $u = 0, v = \partial\phi/\partial x = \tau^x/fH$) for large $|\widehat{y}|$.

This solution, just like the one obtained using a linear drag, has total transport that is independent of the frictional coefficient; that is

$$\int_{-\infty}^{\infty}\widehat{u}\,\mathrm{d}\widehat{y} = \pi \qquad \text{or} \qquad H\int_{-\infty}^{\infty}u\,\mathrm{d}y = -\tau^x\frac{\pi a}{2\Omega}. \tag{22.44}$$

The mathematical derivation of this is left as a (tricky) exercise for the reader (problem 22.1). The integral is in fact *exactly* the same as that obtained using a linear drag, so that the quantitative underestimate of the magnitude of the undercurrent remains. The lack of dependance of the total transport on the viscosity arises because the width of the undercurrent increases with the (one third power of the) horizontal viscosity but the peak value diminishes with the (one third power of the) viscosity. The dependence on the one third power follows from a simple scaling of (22.42): at large $\widehat{y}$ the flow is geostrophic and lateral friction unimportant, whereas at small $\widehat{y}$ the lateral friction is required to remove the equatorial singularity. Thus, the nondimensional width of the undercurrent, $\widehat{L}$ say, is determined by the requirement that the terms on the left-hand side of (22.44) are both important and so that

$$\frac{E_h}{\widehat{L}^2} \sim \widehat{L}. \tag{22.45}$$

Dimensionally, this translates to

$$L \sim E_h^{1/3}a = \left(\frac{\nu_h a}{2\Omega}\right)^{1/3} \sim 100\,\mathrm{km}. \tag{22.46}$$

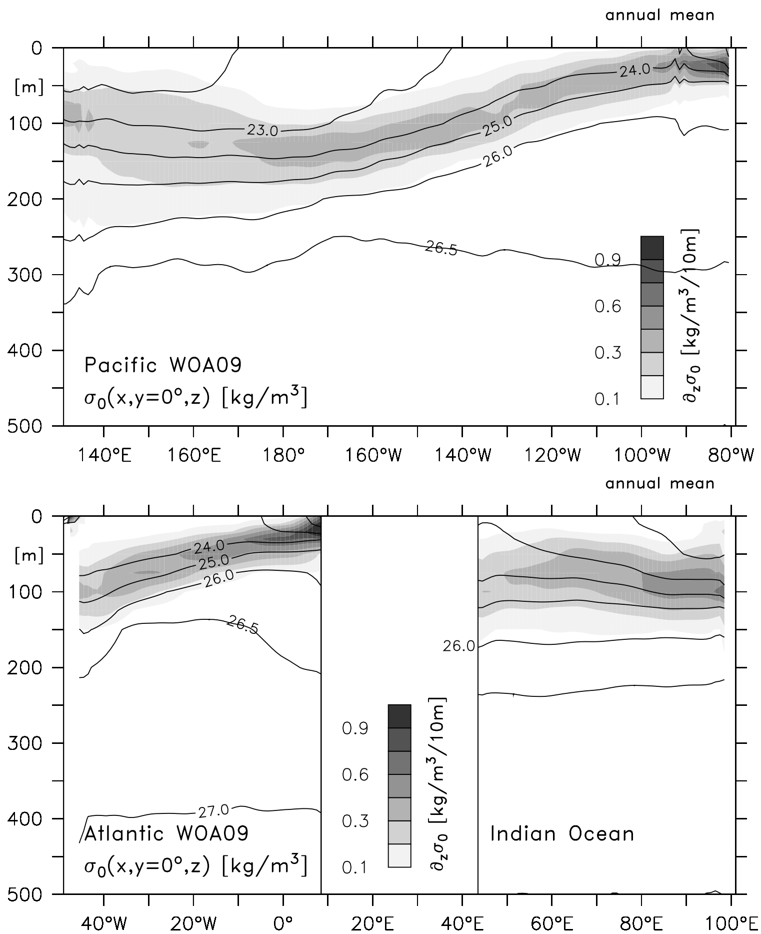


**Fig. 22.7** Zonal sections of annual mean density at the equator in the Pacific (top) and Atlantic and Indian oceans. The contours are of potential density (which is very nearly equal to density itself) and the shading is the vertical derivative of potential density, with data from the World Ocean Atlas.

if $E_h \approx 10^{-6}$ (which implies $\nu_h \approx 10^4\,\mathrm{m^2\,s^{-1}}$, but this value should not be regarded as fundamental).

Horizontal profiles of $u$ and $v$ obtained from (22.43) are plotted in the right-hand panel of Fig. 22.6, and may be compared with the corresponding solutions obtained with a linear drag. The results shown are obtained with $E_h = 2 \times 10^{-6}$ but otherwise the same values as were used with a linear drag, shown in the left-hand panel. Evidently, the results with the two frictional schemes display the same qualitative features, with a peak at the equator and a decay away, and a meridional velocity directed toward the equator in both hemispheres, which gives rise to equatorial upwelling.

## The Equatorial Undercurrent

*What is it?*

The equatorial undercurrent (EUC) is the single most striking feature of the low latitude ocean circulation. It is an eastward flowing subsurface current, mostly confined to depths between about 50 m and 250 m and to latitudes within 2° of the equator, with speeds of up to 1 m s$^{-1}$ (Fig. 22.1). It is sometimes connected to an eastward flowing current a few degrees north or south of the equator. The undercurrent is a permanent feature of the Atlantic and Pacific Oceans, but varies with season in the Indian Ocean because of the monsoon winds.

*What are its dynamics?*

Most models of the equatorial undercurrent tend to lie between two idealized end members that we refer to as the *local theory* and the *inertial theory*.[5]

- The local theory regards the undercurrent as a direct response to the westward winds at the equator. The winds push water westward and create a balancing eastward pressure gradient force. Below a frictional surface layer the influence of the wind stress is small and the pressure gradient leads to an eastward undercurrent.
  - In the frictional surface layer the flow is away from the equator and there is upwelling at the equator. The circulation is closed in the equatorial region. Continuous stratification may be included in the theory, although if there is upwelling through stratified water the diapycnal diffusivity must be non-zero.
  - The dynamics of the simplest models of this ilk are linear, but their quantification relies on the use of somewhat poorly constrained frictional and mixing coefficients.
- In the inertial theory, the equatorial current system is connected to the extra-equatorial region. A subsurface current moves inertially from higher latitudes, conserving its potential vorticity (which includes, crucially, a relative vorticity component) and Bernoulli function into the equatorial region. A pressure head is created in the western equatorial basin, which then pushes the undercurrent along.
  - Even if there were no wind at the equator the theory, in its simplest form, would still predict the presence of an undercurrent.
  - The theory, which is essentially nonlinear, contains parameters that must be specified somewhat arbitrarily, but the results are not especially sensitive to them.
- In reality, the undercurrent contains aspects of both theories, and more. Neither theory can be entirely correct. For example, the local theories do not properly take into account thermodynamic effects and, in contrast to the inertial theory, numerical experiments show that the undercurrent *does* depend on the wind at the equator.
  - Part of the EUC is closed within the equatorial region, and part connect to higher latitudes. A more complete model involves treating the EUC as one branch of a more complex tropical current system.
  - It would be hard, perhaps impossible, to construct a theory of the system that is elegant, complete and correct. But understanding can arise via careful treatments of special cases along with numerical and conceptual models of the areas in between.

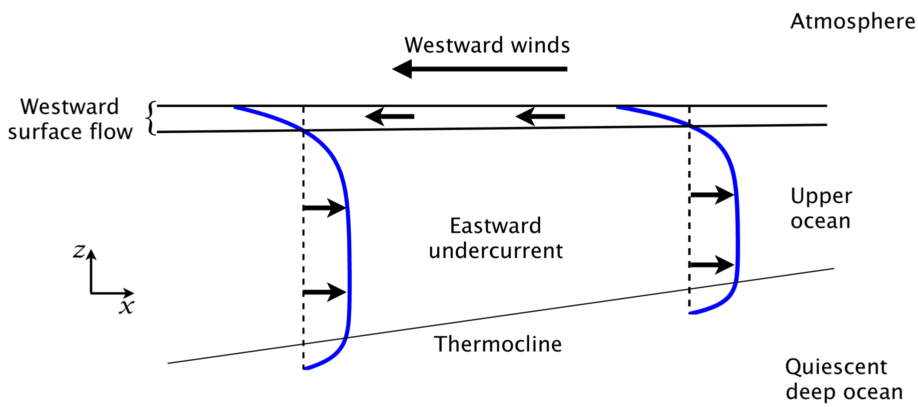


**Fig. 22.8** A schematic zonal section of the local model of the undercurrent and thermocline.

### 22.3.4 Effects of Stratification: A Layered Model of the Undercurrent

One unrealistic aspect of the models described above is that the undercurrent appears to extend all the way to the bottom of the ocean, whereas in reality it is confined to the upper few hundred meters of the ocean, with the deeper fluid being almost quiescent. A potential reason for this discrepancy is that we have neglected stratification, for this tends to limit vertical communication within an ocean column. Let's try to model this with a simple layered model, and for simplicity we revert to the use of linear drag.

Let us suppose that the ocean consists of two homogeneous layers. The continuous, homogeneous model described above describes the solution in the upper layer while the lower layer, of slightly greater density, represents the abyssal ocean and is assumed stationary. The pressure gradient must therefore be zero in the lower level, and we will see that this requires that the interface between the layers must slope, and indeed will usually slope upwards toward the east. The interface is, of course, a crude representation of the equatorial thermocline.

The zonal pressure gradient at the equator at the base of the upper layer is given by (22.16a), namely $\partial\phi/\partial x = (\tau^x - rU)/H$, and this is usually negative. If the upper layer has a density $\rho_1$ and the lower layer has a density $\rho_2$ then, in order for there to be no pressure gradient in the lower layer the interface must slope by an amount

$$s \equiv \frac{\partial z}{\partial x} = -\frac{1}{g'}\frac{\partial \phi}{\partial x} \tag{22.47}$$

where $g' \equiv g\Delta\rho/\rho_1 \equiv g(\rho_2 - \rho_1)/\rho_1$ is the reduced gravity and, as we recall, $\phi \equiv p/\rho_1$. Thus, an estimate of the slope of the thermocline is

$$s \approx \frac{1}{g'H}(\tau^x - rU). \tag{22.48}$$

The quantitative effects of the Sverdrup flow are hard to gauge because of the rather ill-constrained frictional coefficient $r$. The mean wind stress at the equator is westward and about $0.04\,\mathrm{N\,m^{-2}}$, and with $\Delta\rho = \rho_2 - \rho_1 = 3\,\mathrm{kg\,m^{-3}}$ and $H = 200$ m we find

$$s \approx \frac{\tau^{*x}}{g\Delta\rho H} \approx \frac{0.04}{10 \times 3 \times 200} = 6.7 \times 10^{-6}. \tag{22.49}$$

This suggests that over the 15,000 km extent of the equatorial Pacific we might expect the thermocline to shoal upwards toward the east by about 100 m. This slope is comparable to that observed (see Fig. 22.7), although considering the simplicity of the model the agreement is perhaps a little fortuitous. The thermocline slopes up toward the east in both the Atlantic and Pacific, where the prevailing winds are westward, but not in the Indian Ocean where the prevailing winds are seasonally variable because of the monsoons. The undercurrent itself is also a seasonal phenomenon in the Indian Ocean.

Except for the presence of the frictional coefficient, (22.48) is fairly insensitive to the details of the model; a virtually identical expression results if we model the ocean as two immiscible layers of fluid using the shallow water equations. The parameters determining the thermocline slope are just the thickness, $H$, of the upper layer and the density difference, $\Delta\rho$, between the upper and lower layers. A schematic of the flow is given in Fig. 22.8.

## 22.4 AN IDEAL FLUID MODEL OF THE EQUATORIAL UNDERCURRENT

The model of the equatorial undercurrent presented in the previous sections is physically appealing and describes aspects of the underlying dynamical mechanisms in a transparent way. Almost certainly the undercurrent is, at least in part, a consequence of an eastward subsurface pressure force originating from the westward winds in the tropics. However, the model has two potential shortcomings.

*(i)* The detailed results depend on the frictional parameters chosen.

*(ii)* The model makes no connection to the extra-tropical circulation of the ocean. That is, all the dynamics are essentially local.

The second shortcoming would be of no import if we could construct a well-founded model that did involve only purely equatorial dynamics, but the fact that frictional terms of unclear physical origin are of crucial importance suggests that the model may be incomplete. Furthermore, and perhaps more importantly, observations suggest that at least some of the water in the equatorial undercurrent has its origins in the subtropical gyre: temperatures in the core of undercurrent are mostly in the range of 16° C to 22° C, rather lower than the surface temperatures in the equatorial region except at the eastern end of the ocean basins — that is, at the end of the undercurrent. Furthermore, as we see from the upper two panels Fig. 22.1, the current gains strength as it moves eastward, implying that it is drawing water from higher latitudes as it moves.

More modern observational analyses of the equatorial ocean indeed suggest that the equatorial current system is a three-dimensional beast, connecting smoothly with the subtropical current system described in earlier chapters.[7] As subsurface water approaches the equator it largely rises along isopycnal surfaces as it moves eastward, with the cross-isopycnal velocity being only a small fraction of the total vertical velocity. This is in some contrast to the more local picture imagined in section 22.3 in which there is overturning in the vertical-meridional plane, and hence (to the extent that the water is stratified) with cross-isopycnal upwelling at the equator.

The above discussion suggests that it would be useful to construct a model that both connects to the subtropics and does not depend in any essential way or dissipative processes. That is, we should try to construct an ideal fluid model of the equatorial ocean. We'll do this in an way that is analogous to our treatment of the ventilated thermocline in chapter 21. That is, we'll represent the vertical structure of the ocean with a small number (one or two) of immiscible layers, and we'll assume that the subsurface layer conserves its potential vorticity.[8]

### 22.4.1 A simple barotropic model

Suppose that a fluid parcel at some latitude moves toward the equator, preserving its potential vorticity in a shallow-water system. If the fluid parcel originates from a latitude $y_0$ where, we suppose, its relative vorticity is negligible then, as it moves its vorticity, $\zeta$, is determined by

$$\frac{f+\zeta}{h} = \frac{f_0}{h_0} \tag{22.50}$$

where, on the equatorial beta plane, $f = \beta y$, $f_0 = f(y_0) = \beta y_0$ and $h_0$ is the depth of the fluid column at $y_0$. If, simplifying still further, the depth of the fluid column is assumed constant and meridional derivatives are much larger than zonal derivatives so that $\zeta = \partial v/\partial x - \partial u/\partial y \approx -\partial u/\partial y$, we have

$$\beta y - \frac{\partial u}{\partial y} = \beta y_0. \tag{22.51}$$

Integrating this expression, with $u = 0$ at $y = y_0$, gives

$$u = \frac{\beta}{2}(y - y_0)^2. \tag{22.52}$$

Interestingly, at $y = 0$, $u = \beta y_0^2/2$ which is positive. That is, conservation of absolute vorticity has, virtually by itself, produced an eastward flowing current at the equator (Fig. 22.9). Note that the solution is actually the same as the angular momentum conserving solution to the equinoctial Hadley Cell discussed in section 14.3, specifically equation (14.42) but with a different constant of integration: essentially, in the atmospheric case $y_0 = 0$, because the meridionally moving air in the upper branch of the Hadley Cell originates at the equator in the equinoctial case. However, the agreement is a little coincidental because in the oceanic case we do not expect angular momentum to be conserved because of the presence of a zonal pressure gradient, absent in the zonally averaged atmospheric case. Rather, it is absolute vorticity conservation, in its simplest form, that leads to (22.52).

However, from a quantitative standpoint the solution is not very satisfactory. It depends heavily on the value of $y_0$, and for $y_0$ greater than a few degrees the value of the zonal flow at the equator as predicted by the model is far too large, as can be inferred from Fig. 22.9. Also, the model eastward flow at the equator is not as jetlike as the undercurrent in the real ocean (Fig. 22.1). Nevertheless, the qualitative success suggests that it might be useful to proceed with a more complete model, in particular one in which the value of $h$ does vary with latitude, perhaps accounting for a good fraction of the variation of the potential vorticity.

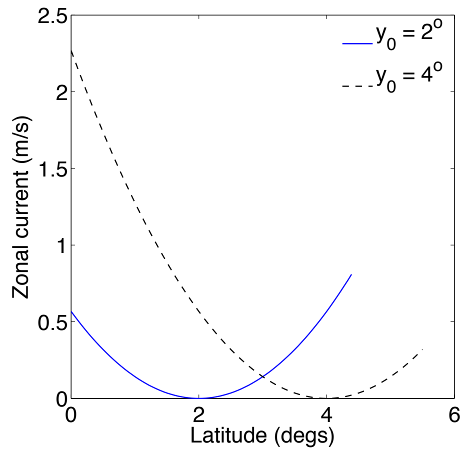


**Fig. 22.9** Zonal current as produced by the absolute vorticity conserving model. Specifically, solutions are plotted of (22.52) with $y_0 = 2°$ and $y_0 = 4°$ ($\times 2\pi a/360$) and $\beta = 2\Omega/a = 2.27 \times 10^{-11}\,\mathrm{m^{-1}\,s^{-1}}$.

### 22.4.2 A two-layer model of the inertial undercurrent

In this section we present an extension of the barotropic model above to two moving layers, with the flow in the lower level presuming to be conserve potential vorticity, and with the height field $h$ determined in a self-consistent fashion rather than being fixed. Thus, the features of the model are as follows.

*(i)* The use of the ideal form (i.e., inviscid, no dissipation) of the two-layer shallow water equations, with the lower level shielded from the wind's influence and conserving potential vorticity and giving rise to the equatorial undercurrent.

*(ii)* At low latitudes the equations are solved in a boundary-layer approximation, with variations in $y$ being much smaller than variations in $x$. Unlike our treatment of the ventilated thermocline in midlatitudes, no assumption is made that the flow satisfies the planetary geostrophic equations. It is the inertial terms that prevent the solution from becoming singular at the equator.

*(iii)* At higher latitudes the solutions are constructed to blend in with the solution of a midlatitude ventilated thermocline model, described in section 21.4. Put another way, the ventilated thermocline provides a high-latitude boundary condition for the model.

See Fig. **??** for a schematic

*Equations of motion*

Our primary concern will be the lower layer (layer 2) for which the momentum and mass continuity equations are, respectively,

$$\frac{\mathrm{D}\boldsymbol{u}_2}{\mathrm{D}t} + \boldsymbol{f} \times \boldsymbol{u}_2 = -\frac{1}{\rho_0}\nabla p_2 = -g_2'\nabla h \tag{22.53a}$$

$$\frac{\partial h_2}{\partial t} + \nabla \cdot (h_2 \boldsymbol{u}_2) = 0, \tag{22.53b}$$

where $h_2$ is the thickness of the layer and $\boldsymbol{u}_2$ the horizontal velocity within it, and $g_2' = g(\rho_3 - \rho_2)/\rho_0$. We remain on the equatorial beta plane so that $\boldsymbol{f} = f\,\mathbf{k} = \beta y\,\mathbf{k}$, where $\mathbf{k}$ is the unit vector in the vertical direction, and we will consider only the steady versions of these equations. We may also write the momentum equation in terms of the Bernoulli function,

$$\frac{\partial u_2}{\partial t} + (f + \zeta_2)v_2 = -\frac{\partial B_2}{\partial x}, \qquad \frac{\partial v_2}{\partial t} + (f + \zeta_2)u_2 = -\frac{\partial B_2}{\partial y} \tag{22.54}$$

where $B_2 = g_2' h + \boldsymbol{u}_2^2/2$.

The above equations conserve potential vorticity, $Q_2 = (f + \zeta)/h_2$ and, because the flow is presumed steady, the Bernoulli function. That is,

$$\boldsymbol{u}_2 \cdot \nabla Q_2 = 0, \qquad \boldsymbol{u}_2 \cdot \nabla B_2 = 0. \tag{22.55a,b}$$

Also, because of the form of the mass continuity equation in the steady state, namely $\nabla \cdot (h_2 \boldsymbol{u}_2) = 0$, we can define a streamfunction, $\psi$, such that

$$h_2 \boldsymbol{u}_2 = \mathbf{k} \times \nabla \psi \qquad \text{or} \qquad h u_2 = -\frac{\partial \psi}{\partial y}, \quad h v_2 = \frac{\partial \psi}{\partial x} \tag{22.56}$$

Using the streamfunction, conservation of potential vorticity and Bernoulli function may be written as

$$J(\psi, Q_2) = 0, \qquad J(\psi, B_2) = 0, \tag{22.57a,b}$$

where $J(a, b) = \partial_x a \, \partial_y b - \partial_y a \, \partial_x b$. Equations (22.57a) and (22.57b) imply, respectively that isolines of $Q_2$ and $\psi$, as well as isolines of $B_2$ and $\psi$, are everywhere parallel to each other. Thus, in general, $Q_2$ is a function of $B$; that is,

$$Q_2 = F(B_2), \tag{22.58}$$

where the function, $F$, is as yet unknown. It is also the case that $Q_2$ is a function of $\psi$; that is, $Q_2 = G(\psi)$ where $G$ is some other function. However, it is not the case that, in general, $Q_2$ is a function of the height field $h$, because $h$ is not proportional to the streamfunction for the flow. This is in contrast to the midlatitude case in which geostrophic balance may be written as $\boldsymbol{u}_2 = (g_2'/f)\mathbf{k} \times \nabla h$, and so the relation $\boldsymbol{u}_2 \cdot \nabla Q_2$ implied that $Q_2$ is a function of $h$ itself. We do not assume geostrophic balance in the equatorial region.

The equations and the properties of the equations so far discussed are quite general (save for the restriction to the beta plane). Let us now consider the equatorial region, and then how it connects to the subtropics.

### *Equatorial dynamics*

Let us now consider the dynamics close to the equator, in the region of the undercurrent. We first derive some elementary scaling relations between the variables.

Consider motion within a narrow strip of distance no more than $L_y$ from the equator where $L_y$ is the characteristic meridional scale of the undercurrent, as yet undetermined. If $L_x$ is the characteristic zonal scale, typically the scale of the ocean basin itself, then $L_y \ll L_x$. We expect that $L_y$ will be the scale over which the relative vorticity becomes comparable to the planetary vorticity, or equivalently the scale such that the beta Rossby number is $\mathcal{O}(1)$. If the scale of the zonal velocity is $U$ then this requirement is $U/(\beta L_y^2) = 1$ or

$$L_y = \left(\frac{U}{\beta}\right)^{1/2} \qquad \text{or} \qquad U = \beta L_y^2. \tag{22.59}$$

The disparity between zonal and meridional scales implies that there will also be a disparity between the zonal and meridional velocities, and in particular from the mass continuity equation we expect that

$$V = \frac{UL_y}{L_x}, \tag{22.60}$$

and so $V \ll U$, where $V$ is the scale of the meridional velocity.

At a (non-zero) distance $L_y$ from the equator the relevant Rossby number in the meridional momentum equation is given by $U/(\beta L_x^2)$, and this remains small. Thus, essentially because $U$ is so much larger than $V$, even very close to the equator the *zonal* flow will be in near geostrophic balance with the meridional pressure gradient. The meridional momentum equation then becomes

$$\beta y u_2 = -g_2' \frac{\partial h}{\partial y}, \tag{22.61}$$

implying the scaling

$$H = \frac{\beta L_y^2 U}{g_2'} = \frac{\beta^2 L_y^4}{g_2'}, \tag{22.62}$$

using (22.59), where $H$ is the scale of the variation of thickness in layer 2.

Now let's consider the equations themselves. As we noted the flow conserves potential vorticity, $Q_2$. Close to the equator $Q_2 \approx (f - \partial u_2/\partial y)/h_2$ so that, using (22.58),

$$\frac{\beta y - \partial u_2/\partial y}{h_2} = F(B_2), \tag{22.63}$$

where $B_2 = g_2' h + u_2^2/2$, noting that $|u_2| \gg |v_2|$ in the equatorial region. There is an obvious similarity between (22.63) and (22.51). Note also that (22.63) and (22.61) are ordinary differential equations, although of course $u_2$ and $h$ do vary in $x$. If we knew the function $F(B_2)$, and we knew the upper layer thickness $h_1$, then the equations would be closed and we could find a solution. For this we turn to the dynamics in the subtropics.

*Extra-equatorial dynamics*

[Need to say why all the equator is not just in the shadow zone.]

The role of the extra-equatorial region in out treatment is to provide a boundary condition for the equatorial dynamics, and to determine the functional relationship between potential vorticity and the Bernoulli function, $F(B_2)$. We will suppose that the fluid obeys the dynamics of the two-layer model of the ventilated thermocline discussed in section 21.4, in which the fluid obey the planetary geostrophic equations. The total depth of the moving fluid, $H$, is given by

$$h^2 = -z_2 = \frac{D_0^2}{[1 + (g_1'/g_2')(1 - f/f_2)^2]}, \tag{22.64}$$

where

$$D_0^2(x, y) = -\frac{2f^2}{\beta g_2'} \int_x^{x_e} w_E(x', y)\,\mathrm{d}x', \tag{22.65}$$

with $w_E = \mathrm{curl}_z(\boldsymbol{\tau}/f)$ being the vertical velocity at the base of the Ekman layer. Assuming the stress is zonal then, at low latitudes, $w_E \approx \beta\tau^x/f^2 = \tau^x/(\beta y^2)$. If the stress is also independent of longitude then we find

$$D_0^2 = \frac{-2(x_e - x)\tau^x}{g_2'} \tag{22.66}$$

so that

$$h^2 = \frac{-2(x_e - x)\tau^x}{g_2'\left[1 + (g_1'/g_2')(1 - f/f_2)^2\right]}. \tag{22.67}$$

The extra-equatorial solution is completed by noting the expressions of the depths of each layer as a function of the total depth, as in (21.57)

$$h_2 = z_1 - z_2 = \frac{f}{f_2}h \qquad \text{and} \qquad h_1 = -z_1 = \left(1 - \frac{f}{f_2}\right)h. \tag{22.68a,b}$$

*Connections*

We now start to connect the extra-equatorial solution to the tropical one. We first note that (22.67) provides a scaling relation for $h$, namely

$$H = \left(\frac{L_x\tau}{g_2'}\right)^{1/2}. \tag{22.69}$$

Now, we are assuming that the equatorial dynamics transition smoothly to the extra-equatorial solution, so that (22.69) must be consistent with (22.62). Taken together they give us estimates for the meridional scale, the zonal velocity and the depth of the moving fluid purely in terms of external parameters, to wit:

$$\boxed{L_y = \left(\frac{L_x\tau g_2'}{\beta^4}\right)^{1/8}, \qquad H = \left(\frac{L_x\tau}{g_2'}\right)^{1/2}, \qquad U = (g_2'\tau L_x)^{1/4}}. \tag{22.70a,b,c}$$

These scalings are important results of the model, just as much as the precise form of the solution discussed below. Note that the scaling for zonal velocity is qualitatively different from that derived earlier using the frictional model — compare (22.70c) with (22.19a) or (22.25), for example. The dependance of $U$ on the wind stress in (22.70c) is perhaps surprisingly weak, although both the layer thickness and the horizontal scale also increase with the wind so that the total transport increases almost linearly with wind stress. To the extent that the thickness of the upper layer stays constant, the transport of the lower layer scales as

$$HU = \left(\frac{L_x^3\tau^3}{g_2'}\right)^{1/4} \qquad \text{and} \qquad HUL = \left(\frac{L_x^7\tau^7}{g_2'\beta^4}\right)^{1/8}. \tag{22.71}$$

We now obtain the functional connection between $Q_2$ and $B_2$ necessary to close (22.63). In the extra-equatorial region, the horizontal shear becomes small compared to the Coriolis term so that (22.63) becomes

$$Q_2 = \frac{\beta y}{h_2} = F(B_2), \tag{22.72}$$

and the Bernoulli function itself, $B_2 = g_2'h + \boldsymbol{u}^2/2$, may be approximated by $B_2 = g_2'h$. Therefore, at the edge of the equatorial region,

$$Q_2 = \frac{f}{h_2} = \frac{f_2}{h}, \tag{22.73}$$

using (22.68a). This functional form holds throughout the equatorial region, and therefore $Q_2(h) = f_2/h$. More generally, $Q_2(\phi) = f_2/\phi$ for any variable $\phi$ and in particular,

$$Q_2(B_2) = \frac{f_2}{B_2} = \frac{f_2}{g_2'h + u_2^2/2} \tag{22.74}$$

Our quest for the solution is now all over bar the shouting, in the sense that we can write down the equations of motion and the boundary conditions. Using geostrophic balance, (22.74) and(22.63) we write down the equations determining the subsurface flow in the equatorial region, namely

$$\frac{\beta y - \partial u_2/\partial y}{h_2} = \frac{f_2}{g_2'h + u_2^2/2}, \tag{22.75}$$

$$\beta y u_2 = -g_2'\frac{\partial h}{\partial y} \tag{22.76}$$

$$h_2 = h - h_1 \tag{22.77}$$

In addition to specifying the value of the upper layer field, $h_1$, we need to specify the boundary conditions and the value of field, $h_1$. At large values of $y$ (i.e., $y/L_y \gg 1$) the value of $h$ should be that given by (22.67). A second boundary condition may be applied at the equator, and if we suppose that the flow is hemispherically symmetric we have

$$v_2 = 0 \qquad \text{at } \ y = 0. \tag{22.78}$$

Taken with (22.54) this equation implies that, for steady flow, $\partial B_2/\partial x = 0$ and so that

$$B_2 = g_2'h + \frac{u_2^2}{2} = B_0 \qquad \text{at } \ y = 0, \tag{22.79}$$

where $B_0$ is a constant. That is to say, the equator is a streamline of the flow. (If there is flow across the equator the problem becomes more complicated, but we leave that for another day.) The value of $B_0$ is plausibly given by supposing it to be the value of $B_2$ at the western edge of the basin just outside the equatorial region, as illustrated in Fig. **??**, but other choices might be made. Finally, we need to specify the field $h_1$, and there are a number of reasonable ways to proceed, although no obviously correct one. One choice would be to suppose that, just as in the extra-tropics, the total thickness of the moving layers is given by Sverdrup balance. If we were to do this we would essentially be extending the ventilated thermocline model all the way to the equator, with the addition of inertial terms. Although Sverdrup balance is qualitatively reasonable in equatorial regions (Fig. 22.2), quantitatively it is not particularly good and a simpler recipe is appropriate. One option is to choose $h_1$ to be a function of $x$ only, such that the value of $h_1$ is equal to the value that it has at the high latitude edge of the equatorial region, at $y = y_n \gg L_y$. Using (22.68b) and (22.64) this gives

$$h_1^2 = \frac{D_0^2(1 - y_n/y_2)^2}{\left[1 + (g_1'/g_2')(1 - y_n/y_2)^2\right]} = \frac{-2(x_e - x)\tau^x(1 - y_n/y_2)^2}{g_2'\left[1 + (g_1'/g_2')(1 - y_n/y_2)^2\right]} \tag{22.80}$$

using (22.66). The choice is simple albeit a little ad hoc, but it turns out that the solution is not especially sensitive to it. That it is a reasonable choice can be seen by noting that for $y_n \ll y_2$,

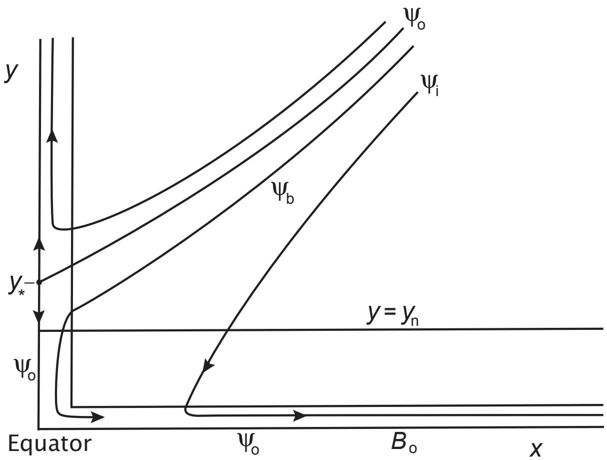


**Fig. 22.10** Schematic of the flow streamlines leading to an equatorial undercurrent. [Add another schematic as well]

$h \to h_1$ so that

$$h \to h_1 = \left[ \frac{-2(x_e - x)\tau^x}{g_2' \left[1 + (g_1'/g_2')\right]} \right]^{1/2}. \tag{22.81}$$

Re-arranging and differentiating this expression with respect to $x$ we obtain

$$\frac{\partial}{\partial x}\left(1 + \frac{g_1'}{g_2'}\right) h_1 = \frac{\tau^x}{h_1}. \tag{22.82}$$

That is, there is a balance between the applied wind stress and the pressure gradient force in the upper layer. [This needs checking xxx]

*The solution and its properties*

The equations of motion and the boundary conditions for this model are summarized in the shaded box on the next page. They are nonlinear and rather complex, and solutions must in general be obtained numerically by an iterative method, and example solutions given later (in Fig. 22.11 and Fig. 22.12). However, some qualitative properties may be deduced from the form of the equations. For clarity we'll use Northern Hemisphere terminology, but the ideas apply equally to the Southern Hemisphere.

The ventilated thermocline gives rise to fluid that, at low latitudes, flows southward and westward. As it flows equatorward, potential vorticity conservation must lead to, in the absence of changes in layer thickness, an increase in relative vorticity — an anticlockwise or cyclonic turning — and the flow veers more southward and then eastward, so giving rise to the equatorial undercurrent. This property is of course present in the barotropic model of absolute vorticity conservation discussed in section 22.4.1. However, the two-layer model differs from the barotropic model in two important regards. First, the layer thickness is allowed to change in a self-consistent fashion. Second, the flow is not particularly sensitive to the matching latitude at which we connect the equatorial equations of motion, (22.63) or (U.1), to the extra-equatorial region using (**??**) or (U.2). This is because the ventilated thermocline model is itself based on conservation of potential vorticity, so that changing the matching latitude will have little effect on the potential vorticity entering the equatorial region. The two layer model

## Equations of Motion for the Inertial Undercurrent

The dimensional form of the equations of motion are

$$\frac{\partial u_2}{\partial y} - \beta y = \frac{f_2 h_2}{h + u_2^2/2}, \qquad (h_2 = h - h_1) \tag{U.1a}$$

$$\beta y u_2 = -g_2' \frac{\partial h}{\partial y}, \tag{U.1b}$$

where $h_1$ is a specified function of $x$. One plausible choice is given by choosing it to be the value of $h_1$ just outside the equatorial region, as given by (22.80).

Two boundary conditions are needed. The first is that in the extra-equatorial region $h$ approaches the value given by the ventilated thermocline model, for example

$$h^2 = \frac{-2(x_e - x)\tau^x}{g_2'\left[1 + (g_1'/g_2')(1 - f/f_2)^2\right]}. \tag{U.2}$$

At the equator the boundary condition on $h$ is obtained by setting $v_2 = 0$ and specifying the Bernoulli function, $B_2$ there. That is, we specify

$$B_2 = g_2' h + \frac{1}{2} u_2^2 = B_0 \qquad \text{at } y = 0. \tag{U.3}$$

Here, $B_0$ is a constant, chosen to be equal to the value of the Bernoulli function on the western edge of the basin just outside the equatorial region (that is, using (U.2) at $x = 0$ and $y = y_n$). There is then a pressure head at the western edge of the equatorial region, and the flow accelerates zonally along the equator preserving its Bernoulli function.

The nondimensional form of (U.1) are

$$\frac{\partial \widehat{u}_2}{\partial \widehat{y}} - \widehat{y} = \frac{-y_2 \widehat{h}_2}{\widehat{h} + \widehat{u}_2^2/2}, \qquad (\widehat{h}_2 = \widehat{h} - \widehat{h}_1), \tag{U.4a}$$

$$y \widehat{u}_2 = -\frac{\partial \widehat{h}}{\partial \widehat{y}}. \tag{U.4b}$$

where $y_2 = f_2/(\beta L_y)$.

does have some parameters that cannot be deduced a priori, in particular the thickness of the top layer and the value of the Bernoulli parameter, but the solutions are not especially sensitive to it. That is, and in common with some other models in geophysical fluid dynamics (for example, the Stommel model of western intensification in a gyre), the behaviour of the solutions is quite robust and transcends the detailed limitations of the model itself.

A numerically obtained solution is illustrated in Fig. 22.11. We see the streamlines sweeping westward and equatorward before taking a sharp equatorward and then eastward turn, with the flow being purely eastward at the equator. The solutions of $u_2$ and $h$ are shown in Fig. 22.12,

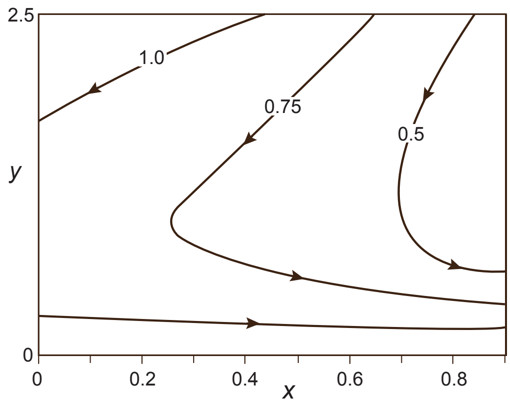


**Fig. 22.11** The streamlines in a solution of the equations for an inertial equatorial undercurrent. The parameters used were $G = 1$, $y_1 = 5$, $y_N = 2.5$. [More info on legend]]

illustrating the formation of the undercurrent and its intensification as it moves eastward at the equator. Note also the latitudinal variation of the layer depth, $h_2$ ($h_2 = h - h_1$). In the extra-tropical ventilated thermocline, the thickness of layer 2 diminishes as we move equatorward, and if the limit of $y \to 0$ were taken the thickness of layer 2 would goes to zero (a consequence of $f$ going to zero and $f/h$ being preserved). However, in the equatorial boundary layer the layer thickness actually increases as the equator is approached, and so the relative vorticity must increase to compensate, because now $(f+\zeta)/h$ is preserved. That is, the cyclonic intensification of the flow is somewhat more pronounced than in the barotropic model. Note also that the layer depth diminishes eastward; that is, the thermocline slopes up toward the east.

## 22.5 RELATION OF INERTIAL AND FRICTIONAL UNDERCURRENTS

In the previous two sections we presented two quite different conceptual models of the equatorial thermocline. The first one is frictional and local, the second one is inertial and remote. In the local model, the westward winds set up a compensating pressure gradient, and below the frictional layer near the surface the pressure gradient dominates leading to an eastward flow. The only cross-latitudinal effects come from the a lateral friction, and if this is replaced with a linear drag the zonal flow at the equator is wholly independent of the dynamics at other latitudes. In contrast, in the inertial model the undercurrent arises as a consequence of potential vorticity conservation of the subsurface flow, with the value of the potential vorticity set in the extra-equatorial region. The undercurrent is fed by extra-equatorial waters at all longitudes and so builds up as it moves eastward (as is observed). There is a 'pressure head' at the western edge of the equatorial basin, so that the flow accelerates eastward *without the need for any winds at all at the equator.* It is the link with the geostrophically balanced motion in the extra-equatorial region that determines the structure of the equatorial undercurrent, not the local winds.

Are these two views of the equatorial undercurrent in complete opposition, to the extent that only one can be true? In fact, the real ocean may have elements of both [more here. Some description of McCreary and Lu? xxx].

## 22.6 AN INTRODUCTION TO EL NIÑO

We talk only about the oceanic aspects of El Niño.

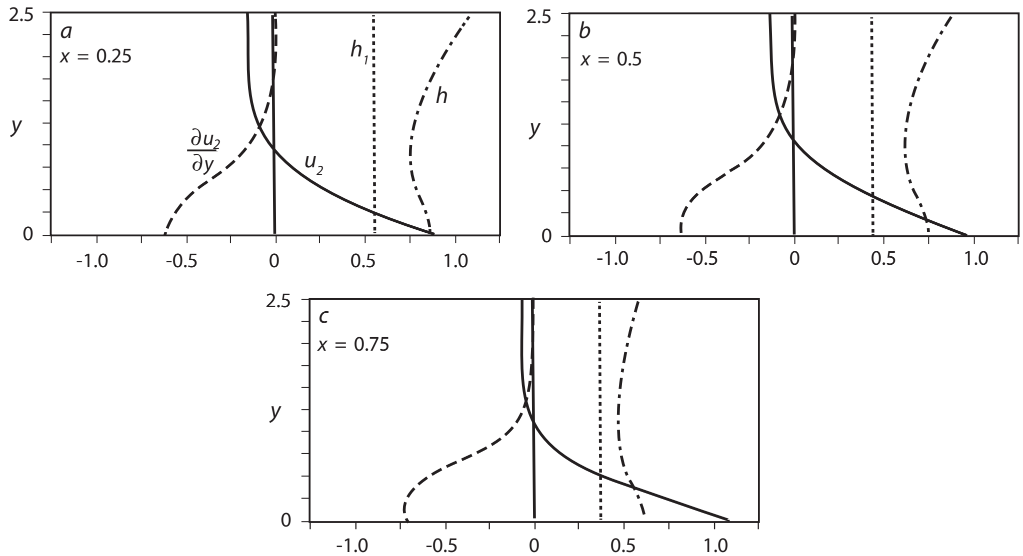


**Fig. 22.12** Numerically obtained solutions for $u_2, \partial u_2/\partial y, h_1$ and $h$ (as labelled in panel a) for the nondimensional equatorial undercurrent equations, for $x = 0.25,\ 0.5$ and $0.75$. The wind stress is constant, $G_{12} = 1$, $y_2 = 5$ and $B_0 = 1.26$.

## Notes

2 Figure kindly made by Neven Fučkar, using a state estimation (i.e., a combination of models and data) from NCEP/GODAS, from http://www.esrl.noaa.gov/psd/data/gridded/data.godas.html. The World Ocean Atlas, another source of observations, can be obtained from http://www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html.

4 Adapted from Kessler *et al.* (2003).

5 The undercurrent itself seems to have been first discovered in the Atlantic by J. Y. Buchanan in the 1880s. He measured a southeastward flowing current with speeds of more than 1 knot (about $0.5\,\mathrm{m\,s^{-1}}$) at depths around 30 fathoms (55 metres) at the equator and 13° W from the steamship *Buccaneer,* which was charted to do a survey prior to the laying of a telegraph cable (Buchanan 1886). The discovery of the undercurrent in the Pacific is sometimes credited to Townsend Cromwell (1922–1958) in the early 1950s, and there the current is often called the Cromwell Current. Cromwell also provided the first credible theoretical model of the undercurrent, as noted below. He tragically died in 1958 in plane crash while en route to an oceanography expedition.

Theories of the equatorial undercurrent tend to fall into or between two camps, which we may call 'local' theories and 'inertial theories'. The local theories began with a description by Cromwell (1953) of the currents produced by a westward wind at the equator and were extended and put into mathematical form by Stommel (1960) with thermal effects added by Veronis (1960) and with a later variation by Robinson (1966). This class of model, which is essentially linear and unavoidably dissipative, was further developed and clarified by Gill (1971), McKee (1973) and Gill (1975) and we mostly follow their treatment. The effects of nonlinearities were looked at first by Charney (1960) and then by McKee (1973) and Cane (1979a,b). The linear model was significantly extended by McCreary (1981) to include the effects of continuous stratification, but a useful discussion of that is beyond our scope.

A different class of model was proposed by Pedlosky (1987b), building on an idea of Fofonoff & Montgomery (1955). In this view the undercurrent is inertial and may be thought of as being pushed by a pressure head that begins in extra-equatorial regions. We describe this model in section 22.4. This viewpoint was significantly extended and, in part, reconciled with the local viewpoint by McCreary & Lu (1994) who considered the equatorial undercurrent as part of a larger and more complex subtropical current system, with both local and inertial aspects. A complete description of these dynamics is perforce in large part numerical, and beyond our scope.

6 The canonical Airy equation is $\partial^2 y/\partial x^2 - xy = 0$. The solution, the Airy function, is discussed in many books on ordinary differential equations and special functions (e.g., Jeffreys & Jeffreys 1946, and Ambramowitz & Stegun 1965) and, perhaps of more relevance to the modern reader, in mathematical software such as Maple. The form of solution we use was presented by McKee (1973). An equivalent form is $Z = u + \mathrm{i}v = \pi C\left[Ai(Cy) - \mathrm{i}\, Gi(Cy)\right]$, where $Ai$ is the standard Airy function, $Gi$ is a particular form of the Airy function introduced by Scorer (1951) and $C = \mathrm{e}^{\mathrm{i}\pi/6} E_h^{-1/2}$.

7 One of the first observational analyses to unambiguously link the equatorial ocean to higher latitudes was Bryden & Brady (1985). More recently, oceanic observations are combined with ocean models to produce state estimates, analogous to the atmospheric re-analyses, that produce more accurate maps of the ocean state than can be produced using observations or models separately.

8 See note 5 above for references.

*Problems*

22.1 Derive or verify the result given in (22.44).

# References


Abarbanel, H. D. I. & Young, W. R., Eds., 1987. *General Circulation of the Ocean*. Springer-Verlag, 291 pp.

Allen, J. S., 1993. Iterated geostrophic intermediate models. *J. Phys. Oceanogr.*, **23**, 2447–2461.

Allen, J. S., Barth, J. A. & Newberger, P. A., 1990a. On intermediate models for barotropic continental shelf and slope flow fields. Part I: formulation and comparison of exact solutions. *J. Phys. Oceanogr.*, **20**, 1017–1042.

Allen, J. S., Barth, J. A. & Newberger, P. A., 1990b. On intermediate models for barotropic continental shelf and slope flow fields. Part II: comparison of numerical model solutions in doubly periodic domains. *J. Phys. Oceanogr.*, **20**, 1043–1082.

Allen, J. S., Holm, D. D. & Newberger, P. A., 2002. Extended-geostrophic Euler–Poincaré models for mesoscale oceanographic flow. In J. Norbury & I. Roulstone, Eds., *Large-scale Atmosphere–Ocean Dynamics I*. Cambridge University Press.

Ambramowitz, M. & Stegun, I. A., 1965. *Handbook of Mathematical Functions*. Dover Publications, 1046 pp.

Andrews, D. G., 1987. On the interpretation of the Eliassen–Palm flux divergence. *Quart. J. Roy. Meteor. Soc.*, **113**, 323–338.

Andrews, D. G., Holton, J. R. & Leovy, C. B., 1987. *Middle Atmosphere Dynamics*. Academic Press, 489 pp.

Angell, J. K. & Korshover, J., 1964. Quasi-biennial variations in temperature, total ozone, and tropopause height. *J. Atmos. Sci.*, **21**, 479–492.

Arbic, B., Flierl, G. & Scott, R., 2005. Cascade inequalities for forced-dissipated geostrophic turbulence. (In review).

Aris, R., 1962. *Vectors, Tensors and the Basic Equations of Fluid Mechanics*. Dover, 286 pp. First published by Prentice-Hall.

Assmann, R., 1902. Über die Existenz eines wärmeren Luftstromes in der Höhe von 10 bis 15 km. (On the existence of a warmer airflow at heights from 10 to 15 km). *Sitzber. Königl. Preuss. Akad. Wiss. Berlin*, **24**, 495–504.

Aubin, D. & Dahan Dalmedico, A., 2002. Writing the history of dynamical systems and chaos: *longue durée* and revolution, disciplines and cultures. *Historia Mathematica*, **29**, 1–67.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K. *et al.*, 2001. The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179–229.

Ball, J. M. & James, R. D., 2002. The scientific life and influence of Clifford Ambrose Truesdell III. *Arch. Rational Mech. Anal.*, **161**, 1–26.

Bannon, P. R., 1995. Potential vorticity conservation, hydrostatic adjustment, and the anelastic approximation. *J. Atmos. Sci.*, **52**, 2301–2312.

Bannon, P. R., 1996. On the anelastic equation for a compressible atmosphere. *J. Atmos. Sci.*, **53**, 3618–3628.

Barnier, B., Hua, B. L. & Provost, C. L., 1991. On the catalytic role of high baroclinic modes in eddy-driven large-scale circulations. *J. Phys. Oceanogr.*, **21**, 976–997.

Bartello, P. & Warn, T., 1996. Self-similarity of decaying two-dimensional turbulence. *J. Fluid Mech.*, **326**, 357–372.

Batchelor, G. K., 1953a. The conditions for dynamical similarity of motions of a frictionless perfect-gas atmosphere. *Quart. J. Roy. Meteor. Soc.*, **79**, 224–235.

Batchelor, G. K., 1953b. *The Theory of Homogeneous Turbulence*. Cambridge University Press, 197 pp.

Batchelor, G. K., 1959. Small-scale variation of convected quantities like temperature in turbulent fluid. Part 1: general discussion and the case of small conductivity. *J. Fluid Mech.*, **5**, 113–133.

Batchelor, G. K., 1967. *An Introduction to Fluid Dynamics*. Cambridge University Press, 615 pp.

Batchelor, G. K., 1969. Computation of the energy spectrum in homogeneous two-dimensional turbulence. *Phys. Fluids Suppl.*, **12**, II–233—II–239.

Batchelor, G. K., Howells, I. D. & Townsend, A. A., 1959. Small-scale variation of convected quantities like temperature in turbulent fluid. Part 2: the case of large conductivity. *J. Fluid Mech.*, **5**, 134–139.

Batchelor, G. K. & Townsend, A. A., 1956. Turbulent diffusion. In G. K. Batchelor & R. M. Davies, Eds., *Surveys in Mechanics*, pp. 352–399. Cambridge University Press.

Baumert, H. Z., Simpson, J. & Sündermann, J., Eds., 2005. *Marine Turbulence: Theories, Observations and Models*. Cambridge University Press, 630 pp.

Bender, C. M. & Orszag, S. A., 1978. *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, 593 pp.

Berrisford, P., Marshall, J. C. & White, A. A., 1993. Quasi-geostrophic potential vorticity in isentropic co-ordinates. *Quart. J. Roy. Meteor. Soc.*, **50**, 778–782.

Birner, T., 2006. Fine-scale structure of the extratropical tropopause region. *J. Geophys. Res.*, **111**, D04104. doi:10.1029/2005JD006301.

Birner, T., Dörnbrack, A. & Schumann, U., 2002. How sharp is the tropopause at midlatitudes? *Geophys. Res. Lett.*, **29**, 1700. doi:10.1029/2002GL015142.

Bjerknes, J., 1919. On the structure of moving cyclones. *Geofys. Publ*, **1** (2), 1–8.

Bjerknes, J., 1937. Die Theorie der aussertropischen Zyklonenbildung (The theory of extratropical cyclone formation). *Meteor. Zeitschr*, **12**, 460–466.

Bjerknes, J., 1959. Atlantic air–sea interaction. In *Advances in Geophysics*, Vol. 10, pp. 1–82. Academic Press.

Bjerknes, J., 1969. Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172.

Bjerknes, V., 1898a. Über die Bildung von Cirkulationsbewegungen und Wirbeln in reibunglosen Flüssigkeiten (On the generation of circulation and vortices in inviscid fluids). *Skr. Nor. Vidensk.-Akad. 1: Mat.-Naturvidensk. Kl.*, **5**, 3–29.

Bjerknes, V., 1898b. Über einen hydrodynamischen Fundamentalsatz und seine Anwendung besonders auf die Mechanik der Atmosphäre und des Weltmeeres (On a fundamental principle of hydrodynamics and its application particularly to the mechanics of the atmosphere and the world's oceans). *Kongl. Sven. Vetensk. Akad. Handlingar*, **31**, 1–35.

Bjerknes, V., 1902. Cirkulation relativ zu der Erde (Circulation relative to the Earth). *Meteor. Z.*, **37**, 97–108.

Bjerknes, V., 1904. Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physic (The problem of weather forecasting as a problem in mathematics and physics). *Meteor. Z.*, January, 1–7. Engl. transl.: Y. Mintz, in Shapiro and Grønas (1999), pp. 1–7.

Boccaletti, G., Pacanowski, R. C., Philander, S. G. H. & Fedorov, A. V., 2004. The thermal structure of the upper ocean. *J. Phys. Oceanogr.*, **34**, 888–902.

Boer, G. J. & Shepherd, T. G., 1983. Large-scale two-dimensional turbulence in the atmosphere. *J. Atmos. Sci.*, **40**, 164–184.

Boussinesq, J., 1903. Théorie analytique de la chaleur (Analytic theory of heat). *Tome, Paris, Gauthier-Villars*, **II**, 170–172.

Boyd, J. P., 1976. The effects of latitudinal shear on equatorial waves, part 1. Theory and method. *J. Atmos. Sci.*, **35**, 2236–2258.

Branscome, L. E., 1983. The Charney baroclinic stability problem: approximate solutions and modal structures. *J. Atmos. Sci.*, **40**, 1393–1409.

Bretherton, C. S. & Schär, C., 1993. Flux of potential vorticity substance: a simple derivation and a uniqueness property. *J. Atmos. Sci.*, **50**, 1834–1836.

Bretherton, F. P., 1964. Low frequency oscillations trapped near the equator. *Tellus*, **16**, 181–185.

Bretherton, F. P., 1966. Critical layer instability in baroclinic flows. *Quart. J. Roy. Meteor. Soc.*, **92**, 325–334.

Brewer, A. W., 1949. Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Meteor. Soc.*, **75**, 251–363.

Brillouin, L., 1926. La mecanique ondulatoire de Schrödinger; une methode generale de resolution par approximations sucessives (The wave mechanics of Schrödinger: a general method of solution by successive approximation). *Comptes Rendus*, **183**, 24–26.

Bryan, F., 1986. High-latitude salinity effects and interhemispheric thermohaline circulations. *Nature*, **323**, 301–304.

Bryden, H. & Brady, E. C., 1985. Diagnostic study of the three-dimensional circulation of the upper equatorialPacific Ocean. *JPO*, **15**, 1255–1273.

Bryden, H. L., 1973. New polynomials for thermal expansion, adiabatic temperature gradient and potential temperature gradient of sea water. *Deep Sea Res.*, **20**, 401–408.

Buchanan, J. Y., 1886. On the similarities in the physical geography of the great oceans. *Proc. Roy. Geogr. Soc.*, **8**, 753–770.

Burger, A., 1958. Scale considerations of planetary motions of the atmosphere. *Tellus*, **10**, 195–205.

Callen, H. B., 1985. *Thermodynamics and an Introduction to Themostatistics*. John Wiley & Sons, 493 pp.

Cane, M. A., 1979a. The response of an equatorial ocean to simple wind stress patterns: I. Model formulation and analytic results. *J. Mar. Res.*, **37**, 232–252.

Cane, M. A., 1979b. The response of an equatorial ocean to simple wind stress patterns: II. Numerical results. *J. Mar. Res.*, **37**, 355–398.

Cessi, P., 2001. Thermohaline circulation variability. In *Conceptual Models of the Climate, Woods Hole Program in Geophysical Fluid Dynamics* (2001). Also available from http://gfd.whoi.edu/proceedings/2001/PDFvol2001.html.

Cessi, P. & Fantini, M., 2004. The eddy-driven thermocline. *J. Phys. Oceanogr.*, **34**, 2642–2658.

Cessi, P. & Young, W. R., 1992. Multiple equilibria in two-dimensional thermohaline circulation. *J. Fluid Mech.*, **241**, 291–309.

Chandrasekhar, S., 1961. *Hydrodynamic and Hydromagnetic Stability*. Oxford University Press, 652 pp. Reprinted by Dover Publications, 1981.

Chapman, D. C., Malanotte-Rizzoli, P. & Hendershott, M., 1989. Wave motions in the ocean. Unpublished. Based on lectures by Myrl Hendershott.

Chapman, S. & Lindzen, R. S., 1970. *Atmospheric Tides*. Gordon and Breach, 200 pp.

Charlton, A. J. & Polvani, L. M., 2007. A new look at stratospheric sudden warmings. part i: Climatology and modeling benchmarks. *J. Climate*, **20**, 449–469.

Charlton, A. J., Polvani, L. M., Perlwitz, J., Sassi, F. *et al.*, 2007. A new look at stratospheric sudden warmings. part i: Evaluation of numerical model simulations. *J. Climate*, **20**, 470–488.

Charney, J. G., 1947. The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135–162.

Charney, J. G., 1948. On the scale of atmospheric motion. *Geofys. Publ. Oslo*, **17** (2), 1–17.

Charney, J. G., 1955. The Gulf Stream as an inertial boundary layer. *Proc. Nat. Acad. Sci.*, **41**, 731–740.

Charney, J. G., 1960. Non-linear theory of a wind-driven homogeneous layer near the equator. *Deep-Sea Res.*, **6**, 303–310.

Charney, J. G., 1971. Geostrophic turbulence. *J. Atmos. Sci.*, **28**, 1087–1095.

Charney, J. G. & Drazin, P. G., 1961. Propagation of planetary scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–109.

Charney, J. G. & Eliassen, A., 1949. A numerical method for predicting the perturbations of the mid-laitude westerlies. *Tellus*, **1**, 38–54.

Charney, J. G. & Eliassen, A., 1964. On the growth of the hurricane depression. *J. Atmos. Sci.*, **21**, 68–75.

Charney, J. G., Fjørtoft, R. & Neumann, J. V., 1950. Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254.

Charney, J. G. & Stern, M. E., 1962. On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19**, 159–172.

Charnock, H., Green, J., Ludlam, F., Scorer, R. & Sheppard, P., 1966. Dr. E. T. Eady, B. A. (Obituary). *Quart. J. Roy. Meteor. Soc.*, **92**, 591–592.

Chasnov, J. R., 1991. Simulation of the inertial-conductive subrange. *Phys. Fluids A*, **3**, 1164–1168.

Chelton, D. B., de Szoeke, R. A., Schlax, M. G., Naggar, K. E. & Siwertz, N., 1998. Geographical variability of the first-baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.*, **28**, 433–460.

Christiansen, B., 1999. Stratospheric vacillations in a general circulation model. *J. Atmos. Sci.*, **56**, 1858–1872.

Christiansen, B., 2000. Chaos, quasiperiodicity, and interannual variability: studies of a stratospheric vacillation model. *J. Atmos. Sci.*, **57**, 3161–3173.

Colin de Verdière, A., 1980. Quasi-geostrophic turbulence in a rotating homogeneous fluid. *Geophys. Astrophys. Fluid Dyn.*, **15**, 213–251.

Colin de Verdière, A., 1989. On the interaction of wind and buoyancy driven gyres. *J. Mar. Res.*, **47**, 595–633.

Conkright, M. E., Antonov, J., Baranova, O., Boyer, T. P. *et al.*, 2001. World ocean database 2001, vol. 1. In S. Levitus, Ed., *NOAA Atlas NESDIS 42*, pp. 167. US Government Printing Office, Washington D. C.

Coriolis, G. G., 1832. Mémoire sur le principe des forces vives dans les mouvements relatifs des machines (On the principle of kinetic energy in the relative movement of machines). *J. Ec. Polytech*, **13**, 268–301.

Coriolis, G. G., 1835. Mémoire sur les équations du mouvement relatif des systèmes de corps (On the equations of relative motion of a system of bodies). *J. Ec. Polytech*, **15**, 142–154.

Corrsin, S., 1951. On the spectrum of isotropic temperature fluctuations in an isotropic turbulence. *J. Appl. Phys.*, **22**, 469–473. Erratum: *J. Appl. Phys.* **22,** 1292, (1951).

Cressman, G. P., 1996. The origin and rise of numerical weather prediction. In J. R. Fleming, Ed., *Historical Essays on Meteorology 1919–1995*, pp. 617. American Meteorological Society.

Cromwell, T., 1953. Circulation in a meridional plane in the central equatorial pacific. *J. Mar. Res.*, **12**, 196–213.

Cushman-Roisin, B., 1994. *Introduction to Geophysical Fluid Dynamics*. Prentice Hall, 320 pp.

Danielsen, E. F., 1990. In defense of Ertel's potential vorticity and its general applicability as a meteorological tracer. *J. Atmos. Sci.*, **47**, 2353–2361.

Danilov, S. & Gryanik, V., 2002. Rhines scale and spectra of the $\beta$-plane turbulence with bottom drag. *Phys. Rev. E*, **65**, 067301–1–067301–3.

Danilov, S. & Gurarie, D., 2001. Quasi-two-dimensional turbulence. *Usp. Fiz. Nauk.*, **170**, 921–968.

Davies-Jones, R., 2003. Comments on "A generalization of Bernoulli's theorem". *J. Atmos. Sci.*, **60**, 2039–2041.

Davis, R. E., de Szoeke, R., Halpern, D. & Niiler, P., 1981. Variability in the upper ocean during MILE. Part I: The heat and momentum balances. *Deep-Sea Res.*, **28**, 1427–1452.

De Morgan, A., 1872. *A Budget of Paradoxes*. Thoemmmes Continuum, 814 pp.

de Szoeke, R., 2004. An effect of the thermobaric nonlinearity of the equation of state: A mechanism for sustaining solitary Rossby waves. *J. Phys. Oceanogr.*, **34**, 2042–2056.

Defant, A., 1921. Die Zirkulation der Atmosphäre in den gemäßigten Breiten der Erde. Grundzüge einer Theorie der Klimaschwankungen (The circulation of the atmosphere in the Earth's midlatitudes. Basic features of a theory of climate fluctuations). *Geograf. Ann.*, **3**, 209–266.

Dewar, W. K. & Huang, R. X., 1995. Fluid flow in loops driven by freshwater and heat fluxes. *J. Fluid Mech.*, **297**, 153–191.

Dewar, W. K., Samelson, R. S. & Vallis, G. K., 2005. The ventilated pool: a model of subtropical mode water. *J. Phys. Oceanogr.*, **35**, 137–150.

Dickinson, R. E., 1968. Planetary Rossby waves propagating vertically through weak westerly wind wave guides. *J. Atmos. Sci.*, **25**, 984–1002.

Dickinson, R. E., 1969. Theory of planetary wave–zonal flow interaction. *J. Atmos. Sci.*, **26**, 73–81.

Dickinson, R. E., 1980. Planetary waves: theory and observation. In *Orographic Effects on Planetary Flows*, Number 23 in GARP Publication Series. World Meteorological Organization.

Dijkstra, H. A., 2002. *Nonlinear Physical Oceanography*. Kluwer, 456 pp.

Dima, I. & Wallace, J. M., 2003. On the seasonality of the Hadley Cell. *J. Atmos. Sci.*, **60**, 1522–1527.

Dobson, G. M. B., 1956. Origin and distribution of the polyatomic molecules in the atmosphere. *Proc. Roy. Soc. Lond. A*, **236**, 187–193.

Döös, K. & Coward, A., 1997. The Southern Ocean as the major upwelling zone of the North Atlantic. *Int. WOCE Newsletter* 27, 3–4.

Drazin, P. G. & Reid, W. H., 1981. *Hydrodynamic Stability*. Cambridge University Press, 527 pp.

Drijfhout, S. S. & Hazeleger, W., 2001. Eddy mixing of potential vorticity versus temperature in an isopycnic ocean model. *J. Phys. Oceanogr.*, **31**, 481–505.

Dunkerton, T., Hsu, C.-P. & McIntrye, M. E., 1981. Some eulerian and lagrangian diagnostics for a model stratosphere warming. *JAS*, **38**, 819–843.

Dunkerton, T. J., 1982. Shear zone asymmetry in the observed and simulated quasi-biennial oscillation. *J. Atmos. Sci.*, **38**, 461–469.

Dunkerton, T. J., 1997. The role of gravity waves in the quasi-biennial oscillation. *JGR*, **102**, 26053–26076.

Dunkerton, T. J., Delisi, D. P. & Baldwin, M. P., 1988. Distribution of major stratospheric warmings in relation to the quasi-biennial oscillation. *Geophys. Res. Lett.*, **115**, 136–139.

Durran, D. R., 1989. Improving the anelastic approximation. *J. Atmos. Sci.*, **46**, 1453–1461.

Durran, D. R., 1990. Mountain waves and downslope winds. *Meteor. Monogr.*, **23**, 59–81.

Durran, D. R., 1993. Is the Coriolis force really responsible for the inertial oscillations? *Bull. Am. Meteor. Soc.*, **74**, 2179–2184.

Durst, C. S. & Sutcliffe, R. C., 1938. The effect of vertical motion on the "geostrophic departure" of the wind. *Quart. J. Roy. Meteor. Soc.*, **64**, 240.

Dutton, J. A., 1986. *The Ceaseless Wind: An Introduction to the Theory of Atmospheric Motion*. Dover Publications, 617 pp.

Eady, E. T., 1949. Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Eady, E. T., 1950. The cause of the general circulation of the atmosphere. In *Cent. Proc. Roy. Meteor. Soc.* (1950), pp. 156–172.

Eady, E. T., 1954. The maintenance of the mean zonal surface currents. *Proc. Toronto Meteor. Conf. 1953*, **138**, 124–128. Royal Meteorological Society.

Eady, E. T., 1957. The general circulation of the atmosphere and oceans. In D. R. Bates, Ed., *The Earth and its Atmosphere*, pp. 130–151. New York, Basic Books.

Eady, E. T. & Sawyer, J. S., 1951. Dynamics of flow patterns in extra-tropical regions. *Quart. J. Roy. Meteor. Soc.*, **77**, 531–551.

Ebdon, R. A., 1960. Notes on the wind flow at 50 mb in tropical and subtropical regions in january 1957 and in 1960. *Quart. J. Roy. Meteor. Soc.*, **86**, 540–542.

Edmon, H. J., Hoskins, B. J. & McIntyre, M. E., 1980. Eliassen–Palm cross sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600–2616.

Egger, J., 1976. Linear response of a two-level primitive equation model to forcing by topography. *Mon. Wea. Rev.*, **104**, 351–364.

Egger, J., 1999. Inertial oscillations revisited. *J. Atmos. Sci.*, **56**, 2951–2954.

Einstein, A., 1916. *Relativity: The Special and the General Theory*. Random House, 188 pp.

Ekman, V. W., 1905. On the influence of the Earth's rotation on ocean currents. *Arch. Math. Astron. Phys.*, **2**, 1–52.

Eluszkiewicz, J., Crisp, D., Grainger, R. G., Lambert, A. *et al.*, 1997. Sensitivity of the residual circulation diagnosed from the UARS data to the uncertainties in the input fields and to the inclusion of aerosols. *J. Atmos. Sci.*, **54**, 1739–1757.

Er-El, J. & Peskin, R., 1981. Relative diffusion of constant-level balloons in the Southern Hemisphere. *J. Atmos. Sci.*, **38**, 2264–2274.

Ertel, H., 1942a. Ein neuer hydrodynamischer Wirbelsatz (A new hydrodynamic eddy theorem). *Meteorolol. Z.*, **59**, 277–281.

Ertel, H., 1942b. Über des Verhältnis des neuen hydrodynamischen Wirbelsatzes zum Zirculationssatz von V. Bjerknes (On the relationship of the new hydrodynamic eddy theorem to the circulation theorem of V. Bjerknes). *Meteorol. Z.*, **59**, 385–387.

Ertel, H. & Rossby, C.-G., 1949a. Ein neuer Erhaltungs-satz der Hydrodynamik (A new conservation theorem of hydrodynamics). *Sitzungsber. d. Deutschen Akad. Wissenschaften Berlin*, **1**, 3–11.

Ertel, H. & Rossby, C.-G., 1949b. A new conservation theorem of hydrodynamics. *Geofis. Pura Appl.*, **14**, 189–193.

Fang, M. & Tung, K. K., 1996. A simple model of nonlinear Hadley circulation with an ITCZ: analytic and numerical solutions. *J. Atmos. Sci.*, **53**, 1241–1261.

Fang, M. & Tung, K. K., 1999. Time-dependent nonlinear Hadley circulation. *J. Atmos. Sci.*, **56**, 1797–1807.

Farrell, B., 1984. Modal and non-modal baroclinic waves. *J. Atmos. Sci.*, **41**, 668–673.

Farrell, B. F. & Ioannou, P. J., 1996. Generalized stability theory. Part I: autonomous operators. *J. Atmos. Sci.*, **53**, 2025–2040.

Feistel, R., 2003. A new extended thermodynamic potential of seawater. *Prog. Oceanogr.*, **58**, 41–114.

Feistel, R. & Hagen, E., 1995. On the Gibbs thermodynamic potential of seawater. *Prog. Oceanogr.*, **36**, 249–347.

Fels, S., 1985. Radiative-dynamical interactions in the middle atmosphere. In S. Manabe, Ed., *Adv. Geophys. Vol 28, Part A, Issues in Atmospheric and Oceanic Modeling*, pp. 27–300. Academic Press.

Ferrari, R. & McWilliams, J. C., 2006. Parameterization of eddy fluxes near ocean boundaries. *Ocean Modelling.* (In review).

Ferrel, W., 1856a. An essay on the winds and currents of the ocean. *Nashville J. Med. & Surg.*, **11**, 287–301.

Ferrel, W., 1856b. The problem of the tides. *Astron. J*, **4**, 173–176.

Ferrel, W., 1858. The influence of the Earth's rotation upon the relative motion of bodies near its surface. *Astron. J.*, **V**, No. 13 (109), 97–100.

Ferrel, W., 1859. The motion of fluids and solids relative to the Earth's surface. *Math. Monthly*, **1**, 140–148, 210–216, 300–307, 366–373, 397–406.

Fjørtoft, R., 1950. Application of integral theorems in deriving criteria for laminar flows and for the baroclinic circular vortex. *Geophys. Publ.*, **17**, 1–52.

Fjørtoft, R., 1953. On the changes in the spectral distribution of kinetic energy for two-dimensional nondivergent flow. *Tellus*, **5**, 225–230.

Fleming, E. L., Chandra, S., Schoeberl, M. R. & Barnett, J. J., 1988. Monthly mean global climatology of temperature, wind, geopotential height, and pressure for 0-120 km. Technical report, NASA/Goddard Space Flight Center, Greenbelt, MD. NASA Tech. Memo. 100697.

Fofonoff, N. P., 1954. Steady flow in a frictionless homogeneous ocean. *J. Mar. Res.*, **13**, 254–262.

Fofonoff, N. P., 1985. Physical properties of seawater: a new salinity scale and equation of state for seawater. *J. Geophys. Res.*, **90** (C), 3332–3342.

Fofonoff, N. P. & Montgomery, R. B., 1955. The equatorial undercurrent in the light of the vorticity equation. *Tellus*, **7**, 518–521.

Foster, T. D., 1972. An analysis of the cabbeling instability in sea water. *J. Phys. Oceanogr.*, **2**, 294–301.

Fox-Kemper, B. & Pedlosky, J., 2004. Wind-driven barotropic gyre I: circulation control by eddy vorticity fluxes to an enhanced removal region. *J. Mar. Res.*, **62**, 169–193.

Franklin, W. S., 1898. Review of P. Duhem, *Traité Elementaire de Méchanique Chimique fondée sur la Thermodynamique,* Two volumes. Paris, 1897. *Phys. Rev.*, **6**, 170–175.

Friedman, R. M., 1989. *Appropriating the Weather: Vilhelm Bjerknes and the Construction of a Modern Meteorology*. Cornell University Press, 251 pp.

Frisch, U., 1995. *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 296 pp.

Fu, L. L. & Flierl, G. R., 1980. Nonlinear energy and enstrophy transfers in a realistically stratified ocean. *Dyn. Atmos. Oceans*, **4**, 219–246.

Gage, K. S. & Nastrom, G. D., 1986. Theoretical interpretation of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft during GASP. *J. Atmos. Sci.*, **43**, 729–740.

Galperin, B., Sukoriansky, S., Dikovskaya, N., Read, P. *et al.*, 2006. Anisotropic turbulence and zonal jets in rotating flows with a $\beta$-effect. *Nonlinear Proc. Geophys.*, **13**, 83–98.

Garcia, R. R., 1987. On the mean meridional circulation of the stratosphere. *J. Atmos. Sci.*, **44**, 2599–3609.

Gardiner, C. W., 1985. *Handbook of Stochastic Methods*. Springer-Verlag, 442 pp.

Gent, P. R. & McWilliams, J. C., 1990. Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.*, **20**, 150–155.

Gent, P. R., Willebrand, J., McDougall, T. J. & McWilliams, J. C., 1995. Parameterizing eddy-induced transports in ocean circulation models. *J. Phys. Oceanogr.*, **25**, 463–474.

Gierasch, P. J., 1975. Meridional circulation and the maintenance of the Venus atmospheric rotation. *J. Atmos. Sci.*, **32**, 1038–1044.

Gill, A. E., 1971. The equatorial current in a homogeneous ocean. *Deep-Sea Res.*, **18**, 421–431.

Gill, A. E., 1975. Models of equatorial currents. In *Proceedings of Numerical Models of Ocean Circulation*, pp. 181–203. National Academy of Science.

Gill, A. E., 1980. Some simple solutions for heat induced tropical circulation. *Quart. J. Roy. Meteor. Soc.*, **106**, 447–462.

Gill, A. E., 1982. *Atmosphere–Ocean Dynamics*. Academic Press, 662 pp.

Gill, A. E. & Clarke, A. J., 1974. Wind-induced upwelling, coastal currents and sea-level changes. *Deep-Sea Res.*, **21**, 325–345.

Gill, A. E., Green, J. S. A. & Simmons, A. J., 1974. Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies. *Deep-Sea Res.*, **21**, 499–528.

Gille, S. T., 1997. The Southern Ocean momentum balance: evidence for topographic effects from numerical model output and altimeter data. *J. Phys. Oceanogr.*, **27**, 2219–2232.

Gilman, P. A. & Glatzmaier, G. A., 1981. Compressible convection in a rotating spherical shell. I. Anelastic equations. *Astrophys. J. Suppl. Ser.*, **45**, 335–349.

Giorgetta, M. A., Manzini, E. & Roeckner, E., 2002. Forcing of the quasi-biennial oscillation from a broad spectrum of atmospheric waves. *Geophys. Res. Lett.*, **29**, 1245.

Gnanadesikan, A., 1999. A simple predictive model for the structure of the oceanic pycnocline. *Science*, **283**, 2077–2079.

Godske, C. L., Bergeron, T., Bjerknes, J. & Bundgaard, R. C., 1957. *Dynamic Meteorology and Weather Forecasting*. American Meteorological Society, 864 pp.

Gough, D. O., 1969. The anelastic approximation for thermal convection. *J. Atmos. Sci.*, **216**, 448–456.

Grant, H. L., Hughes, B. A., Vogel, W. M. & Moilliet, A., 1968. The spectrum of temperaure fluctuation in turbulent flow. *J. Fluid Mech.*, **344**, 423–442.

Grant, H. L., Stewart, R. W. & Moilliet, A., 1962. Turbulent spectra from a tidal channel. *J. Fluid Mech.*, **12**, 241–268.

Gray, L. J., 2010. Stratospheric equatorial dynamics. In *The Stratosphere: Dynamics, Transport, and Chemistry*, Geophys. Monogr. Ser, Vol. 190 (2010).

Gray, L. J., Crooks, S., C.Pascoe & Palmer, M., 2001. Solar and QBO influences on the timing of stratospheric sudden warmings. *J. Atmos. Sci.*, **61**, 2777–2796.

Gray, L. J., Phipps, S. J., Dunkerton, T. J., Baldwin, M. P. *et al.*, 2001. A data study of the influence of the upper stratosphere on northern hemisphere stratospheric warmings. *Quart. J. Roy. Meteor. Soc.*, **127**, 1985–2003.

Greatbatch, R. J., 1998. Exploring the relationship between eddy-induced transport velocity, vertical momentum transfer, and the isopycnal flux of potential vorticity. *J. Phys. Oceanogr.*, **28**, 422–432.

Green, G., 1837. On the motion of waves in a variable canal of small depth and width. *Trans. Camb. Phil. Soc.*, **6**, 457–462.

Green, J. S. A., 1960. A problem in baroclinic stability. *Quart. J. Roy. Meteor. Soc.*, **86**, 237–251.

Green, J. S. A., 1970. Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **96**, 157–185.

Green, J. S. A., 1977. The weather during July 1976: some dynamical considerations of the drought. *Weather*, **32**, 120–128.

Green, J. S. A., 1999. *Atmospheric Dynamics*. Cambridge University Press, 213 pp.

Greenspan, H., 1962. A criterion for the existence of inertial boundary layers in oceanic circulation. *Proc. Nat. Acad. Sci.*, **48**, 2034–2039.

Gregg, M. C., 1998. Estimation and geography of diapycnal mixing in the stratified ocean. In J. Imberger, Ed., *Physical Processes in Lakes and Oceans*, pp. 305–338. American Geophysical Union.

Griannik, N., Held, I., Smith, K. & Vallis, G. K., 2004. Effect of nonlinear drag on the inverse cascade. *Phys. Fluids*, **16**, 73–78.

Griffies, S. M., 1998. The Gent–McWilliams skew flux. *J. Phys. Oceanogr.*, **28**, 831–841.

Griffies, S. M., 2004. *Fundamentals of Ocean Climate Models*. Princeton University Press, 518 pp.

Grose, W. & Hoskins, B., 1979. On the influence of orography on large-scale atmospheric flow. *J. Atmos. Sci.*, **36**, 223–234.

Hadley, G., 1735. Concerning the cause of the general trade-winds. *Phil. Trans. Roy. Soc.*, **29**, 58–62.

Haine, T. W. N. & Marshall, J., 1998. Gravitational, symmetric, and baroclinic instability of the ocean mixed layer. *J. Phys. Oceanogr.*, **28**, 634–658.

Hallberg, R. & Gnanadesikan, A., 2001. An exploration of the role of transient eddies in determining the transport of a zonally reentrant current. *J. Phys. Oceanogr.*, **31**, 3312–3330.

Hamilton, K., Wilson, R. J. & Hemler, R. S., 2001. Spontaneous QBO-like oscillations simulated by the GFDL SKYHI general circulation mode. *J. Atmos. Sci.*, **58**, 3271–3292.

Hamilton, K. P., 1998. Dynamics of the tropical middle atmosphere: a tutorial review. *Atmosphere–Ocean*, **36**, 319–354.

Haney, R. L., 1971. Surface thermal boundary condition for ocean circulation models. *J. Phys. Oceanogr.*, **1**, 241–248.

Hayes, M., 1977. A note on group velocity. *Proc. Roy. Soc. Lond. A*, **354**, 533–535.

Haynes, P., 2005. Stratospheric dynamics. *Ann. Rev. Fluid Mech.*, **37**, 263–293.

Haynes, P. H. & McIntyre, M. E., 1987. On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *J. Atmos. Sci.*, **44**, 828–841.

Haynes, P. H. & McIntyre, M. E., 1990. On the conservation and impermeability theorem for potential vorticity. *J. Atmos. Sci.*, **47**, 2021–2031.

Heifetz, E. & Caballero, R., 2014. Revisiting the rossby wave propagation mechanism. *Tellus A*. Submitted.

Held, I. M., 1982. On the height of the tropopause and the static stability of the troposphere. *J. Atmos. Sci.*, **39**, 412–417.

Held, I. M., 1983. Stationary and quasi-stationary eddies in the extratropical troposphere: theory. In B. Hoskins & R. P. Pearce, Eds., *Large-Scale Dynamical Processes in the Atmosphere*, pp. 127–168. Academic Press.

Held, I. M., 2000. The general circulation of the atmosphere. In *Woods Hole Program in Geophysical Fluid Dynamics* (2000), pp. 66.

Held, I. M. & Hou, A. Y., 1980. Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J. Atmos. Sci.*, **37**, 515–533.

Held, I. M. & Larichev, V. D., 1996. A scaling theory for horizontally homogeneous, baroclinically unstable flow on a beta-plane. *J. Atmos. Sci.*, **53**, 946–952.

Held, I. M., Tang, M. & Wang, H., 2002. Northern winter stationary waves: theory and modeling. *J. Climate*, **15**, 2125–2144.

Helmholtz, H., 1858. Über Integrale der hydrodynamischen Gleichungen welche den Wirbelbewengungen entsprechen (On the integrals of the hydrodynamic equations that correspond to eddy motion). *J. Reine Angew. Math*, **25**, 25–55. Engl. transl.: C. Abbe, *Smithson. Misc. Collect.*, no. 34, pp. 78–93, Smithsonian Institution, Washington D. C., 1893.

Helmholtz, H., 1868. Über discontinuirliche Flüssigkeitsbewegungen (On discontinuous liquid motion). *Monats. Königl. Preuss. Akad. Wiss. Berlin*, **23**, 215–228. Engl. trans.: F. Guthrie: On discontinuous movements of fluids. *Phil. Mag.*, **36**, 337–346 (1868).

Hendershott, M., 1987. Single layer models of the general circulation. In H. Abarbanel & W. R. Young, Eds., *General Circulation of the Ocean*, pp. 202–267. Springer-Verlag.

Henning, C. C. & Vallis, G. K., 2004. The effect of mesoscale eddies on the main subtropical thermocline. *J. Phys. Oceanogr.*, **34**, 2428–2443.

Henning, C. C. & Vallis, G. K., 2005. The effects of mesoscale eddies on the stratification and transport of an ocean with a circumpolar channel. *J. Phys. Oceanogr.*, **35**, 880–896.

Hide, R., 1969. Dynamics of the atmospheres of major planets with an appendix on the viscous boundary layer at the rigid boundary surface of an electrically conducting rotating fluid in the presence of a magnetic field. *J. Atmos. Sci.*, **26**, 841–853.

Hilborn, R. C., 2004. Sea gulls, butterflies, and grasshoppers: a brief history of the butterfly effect in nonlinear dynamics. *Am. J. Phys.*, **72**, 425–427.

Hockney, R., 1970. The potential calculation and some applications. In *Methods of Computational Physics*, Vol. 9, pp. 135–211. Academic Press.

Hogg, N., 2001. Quantification of the deep circulation. In G. Siedler, J. Church, & J. Gould, Eds., *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, pp. 259–270. Academic Press.

Hoinka, K. P., 1997. The tropopause: discovery, definition and demarcation. *Meteorol. Z.*, **6**, 281–303.

Holland, W. R., Keffer, T. & Rhines, P. B., 1984. Dynamics of the oceanic circulation: The potential vorticity field. *Nature*, **308**, 698–705.

Holloway, G. & Hendershott, M. C., 1977. Stochastic closure for nonlinear Rossby waves. *J. Fluid Mech.*, **82**, 747–765.

Holmes, M. H., 2013. *Introduction to Perturbation Methods*. 2nd edn. Springer, 436 pp.

Holton, J. R., 1992. *An Introduction to Dynamic Meteorology*. 3rd edn. Academic Press, 507 pp.

Holton, J. R., Haynes, P. R., McIntyre, M. E., Douglass, A. R. *et al.*, 1995. Stratosphere-troposphere exchange. *Rev. Geophys.*, **33**, 403–439.

Holton, J. R. & Lindzen, R. S., 1972. An updated theory for the quasi-biennial cycle of the tropical stratosphere. *J. Atmos. Sci.*, **29**, 1076–1080.

Holton, J. R. & Mass, C., 1976. Stratospheric vacillation cycles. *J. Atmos. Sci.*, **33**, 2218–2215.

Holton, J. R. & Tan, H.-C., 1980. The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *J. Atmos. Sci.*, **37**, 2200–2208.

Holton, J. R. & Tan, H.-C., 1982. The quasi-biennial oscillation in the northern hemisphere lower stratosphere. *J. Meteor. Soc. Japan*, **60**, 140–18.

Hoskins, B. J. & Karoly, D. J., 1981. The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196.

Hough, S. S., 1897. On the application of harmonic analysis to the dynamical theory of the tides. Part I: On Laplace's "oscillations of the first species", and on the dynamics of ocean currents. *Phil. Trans. (A)*, **189 (IX)**, 201–258.

Hough, S. S., 1898. On the application of harmonic analysis to the dynamical theory of the tides. Part II: on the general integration of Laplace's dynamical equations. *Phil. Trans. (A)*, **191 (V)**, 139–186.

Hua, B. L. & Haidvogel, D. B., 1986. Numerical simulations of the vertical structure of quasi-geostrophic turbulence. *J. Atmos. Sci.*, **43**, 2923–2936.

Huang, R. X., 1988. On boundary value problems of the ideal-fluid thermocline. *J. Phys. Oceanogr.*, **18**, 619–641.

Huang, R. X., 1998. Mixing and available potential energy in a Boussinesq ocean. *J. Phys. Oceanogr.*, **28**, 669–678.

Huang, R. X., 1999. Mixing and energetics of the oceanic thermohaline circulation. *J. Phys. Oceanogr.*, **29**, 727–746.

Hughes, C. W., 2002. Sverdrup-like theories of the Antarctic Circumpolar Current. *J. Mar. Res.*, **60**, 1–17.

Hughes, C. W. & de Cuevas, B., 2001. Why western boundary currents in realistic oceans are inviscid: a link between form stress and bottom pressure torques. *J. Phys. Oceanogr.*, **31**, 2871–2885.

Hughes, G. O. & Griffiths, R. W., 2008. Horizontal convection. *Ann. Rev. Fluid Mech.*, **185–2008**, 40.

Ierley, G. R. & Ruehr, O. G., 1986. Analytic and numerical solutions of a nonlinear boundary value problem. *Stud. Appl. Math.*, **75**, 1–36.

Il'in, A. M. & Kamenkovich, V. M., 1964. The structure of the boundary layer in the two-dimensional theory of ocean currents (in Russian). *Okeanologiya*, **4** (5), 756–769.

Ingersoll, A. P., 2005. Boussinesq and anelastic approximations revisited: potential energy release during thermobaric instability. *J. Phys. Oceanogr.*, **35**, 1359–1369.

Iwayama, T., Shepherd, T. G. & Watanabe, T., 2002. An 'ideal' form of decaying two-dimensional turbulence. *J. Fluid Mech.*, **456**, 183–198.

Jackett, J. & McDougall, T., 1995. Minimal adjustment of hydrographic profiles to achieve static stability. *J. Atmos. Ocean. Tech.*, **12**, 381–389.

Jackson, L., Hughes, C. W. & Williams, R. G., 2006. Topographic control of basin and channel flows: the role of bottom pressure torques and friction. *J. Phys. Oceanogr.*, **36**, 1786–1805.

James, I. N., 1994. *Introduction to Circulating Atmospheres*. Cambridge University Press, 422 pp.

Jeffreys, H., 1924. On certain approximate solutions of linear differential equations of the second order. *Proc. London Math. Soc. (2)*, **23**, 428–436.

Jeffreys, H., 1926. On the dynamics of geostrophic winds. *Quart. J. Roy. Meteor. Soc.*, **51**, 85–104.

Jeffreys, H. & Jeffreys, B. S., 1946. *Methods of Mathematical Physics*. Cambridge University Press, 728 pp.

Johnson, G. C. & Bryden, H. L., 1989. On the size of the Antarctic Circumpolar Current. *Deep-Sea Res.*, **36**, 39–53.

Jones, D. B. A., Schneider, H. R. & McElroy, M. B., 1998. Effects of the quasi-biennial oscillation on the zonally averaged transport of tracer. *J. Geophys. Res.*, **103**, 11235–11249.

Jucker, M., Fueglistaler, S. & Vallis, G. K., 2013. Stratospheric sudden warmings in an idealized gcm. *J. Geophys. Res.*. Submitted.

Juckes, M. N., 2000. The static stability of the midlatitude troposphere: the relevance of moisture. *J. Atmos. Sci.*, **57**, 3050–3057.

Juckes, M. N., 2001. A generalization of the transformed Eulerian-mean meridional circulation. *Quart. J. Roy. Meteor. Soc.*, **127**, 147–160.

Kalnay, E., 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

Karsten, R., Jones, H. & Marshall, J., 2002. The role of eddy transfer in setting the stratification and transport of a circumpolar current. *J. Phys. Oceanogr.*, **32**, 39–54.

Keffer, T., 1985. The ventilation of the world's oceans: maps of potential vorticity. *J. Phys. Oceanogr.*, **15**, 509–523.

Kessler, W. S., Johnson, G. C. & Moore, D. W., 2003. Sverdrup and nonlinear dynamics of the Pacific equatorial currents. *J. Phys. Oceanogr.*, **33**, 994–1008.

Kevorkian, J. & Cole, J. D., 2011. *Multiple Scale and Singular Perturbation Methods*. Springer-Verlag, 648 pp.

Kibel, I., 1940. Priloozhenie k meteorogi uravnenii mekhaniki baroklinnoi zhidkosti (Application of baroclinic fluid dynamic equations to meteorology). *SSSR Ser. Geogr. Geofiz.*, **5**, 627–637.

Killworth, P. D., 1987. A continuously stratified nonlinear ventilated thermocline. *J. Phys. Oceanogr.*, **17**, 1925–1943.

Killworth, P. D., 1997. On the parameterization of eddy transfer. Part I: theory. *J. Marine Res.*, **55**, 1171–1197.

Kimoto, M. & Ghil, M., 1993. Multiple flow regimes in the northern hemisphere winter. Part 1: methodology and hemispheric regimes. *J. Atmos. Sci.*, **50**, 2625–2643.

Kolmogorov, A. N., 1941. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Acad. Sci. USSR*, **30**, 299–303.

Kolmogorov, A. N., 1962. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds numbers. *J. Fluid Mech.*, **13**, 82–85.

Kraichnan, R., 1967. Inertial ranges in two-dimensional turbulence. *Phys. Fluids*, **10**, 1417–1423.

Kraichnan, R. & Montgomery, D., 1980. Two-dimensional turbulence. *Rep. Prog. Phys.*, **43**, 547–619.

Kramers, H. A., 1926. Wellenmechanik und halbzahlige Quantisierung (Wave mechanics and semi-integral quantization). *Zeit. fur Physik A*, **39**, 828–840.

Kundu, P., Allen, J. S. & Smith, R. L., 1975. Modal decomposition of the velocity field near the Oregon coast. *J. Phys. Oceanogr.*, **5**, 683–704.

Kundu, P. & Cohen, I. M., 2002. *Fluid Mechanics*. Academic Press, 730 pp.

Kuo, H.-l., 1949. Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteorol.*, **6**, 105–122.

Kuo, H.-l., 1951. Vorticity transfer as related to the development of the general circulation. *J. Meteorol.*, **8**, 307–315.

Kushner, P. J., 2010. Annular modes of the troposphere and stratosphere. In *The Stratosphere: Dynamics, Transport, and Chemistry*, Geophys. Monogr. Ser (2010).

Kushnir, Y., Robinson, W. A., Bladé, I., Hall, N. M. J. *et al.*, 2002. Atmospheric GCM response to extratropical SST anomalies: synthesis and evaluation. *J. Climate*, **15**, 2233–2256.

Labitzke, K., Kunze, M. & Bronnimann, S., 2006. Sunspots, the qbo and the stratosphere in north polar regions —- 20 years later. *Meteor. Z.*, **15**, 355–363.

LaCasce, J. H. & Ohlmann, C., 2003. Relative dispersion at the surface of the Gulf of Mexico. *J. Mar. Res.*, **65**, 285–312.

Lait, L. R., 1994. An alternative form for potential vorticity. *J. Atmos. Sci.*, **51**, 1754–1759.

Lanczos, C., 1970. *The Variational Principles of Mechanics*. University of Toronto Press, Reprinted by Dover Publications 1980, 418 pp.

Landau, L. D., 1944. On the problem of turbulence. *Dokl. Akad. Nauk SSSR*, **44**, 311–314.

Landau, L. D. & Lifshitz, E. M., 1987. *Fluid Mechanics* (Course of Theoretical Physics, v. 6). 2nd edn. Pergamon Press, 539 pp.

Larichev, V. D. & Held, I. M., 1995. Eddy amplitudes and fluxes in a homogeneous model of fully developed baroclinic instability. *J. Phys. Oceanogr.*, **25**, 2285–2297.

LeBlond, P. H. & Mysak, L. A., 1980. *Waves in the Ocean*. Elsevier, 616 pp.

Ledwell, J., Watson, A. & Law, C., 1998. Mixing of a tracer released in the pycnocline. *J. Geophys. Res.*, **103**, 21499–21529.

Lee, M.-M., Marshall, D. P. & Williams, R. G., 1997. On the eddy transfer of tracers: advective or diffusive? *J. Mar. Res.*, **55**, 483–595.

Lee, T. D., 1951. Difference between turbulence in a two-dimensional fluid and in a three-dimensional fluid. *J. Appl. Phys.*, **22**, 524.

Leetmaa, A., Niiler, P. & Stommel, H., 1977. Does the Sverdrup relation account for the mid-Atlantic circulation? *J. Mar. Res.*, **35**, 1–10.

Leith, C. E., 1968. Diffusion approximation for two-dimensional turbulence. *Phys. Fluids*, **11**, 671–672.

Lesieur, M., 1997. *Turbulence in Fluids: Third Revised and Enlarged Edition*. Kluwer, 515 pp.

Levinson, N., 1950. The 1st boundary value problem for $\epsilon \Delta U + A(x,y)U_x + B(x,y)U_y + C(x,y)U = D(x,t)$ for small epsilon. *Ann. Math.*, **51**, 429–445.

Lewis, R., Ed., 1991. *Meteorological Glossary*. 6th edn. Her Majesty's Stationery Office, 335 pp.

Lighthill, J., 1978. *Waves in Fluids*. Cambridge University Press, 504 pp.

Lighthill, M. J., 1965. Group velocity. *J. Inst. Math. Appl.*, **1**, 1–28.

Lilly, D. K., 1969. Numerical simulation of two-dimensional turbulence. *Phys. Fluid Suppl. II*, **12**, 240–249.

Lilly, D. K., 1996. A comparison of incompressible, anelastic and Boussinesq dynamics. *Atmos. Res*, **40**, 143–151.

Limpasuvan, V., Thompson, D. W. J. & Hartmann, D. L., 2000. The life cycle of the Northern Hemisphere sudden stratospheric warmings. *J. Climate*, **17**, 2584–2596.

Lindborg, E., 1999. Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259–288.

Lindborg, E. & Alvelius, K., 2000. The kinetic energy spectrum of the two-dimensional enstrophy turbulence cascade. *Phys. Fluids*, **12**, 945–947.

Lindzen, R. S. & Farrell, B., 1980. The role of the polar regions in global climate, and a new parameterization of global heat transport. *Mon. Wea. Rev.*, **108**, 2064–2079.

Lindzen, R. S. & Holton, J. R., 1968. A theory of the quasi-biennial oscillation. *J. Atmos. Sci.*, **25**, 1095–1107.

Lindzen, R. S. & Hou, A. Y., 1988. Hadley circulation for zonally averaged heating centered off the equator. *J. Atmos. Sci.*, **45**, 2416–2427.

Lindzen, R. S., Lorenz, E. N. & Plazman, G. W., Eds., 1990. *The Atmosphere — a Challenge: the Science of Jule Gregory Charney*. American Meteorological Society, 321 pp.

Liouville, J., 1837. Sur le développement des fonction ou parties de fonction en séries (On the development of functions of parts of functions in series). *J. Math. Pures Appl.*, **2**, 16–35.

Lipps, F. B. & Hemler, R. S., 1982. A scale analysis of deep moist convection and some related numerical calculations. *J. Atmos. Sci.*, **39**, 2192–2210.

Longuet-Higgins, M. S., 1964. Planetary waves on a rotating sphere, I. *Proc. Roy. Soc. Lond. A*, **279**, 446–473.

Lorenz, E. N., 1955. Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 157–167.

Lorenz, E. N., 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Lorenz, E. N., 1967. *The Nature and the Theory of the General Circulation of the Atmosphere*. WMO Publications, Vol. 218, World Meteorological Organization.

Lozier, S., Owens, W. B. & Curry, R. G., 1996. The climatology of the North Atlantic. *Prog. Oceanog.*, **36**, 1–44.

Luyten, J. R., Pedlosky, J. & Stommel, H., 1983. The ventilated thermocline. *J. Phys. Oceanogr.*, **13**, 292–309.

Mahrt, L., 1986. On the shallow motion approximations. *J. Atmos. Sci.*, **43**, 1036–1044.

Majda, A. J., 2003. *Introduction to PDEs and Waves for the Atmosphere and Ocean*. American Mathematical Society, 234 pp.

Maltrud, M. E. & Vallis, G. K., 1991. Energy spectra and coherent structures in forced two-dimensional and beta-plane turbulence. *J. Fluid Mech.*, **228**, 321–342.

Manabe, S. & Stouffer, R. J., 1988. Two stable equilibria of a coupled ocean–atmosphere model. *J. Climate*, **1**, 841–866.

Margules, M., 1903. Über die Energie der Stürme (On the energy of storms). *Jahrb. Kais.-kön Zent. für Met. und Geodynamik, Vienna*, 26 pp. Engl. transl.: C. Abbe, *Smithson. Misc. Collect.*, no. 51, pp. 533-595, Smithsonian Institution, Washington D. C., 1910.

Marotzke, J., 1989. Instabilities and multiple steady states of the thermohaline circulation. In D. L. T. Anderson & J. Willebrand, Eds., *Oceanic Circulation Models: Combining Data and Dynamics*, pp. 501–511. NATO ASI Series, Kluwer.

Marshall, D. P., Williams, R. G. & Lee, M.-M., 1999. The relation between eddy-induced transport and isopycnic gradients of potential vorticity. *J. Phys. Oceanogr.*, **29**, 1571–1578.

Marshall, J. C., 1981. On the parameterization of geostrophic eddies in the ocean. *J. Phys. Oceanogr.*, **11**, 257–271.

Marshall, J. C. & Nurser, A. J. G., 1992. Fluid dynamics of oceanic thermocline ventilation. *J. Phys. Oceanogr.*, **22**, 583–595.

Marshall, J. C. & Radko, T., 2003. Residual-mean solutions for the Antarctic Circumpolar Current and its associated overturning circulation. *J. Phys. Oceanogr.*, **22**, 2341–2354.

Marshall, J. C. & Schott, F., 1999. Open-ocean convection: observations, theory, and models. *Rev. Geophys.*, **37**, 1–64.

Matsuno, T., 1966. Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44**, 25–43.

Matsuno, T., 1971. A dynamical model of the sudden stratospheric warming. *J. Atmos. Sci.*, **28**, 1479–1494.

McComb, W. D., 1990. *The Physics of Fluid Turbulence*. Clarendon Press, 572 pp.

McCreary, J. P., 1981. A linear stratified ocean model of the equatorial undercurrent. *Phil. Trans. Roy. Soc. Lond.*, **A298**, 603–635.

McCreary, J. P., 1985. Modeling equatorial ocean circulation. *Ann. Rev. Fluid Mech.*, **17**, 359–407.

McCreary, J. P. & Lu, P., 1994. Interaction between the subtropical and equatorial ocean circulations: the subtropical cell. *J. Phys. Oceanogr.*, **24**, 466–497.

McDougall, T. J., 1998. Three-dimensional residual mean theory. In E. P. Chassignet & J. Verron, Eds., *Ocean Modeling and Parameterization*, pp. 269–302. Kluwer Academic.

McDougall, T. J., 2003. Potential enthalphy: a conservative oceanic variable for evaluating heat content and heat fluxes. *J. Phys. Oceanogr.*, **33**, 945–963.

McDougall, T. J., Jackett, D. R., Wright, D. G. & Feistel, R., 2002. Accurate and computationally efficient formulae for potential temperature and density of seawater. *J. Atmos. Ocean. Tech.*, **20**, 730–741.

McIntyre, M. E. & Norton, W. A., 1990. Dissipative wave–mean interactions and the transport of vorticity or potential vorticity. *J. Fluid Mech.*, **212**, 403–435.

McIntyre, M. E. & Norton, W. A., 2000. Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.*, **57**, 1214–1235.

McKee, W. D., 1973. The wind-driven equatorial circulation in a homogeneous ocean. *Deep-Sea Res.*, **20**, 889–899.

McWilliams, J. C., 1984. The emergence of isolated coherent vortices in turbulent flow. *J. Fluid Mech.*, **146**, 21–43.

Mellor, G. L., 1991. An equation of state for numerical models of oceans and estuaries. *J. Atmos. Ocean. Tech.*, **8**, 609–611.

Mihaljan, J. M., 1962. A rigorous exposition of the Boussinesq approximations applicable to a thin layer of fluid. *Astrophysical J.*, **136**, 1126–1133.

Morel, P. & Larcheveque, M., 1974. Relative dispersion of constant-level balloons in the 200 mb general circulation. *J. Atmos. Sci.*, **31**, 2189–2196.

Mundt, M., Vallis, G. K. & Wang, J., 1997. Balanced models for the large- and meso-scale circulation. *J. Phys. Oceanogr.*, **27**, 1133–1152.

Munk, W. H., 1950. On the wind-driven ocean circulation. *J. Meteorol*, **7**, 79–93.

Munk, W. H., 1966. Abyssal recipes. *Deep-Sea Res.*, **13**, 707–730.

Munk, W. H., 1981. Internal waves and small-scale processes. In B. A. Warren & C. Wunsch, Eds., *Evolution of Physical Oceanography*, pp. 264–291. The MIT Press.

Munk, W. H., Groves, G. W. & Carrier, G. F., 1950. Note on the dynamics of the Gulf Stream. *J. Mar. Res.*, **9**, 218–238.

Munk, W. H. & Palmén, E., 1951. Note on dynamics of the Antarctic Circumpolar Current. *Tellus*, **3**, 53–55.

Munk, W. H. & Wunsch, C., 1998. Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res.*, **45**, 1976–2009.

Namias, J., 1959. Recent seasonal interaction between North Pacific waters and the overlying atmospheric circulation. *J. Geophys. Res.*, **64**, 631–646.

Newell, A. C., 1969. Rossby wave packet interactions. *J. Fluid Mech.*, **35**, 255–271.

Nicholls, S., 1985. Aircraft observations of the Ekman layer during the joint air-sea interaction experiment. *Quart. J. Roy. Meteor. Soc.*, **111**, 391–426.

Nof, D., 2003. The Southern Ocean's grip on the northward meridional flow. In *Progress in Oceanography*, Vol. 56, pp. 223–247. Pergamon.

Novikov, E. A., 1959. Contributions to the problem of the predictability of synoptic processes. *Izv. An. SSSR Ser. Geophys.*, **11**, 1721. Eng. transl.: *Am. Geophys. U. Transl.*, 1209-1211.

Oberbeck, A., 1879. Über die Wärmeleitung der Flüssigkeiten bei Berücksichtigung der Strömungen infolge vor Temperaturdifferenzen (On the thermal conduction of liquids taking into account flows due to temperature differences). *Ann. Phys. Chem., Neue Folge*, **7**, 271–292.

Oberbeck, A., 1888. Über die Bewegungserscheinungen der Atmosphäre (On the phenomena of motion in the atmosphere). *Sitzb. K. Preuss. Akad. Wiss*, **7**, 383–395 and 1129–1138. Engl trans.: in B. Saltzman, Ed.,*Theory of Thermal Convection*, Dover, 162–183.

Obukhov, A. M., 1941. Energy distribution in the spectrum of turbulent flow. *Izv. Akad. Nauk. SSR, Ser. Geogr. Geofiz.*, **5**, 453–466.

Obukhov, A. M., 1949. Structure of the temperature field in turbulent flows. *Izv. Akad. Nauk. SSR, Ser. Geogr. Geofiz.*, **13**, 58–63.

Obukhov, A. M., 1962. On the dynamics of a stratified liquid. *Dokl. Akad. Nauk SSSR*, **145**, 1239–1242. Engl. transl.: *Soviet Physics–Dokl.* 7, 682–684.

Oetzel, K. & Vallis, G. K., 1997. Strain, vortices, and the enstrophy inertial range in two-dimensional turbulence. *Phys. Fluids*, **9**, 2991–3004.

O'Gorman, P. A. & Pullin, D. I., 2005. Effect of Schmidt number of the velocity-scaler cospectrum in isotropic turbulence with a mean scalar gradient. *J. Fluid Mech.*, **532**, 111–140.

Ogura, Y. & Phillips, N. A., 1962. Scale analysis of deep and shallow convection in the atmosphere. *J. Atmos. Sci.*, **19**, 173–179.

Olbers, D., Borowski, D., Völker, C. & Wolff, J.-O., 2004. The dynamical balance, transport and circulation of the Antarctic Circumpolar Current. *Antarctic Science*, **16**, 439–470.

Ollitrault, M., Gabillet, C. & Colin de Verdière, A., 2005. Open ocean regimes of relative dispersion. *J. Fluid Mech.*, **533**, 381–407.

Onsager, L., 1949. Statistical hydrodynamics. *Nuovo Cim. (Suppl.)*, **6**, 279–287.

Paldor, N., Rubin, S. & Mariono, A. J., 2007. A consistent theory for linear waves of the shallow-water equations on a rotating plane in midlatitudes. *J. Atmos. Sci.*, **37**, 115–128.

Palmer, T. N., 1981. Diagnostic study of a wavenumber-2 stratospheric sudden warming in a transformed Eulerian-mean formalism. *J. Atmos. Sci.*, **38**, 844–855.

Palmer, T. N., 1997. A nonlinear dynamical perspective on climate prediction. *J. Climate*, **12**, 575–591.

Paparella, F. & Young, W. R., 2002. Horizontal convection is non-turbulent. *J. Fluid Mech.*, **466**, 205–214.

Pascoe, C. L., Gray, L. J., Crooks, S. A., Juckes, M. N. & Baldwin, M. P., 2005. The quasi-biennial oscillation: Analysis using ERA-40 data. *J. Geophys. Res.*, **110**, D08105.

Pedlosky, J., 1964. The stability of currents in the atmosphere and ocean. Part I. *J. Atmos. Sci.*, **21**, 201–219.

Pedlosky, J., 1987a. *Geophysical Fluid Dynamics*. 2nd edn. Springer-Verlag, 710 pp.

Pedlosky, J., 1987b. An inertial theory of the equatorial undercurrent. *J. Phys. Oceanogr.*, **17**, 1978–1985.

Pedlosky, J., 1996. *Ocean Circulation Theory*. Springer-Verlag, 453 pp.

Pedlosky, J., 2003. *Waves in the Ocean and Atmosphere: Introduction to Wave Dynamics*. Springer-Verlag, 260 pp.

Peixoto, J. P. & Oort, A. H., 1992. *Physics of Climate*. American Institute of Physics, 520 pp.

Peltier, W. R. & Stuhne, G., 2002. The upscale turbulence cascade: shear layers, cyclones and gas giant bands. In R. P. Pearce, Ed., *Meteorology at the Millennium*, pp. 43–61. Academic Press.

Persson, A., 1998. How do we understand the Coriolis force? *Bull. Am. Meteor. Soc.*, **79**, 1373–1385.

Philander, S. G., 1990. *El Niño, La Niña, and the Southern Oscillation*. Academic Press, 289 pp.

Phillips, N. A., 1954. Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus*, **6**, 273–286.

Phillips, N. A., 1956. The general circulation of the atmosphere: a numerical experiment. *Quart. J. Roy. Meteor. Soc.*, **82**, 123–164.

Phillips, N. A., 1963. Geostrophic motion. *Rev. Geophys.*, **1**, 123–176.

Phillips, N. A., 1966. The equations of motion for a shallow rotating atmosphere and the traditional approximation. *J. Atmos. Sci.*, **23**, 626–630.

Phillips, N. A., 1973. Principles of large-scale numerical weather prediction. In P. Morel, Ed., *Dynamic Meteorology*, pp. 1–96. Riedel.

Phillips, N. A., 2000. Explication of the Coriolis effect. *Bull. Am. Meteor. Soc.*, **81**, 299–303.

Pierrehumbert, R. T. & Swanson, K. L., 1995. Baroclinic instability. *Ann. Rev. Fluid Mech.*, **27**, 419–467.

Plumb, R. A., 1977. The interaction of two internal waves with the mean flow: Implications for the theory of the Quasi-Biennial Oscillation. *J. Atmos. Sci.*, **34**, 1847–1858.

Plumb, R. A., 1979. Eddy fluxes of conserved quantities by small-amplitude waves. *J. Atmos. Sci.*, **36**, 1699–1704.

Plumb, R. A., 1984. The quasi-biennial oscillation. In J. R. Holton & T. Matsuno, Eds., *Dynamics of the Middle Atmosphere*, pp. 217–251. Terra Scientific Publishing.

Plumb, R. A., 2002. Stratospheric transport. *J. Meteor. Soc. Japan*, **80**, 793–809.

Plumb, R. A. & Bell, R. C., 1982. An analysis of the quasi-biennial oscillation on an equtorial beta-plane. *Quart. J. Roy. Meteor. Soc.*, **108**, 335–352.

Poincaré, H., 1893. *Théorie des Tourbillons (Theory of Vortices [literally, Swirls])*. Georges Carré, Éditeur. Reprinted by Éditions Jacques Gabay, 1990, 211 pp.

Poincaré, H., 1908. *Science and Method*. T. Nelson and Sons. Engl. transl.: F. Maitland. Reprinted in *The Value of Science: Essential Writings of Henri Poincaré*, Ed. S. J. Gould, Random House, 584 pp.

Polzin, K. L., Toole, J. M., Ledwell, J. R. & Schmidt, R. W., 1997. Spatial variability of turbulent mixing in the abyssal ocean. *Science*, **276**, 93–96.

Price, J. F., Weller, R. A. & Schudlich, R. R., 1987. Wind-driven ocean currents and Ekman transport. *Science*, **238**, 1534–1538.

Proudman, J., 1916. On the motion of solids in liquids. *Proc. Roy. Soc. Lond. A*, **92**, 408–424.

Queney, P., 1948. The problem of air flow over mountains: A summary of theoretical results. *BAMS*, **29**, 16–26.

Quon, C. & Ghil, M., 1992. Multiple equilibria in thermosolutal convection due to salt-flux boundary conditions. *J. Fluid Mech.*, **245**, 449–484.

Randel, W. J., Wu, F., Swinbank, R., Nash, J. & O'Neill, A., 1999. Global QBO circulation derived from UKMO stratospheric analyse. *J. Atmos. Sci.*, **56**, 457–474.

Rayleigh, Lord, 1880. On the stability, or instability, of certain fluid motions. *Proc. London Math. Soc.*, **11**, 57–70.

Rayleigh, Lord, 1894. *The Theory of Sound, Volume II*. 2nd edn. Macmillan, 522 pp. Reprinted by Dover Publications, 1945.

Rayleigh, Lord, 1912. On the propagation of waves through a stratified medium, with special reference to the question of reflection. *Proc. Roy. Soc. Lond. A*, **86**, 207–226.

Read, P. L., 2001. Transition to geostrophic turbulence in the laboratory, and as a paradigm in atmospheres and oceans. *Surveys Geophys.*, **33**, 265–317.

Reed, R. J., 1960. The structure and dynamics of the 26-month oscillation. Paper presented at the 40th anniversary meeting of the Am. Meter. Soc., Boston.

Reed, R. J., Campbell, W. J., Rasmussen, L. A. & Rogers, D. G., 1961. Evidence of a downward-propagating annual wind reversal in the equatorial stratosphere. *J. Geophys. Res.*, **66**, 813–818.

Reif, F., 1965. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill, 651 pp.

Rhines, P. B., 1975. Waves and turbulence on a $\beta$-plane. *J. Fluid. Mech.*, **69**, 417–443.

Rhines, P. B., 1977. The dynamics of unsteady currents. In E. A. Goldberg, I. N. McCane, J. J. O'Brien, & J. H. Steele, Eds., *The Sea*, Vol. 6, pp. 189–318. J. Wiley and Sons.

Rhines, P. B., 1979. Geostrophic turbulence. *Ann. Rev. Fluid Mech.*, **11**, 401–441.

Rhines, P. B. & Holland, W. R., 1979. A theoretical discussion of eddy-driven mean flows. *Dyn. Atmos. Oceans*, **3**, 289–325.

Rhines, P. B. & Young, W. R., 1982a. Homogenization of potential vorticity in planetary gyres. *J. Fluid Mech.*, **122**, 347–367.

Rhines, P. B. & Young, W. R., 1982b. A theory of wind-driven circulation. I. Mid-ocean gyres. *J. Mar. Res. (Suppl)*, **40**, 559–596.

Richardson, L. F., 1920. The supply of energy from and to atmospheric eddies. *Proc. Roy. Soc. Lond. A*, **97**, 354–373.

Richardson, L. F., 1922. *Weather Prediction by Numerical Process*. Cambridge University Press, 236 pp. Reprinted by Dover Publications.

Richardson, L. F., 1926. Atmospheric diffusion on a distance-neighbour graph. *Proc. Roy. Soc. Lond. A*, **110**, 709–737.

Richardson, P. L., 1983. Eddy kinetic-energy in the North Atlantic from surface drifters. *J. Geophys. Res.*, **88**, NC7, 4355–4367.

Rintoul, S. R., Hughes, C. & Olbers, D., 2001. The Antarctic Circumpolar Current system. In G. Siedler, J. Church, & J. Gould, Eds., *Ocean Circulation and Climate*, pp. 271–302. Academic Press.

Robinson, A. R., 1966. An investigation into the wind as the cause of the equatorial undercurrent. *J. Mar. Res.*, **24**, 179–204.

Robinson, A. R., Ed., 1984. *Eddies in Marine Science*. Springer-Verlag, 609 pp.

Robinson, A. R. & McWilliams, J. C., 1974. The baroclinic instability of the open ocean. *J. Phys. Oceanogr.*, **4**, 281–294.

Robinson, A. R. & Stommel, H., 1959. The oceanic thermocline and the associated thermohaline circulation. *Tellus*, **11**, 295–308.

Rooth, C., 1982. Hydrology and ocean circulation. *Prog. Oceanogr.*, **11**, 131–149.

Rossby, C.-G., 1936. Dynamics of steady ocean currents in the light of experimental fluid dynamics. *Papers Phys. Oceanog. Meteor.*, **5**, 1–43.

Rossby, C.-G., 1938. On the mutual adjustment of pressure and velocity distributions in certain simple current systems, II. *J. Mar. Res.*, **5**, 239–263.

Rossby, C.-G., 1939. Relations between variation in the intesity of the zonal circulation and the displacements of the semi-permanent centers of action. *J. Marine Res.*, **2**, 38–55.

Rossby, C.-G., 1940. Planetary flow patterns in the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **66,** suppl., 68–87.

Rossby, C.-G., 1949. On the nature of the general circulation of the lower atmosphere. In G. P. Kuiper, Ed., *The Atmospheres of the Earth and Planets*, pp. 16–48. University of Chicago Press.

Rossby, H. T., 1965. On thermal convection driven by non-uniform heating from below: an experimental study. *Deep-Sea Res.*, **12**, 9–16.

Rossby, H. T., 1998. Numerical experiments with a fluid heated non-uniformly from below. *Tellus A*, **50**, 242–257.

Rudnick, D. L. & Weller, R. A., 1993. Observations of superinertial and near-inertial wind-driven flow. *J. Phys. Oceanogr.*, **23**, 2351–2359.

Ruelle, D. & Takens, F., 1971. On the nature of turbulence. *Commun. Math. Phys.*, **20**, 167–192.

Salmon, R., 1980. Baroclinic instability and geostrophic turbulence. *Geophys. Astrophys. Fluid Dyn.*, **10**, 25–52.

Salmon, R., 1983. Practical use of Hamilton's principle. *J. Fluid Mech.*, **132**, 431–444.

Salmon, R., 1990. The thermocline as an internal boundary layer. *J. Mar. Res.*, **48**, 437–469.

Salmon, R., 1998. *Lectures on Geophysical Fluid Dynamics*. Oxford University Press, 378 pp.

Saltzman, B., 1962. Finite amplitude free convection as an initial value problem. *J. Atmos. Sci.*, **19**, 329–341.

Samelson, R. M., 1999a. Geostrophic circulation in a rectangular basin with a circumpolar connection. *J. Phys. Oceanogr.*, **29**, 3175–3184.

Samelson, R. M., 1999b. Internal boundary layer scaling in 'two-layer' solutions of the thermocline equations. *J. Phys. Oceanogr.*, **29**, 2099–2102.

Samelson, R. M., 2004. Simple mechanistic models of middepth meridional overturning. *J. Phys. Oceanogr.*, **34**, 2096–2103.

Samelson, R. M. & Vallis, G. K., 1997. Large-scale circulation with small diapycnal diffusion: the two-thermocline limit. *J. Mar. Res.*, **55**, 223–275.

Sandström, J. W., 1908. Dynamische Versuche mit Meerwasser (Dynamical experiments with seawater). *Annal. Hydrogr. Marit. Meteorol.*, **36**, 6–23.

Sandström, J. W., 1916. Meteologische Studien im Schwedischen Hochgebirge (Meteological studies in the Swedish high mountains). *Goteborgs Kungl. Vetenskaps-och Vitterhets-Samhalles, Handingar*, **27**, 1–48.

Sarkisyan, A. & Ivanov, I., 1971. Joint effect of baroclinicity and relief as an important factor in the dynamics of sea currents. *Izv. Akad. Nauk Atmos. Ocean. Phys.*, **7**, 116–124.

Scaife, A., Butchart, N., Warner, C. D. & Swinbank, R., 2002. Impact of a spectral gravity wave parameterization on the stratosphere in the Met Office Unified Model. *J. Atmos. Sci.*, **59**, 1473–1489.

Scaife, A. & James, I. N., 2000. Response of the stratosphere to interannual variability of tropospheric planetary waves. *Quart. J. Roy. Meteor. Soc.*, **126**, 275–297.

Schär, C., 1993. A generalization of Bernoulli's theorem. *J. Atmos. Sci.*, **50**, 1437–1443.

Schmitz, W. J., 1995. On the interbasin-scale thermohaline circulation. *Rev. Geophysics*, **33**, 151–173.

Schneider, E. K., 1977. Axially symmetric steady-state models of the basic state for instability and climate studies. Part II: nonlinear calculations. *J. Atmos. Sci.*, **34**, 280–297.

Schneider, E. K. & Lindzen, R. S., 1977. Axially symmetric steady-state models of the basic state for instability and climate studies. Part I: linearized calculations. *J. Atmos. Sci.*, **34**, 263–279.

Schneider, T., Held, I. & Garner, S. T., 2003. Boundary effects in potential vorticity dynamics. *J. Atmos. Sci.*, **60**, 1024–1040.

Schneider, T. & Sobel, A., Eds., 2007. *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*. Princeton University Press. To appear.

Schneider, T. & Walker, C. C., 2006. Self-organization of atmospheric macroturbulence into critical states of weak nonlinearity. *J. Atmos. Sci.*, **63**, 1569–1586.

Schoeberl, M. R., 1978. Stratospheric warmings: observations and theory. *Revs. Geophys. Space Phys.*, **16**, 521–538.

Schubert, W. H., Hausman, S. A., Garcia, M., Ooyama, K. V. & Kuo, H.-C., 2001. Potential vorticity in a moist atmosphere. *J. Atmos. Sci.*, **58**, 3148–3157.

Schubert, W. H., Ruprecht, E., Hertenstein, R., Nieto Ferreira, R. *et al.*, 2004. English translations of twenty-one of Ertel's papers on geophysical fluid dynamics. *Meteor. Z.*, **13**, 527–576.

Scinocca, J. F. & Shepherd, T. G., 1992. Nonlinear wave-activity conservation laws and Hamiltonian structure for the two-dimensional anelastic equations. *J. Atmos. Sci.*, **49**, 5–28.

Scorer, R. S., 1951. Numerical evaluation of integrals of the form $I = \int f(x)\mathrm{e}^{\mathrm{i}\theta(x)}\,\mathrm{d}x$ and tabulation of the function $Gi(x)$. *Q. J. Mech. Appl. Maths*, **3**, 107–112.

Scott, R. B., 2001. Evolution of energy and enstrophy containing scales in decaying, two-dimensional turbulence with friction. *Phys. Fluids*, **13**, 2739–2742.

Scott, R. K. & Haynes, P. H., 1998a. Internal interannual variability of the extratropical stratospheric circulation: The low latitude flywheel. *Quart. J. Roy. Meteor. Soc.*, **124**, 2149–2173.

Scott, R. K. & Haynes, P. H., 1998b. Internal vacillations in stratosphere-only models. *J. Atmos. Sci.*, **57**, 2333–2350.

Shapiro, M. & Grønas, S., Eds., 1999. *The Life Cycles of Extratropical Cyclones*. American Meteorological Society, 359 pp.

Shepherd, T. G., 1987. A spectral view of nonlinear fluxes and stationary-transient interaction in the atmosphere. *J. Atmos. Sci.*, **44**, 1166–1179.

Shepherd, T. G., 1993. A unified theory of available potential energy. *Atmosphere–Ocean*, **31**, 1–26.

Shepherd, T. G., 2002. Issues in stratosphere–troposphere coupling. *J. Meteor. Soc. Japan*, **80**, 769–792.

Shutts, G. J., 1983. Propagation of eddies in diffluent jet streams: eddy vorticity forcing of blocking flow fields. *Quart. J. Roy. Meteor. Soc.*, **109**, 737–761.

Siedler, G., Church, J. & Gould, J., 2001. *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*. Academic Press, 715 pp.

Silberstein, L., 1896. O tworzeniu sie wirow, w plynie doskonalym (On the creation of eddies in an ideal fluid). *W Krakaowie Nakladem Akademii Umiejetnosci (Proc. Cracow Acad. Sci.)*, **31**, 325–335. Also published as: Über die Entstehung von Wirbelbewegungen in einer reibungslosen Flüssigkeit in *Bull. Int. l'Acad. Sci. Cracovie, Compte Rendue Scéances Année*, 280–290, 1896. Engl. transl. by M. Ziemiánski available from the author.

Simmonds, J. G. & Mann, J. E., 1998. *A First Look at Perturbation Theory*. Dover Publications, 139 pp.

Simmons, A. & Hoskins, B., 1978. The life-cycles of some nonlinear baroclinic waves. *J. Atmos. Sci.*, **35**, 414–432.

Smagorinsky, J., 1953. The dynamical influences of large-scale heat sources and sinks on the quasi-stationary mean motions of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **79**, 342–366.

Smagorinsky, J., 1969. Problems and promises of deterministic extended range forecasting. *Bull. Am. Meteor. Soc.*, **50**, 286–311.

Smith, K. S., Boccaletti, G., Henning, C. C., Marinov, I. *et al.*, 2002. Turbulent diffusion in the geostrophic inverse cascade. *J. Fluid Mech.*, **469**, 13–48.

Smith, K. S. & Vallis, G. K., 1998. Linear wave and instability properties of extended range geostrophic models. *J. Atmos. Sci.*, **56**, 1579–1593.

Smith, K. S. & Vallis, G. K., 2001. The scales and equilibration of mid-ocean eddies: freely evolving flow. *J. Phys. Oceanogr.*, **31**, 554–571.

Smith, K. S. & Vallis, G. K., 2002. The scales and equilibration of mid-ocean eddies: forced-dissipative flow. *J. Phys. Oceanogr.*, **32**, 1669–1721.

Smith, L. M. & Waleffe, F., 1999. Transfer of energy to two-dimensional large scales in forced, rotating three-dimensional turbulence. *Phys. Fluids*, **11**, 1608–1622.

Smith, R. B., 1979. The influence of mountains on the atmosphere. In B. Saltzman, Ed., *Advances in Geophysics*, vol. 21, pp. 87–230. Academic Press.

Spall, M. A., 2000. Generation of strong mesoscale eddies by weak ocean gyres. *J. Mar. Res.*, **58**, 97–116.

Spiegel, E. A. & Veronis, G., 1960. On the Boussinesq approximation for a compressible fluid. *Astrophys. J.*, **131**, 442–447. (Correction: *Astrophys. J.*, **135,** 655–656).

Squire, H., 1933. On the stability of three-dimensional disturbances of viscous flow between parallel walls. *Proc. Roy. Soc. Lond. A*, **142**, 621–628.

Stammer, D., 1997. Global characteristics of ocean variability estimated from regional TOPEX/Poseidon altimeter measurements. *J. Phys. Oceanogr.*, **27**, 1743–1769.

Starr, V. P., 1948. An essay on the general circulation of the Earth's atmosphere. *J. Meteor.*, **78,** 39–43.

Starr, V. P., 1968. *Physics of Negative Viscosity Phenomena*. McGraw-Hill, 256 pp.

Steers, J. A., 1962. *An Introduction to the Study of Map Projections*. Univ. of London Press, 288 pp.

Stern, M. E., 1963. Trapping of low frequency oscillations in an equatorial boundary layer. *Tellus*, **15**, 246–250.

Stevens, D. P. & Ivchenko, V. O., 1997. The zonal momentum balance in an eddy-resolving general-circulation model of the southern ocean. *Quart. J. Roy. Meteor. Soc.*, **123**, 929–951.

Stewart, G. R., 1941. *Storm*. Random House, 349 pp.

Stips, A., 2005. Dissipation measurement: theory. In H. Z. Baumert, J. Simpson, & J. Sündermann, Eds., *Marine Turbulence*. Cambridge University Press.

Stommel, H., 1958. The abyssal circulation. *Deep-Sea Res.*, **5**, 80–82.

Stommel, H., 1960. Wind-drift near the equator. *Deep Sea Research Part*, **6**, 298–302.

Stommel, H., 1961. Thermohaline convection with two stable regimes of flow. *Tellus*, **13**, 224–230.

Stommel, H. & Arons, A. B., 1960. On the abyssal circulation of the world ocean—I. Stationary planetary flow patterns on a sphere. *Deep-Sea Res.*, **6**, 140–154.

Stommel, H., Arons, A. B. & Faller, A. J., 1958. Some examples of stationary planetary flow patterns in bounded basins. *Tellus*, **10**, 179–187.

Stommel, H. & Moore, D. W., 1989. *An Introduction to the Coriolis Force*. Columbia University Press, 297 pp.

Stommel, H. & Webster, J., 1963. Some properties of the thermocline equations in a subtropical gyre. *J. Mar. Res.*, **44**, 695–711.

Stone, P. H., 1972. A simplified radiative-dynamical model for the static stability of rotating atmospheres. *J. Atmos. Sci.*, **29**, 405–418.

Stone, P. H., 1978. Baroclinic adjustment. *J. Atmos. Sci.*, **35**, 561–571.

Stone, P. H. & Nemet, B., 1996. Baroclinic adjustment: a comparison between theory, observations, and models. *J. Atmos. Sci.*, **53**, 1663–1674.

Straub, D. N., 1993. On the transport and angular momentum balance of channel models of the Antarctic Circumpolar Current. *J. Phys. Oceanogr.*, **23**, 776–782.

Sutcliffe, R. C., 1939. Cyclonic and anticylonic development. *Quart. J. Roy. Meteor. Soc.*, **65**, 518–524.

Sutcliffe, R. C., 1947. A contribution to the problem of development. *Quart. J. Roy. Meteor. Soc.*, **73**, 370–383.

Sutherland, B., 2010. *Internal Gravity Waves*. Cambridge University Press, 394 pp. ISBN: 978-0521839150.

Swallow, J. C. & Worthington, V., 1961. An observation of a deep countercurrent in the western North Atlantic. *Deep-Sea Res.*, **8**, 1–19.

Talley, L., 1988. Potential vorticity distribution in the North Pacific. *J. Phys. Oceanogr.*, **18**, 89–106.

Taylor, G. I., 1921a. Diffusion by continuous movements. *Proc. London Math. Soc.*, **2** (**20**), 196–211.

Taylor, G. I., 1921b. Experiments with rotating fluids. *Proc. Roy. Soc. Lond. A*, **100**, 114–121.

Tennekes, H. & Lumley, J. L., 1972. *A First Course in Turbulence*. The MIT Press, 330 pp.

Tesserenc De Bort, L. P., 1902. Variations de la température de l'air libre dans la zone comprise 8 km et 13 km d'altitude (Variations in the temperature of the free air in the zone between 8 km and 13 km of altitude). *C. R. Hebd. Séances Acad. Sci.*, **134**, 987–989.

Thompson, A. F. & Young, W. R., 2006. Scaling baroclinic eddy fluxes: vortices and energy balance. *J. Phys. Oceanogr.*, **36**, 720–738.

Thompson, P. D., 1957. Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275–295.

Thompson, R. O. R. Y., 1971. Why there is an intense eastward current in the North Atlantic but not in the South Atlantic. *J. Phys. Oceanogr.*, **1**, 235–237.

Thompson, R. O. R. Y., 1980. A prograde jet driven by Rossby waves. *J. Atmos. Sci.*, **37**, 1216–1226.

Thomson, J., 1892. Bakerian lecture. On the grand currents of atmospheric circulation. *Phil. Trans. Roy. Soc. Lond. A*, **183**, 653–684.

Thomson, W. (Lord Kelvin), 1869. On vortex motion. *Trans. Roy. Soc. Edinburgh*, **25**, 217–260.

Thomson, W. (Lord Kelvin), 1871. Hydrokinetic solutions and observations. *Phil. Mag. and J. Science*, **42**, 362–377.

Thomson, W. (Lord Kelvin), 1879. On gravitational oscillations of rotating water. *Proc. Roy. Soc. Edinburgh*, **10**, 92–100.

Thorncroft, C. D., Hoskins, B. J. & McIntyre, M. E., 1993. Two paradigms of baroclinic-wave life-cycle behaviour. *Quart. J. Roy. Meteor. Soc.*, **119**, 17–55.

Thorpe, A. J., Volkert, H. & Ziemianski, M. J., 2003. The Bjerknes' circulation theorem: a historical perspective. *Bull. Am. Meteor. Soc.*, **84**, 471–480.

Thual, O. & McWilliams, J. C., 1992. The catastrophe structure of thermohaline convection in a two-dimensional fluid model and a comparison with low-order box models. *Geophys. Astrophys. Fluid Dyn.*, **64**, 67–95.

Thuburn, J. & Craig, G. C., 1997. GCM tests of theories for the height of the tropopause. *J. Atmos. Sci.*, **54**, 869–882.

Toggweiler, J. R. & Samuels, B., 1995. Effect of Drake Passage on the global thermohaline circulation. *Deep-Sea Res.*, **42**, 477–500.

Toggweiler, J. R. & Samuels, B., 1998. On the ocean's large-scale circulation in the limit of no vertical mixing. *J. Phys. Oceanogr.*, **28**, 1832–1852.

Toole, J. M., Polzin, K. L. & Schmitt, R. W., 1994. Estimates of diapycnal mixing in the abyssal ocean. *Science*, **264**, 1120–1123.

Tréguier, A. M., Held, I. M. & Larichev, V. D., 1997. Parameterization of quasi-geostrophic eddies in primitive equation ocean models. *J. Phys. Oceanogr.*, **29**, 567–580.

Trenberth, K. E. & Caron, J. M., 2001. Estimates of meridional atmosphere and ocean heat transports. *J. Climate*, **14**, 3433–3443.

Tritton, D. J., 1988. *Physical Fluid Dynamics*. Oxford University Press, 519 pp.

Truesdell, C., 1951. Proof that Ertel's vorticity theorem holds in average for any medium suffering no tangential acceleration on the boundary. *Geofis Pura Appl.*, **19**, 167–169.

Truesdell, C., 1954. *The Kinematics of Vorticity*. Indiana University Press, 232 pp.

Tung, K. K., 1979. A theory of stationary long waves. Part III: quasi-normal modes in a singular wave guide. *Mon. Wea. Rev.*, **107**, 751–774.

UNECSO, 1981. The practical salinity scale 1978 and the international equation of state of seawater 1980. Tenth report of the joint panel on oceanographic tables and standards. Technical report, UNESCO Technical Papers in Marine Science No. 36, Paris.

Valdes, P. J. & Hoskins, B. J., 1989. Baroclinic instability of the zonally averaged flow with boundary layer damping. *J. Atmos. Sci.*, **45**, 1584–1593.

Vallis, G. K., 1982. A statistical dynamical climate model with a simple hydrology cycle. *Tellus*, **34**, 211–227.

Vallis, G. K., 1988. Numerical studies of eddy transport properties in eddy-resolving and parameterized models. *Quart. J. Roy. Meteor. Soc.*, **114**, 183–204.

Vallis, G. K., 1996. Potential vorticity and balanced equations of motion for rotating and stratified flows. *Quart. J. Roy. Meteor. Soc.*, **122**, 291–322.

Vallis, G. K., 2000. Large-scale circulation and production of stratification: effects of wind, geometry and diffusion. *J. Phys. Oceanogr.*, **30**, 933–954.

Vallis, G. K. & Maltrud, M. E., 1993. Generation of mean flows and jets on a beta plane and over topography. *J. Phys. Oceanogr.*, **23**, 1346–1362.

Vanneste, J. & Shepherd, T. G., 1998. On the group-velocity property for wave-activity conservation laws. *J. Atmos. Sci.*, **55**, 1063–1068.

Verkley, W. T. M. & van der Velde, I. R., 2010. Balanced dynamics in the tropics. *QJRMS*, **136**, 41–49.

Veronis, G., 1960. An approximate theoretical analysis of the equatorial undercurrent. *Deep-Sea Res.*, **6**, 318–327.

Veronis, G., 1966a. Wind-driven ocean circulation – Part 1: linear theory and perturbation analysis. *Deep-Sea Res.*, **13**, 17–29.

Veronis, G., 1966b. Wind-driven ocean circulation – Part 2: numerical solutions of the non-linear problem. *Deep-Sea Res.*, **13**, 30–55.

Veronis, G., 1969. On theoretical models of the thermocline circulation. *Deep-Sea Res.*, **31** Suppl., 301–323.

Veryard, R. G. & Ebdon, R. A., 1961. Fluctuations in tropical stratospheric winds. *Meteor. Mag*, **90**, 125–143.

Visbeck, M., Marshall, J., Haine, T. & Spall, M., 1997. Specification of eddy transfer coefficients in coarse-resolution ocean circulation models. *J. Phys. Oceanogr.*, **27**, 381–402.

Walker, C. & Schneider, T., 2005. Response of idealized Hadley circulations to seasonally varying heating. *Geophys. Res. Lett.*, **32**, L06813. doi:10.1029/2004GL022304.

Wallace, J. M., 1973. General circulation of the tropical lower stratosphere. *Rev. Geophys.*, **11**, 191–222.

Wallace, J. M., 1983. The climatological mean stationary waves: observational evidence. In B. Hoskins & R. P. Pearce, Eds., *Large-Scale Dynamical Processes in the Atmosphere*, pp. 27–63. Academic Press.

Wallace, J. M. & Holton, J. R., 1968. A diagnostic numerical model of the quasi-biennial oscillation. *J. Atmos. Sci.*, **25**, 280–292.

Warn, T., Bokhove, O., Shepherd, T. G. & Vallis, G. K., 1995. Rossby number expansions, slaving principles, and balance dynamics. *Quart. J. Roy. Meteor. Soc.*, **121**, 723–739.

Warren, B. A., 1981. Deep circulation of the world ocean. In B. A. Warren & C. Wunsch, Eds., *Evolution of Physical Oceanography*, pp. 6–41. The MIT Press.

Warren, B. A., LaCasce, J. H. & Robbins, P. E., 1996. On the obscurantist physics of form drag in theorizing about the circumpolar current. *J. Phys. Oceanogr.*, **26**, 2297–2301.

Wasow, W., 1944. Asymptotic solution of boundary value problems for the differential equation $\Delta U + \lambda(\partial/\partial x)U = \lambda f(x, y)$. *Duke Math J.*, **11**, 405–415.

Watanabe, T., Iwayama, T. & Fujisaka, H., 1998. Scaling law for coherent vortices in decaying drift Rossby wave turbulence. *Phys. Rev. E*, **57**, 1636–1643.

Webb, D. J. & Suginohara, N., 2001. Vertical mixing in the ocean. *Nature*, **409**, 37.

Weinstock, R., 1952. *Calculus of Variations*. McGraw-Hill. Reprinted by Dover Publications, 1980, 328 pp.

Welander, P., 1959. An advective model of the ocean thermocline. *Tellus*, **11**, 309–318.

Welander, P., 1968. Wind-driven circulation in one- and two-layer oceans of variable depth. *Tellus*, **20**, 1–15.

Welander, P., 1971a. Some exact solutions to the equations describing an ideal-fluid thermocline. *J. Mar. Res.*, **29**, 60–68.

Welander, P., 1971b. The thermocline problem. *Phil. Trans. Roy. Soc. Lond. A*, **270**, 415–421.

Welander, P., 1973. Lateral friction in the ocean as an effect of potential vorticity mixing. *Geophys. Fluid Dyn.*, **5**, 101–120.

Welander, P., 1986. Thermohaline effects in the ocean circulation and related simple models. In J. Willebrand & D. L. T. Anderson, Eds., *Large-scale Transport Processes in Oceans and Atmospheres*, pp. 163–200. Reidel.

Wentzel, G., 1926. Eine Verallgemeinerung der Quantenbedingungen für die Zwecke der Wellenmechanik (A generalization of the quantum conditions for the purposes of wave mechanics). *Zeit. fur Physic A*, **38**, 518–529.

White, A. A., 1977. Modified quasi-geostrophic equations using geometric height as vertical co-ordinate. *Quart. J. Roy. Meteor. Soc.*, **103**, 383–396.

White, A. A., 2002. A view of the equations of meteorological dynamics and various approximations. In J. Norbury & I. Roulstone, Eds., *Large-Scale Atmosphere-Ocean Dynamics I*, pp. 1–100. Cambridge University Press.

White, A. A., 2003. The primitive equations. In J. Holton, J. Pyle, & J. Curry, Eds., *Encyclopedia of Atmospheric Science*, pp. 694–702. Academic Press.

Whitehead, J. A., 1975. Mean flow generated by circulation on a beta-plane: An analogy with the moving flame experiment. *Tellus*, **27**, 358–364.

Whitehead, J. A., 1995. Thermohaline ocean processes and models. *Ann. Rev. Fluid Mech.*, **27**, 89–113.

Whitham, G. B., 1974. *Linear and Nonlinear Waves*. Wiley-Interscience, 656 pp.

Williams, G. P., 1978. Planetary circulations: 1. Barotropic representation of Jovian and terrestrial turbulence. *J. Atmos. Sci.*, **35**, 1399–1426.

World Meteorological Organization, 1957. Definition of the tropopause. *WMO Bulletin*, **6**, 136.

Wright, D. G., 1997. An equation of state for use in ocean models: Eckart's formula revisited. *J. Atmos. Ocean. Tech.*, **14**, 735–740.

Wunsch, C., 2002. What is the thermohaline circulation? *Science*, **298**, 1179–1180.

Wunsch, C. & Ferrari, R., 2004. Vertical mixing, energy, and the general circulation of the oceans. *Ann. Rev. Fluid Mech.*, **36**, 281–314.

Wunsch, C. & Roemmich, D., 1985. Is the North Atlantic in Sverdrup balance? *J. Phys. Oceanogr.*, **15**, 1876–1880.

Wyrtki, K., Magaard, L. & Hager, J., 1976. Eddy energy in oceans. *J. Geophys. Res.*, **81**, 15, 2641–2646.

Yaglom, A. M., 1994. A. N. Kolmogorov as a fluid mechanician and founder of a school in turbulence research. *Ann. Rev. Fluid Mech.*, **26**, 1–22.

Yoden, S., 1987. Bifurcation properties of a stratospheric vacillation model. *J. Atmos. Sci.*, **44**, 1723–1733.

Yoden, S., 1990. An illustrative model of sea- sonal and interannual variations of the stratospheric circulation. *JAS*, **47**, 1845–1853.

Young, W. R., 2006. Boussinesq notes. Unpublished.

Young, W. R. & Rhines, P. B., 1982. A theory of the wind-driven circulation II. Gyres with western boundary layers. *J. Mar. Res.*, **40**, 849–872.

Zdunkowski, W. & Bott, A., 2003. *Dynamics of the Atmosphere: A Course in Theoretical Meteorology*. Cambridge University Press, 719 pp.

Zhang, R. & Vallis, G. K., 2007. The role of the bottom vortex stretching on the path of the North Atlantic western boundary current and on the northern recirculation gyre. *J. Phys. Oceanogr.*, **37**, 2053–2080.

Zurita-Gotor, P. & Lindzen, R., 2007. Theories of baroclinic adjustment and eddy equilibration. In T. Schneider & A. Sobel, Eds., *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*. Princeton University Press.

# Index

**Bold face** denotes a primary entry or an extended discussion.

## Q