Instability and nonlinear equilibration of baroclinic flows

Benjamin Thomas Willcocks

DEPARTMENT OF MATHEMATICS
UNIVERSITY COLLEGE, LONDON

A THESIS PRESENTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SUPERVISOR
Dr J. Gavin Esler

APRIL 2012

I, Benjamin Thomas Willcocks, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

SIGNED

DISCLAIMER

In certain parts of the thesis, the author has benefited from an exchange of notes with his supervisor, Dr J. Gavin Esler. Specifically, sections 2.4 and 2.5 (including appendices 2.A, 2.B and 2.C) should be regarded as joint work.

Elements of chapter 2 form part of a manuscript (Esler & Willcocks 2011 [31]) accepted for publication in *Geophysical and Astrophysical Fluid Dynamics* and chapter 4 forms the main part of a manuscript (Willcocks & Esler 2011 [138]) accepted for publication in the *Journal of Physical Oceanography*.

To my parents

Acknowledgements

The author would like to thank his supervisor, Dr. J. Gavin Esler, for his patient advice and constant encouragement, without which this thesis would never have been completed.

Thanks also to examiners Prof. Peter Read and Dr. James Cho for their many useful comments, which have greatly improved the script.

The Engineering and Physical Sciences Research Council and UCL mathematics department are also gratefully acknowledged for their financial support.

Abstract

Baroclinic instability, the fundamental mechanism underlying the generation of baroclinic eddies in the atmosphere and ocean is investigated in the two-layer, quasi-geostrophic model. The aim is to bridge the gap in understanding between analytical theories and high resolution numerical simulations of more realistic flows.

In chapter 1 the physical motivation for the problems, two-layer model and numerical scheme are introduced. In chapter 2, the instability of a uniform flow profile without Ekman friction is investigated. The success of a weakly nonlinear theory due to Warn & Gauthier at finite criticality is assessed over the full parameter space. The relevance of nonlinear bounds on wave amplitude and perturbation energy due to Shepherd is also evaluated.

Chapters 3 and 4 investigate the Holopainen instability, whereby a uniform flow profile, otherwise stable in frictionless flow, is destabilized by the addition of a small amount of Ekman friction. In chapter 3, the physical mechanisms of the baroclinic and Holopainen instabilities are contrasted in terms of potential vorticity disturbances. The instability of the Eady model is also discussed.

In chapter 4, a weakly nonlinear theory due to Romea is shown to be accurate for flows unstable to the Holopainen instability and flows unstable to baroclinic instability in the presence of significant Ekman friction. An intermediate flow region is found where Warn & Gauthier's theory is accurate at early times, but the final state is well predicted by Romea's theory.

The equilibration of an unstable baroclinic jet is investigated in chapter 5. A predictive theory due to Esler based on global constraints is extended to test two new hypotheses, which are also shown to be successful in predicting the equilibrated flow profile of initially symmetric jets. The theory is adapted to include asymmetric initial jets where each hypothesis is found to have limited quantitative success.

Contents

Chapter 1

Introduction

1.1 Physical problems to be addressed and introduction to the two-layer, quasi-geostrophic model

An important question in atmospheric science is what determines the temperature difference between the tropics and high latitudes in the terrestrial atmosphere? Of great importance to this question is the study of baroclinic instability, the key process through which available potential energy stored in the pole-equator temperature gradient is released to generate synoptic scale[1] eddies [98]. Baroclinic eddies on the synoptic scale transport heat polewards and act to reduce the pole-equator temperature gradient. These eddies are primarily confined to so-called storm tracks [127] in the mid-latitudes of the

---

[1] Synoptic scale refers to large scale phenomena in the latitudinal/longitudinal directions such as extra-tropical cyclones or jet streams, typically on scales $\sim 1000km$.

Earth where they are associated with jet formation [101] and play a key role in momentum transport and precipitation patterns. It is impossible for a complete description of the terrestrial climate to be made without understanding baroclinic eddies, for which the theory of baroclinic instability is a key component.

Baroclinic eddies are also observed in the mid- to high-latitude oceans [107], particularly in the vicinity of intense currents such as the Gulf Stream and the Arctic Circumpolar Current [98, 133]. Baroclinically unstable flows can be generated in rotating annulus experiments [6, 50, 51]. Such experiments provide a simple analogue to the large-scale circulation in the Earth's atmosphere or oceans and re-create many observable features such as jet streams and cyclone development [102]. Some authors have argued that baroclinic eddies are present in the atmosphere of other planets such as Jupiter [72], though this is not conclusive.

The initial aim of baroclinic instability theory was to determine and quantify the mechanism by which baroclinic waves grow. The first such studies were the classic linear models of Charney (1947) [17] and Eady (1949) [24], which examined the growth of an initially small disturbance and uncovered the basic mechanism of instability. The two-layer model of Phillips (1951,1954) [95, 96] provided another simple model of linear baroclinic instability. The linear behaviour of baroclinic waves is generally well-understood and a natural extension is to study nonlinear effects, in particular how baroclinic eddies force the background flow and bring about an equilibrated state or 'climate' [98]. An early example of a nonlinear model of the large scale atmospheric circulation was provided by Phillips [97] who studied a baroclinic jet in the two-layer model.

Simplified models of geophysical flows are of great use in the assessment of analytical theories. The most complex models of atmospheric circulation are computationally expensive [35] and tend to 'fit' physical parameterizations of small scale processes, which cannot be resolved, to observations [81]. A consequence of this is that if a physical theory is successful in predicting aspects of the climate, one cannot necessarily be sure it is accurate for the right reasons, i.e. that it is adequately capturing the governing physical processes at work. It is therefore instructive to look at simpler models where the aim is not so much to attain quantitative agreement with observations but to obtain a better qualitative understanding of the governing physics [146]. In this way, as eloquently argued by Held [45], a hierarchy of models of increasing complexity can be built up and reconciled with observational data to capture the essential characteristics of the global circulation [54].

Arguably the simplest model of a global climate is the two-layer, quasi-geostrophic (QG) model [95, 96], which has been used widely to study atmospheric and oceanic flows. A useful review of the literature was made by Held [46] who emphasises the role such 'toy' models have in testing new ideas. The model consists of two layers of inviscid, immiscible fluid, one lying above the other. The fluid is stably stratified so that the interface between the layers is analogous to the sloping isentropic surfaces in the troposphere which tilt upwards with latitude towards the poles. The two-layer model retains only the essential physics of the large-scale circulation in the atmosphere or oceans, which enables mathematical progress to be made for a relatively simple perspective of key dynamical processes.

Clearly, the two-layer QG model is limited in its ability to predict the real climate, for example it is not clear where the two layers would lie in the atmosphere, whether at the tropopause, or in the lower troposphere next to the atmospheric boundary layer. A significant drawback of the two-layer model (as with all quasi-geostrophic models) is that it assumes a fixed static stability as the potential temperature difference between the two layers is constant [114]. It is impossible to apply this model to determine the Earth's thermal stratification and the model cannot describe tropical flows for which convective processes are important. It should be kept in mind that the two-layer model is not intended to be a highly accurate model of any observed geophysical flows. However, it would be premature to discard the model for this reason. There are cases of phenomena being observed in the two-layer model, which have later been observed in the atmosphere [68].

Numerical models are used in a variety of ways to test analytical theories. One avenue of study is the numerical simulation of baroclinic lifecycles [119, 34, 132] in which one perturbs an initially unstable basic state and follows the linear growth of the perturbation to finite amplitude. The finite-amplitude disturbance modifies the basic flow and typically, in the absence of (external) forcing, nonlinear eddy forcing renders the flow stable to further instability. Baroclinic lifecycles play a useful role in addressing the question of how baroclinic eddies force the mean flow and have had success in describing observed states of mid-latitude dynamics [56].

In reality, baroclinic eddies do not proceed from an initially small disturbance of the kind used to initialize baroclinic lifecycles and are subject to dissipation and forcing. It is expected that non-conservative processes play an important role in determining the climate of the mid-latitude atmosphere [146]. A second area of research is the study of the statistical equilibrium of so-called forced-dissipative baroclinic systems (baroclinic turbulence). These studies allow an externally forced flow, in the presence of dissipative terms to relax to a statistically steady equilibrium [47, 89]. The aim is to assess the importance of non-conservative processes in determining the Earth's climate and especially dissipative effects in the planetary boundary layer [129, 108].

Both baroclinic lifecycle and forced-dissipative studies are used to assess theories of baroclinic equilibration. The object of an analytical theory is to understand and quantify forcing by baroclinic eddies and therefore predict the equilibrated state of the flow. One such simple predictive theory is known as baroclinic adjustment [124]. Baroclinic adjustment exploits the idea that baroclinic eddies have a strong feedback on the background flow and act on short enough timescales to rapidly bring the flow to a state that is neutral to further instability (or at least to a state that is only marginally unstable). The hypothesis of baroclinic adjustment theories is that the eddy transport is just sufficient to bring about the adjusted state. Notably however, there is no description for the route by which the neutral state is reached, nor is there provision for the eddy transport of heat and momentum observed in real flows [85]. Baroclinic adjustment is a global theory of equilibration since it relies on non-local information. For example, the eddy flux required to bring about a stable state of a given system requires knowledge of the width of the baroclinically unstable zone.

An example of a local predictive theory is provided by the parameterization of the diffusive effects of eddies from the properties of the local flow [46]. The parameterization of eddy fluxes by local mean quantities cannot be exact and so averaging over regions of eddy growth and decay is required for diffusive theories to be applicable. Theories of eddy diffusion rely on calculation of an eddy diffusivity (with units $length^2/time$) based on quantities such as the eddy length and time scales and estimates of the eddy velocity scale [123]. Diffusive theories have been most widely tested in models of homogeneous two-dimensional baroclinic turbulence [47, 65, 131] (two-layer models with doubly periodic boundary conditions) where there is a cascade of energy to larger scales. The eddy diffusivity is estimated from the rate of transfer of energy and the length scale at which energy flux is transferred to the zonal part of the flow (resulting in the reorganization of the flow into alternating zonal jets, see e.g. Sukoriansky et al. [126])

A further area of research has been the development of weakly nonlinear theories of baroclinic instability, for example, by Drazin (1970) [21] and Pedlosky (1970) [91]. These two works in particular stimulated a lot of research into weakly nonlinear theories [92, 93, 109, 134]. Not only are these theories attractive mathematically but, according to Pierrehumbert & Swanson [98] provide *'an island of certainty in a bewildering sea of nonlinear behaviors'*.

This work focuses on the two-layer Phillips model, which is used to investigate linear, weakly nonlinear and fully nonlinear methods to explain the growth and subsequent equilibration of a variety of flow configurations. The rest of this chapter is set out as follows: in section 1.2, for completeness, the equations of motion for the two-layer, quasi-geostrophic model are derived from the incompressible Euler equations. The two-layer model will be used to study a variety of flow configurations, which are introduced in section 1.3. The relevance of each of these configurations to real geophysical flows will also be discussed. In section 1.4 a numerical scheme to integrate the fully nonlinear equations is reviewed and finally section 1.5 provides an outline of the research carried out the remainder of this thesis.

1.2 Derivation of the two-layer, quasi-geostrophic equations

The derivation of the two-layer, quasi-geostrophic model is carried out with the atmosphere in mind, however the non-dimensionalized equations derived are also appropriate for flows in the ocean and annulus experiments.

1.2.1 Introduction and conventions

The physical situation to be modelled is that of two shallow layers of inviscid, incompressible and immiscible fluid, one lying above the other on the surface of a rotating sphere. At the bottom of the lower layer is a flat rigid surface, and at the top of the upper layer is a free surface (or rigid lid[2]). In the

---

[2] The free surface at the top of the upper layer will be shown to be equivalent to a rigid lid in the derivation which follows.

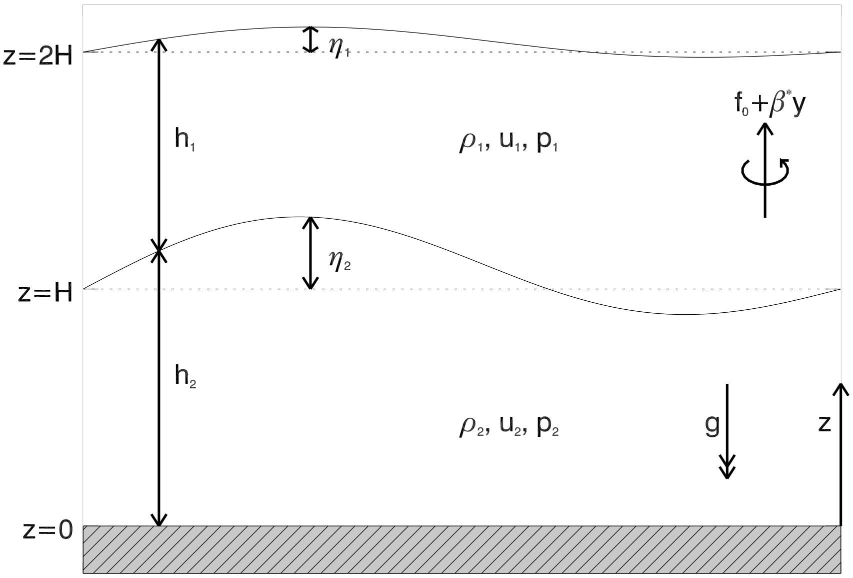

Figure 1.2.1: Schematic of the two-layer model with a flat rigid boundary at the bottom of the lower layer $z = 0$ and a free surface at the top of the upper layer $z = h_1 + h_2$. η_1 and η_2 are the perturbations of the upper and lower layers about the rest depths of the fluid.

frame of reference rotating with the sphere, using Cartesian coordinates, the (x, y)-plane is taken to be tangent to the surface of the sphere (x being the longitudinal and y the meridional/latitudinal direction). The positive y-direction corresponds to north when dealing with the northern hemisphere, which will be the case throughout this work. The z-direction (altitude) is perpendicular to the (x, y)- plane and the usual Cartesian vector convention is adopted such that $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ are the unit vectors in the $(x, y, z)-$directions. A schematic of the two layer model is shown in Fig. 1.2.1.

The following conventions are used for the upper (subscript $i = 1$) and lower (subscript $i = 2$) layers[3]

---

[3] Throughout this work a subscript 1 will refer to the upper layer and a subscript 2 to the lower layer. A subscript i will be used to denote both layers and this should not be confused with the imaginary number $i = \sqrt{-1}$.

- $\mathbf{u}_i(x, y, z, t) = (u_i, v_i, w_i)^T$ is the velocity field.
- $p_i(x, y, z, t)$ is the pressure field.
- $\mathbf{g}_{grav}$ is the gravitational acceleration directed towards the Earth's centre. Gravity varies as the inverse square of the distance to the centre of the Earth, where the radius of the Earth is $a \sim 6.4 \times 10^6 m$ (e.g. Vallis [133] p.56). Motions are over the depth of the troposphere $h \sim 8 - 16 \times 10^3 m$, so that the variation of gravity over the depth of the troposphere is very small and taking the 'shallow atmosphere approximation', g is constant.

At rest each layer has depth H and horizontal length scale L. The two layers are shallow in the sense that the aspect ratio α is small:

$$\alpha = \frac{H}{L} \ll 1. \tag{1.2.1}$$

The Earth rotates with angular velocity $2\boldsymbol{\Omega}$ about an axis through the poles. In the frame of reference rotating with the Earth, there are two apparent (or virtual) forces that appear to affect the motion of the fluid, the Coriolis and Centrifugal forces, which are $\boldsymbol{\Omega} \times \mathbf{u}$ and $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r})$ respectively.

At any point on the surface of the sphere, the centrifugal force can be written in terms of a scalar potential $\Phi_{ce} = \frac{1}{2}|\boldsymbol{\Omega}|^2|\mathbf{r}_\perp|^2$ where $\mathbf{r}_\perp$ is the perpendicular distance to the axis of rotation. An 'effective gravity' $\mathbf{g}$ may then be defined as the sum of the true gravity and the centrifugal force

$$\mathbf{g} = \mathbf{g}_{grav} + |\boldsymbol{\Omega}|^2\mathbf{r}_\perp.$$

Alternatively the effective gravity is written in terms of the geopotential Φ as

$$\mathbf{g} = -\nabla\Phi.$$

Since $\mathbf{r}_\perp$ and therefore the centrifugal force varies with latitude, the surfaces of constant geopotential Φ, like the Earth itself, are not truly spherical but ellipsoids. If spherical coordinates are used then centrifugal terms appear in both the horizontal and vertical momentum equations of the tangent plane. To avoid this, the vertical direction of the tangent plane is defined as the gradient of geopotential and then geometrically all constant geopotential surfaces are assumed to be spherical. Due to the oblateness of the Earth the effective gravity acts in a direction very close to normal to the surface and the geopotential is a function of height alone $\Phi = gz$ (using the scalar effective gravity g where $\mathbf{g} = g\mathbf{k}$).

The densities of the upper and lower layer are ρ_1 and ρ_2 respectively and the system is stably stratified ($\rho_1 < \rho_2$). The variation in density between the two layers is small so that

$$\Delta = \frac{(\rho_2 - \rho_1)}{\overline{\rho}} \ll 1 \qquad \text{where} \qquad \overline{\rho} = \frac{\rho_1 + \rho_2}{2}. \tag{1.2.2}$$

Under the Boussinesq approximation the difference in density is only appreciable under the action of gravity and a reduced gravity g' acts on the interface between the two layers:

$$g' = \Delta g. \tag{1.2.3}$$

Effectively, this means that the density contrast is concentrated at the interface between the two layers.

The presence of gravity allows the propagation of inertia-gravity waves in the fluid system. In the single layer shallow water model, inertia-gravity waves travel with phase speed $\sqrt{gH}$. In the two-layer model with the Boussinesq approximation the relevant scaling for the wave speed is $c = \sqrt{g'H}$ and the ratio of c to a typical flow speed U is known as the Froude number Fr. It is assumed that the speed of propagation of gravity waves is a lot greater than typical flow velocities and so the Froude number is small

$$Fr = \frac{\sqrt{2}U}{\sqrt{g'H}} \ll 1. \tag{1.2.4}$$

The extra factor $\sqrt{2}$ is included in this definition as it will simplify the equations of motion later in the derivation.

The component of the Earth's Coriolis parameter normal to the geopotential tangent plane is $f = 2\mathbf{\Omega} \cdot \mathbf{k} = 2|\mathbf{\Omega}| \sin\theta$ where θ is latitude. The variation of the Coriolis parameter with latitude is modelled by making the approximation $L \ll a$ (physically, horizontal length scales are smaller than the Earth's radius) and expanding f about a fixed latitude θ_0 so that

$$f = 2\Omega \sin\theta_0 + \frac{2\Omega}{a} y \cos\theta_0,$$

where $\Omega = |\mathbf{\Omega}$, or alternatively, writing $f_0 = 2\Omega \sin\theta_0$ and $\beta^* = (2\Omega/a)\cos\theta_0$,

$$f = f_0 + \beta^* y, \tag{1.2.5}$$

where the difference in the Coriolis force over a typical latitudinal distance L is a lot smaller than its total magnitude

$$\frac{\beta^* L}{f_0} \ll 1. \tag{1.2.6}$$

Equations (1.2.5-1.2.6) are known as the β-plane approximation first used by Rossby [112] in 1939. The β-plane captures the leading order effect of the sphericity of the Earth on the dynamics of the flow while allowing a simplification of the geometry.

The Rossby number Ro is now introduced as the ratio of the Coriolis to inertial forces

$$Ro = \frac{U}{f_0 L} \ll 1. \tag{1.2.7}$$

An alternative interpretation (e.g. Vallis [133] p.85-86) is to think of the Rossby number as the ratio of the period of rotation of the Earth to the advective timescale of a typical flow . Thus for phenomena that evolve over a number of days such as the development of mid-latitude storm systems and the motion of the Gulf Stream, the Rossby number is small.

The Boussinesq approximation (1.2.2), Froude number (1.2.4) and Rossby number (1.2.7) each define a small parameter and it is important to state the relative magnitudes of each as

$$\Delta \ll Ro, Fr \ll 1. \tag{1.2.8}$$

The limit of small Rossby and Froude number $Ro, Fr \to 0$ is taken such that $Ro/Fr = 1$. This ratio is, alternatively, the ratio of horizontal length scale L to Rossby deformation radius L_D, the latter defined as

$$L_D = \frac{\sqrt{g' H}}{\sqrt{2} f_0}. \tag{1.2.9}$$

The Rossby radius of deformation is the distance over which the tendency of the Coriolis force to deform the interface is balanced by the gravitational forces which flatten it. The Rossby radius of deformation arises naturally from this balance and is a fundamental length scale of the flow. The choice of

$$\frac{Ro}{Fr} = 1 \tag{1.2.10}$$

therefore corresponds to non-dimensionalizing horizontal length scales using the Rossby radius of deformation ($L_D = L$). This is justified on the grounds that the motions in which we are interested in an atmospheric or oceanic context typically have lengthscales of order L_D [4]..

1.2.2 Incompressible Euler equations

The incompressible Euler equations (e.g. Batchelor [7], section 3.6, Acheson [2], section 1.3) for an inviscid fluid of constant density, in a rotating frame of reference with effective gravity $\mathbf{g} = g\mathbf{k}$ consist of the momentum equations in each layer

$$\rho_i \left(\frac{D\mathbf{u}_i}{Dt} + \mathbf{f} \times \mathbf{u}_i \right) = -\nabla p_i - \rho_i g \mathbf{k}, \tag{1.2.11}$$

and incompressibility equations

$$\nabla \cdot \mathbf{u}_i = 0, \tag{1.2.12}$$

where $i = 1, 2$ denote the upper and lower layer respectively.

1.2.3 Boundary conditions at the interface, free surface and bottom of the channel

As illustrated in Fig. 1.2.1 there is a flat, rigid bottom at $z = 0$ and the depths of the upper and lower layers above this are $h_1(x, y, t)$ and $h_2(x, y, t)$ respectively. At the top of the lower layer $z = h_2$ is the interface between the layers and at the top of the upper layer $z = h_1 + h_2$ there is a free surface. There is no flow across the solid bottom $z = 0$ and so the kinematic boundary

---

[4] An alternative non-dimensionalization (e.g. Pedlosky [94]) is to scale horizontal length scales on the channel width in which case the ratio Ro/Fr is a free parameter of the system. Under our choice of non-dimensionalization the channel width W is a parameter of the system instead.

condition here is

$$w_2 = 0 \qquad \text{on} \qquad z = 0. \tag{1.2.13}$$

At $z = h_1 + h_2$ the vertical velocity matches that of the free surface:

$$w_1 = \frac{D}{Dt}(h_1 + h_2) = \frac{D}{Dt}(\eta_1 + \eta_2) \qquad \text{on} \qquad z = h_1 + h_2 \tag{1.2.14}$$

and at $z = h_2$ the vertical velocity matches that of the interface

$$w_1 = w_2 = \frac{Dh_2}{Dt} = \frac{D\eta_2}{Dt} \qquad \text{on} \qquad z = h_2. \tag{1.2.15}$$

For inviscid flow there are no conditions on the horizontal velocity at the bottom of the channel and free surface at the top of the upper layer. When Ekman layers are introduced in section 1.2.8 no-slip boundary conditions at the top and bottom of the channel are used to calculate the effect of Ekman layers on the interior flow.

The dynamic boundary condition at the free surface of an inviscid fluid is simply that pressure is continuous

$$p_1 = p_0 \qquad \text{on} \qquad z = h_1 + h_2, \tag{1.2.16}$$

where p_0 is the atmospheric pressure at $z = h_1 + h_2$ due to the overlying fluid and is assumed constant. At the interface $z = h_2$ the pressure must also be continuous

$$p_1 = p_2 \qquad \text{on} \qquad z = h_2. \tag{1.2.17}$$

1.2.4 Scaling

Continuing with the convention of taking U as a typical horizontal velocity, H as the undisturbed depth of each layer and the deformation radius L_D as a typical horizontal length scale, dimensionless variables are introduced as follows

$$\begin{aligned}
\mathbf{u}_i = (u_i, v_i)^T &= U(\hat{u}_i, \hat{v}_i)^T, & t &= \tfrac{L_D}{U}\hat{t}, \\
(x, y) &= L_D(\hat{x}, \hat{y}), & z &= H\hat{z}, \\
h_i &= H\hat{h}_i, & \eta_i &= d_i H\hat{\eta}_i
\end{aligned} \tag{1.2.18}$$

where a hat denotes a dimensionless variable. Note that the non-dimensional velocity $\hat{\mathbf{u}}_i$ vector is in the horizontal direction only. The dimensionless vertical velocity is given below. There are also non-dimensional scalars d_1 and d_2, which denote the order of magnitude of the displacements η_1 and η_2 respectively. These are also to be determined below.

The scalings (1.2.18) fix the thicknesses of each layer at rest to be equal ($= H$). This choice is motivated by the observation that little interesting dynamics emerges by allowing the layer depths to be different (see e.g. Esler [27]). This aspect of the model is therefore omitted for the sake of brevity and simplification of the equations of motion.

The scaling for the pressure is $P \sim \overline{\rho} U f_0 L_D$ and comes from geostrophic balance in the leading order horizontal momentum equations as will be described below.

$$p = P\hat{p} = \overline{\rho} f_0 U L_D \hat{p}. \tag{1.2.19}$$

The scaling for the vertical velocity comes from the incompressibility equation (1.2.12)

$$w = W\hat{w} = \frac{H}{L_D} U\hat{w} = \alpha U\hat{w}. \tag{1.2.20}$$

The scalings (1.2.18-1.2.20) are substituted into (1.2.12) and (1.2.11) to give the non-dimensional incompressibility equation

$$\nabla \cdot \hat{\mathbf{u}}_i = -\frac{\partial \hat{w}_i}{\partial \hat{z}}, \tag{1.2.21}$$

and non-dimensionalized momentum equations

$$\begin{aligned}
\left(1+(-1)^i\frac{\Delta}{2}\right)\left(Ro\frac{D\hat{\mathbf{u}}_i}{D\hat{t}} + (1+Ro\beta\hat{y})\,\mathbf{k}\times\hat{\mathbf{u}}_i\right) &= -\nabla\hat{p}_i, && (1.2.22)\\
\left(1+(-1)^i\frac{\Delta}{2}\right)\left(\frac{\Delta Fr^2\alpha^2}{2}\right)\frac{D\hat{w}_i}{D\hat{t}} &= -\frac{\Delta Fr^2}{2Ro}\frac{\partial\hat{p}_i}{\partial\hat{z}} - \left(1+(-1)^i\frac{\Delta}{2}\right). && (1.2.23)
\end{aligned}$$

The gradient operator ∇ from this point on is defined 'on the horizontal plane' so that $\nabla = (\partial/\partial\hat{x}, \partial/\partial\hat{y}, 0)^T$. The value β is an $O(1)$ non-dimensional parameter called the inverse criticality given by

$$\beta = \frac{\beta^* L_D/f_0}{U/f_0 L_D} = \frac{\beta^* L_D^2}{U}, \tag{1.2.24}$$

a ratio of the two small parameters (1.2.6) and (1.2.7). The parameter β is a dimensionless measure of the relative and planetary vorticity gradients. It has been named inverse criticality following Esler [29, 28] where instability occurs for $\beta < 0.5$. The inverse criticality is a measure of the degree to which the flow is unstable (inverse in the sense that β decreases away from the stability threshold). Throughout this work, β fulfils the same role (though the stability threshold is not $\beta = 0.5$ in every flow considered) and as such the name 'inverse criticality' is retained for continuity with previous works [31, 138].

1.2.5 Boussinesq and hydrostatic approximations

Fig. 1.2.1 shows the depth of each layer written in terms of a constant rest depth and displacements about this, $\hat{\eta}_1(x,y,t)$ and $\hat{\eta}_2(x,y,t)$ such that

$$\begin{aligned} \hat{h}_1 &= 1 + d_1\hat{\eta}_1 - d_2\hat{\eta}_2, \\ \hat{h}_2 &= 1 + d_2\hat{\eta}_2. \end{aligned} \tag{1.2.25}$$

Considering the vertical momentum equation (1.2.23), the inertia term on the left hand side can be neglected due to the shallow water approximation (1.2.1) and the leading order balance is between the vertical pressure gradient and gravity

$$\begin{gathered} 0 = -\left(\Delta Ro\right)\frac{\partial \hat{p}_1}{\partial \hat{z}} - \left(2 - \Delta\right), \\ 0 = -\left(\Delta Ro\right)\frac{\partial \hat{p}_2}{\partial \hat{z}} - \left(2 + \Delta\right), \end{gathered} \tag{1.2.26}$$

where the ratio $Ro/Fr = 1$ has been invoked. The small parameter Δ is left in equations (1.2.26) as it will later reveal information about the magnitudes of the layer displacements d_1 and d_2.

Equations (1.2.26) are integrated with respect to z and use of the dynamic boundary conditions at the free surface (1.2.16) and interface (1.2.17) gives the hydrostatic balance equations

$$\begin{aligned} \left(\Delta Ro\right)\hat{p}_1 &= \left(\Delta Ro\right)\hat{p}_0 + \left(2 - \Delta\right)\left(\hat{h}_1 + \hat{h}_2 - \hat{z}\right), \\ \left(\Delta Ro\right)\hat{p}_2 &= \left(\Delta Ro\right)\hat{p}_0 + 2\left(\hat{h}_1 + \hat{h}_2 - \hat{z}\right) + \Delta\left(\hat{h}_2 - \hat{z}\right). \end{aligned} \tag{1.2.27}$$

The upper layer horizontal pressure gradient is then expressed in terms of the depth perturbations $\hat{\eta}_1$ and $\hat{\eta}_2$ by substitution of (1.2.25) into (1.2.27),

$$\left(\Delta Ro\right)\nabla\hat{p}_1 = \left(2d_1\right)\nabla\hat{\eta}_1. \tag{1.2.28}$$

For these to balance, $d_1 = \Delta Ro$, i.e. the perturbation in the upper layer depth $\hat{\eta}_1$ is $O(\Delta)$. The gradient of the lower layer pressure is

$$(\Delta Ro)\,\nabla\hat{p}_2 = (2\Delta Ro)\,\nabla\hat{\eta}_1 + (2\Delta d_2)\,\nabla\hat{\eta}_2. \tag{1.2.29}$$

The balance in equation (1.2.29) indicates that $d_2 = Ro$, i.e. the interface displacement η_2 is $O(1)$, or $O(\Delta^{-1})$ larger than the free surface.

The difference in the velocity between layers depends on the sloping interface where the density gradient is effectively concentrated. The upper layer free surface displacement η_1 is $O(\Delta)$ due to the $O(1)$ density jump between the upper layer and the effectively zero density region above it. In comparison, the interface displacement η_2 is $O(1)$ as the density jump between the two layers is $O(\Delta)$. The motion of the free surface η_1 is negligible under the Boussinesq approximation compared to the motion of the interface η_2 and the upper surface therefore has the same boundary conditions as a rigid lid.

1.2.6 Geostrophic balance

To proceed, the non-dimensionalized physical fields may be expanded in powers of the Rossby number

$$(\hat{u}_i, \hat{v}_i, \hat{p}_i) = (u_i^{(0)}, v_i^{(0)}, p_i^{(0)}) + Ro(u_i^{(1)}, v_i^{(1)}, p_i^{(1)}) + \dots \tag{1.2.30}$$

When Δ is neglected, the $O(1)$ horizontal momentum equations (1.2.22) are

$$u_i^{(0)} = -\frac{\partial p_i^{(0)}}{\partial y}, \qquad v_i^{(0)} = \frac{\partial p_i^{(0)}}{\partial x}, \tag{1.2.31}$$

or $\mathbf{u}_i^{(0)} = -\mathbf{k}\times\nabla\psi_i^{(0)}$. The leading order balance is between the Coriolis force and horizontal pressure gradients. It is from this geostrophic balance that the scaling for the pressure ($P \sim \overline{\rho} f_0 U L_D$ in (1.2.19)) is derived. The geostrophic velocity can be represented by the geostrophic streamfunctions $\psi_i^{(0)} = p_i^{(0)}$, so that

$$u_i^{(0)} = -\frac{\partial \psi_i^{(0)}}{\partial y}, \qquad v_i^{(0)} = \frac{\partial \psi_i^{(0)}}{\partial x}. \tag{1.2.32}$$

The geostrophic streamfunctions $\psi_i^{(0)} = p_i^{(0)}$ may be re-written using equations (1.2.28-1.2.29) in terms of the displacement of the interface and free surface

$$\psi_1^{(0)} = 2\hat{\eta}_1, \qquad \psi_2^{(0)} = 2\left(\hat{\eta}_1 + \hat{\eta}_2\right). \tag{1.2.33}$$

The leading order geostrophic balance also means that, to leading order, the horizontal divergence vanishes

$$\nabla \cdot \mathbf{u}_i^{(0)} = 0.$$

Upon substitution into the $O(1)$ incompressibility equation (1.2.21) this shows that the leading order vertical velocity in each layer $w_i^{(0)}$ is independent of height

$$\frac{\partial w_i^{(0)}}{\partial z} = 0. \tag{1.2.34}$$

In fact, $w_i^{(0)} = 0$ throughout each layer due to the boundary condition at the flat bottom ($w_2^{(0)} = 0$) and at the interface ($w_1^{(0)} = w_2^{(0)}$). The kinematic boundary conditions at the free surface (1.2.14) and interface (1.2.15) are then $O(Ro)$, given by

$$\begin{aligned} w_1^{(1)}(z = h_1 + h_2) &= \Delta \frac{D\hat{\eta}_1}{Dt} = 0, & (1.2.35)\\ w_2^{(1)}(z = h_1) &= \frac{D\hat{\eta}_2}{Dt}. & (1.2.36) \end{aligned}$$

The Boussinesq approximation $\Delta << 1$ has been used in (1.2.35).

1.2.7 Quasi-geostrophic equations

Now proceed to the $O(Ro)$ equations to determine the non-geostrophic motion of the flow. At $O(Ro)$ the horizontal momentum equations (1.2.22) are

$$\frac{D_i \mathbf{u}_i^{(0)}}{Dt} + \mathbf{k} \times \mathbf{u}_i^{(1)} + \beta y \mathbf{k} \times \mathbf{u}_i^{(0)} = -\nabla p_i^{(1)}, \tag{1.2.37}$$

where the material derivative is

$$\frac{D_i}{Dt} \equiv \frac{\partial}{\partial t} - \frac{\partial \psi_i^{(0)}}{\partial y}\frac{\partial}{\partial x} + \frac{\partial \psi_i^{(0)}}{\partial x}\frac{\partial}{\partial y}.$$

Taking the curl of (1.2.37) gives the vertical component $\zeta_i^{(0)} = \nabla \times \mathbf{u}_i^{(0)} = \nabla^2 \psi_i^{(0)}$ of the lowest order vorticity equation

$$\frac{D_i}{Dt}\left(\zeta_i^{(0)} + \beta y\right) = -\nabla \cdot \mathbf{u}_i^{(1)}. \tag{1.2.38}$$

It has been shown through (1.2.32) that the geostrophic velocity $\mathbf{u}_i^{(0)}$ and therefore geostrophic vorticity $\zeta_i^{(0)}$ are independent of z. Equation (1.2.38) may therefore be integrated with respect to z

$$\begin{aligned} \frac{D_1}{Dt}\left(\zeta_1^{(0)} + \beta y\right) &= -\int_{h_2}^{h_1+h_2} \nabla \cdot \mathbf{u}_1^{(1)} dz, \\ \frac{D_2}{Dt}\left(\zeta_2^{(0)} + \beta y\right) &= -\int_{0}^{h_2} \nabla \cdot \mathbf{u}_2^{(1)} dz. \end{aligned} \tag{1.2.39}$$

The $O(Ro)$ mass equations (1.2.21) are also integrated with respect to z, which give

$$\begin{aligned} \int_{h_2}^{h_1+h_2} \nabla \cdot \mathbf{u}_1^{(1)} dz &= -\left[w_1^{(1)}\right]_{h_2}^{h_1+h_2} = \frac{D_1 \hat{\eta}_2}{Dt}, \\ \int_{0}^{h_2} \nabla \cdot \mathbf{u}_2^{(1)} dz &= -\left[w_2^{(1)}\right]_{0}^{h_2} = -\frac{D_2 \hat{\eta}_2}{Dt}. \end{aligned} \tag{1.2.40}$$

Eliminating the right-hand side of (1.2.38) using (1.2.39-1.2.40) leads to

$$\frac{D_i}{Dt}\left(\zeta_i^{(0)} + \beta y - (-1)^i \hat{\eta}_2\right) = 0.$$

Alternatively, writing this in terms of the geostrophic streamfunctions $\psi_i^{(0)}$, $\zeta_i^{(0)} = \nabla^2\psi_i^{(0)}$ and $\hat{\eta}_2 = \frac{1}{2}(\psi_2^{(0)} - \psi_1^{(0)})$ gives the two-layer, quasi-geostrophic equation:

$$\begin{aligned} \frac{D_i q_i}{Dt} &= 0 && (1.2.41) \\ q_i &= \beta y + \nabla^2\psi_i + (-1)^i\left(\frac{\psi_1 - \psi_2}{2}\right), && (1.2.42) \end{aligned}$$

where the scalar field q_i is known as the potential vorticity and the superscript (0) has been dropped from the geostrophic streamfunctions.

Equation (1.2.41) governs the advection of potential vorticity in each layer. At any instant in time the PV field may be inverted using (1.2.42) together with suitable boundary conditions in the x- and y-directions to calculate the streamfunction ψ_i, from which physical fields such as pressure and velocity can be found. Crucially, one needs only to consider the evolution of this single scalar field from which all other physical quantities can be calculated via PV inversion.

The two-layer shallow water potential vorticity (1.2.41) is a specific example of PV conservation, which is used more widely in geophysical flows. The conservation of PV was first derived for the shallow water equations by Rossby (1936) [110] and later for continuously stratified models by Rossby and Ertel [111, 113, 26]. Today the quantity of use in the atmospheric sciences is known as the Rossby-Ertel potential vorticity and a comprehensive review of its use is made by Hoskins et al. [55].

1.2.8 Ekman friction

At the top and bottom of the domain are flat, rigid boundaries near which the effects of friction are important and the dynamics of the flow may change rapidly from the interior within a thin boundary layer. In small scale fluid flows, the viscous effects in the boundary layer may be due to molecular viscosity on length scales less than a millimetre thick. However in geophysical flows boundary layers owe their existence to the effects of small scale turbulence, for example due to the effect of the Earth's surface in the atmosphere (of the order of $10^2 - 10^3 m$ thick) or wind stress at the ocean's surface (of the order $10^1 m$ thick, Vallis [133]).

Within these boundary layers, known as Ekman layers [25], the dominant balance is between the Coriolis force, horizontal pressure gradients and frictional terms. It is impossible to precisely model the turbulent Ekman layers and so a degree of parameterization is needed. The aim is to deduce the general properties of the Ekman layer irrespective of the type of friction present. To achieve this, the frictional stress exerted by the Ekman layer is parameterized by an eddy diffusion, which is then balanced by the Coriolis force. The Ekman layer at the bottom of the lower layer will be considered in the derivation below, but the treatment of a rigid boundary at the top of the domain is similar [5].

The flow within the Ekman layer satisfies a no-slip condition on the boundary and is matched to the interior geostrophic flow $\mathbf{u}_g$ as $z \to \infty$ [6] These

---

[5] In fact, Ekman layers can also form with a stress-free surface (see Hide [49])

[6] Here, for simplicity, $\mathbf{u}_g = U\mathbf{u}_i^{(0)}$ is the dimensional geostrophic velocity (recall $\mathbf{u}_i^{(0)} =$

boundary conditions on the flow in the Ekman layer $\mathbf{u}_e$ may be written in the new z coordinate as

$$\begin{aligned} \mathbf{u}_e &= \mathbf{0} \qquad \text{on} \qquad z = 0, \\ \mathbf{u}_e &= \mathbf{u}_g \qquad \text{as} \qquad z \to \infty. \end{aligned} \tag{1.2.43}$$

The leading order balance of the non-geostrophic flow is between the Coriolis force, horizontal pressure gradients and frictional terms, which may be written as [7] (Vallis [133], p.109)

$$\mathbf{f}_0 \times (\mathbf{u}_e - \mathbf{u}_g) = A \frac{\partial^2 \mathbf{u}_e}{\partial z^2}, \tag{1.2.44}$$

where the frictional stress has been modelled as an eddy viscosity with kinematic diffusivity A. Note that only the vertical derivatives are included as over the thin boundary layer, vertical derivatives are larger than the horizontal derivatives due to the small aspect ratio. Equation (1.2.44) may be non-dimensionalized using (1.2.18-1.2.20) to give

$$\mathbf{k} \times \left(\hat{\mathbf{u}}_e - \mathbf{u}_2^{(0)} \right) = \frac{E_k}{2} \frac{\partial^2 \hat{\mathbf{u}}_e}{\partial \hat{z}^2}, \tag{1.2.45}$$

where the Ekman number E_k is given by

$$E_k = \frac{2A}{f_0 H^2}. \tag{1.2.46}$$

The Ekman number indicates the importance of the frictional terms in the flow. Alternatively, one may define the length scale

$$\delta_e = \left(\frac{2A}{f_0} \right)^{1/2}, \tag{1.2.47}$$

$-\mathbf{k} \times \nabla \psi_i^{(0)}$).

[7] Horizontal pressure gradients are implicitly included in equation (1.2.44) with the presence of the geostrophic wind term $\mathbf{u}_g$.

which is an order of magnitude estimate of the thickness of the boundary layer. The Ekman number (1.2.46) is then a ratio of the boundary layer thickness to the lower layer thickness so that $E_k \ll 1$.

To proceed, note that by (1.2.34) the geostrophic velocity is independent of $\hat{z}$ and and so equation (1.2.45) may be solved straightforwardly to give

$$\begin{aligned} \hat{u}_e &= u_2^{(0)} - e^{-\hat{z}/\sqrt{E_k}} \left(u_2^{(0)} \cos\left(\frac{\hat{z}}{\sqrt{E_k}}\right) + v_2^{(0)} \sin\left(\frac{\hat{z}}{\sqrt{E_k}}\right)\right) \\ \hat{v}_e &= v_2^{(0)} + e^{-\hat{z}/\sqrt{E_k}} \left(u_2^{(0)} \sin\left(\frac{\hat{z}}{\sqrt{E_k}}\right) - v_2^{(0)} \cos\left(\frac{\hat{z}}{\sqrt{E_k}}\right)\right) \quad (1.2.48) \end{aligned}$$

Equations (1.2.48) are integrated with respect to $\hat{z}$ to find the mass transport within the Ekman layer

$$\hat{\mathbf{M}}_e = \int_0^\infty (\hat{\mathbf{u}}_e - \mathbf{u}_2^{(0)}) d\hat{z} = \frac{\sqrt{E_k}}{2} \begin{pmatrix} -u_2^{(0)} - v_2^{(0)} \\ u_2^{(0)} - v_2^{(0)} \end{pmatrix}. \quad (1.2.49)$$

The incompressibility equation (1.2.12) is then integrated with respect to $\hat{z}$, recalling that the geostrophic velocity is independent of $\hat{z}$

$$\begin{aligned} [\hat{w}_e]_0^\infty &= -\int_0^\infty \nabla \cdot \hat{\mathbf{u}}_e, \\ \hat{w}_e|_\infty &= -\nabla \cdot \hat{\mathbf{M}}_e, \\ &= \frac{\sqrt{E_k}}{2} \zeta_2^{(0)}. \quad (1.2.50) \end{aligned}$$

Therefore, the Ekman layer acts as a pump into the interior fluid and the kinematic boundary condition on the rigid bottom (1.2.13) may be replaced with (1.2.50) written in terms of the geostrophic streamfunction

$$w_2^{(1)} = \kappa_2 \nabla^2 \psi_2^{(0)} \qquad \text{on} \qquad \hat{z} = 0, \quad (1.2.51)$$

where

$$\kappa_2 = \frac{\sqrt{E_k}}{2Ro} = \frac{A^{1/2} f_0^{1/2} L_D}{2UH} \quad (1.2.52)$$

is a non-dimensional parameter measuring the strength of the Ekman pumping. Throughout the rest of this work, κ_i will be referred to as the quasi-geostrophic (QG) Ekman number in layer i. The boundary condition (1.2.51) is on the $O(Ro)$ vertical velocity $w_1^{(1)}$ since, by equation (1.2.34), the leading order vertical velocity $w_2^{(0)}$ is zero.

If an Ekman layer is also present at the top of the domain, the boundary condition at the top of the upper layer is

$$w_1^{(1)} = -\kappa_1 \nabla^2 \psi_1^{(0)} \qquad \text{on} \qquad \hat{z} = 2. \tag{1.2.53}$$

The new boundary conditions (1.2.51) and (1.2.53) are then used in (1.2.39-1.2.40) to give the two-layer quasi-geostrophic equations in the presence of Ekman friction at the top and bottom of the domain.

$$\begin{aligned} \frac{D_i q_i}{Dt} &= -\kappa_i \nabla^2 \psi_i, \\ q_i &= \nabla^2 \psi_i + \beta y + (-1)^i \left(\frac{\psi_1 - \psi_2}{2} \right). \end{aligned} \tag{1.2.54}$$

These equations replace (1.2.41-1.2.42) when Ekman friction is to be included in the model.

1.3 Two model configurations

The two-layer, quasi-geostrophic equations (1.2.54) will be used in two different flow configurations, which are now described.

1.3.1 The two-layer Phillips model: uniform flow

The two-layer Phillips model [95, 96] consists of flow between two rigid sidewalls in the latitudinal direction and a re-circulating channel in the longitudinal direction. The Phillips model has a uniform flow in each layer and is useful for a variety of reasons:

- The Phillips model has great pedagogical value in the description of baroclinic instability, as two layers of immiscible fluid is arguably the simplest way to include vertical structure in a model. Therefore, the Phillips model could be described as a 'foundation stone' in the hierarchy of models suggested by I.M Held [45].
- As a benchmark experiment, the Phillips model brings together investigations over a wide range of the atmospheric, oceanic and theoretical fluid dynamics literature. The model therefore provides a point of reference in which ideas from each of these fields may be compared.
- The mechanism by which baroclinic waves grow and equilibrate may be fruitfully studied and precise mathematical solutions obtained due to the simplicity of the model.
- The re-circulating channel can be used as a simple model of a wind driven oceanic flow such as the Antarctic Circumpolar Current.
- Results from the two-layer model with rigid upper and lower boundaries may be expected to be relevant to mechanically driven, two-layer annulus experiments of the type carried out by Hart [38, 39], King [61] and

Bradford et al. [13], or more recently, Ohlsen et al. [87], Lovegrove et al. [74] and Williams et al. [141, 140].

The flow is in a recirculating channel of non-dimensional length L and width W. The channel is periodic in the x-direction with $x \in [0, L]$ such that

$$\psi_i(x, y, t) = \psi_i(x + L, y, t). \tag{1.3.55}$$

for $i = 1, 2$. There are rigid sidewalls at $y = \pm W/2$ and there is no flow perpendicular to the sidewalls. The first boundary conditions on the channel walls are therefore

$$v_i = \frac{\partial \psi_i}{\partial x} = 0 \qquad \text{on} \qquad y = \pm\frac{W}{2}, \tag{1.3.56}$$

for $i = 1, 2$.

To derive a second boundary condition, first used by Phillips (1954) [96], consider the x-momentum equation (1.2.22) on $y = \pm W/2$:

$$-v_i^{(0)} + \frac{\partial p_i^{(0)}}{\partial x} + Ro\left(\frac{\partial u_i^{(0)}}{\partial t} + U_i\frac{\partial u_i^{(0)}}{\partial x} + u_i^{(0)}\frac{\partial u_i^{(0)}}{\partial x} + v_i^{(0)}\frac{\partial u_i^{(0)}}{\partial y} + \beta y v_i^{(0)} + \frac{\partial p_i^{(1)}}{\partial x}\right) = -Ro\kappa_i u_i^{(0)}$$

Upon use of the first boundary condition (1.3.56) and rearrangement this becomes

$$Ro\left(\frac{\partial u_i^{(0)}}{\partial t} + \kappa_i u_i^{(0)}\right) + RoU_i\frac{\partial u_i^{(0)}}{\partial x} + \frac{\partial}{\partial x}\left(p_i^{(0)} + Ro\left[\frac{1}{2}(u_i^{(0)})^2 + p_i^{(1)}\right]\right) = 0,$$

When integrated over the length of the channel the x-derivatives vanish and the result is the Phillips boundary condition

$$\frac{\partial \overline{u}_i^{(0)}}{\partial t} + \kappa_i \overline{u}_i^{(0)} = 0 \qquad \text{on} \qquad y = \pm\frac{W}{2}. \tag{1.3.57}$$

where, throughout this work, the over-line denotes a zonal average, i.e. for a scalar field $f(x,y,t)$

$$\overline{f}(y,t) = \frac{1}{L}\int_0^L f(x,y,t)dx. \qquad (1.3.58)$$

Equation (1.3.57) may be written in terms of the geostrophic streamfunction ψ_i as

$$\frac{\partial^2\overline{\psi}_i}{\partial t\partial y} + \kappa_i\frac{\partial\overline{\psi}_i}{\partial y} = 0 \qquad \text{on} \qquad y = \pm\frac{W}{2}. \qquad (1.3.59)$$

All the flows to be considered are written in terms of a prescribed zonal background flow whose potential vorticity and streamfunction are written as $Q_i(y)$ and $\Psi_i(y)$ respectively and small perturbations about it $q_i'(x,y,t)$ and $\psi_i'(x,y,t)$,

$$\psi_i = \Psi_i(y) + \psi_i'(x,y,t), \qquad q_i = Q_i(y) + q_i'(x,y,t), \qquad (1.3.60)$$

for $i = 1,2$ and

$$q_i' = \nabla^2\psi_i' + (-1)^i\left(\frac{\psi_1' - \psi_2'}{2}\right). \qquad (1.3.61)$$

For the Phillips model, the uniform flow profile is independent of y and the upper and lower layer background velocities, U_1 and U_2 respectively, are set to be

$$U_1 = -\frac{\partial\Psi_1}{\partial y} = 1, \qquad U_2 = -\frac{\partial\Psi_2}{\partial y} = 0. \qquad (1.3.62)$$

The background potential vorticities are then

$$Q_1(y) = \left(\beta + \frac{1}{2}\right)y, \qquad Q_2(y) = \left(\beta - \frac{1}{2}\right)y. \qquad (1.3.63)$$

Substituting (1.3.60), (1.3.62) and (1.3.63) into the equations of motion (1.2.54) gives

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\left(\nabla^2\psi_1'-\frac{\psi_1'-\psi_2'}{2}\right)+\left(\beta+\frac{1}{2}\right)\frac{\partial\psi_1'}{\partial x}+J\left(\psi_1',q_1'\right) = -\kappa_1\nabla^2\psi_1', \tag{1.3.64}$$

$$\left(\frac{\partial}{\partial t}\right)\left(\nabla^2\psi_2'+\frac{\psi_1'-\psi_2'}{2}\right)+\left(\beta-\frac{1}{2}\right)\frac{\partial\psi_2'}{\partial x}+J\left(\psi_2',q_2'\right) = -\kappa_2\nabla^2\psi_2', \tag{1.3.65}$$

where $J(f,g)=\frac{\partial f}{\partial x}\frac{\partial g}{\partial y}-\frac{\partial f}{\partial y}\frac{\partial g}{\partial x}$ is the Jacobian operator. The appropriate boundary conditions are (1.3.55), (1.3.56) and (1.3.59), which are

$$\psi_i'(x,y,t) = \psi_i'(x+L,y,t). \tag{1.3.66}$$

$$\frac{\partial\psi_i'}{\partial x} = 0 \qquad \text{on} \qquad y=\pm\frac{W}{2}, \tag{1.3.67}$$

$$\frac{\partial^2\overline{\psi}_i'}{\partial t\partial y}+\kappa_i\frac{\partial\overline{\psi}_i'}{\partial y} = 0 \qquad \text{on} \qquad y=\pm\frac{W}{2}. \tag{1.3.68}$$

By fixing the background velocities the vertical shear is controlled entirely by the inverse criticality β. The system can be uniquely defined by the parameter set $(\beta,\kappa_1,\kappa_2)$, the dimensions of the channel (L,W) and a suitably defined set of initial conditions.

1.3.2 The two-layer β-channel model with an isolated jet

The uniform flow profile in the upper layer of section 1.3.1 is replaced with a single jet similar to the first ‘climate model’ formulated by Phillips (1956) [97]. The isolated jet profile is valuable for a number of reasons:

- The jet undergoes a baroclinic lifecycle which provides an idealised problem of quasi-geostrophic turbulence. The mechanism by which eddies act to turbulently equilibrate the flow may be studied.

- The jet profile simulates the mean state of a single jet stream in the extra-tropical troposphere of the Earth [45, 145] provided that the channel is sufficiently wide to ignore side-wall effects.

- The jet profile is also appropriate for oceanic circulations such as the Gulf stream or the Antarctic Circumpolar Current [3, 79, 103]. It is more relevant than the uniform flow profile of section 1.3.1.

- Jet systems are also observed in rotating annulus experiments [102] where the two-layer model with a background jet profile may provide useful insights into the resulting flow behaviour.

Following e.g. Pavan and Held [89] the jet is defined as having a background velocity in each layer given by

$$U_1(y) = \mathrm{sech}^2\left(\frac{y}{\sigma}\right), \qquad U_2(y) = 0, \tag{1.3.69}$$

where σ is the half-width of the jet. For this flow profile, no Ekman friction will be considered and so the system is entirely determined by the two parameters (β, σ), the dimensions of the channel (L, W) and the initial conditions. The background potential vorticity for this flow is calculated from (1.3.69) as

$$\begin{aligned} Q_1(y) &= \beta y + \left(\frac{2}{\sigma}\mathrm{sech}^2\left(\frac{y}{\sigma}\right) + \frac{\sigma}{2}\right)\tanh\left(\frac{y}{\sigma}\right), \\ Q_2(y) &= \beta y - \frac{\sigma}{2}\tanh\left(\frac{y}{\sigma}\right), \end{aligned} \tag{1.3.70}$$

and the potential vorticity gradients are

$$\begin{aligned}\frac{dQ_1}{dy} &= \beta + \left(\frac{\sigma^2 - 8}{2\sigma^2}\right)\operatorname{sech}^2\left(\frac{y}{\sigma}\right) + \frac{6}{\sigma^2}\operatorname{sech}^4\left(\frac{y}{\sigma}\right), \\ \frac{dQ_2}{dy} &= \beta - \frac{1}{2}\operatorname{sech}^2\left(\frac{y}{\sigma}\right). \qquad (1.3.71)\end{aligned}$$

The background fields (1.3.69-1.3.71) are used in (1.2.41-1.2.42) to give the equations of motion

$$\begin{aligned}\left(\frac{\partial}{\partial t} + U_1\frac{\partial}{\partial x}\right)q_1' + \frac{dQ_1}{dy}\frac{\partial\psi_1'}{\partial x} + J\left(\psi_1', q_1'\right) &= 0, \qquad (1.3.72) \\ \left(\frac{\partial}{\partial t} + U_2\frac{\partial}{\partial x}\right)q_2' + \frac{dQ_2}{dy}\frac{\partial\psi_2'}{\partial x} + J\left(\psi_2', q_2'\right) &= 0, \qquad (1.3.73) \\ & \qquad (1.3.74)\end{aligned}$$

with the boundary conditions (1.3.55), (1.3.56) and (1.3.59)

$$\begin{aligned}\psi_i(x,y,t) &= \psi_i(x+L,y,t), \qquad (1.3.75) \\ \frac{\partial\psi_i}{\partial x} &= 0 \quad \text{on} \quad y = \pm\frac{W}{2}, \qquad (1.3.76) \\ \frac{\partial\overline{\psi}_i}{\partial y} &= 0 \quad \text{on} \quad y = \pm\frac{W}{2}, \qquad (1.3.77)\end{aligned}$$

for $i = 1, 2$.

1.3.3 Physical invariants for frictionless flow

The Phillips model in the absence of Ekman friction has a number of physical invariants. First of all total momentum M is conserved throughout the evolution of the flow,

$$M = M_0 = \int_D u_1 + u_2 d^2\mathbf{x}, \qquad (1.3.78)$$

where integration is over the channel $D = \{x \in [0, L], y \in [-W/2, W/2]\}$ and M_0 is the initial momentum. Closely related to momentum is another conserved quantity, the x-component of Kelvin's impulse given by

$$M = \int_D y\,(q_1 + q_2)\,d^2\mathbf{x}. \tag{1.3.79}$$

Kelvin's impulse (1.3.79) is the first moment of vorticity integrated across the channel and differs from the total momentum if there is a non-zero circulation on the boundary of the domain.

The total energy of the system E is also conserved for frictionless flows

$$\begin{aligned} E &= \frac{1}{2}\int_D |\nabla\psi_1|^2 + |\nabla\psi_2|^2 + \frac{1}{2}(\psi_1 - \psi_2)^2\,d^2\mathbf{x}, \\ &= E_0. \end{aligned} \tag{1.3.80}$$

where E_0 is the initial energy. Via an integration by parts and use of (1.3.75-1.3.77), the energy may alternatively be written as

$$E = \int_0^L \left[\psi_1\frac{\partial\psi_1}{\partial y} + \psi_1\frac{\partial\psi_2}{\partial y}\right]_{-W/2}^{W/2} dx - \frac{1}{2}\int_D \psi_1\,(q_1 - \beta y) + \psi_2\,(q_2 - \beta y)\,d^2\mathbf{x}, \tag{1.3.81}$$

where the value of the first term is determined by the flow on the channel sidewalls.

It is also useful at this point to define the available potential energy in the system V, which is not a conserved quantity. The available potential energy is stored in the sloping interface, the displacement of which is $(\psi_1 - \psi_2)/2$.

$$V = \frac{1}{4}\int_D (\psi_1 - \psi_2)^2\,d^2\mathbf{x}. \tag{1.3.82}$$

Another set of invariant integral quantities is the family of Casimir invariants, which in the two-layer quasi-geostrophic model are

$$\mathcal{C}[q_1, q_2] = \int_D C_1(q_1) + C_2(q_2) d^2\mathbf{x}, \tag{1.3.83}$$

and hold for arbitrary functions C_1 and C_2. Casimir invariants are functionals that arise in Hamiltonian systems (e.g. Glendinning [37] p.173). In quasi-geostrophic systems, families of Casimir invariants such as (1.3.83) arise from the material conservation of potential vorticity (1.2.41). Any state that the system reaches must be a conservative rearrangement of the initial PV distribution (Casimir invariants in fluid systems are reviewed by Shepherd [117]).

Two important Casimir invariants are the upper and lower layer planetary[8] enstrophy, Z_1 and Z_2 respectively, given by

$$Z_i = \frac{1}{2} \int_D q_i^2 d^2\mathbf{x} \quad \text{for} \quad i = 1, 2. \tag{1.3.84}$$

The total enstrophy is then $Z = Z_1 + Z_2$.

1.4 Numerical integration of the β-channel model

The numerical scheme used to integrate the β-channel models described in sections 1.3.1 and 1.3.2 is a pseudo-spectral method similar to that used by Esler and Haynes (1999) [30], which will now be discussed.

To integrate the fully nonlinear quasi-geostrophic equations (1.2.54) using a spectral scheme, the channel must be discretized: a spectral representation, truncated at some finite number of modes, is used in the x-direction and a grid-point representation is used in the y-direction. One issue with the use of a

---

[8] Planetary enstrophy since they include the inverse criticality β.

spectral method to find a numerical solution to nonlinear differential equations is the phenomenon known as spectral blocking (e.g. Boyd [12] section 11, Vallis [133] p.364). Enstrophy is transferred nonlinearly in spectral space to higher wavenumbers (smaller scales). Since the number of Fourier modes must be finite in any numerical algorithm, spectral space is truncated at some maximum wavenumber. The cascade of enstrophy to small scales halts at the grid-scale where it cannot move to sub-grid scales, so that artificial noise builds up on the grid-scale. There is a finite (rather than infinitesimal) error on modes at this scale, which may gradually lead to a catastrophic numerical instability where the finite error spreads to larger scales throughout wavenumber space.

A common method employed to avoid the build up of enstrophy on grid-scales is the introduction of artificial diffusion (e.g. [29, 66, 131]). In this work, two types of artificial diffusion will be used, which in the equations of motion (1.2.54) are

$$\begin{aligned} \frac{D_i q_i}{Dt} &= -\kappa_i \nabla^2 \psi_i + \nu_4 \nabla^4 \psi_i + \nu_q \nabla^2 q_i, \\ q_i &= \beta y + \nabla^2 \psi_i + (-1)^i \left(\frac{\psi_1 - \psi_2}{2} \right). \end{aligned} \tag{1.4.85}$$

The parameters ν_4 and ν_q are the non-dimensionalized diffusivities, the former a hyperdiffusion applied to the perturbation relative vorticity $\nabla^2 \psi_i'$ and the latter is a diffusion applied to the perturbation potential vorticity q_i'. Only one type of artificial diffusion is employed for each numerical simulation. The value of the diffusivity (ν_4 or ν_q) may be estimated on scaling grounds but is ultimately chosen using trial and error for each simulation. It is chosen to be as small as possible to ensure numerical stability. Note that the numerical diffusivities must be much larger than actual viscosities found in geophysical flows for the resolutions allowed by the computing speeds of the machines used in this work (or any modern computer).

- $\nu_4 \nabla^4 \psi_i$ is a hyperdiffusion widely used in numerical studies of the two-layer, quasi-geostrophic equations to remove enstrophy at grid-scales (e.g. [66, 131]). Hyperdiffusion terms generally take the form $\nu \nabla^{2n} \psi_i$ and the form used here is employed in chapter 5 to provide continuity with previous studies [29, 28] on which the work is based.

- $\nu_q \nabla^2 q_i$ is an artificial diffusion applied to the perturbation potential vorticity. This is necessary for the uniform flow scenario described in section 1.3.1 where the stability of a wave is governed by a marginal stability curve dependent on the wave number of the disturbance and QG Ekman number (see equation (2.3.16) in chapter 2 and (3.2.7) in chapter 4). In these cases PV diffusion is applied, since in the limit $\nu_q \rightarrow 0$, the position of the marginal stability curve is unchanged in the parameter space. If hyperdiffusion were to be applied for a frictionless flow ($\kappa = 0$), then the marginal stability curve would be shifted by an $O(1)$ amount in the parameter space, destroying the stability properties of the flow. The choice of artificial diffusion $\nu_q \nabla^2 q_i$ follows Nakamura [85] Although a higher order diffusion applied to the perturbation PV could also be employed it is not necessary so long as numerical simulations can be shown to converge with decreasing diffusivity ν_q. Tests of this convergence are discussed further below.

For each simulation, the grid-spacing in the x- and y- direction, Δx and Δy

respectively, must be specified and then the time-step Δt is chosen to satisfy the Courant-Friedrichs-Lewy (CFL) stability criterion, which for a general grid-spacing Δz, may be stated as

$$\mathrm{Co} = \frac{U\Delta t}{\Delta z} < 1. \tag{1.4.86}$$

U is a typical velocity across the grid. The CFL criterion (1.4.86) is explicitly verified for every numerical experiment.

The first time step is made using an Euler scheme, followed by a second order Adams-Bashforth scheme for all subsequent time-steps. The disturbance PV field (in spectral space in the x-direction) is stepped forward in time and then inverted by solving a discretized version of (1.2.42) with a tridiagonal matrix equation to find the geostrophic streamfunction. The linear terms in the PV equation (1.4.85), including the interaction of disturbance fields with the background flow $J(q_i', \Psi_i)$ and $J(Q_i, \psi_i')$, the Ekman friction terms $\kappa_i \nabla^2 \psi_i$ and artificial diffusion terms are applied implicitly. The nonlinear self-advection of the disturbance fields is calculated once a Fast Fourier transform has been applied (in the x-direction) into real space.

Details of the number of Fourier modes, number of grid-points in the y-direction, grid-point spacings Δx, Δy, time-step Δt and artificial diffusion ν_4 or ν_q will be given for each set of numerical simulations (see tables 2.1 and 5.1). A higher numerical resolution has a smaller grid scale $\Delta x, \Delta y$ and time-step Δt and requires weaker artificial hyperdiffusion for the solution to converge. The accuracy of the numerical model is then tested by ensuring that as the resolution is increased and artificial diffusion is decreased, solutions converge. The enstrophy spectrum in the simulation is also monitored throughout and simulations are discarded if, at any point in the evolution in the flow, the enstrophy does not decay super-algebraically within a band of wavenumbers closest to the largest resolved wavenumber.

1.5 Outline of research

Chapter 2 investigates baroclinic lifecycles in the frictionless Phillips model. A weakly nonlinear analytic solution for flows near marginal stability due to Warn & Gauthier [134] is reviewed. The relevance of this analytical theory at finite criticality is assessed for the first time using high resolution numerical simulations over a range of the (β, W) parameter space. Two different equilibration mechanisms are observed, which depend on the width of the channel and the inverse criticality. The numerical simulations are also used to investigate the success of pseudomomentum and pseudoenergy bounds due to Shepherd [116, 118] and an improved pseudoenergy bound due to Esler [31] over the parameter space.

In chapter 3 the mechanism of baroclinic instability in the Phillips model is described in terms of potential vorticity disturbances and counter-propagating Rossby waves (CRWs), a description first put forward by Bretherton [14]. As remarked by Holopainen [53], the Phillips model may be destabilized by the addition of a small amount of Ekman friction. The physical mechanism of destabilization by Ekman friction, referred to as the 'Holopainen instability mechanism', is investigated using the linearized equations of motion. A novel description of this instability is put forward, which uses the same 'PV-thinking' approach as for the CRW description of baroclinic instability. The governing dynamical systems for the Holopainen instability and baroclinic instability are shown to be fundamentally different and a description of how dissipative destabilization occurs is made using phase portrait representations. The Eady model [24] with sloping vertical boundaries is also shown to be susceptible to dissipative destabilization by Ekman friction, the mechanism of which is governed by an equivalent dynamical system to that in the two-layer model.

Chapter 4 investigates baroclinic lifecycles with a uniform flow in the two-layer channel model of chapter 2, but with equal Ekman friction in each layer. A weakly nonlinear theory by Romea [109] is derived using a slightly more modern mathematical method. Simplified versions of his equations are obtained, which show more clearly the parameter dependencies of the solution. Romea's theory is assessed across a wide range of the parameter space for a channel of fixed width using high resolution numerical simulations. In particular, Romea's theory is shown to work in the region of parameter space where waves are destabilized by the Holopainen mechanism, a result suggested by Romea. However, it is also shown to be successful for much stronger Ekman friction, a result not previously confirmed. A new region of the parameter space is found for waves which are marginally unstable in the frictionless problem, but with a small amount of Ekman friction present. In these cases, the Warn & Gauthier nonlinear theory accurately predicts the early time evolution of the wave including the initial linear growth rate and maximum wave amplitude reached, however the final equilibrated amplitude is well predicted by Romea's theory.

Chapter 5 departs from uniform flow profiles to extend a theory due to Esler 2008 [29], which predicts the equilibrated state of an initially unstable symmetric jet in the two-layer model. The theory, known as 'Equilibration via Potential Vorticity Homogenization' (EPVH-A) predicts the final, steady state of the jet using a variational principle which minimizes the available potential energy of the flow under the constraints of energy and momentum conservation. Two new hypotheses that replace the minimization of available potential energy are advanced in this chapter: (i) the flow acts to maximize the area where turbulent mixing takes place (EPVH-B), and (ii) the flow acts to minimize the zonal enstrophy of the flow (EPVH-C). These hypotheses are compared by re-examining the (β, σ) parameter space of Esler. The condition of symmetry is then relaxed as a background, linear, latitudinal shear is introduced to the jet and the EPVH theory is reformulated to predict asymmetric equilibrated states. The success of each of the three hypotheses are examined for asymmetric jets and the hypotheses are found to be less successful than the predictions for symmetric jets.

Chapter 2

Nonlinear baroclinic equilibration at finite supercriticality

2.1 Introduction

The current chapter will investigate, in the context of Phillips' two-layer model, the nonlinear equilibration of a single wave mode in the absence of Ekman friction. An unstable wave initially undergoes linear growth and when it is of sufficiently large magnitude interacts with the background flow. The nonlinear interaction of the wave field with the mean flow brings the background flow to a state that is baroclinically stable and prevents further growth. The aim of an analytic theory is to predict the adjustment of the mean flow by the wave and hence the finite amplitude at which the wave equilibrates.

An initially unstable baroclinic wave in the two-layer model was first studied by Phillips in 1951 [95] who solved the linear equations of motion. Later, the hugely influential, weakly nonlinear theory of Pedlosky (1970) [91] found that the wave reaches a state that undergoes a time-periodic nonlinear oscillation where eddy energy is extracted from and returned to the mean flow. Numerical experiments carried out by Boville (1981) [11] showed a discrepancy in the equilibrated amplitude with Pedlosky's solution near minimum critical shear. The equilibrated wave amplitude in the numerical simulations was also sensitive to the number of Fourier modes included in the integration. The discrepancy was later explained by Pedlosky (1982) [92] as being the result of the resonant excitement of higher harmonics of the fundamental at minimum critical shear. The higher harmonics interact with the fundamental wave at leading order in the analytical solution, which invalidates the asymptotic expansion exploited in Pedlosky's (1970) theory.

A physical explanation for the resonant excitation of higher harmonics is given by the observation that in the lower layer the frequency of the fundamental wave vanishes relative to the background flow and the entire layer is a Rossby wave critical layer (see e.g. Maslowe [76] for a review of critical layers in shear flows). Pedlosky (1982) [93] derived an infinite-dimensional system governing the evolution of the critical layer, which included all harmonics of the fundamental. The truncated dynamical system was solved numerically (with artificial dissipation) to describe some features of the flow. In the lower layer, it was found that the initial potential vorticity gradient is completely homogenized in the equilibrated state of the system. The important observation of PV homogenization allowed a prediction to be made for the equilibrated wave amplitude of the fundamental in the limit of vanishing dissipation.

Warn & Gauthier (1989) [134] (WG89 hereafter) derived a more elegant analytical solution, recasting the unwieldy Pedlosky system of equations as a pair of integro-differential equations, which are integrable. This analytical solution will henceforth be referred to as the Warn-Gauthier-Pedlosky (WGP) solution. In the WGP solution, the evolution of the lower layer PV field is forever transient, but the fundamental wave approaches an equilibrated amplitude. For these exact equations, there is a cascade of PV to small scales, while the upper layer wave 'sees' only the large scale or 'coarse-grain' PV in the lower layer. Equilibration is possible as the upper layer wave attains a steady amplitude while the lower layer PV remains transient.

The WGP theory is formally valid for flows that are weakly unstable in the limit of infinitesimal supercriticality. Shepherd pioneered an alternative analytical approach for flows at finite supercriticality by adapting the nonlinear stability theorems of Arnol'd [4] to unstable flow in Phillips' model. Bounds were obtained on perturbation enstrophy [116] using the momentum-Casimir method, and perturbation energy [118] using the energy-Casimir method (see Shepherd [117], section 6 for a review of Casimir methods).

It is unknown whether the analytical results are relevant to experimental flows or to the atmosphere and ocean. This chapter assesses the relevance of the analytical WGP solution and Shepherd bounds to realizable flows in the form of high resolution numerical simulations, a systematic investigation of which has not been carried out before. The following questions will be addressed:

- How relevant is the WGP solution at finite supercriticality? Gauthier and Shepherd [36, 118] have argued that the Rossby wave critical layer will not form or will only occupy part of the channel at finite criticality and therefore the wave behaviour may be more accurately described by the original Pedlosky (1970) [91] analytical solution. A suitable question is how accurate are WGP predictions of the maximum and equilibrated wave amplitude at finite criticality?
- Is the WGP solution stable on all timescales or subject to a secondary instability? e.g. of the form described by Killworth & McIntyre [60] for Rossby wave critical layers and confirmed numerically by Haynes [40, 41].
- Is the homogenization of coarse-grain potential vorticity, the equilibration mechanism in the WGP solution, unaltered at finite criticality?
- A bound on the amplitude of the fundamental wave in the upper layer, which is independent of the channel width W, can be calculated from Shepherd's bound on perturbation enstrophy. How closely is this bound attained in numerical simulations? Is the maximum wave amplitude affected by the width of the channel?
- To what extent are Shepherd's perturbation energy bounds attained by the disturbances? The perturbation energy bound has a non-trivial dependence on W, does this indicate how flows are affected by variations in W?

The content of this chapter is as follows: in section 2.2, the model and numerical set-up are reviewed. The linear and WGP weakly nonlinear theories will be derived in sections 2.3 and 2.4 respectively. Nonlinear bounds on the growing wave using Shepherd's arguments are calculated in section 2.5. The results of a new suite of high resolution numerical simulations is reported in section 2.6 and these are used to assess Shepherd's bounds in section 2.7. Finally, conclusions are drawn in section 2.8.

2.2 Introduction to the model

2.2.1 The two-layer quasi-geostrophic model

The two-layer, quasi-geostrophic model in a channel with uniform vertical shear was described in section 1.3.1. The background velocity in each layer is given by (1.3.62) as

$$U_1 = -\frac{\partial \Psi_1}{\partial y} = 1, \qquad U_2 = -\frac{\partial \Psi_2}{\partial y} = 0, \tag{2.2.1}$$

and the background potential vorticities from (1.3.63) are

$$Q_1(y) = \left(\beta + \frac{1}{2}\right) y, \qquad Q_2(y) = \left(\beta - \frac{1}{2}\right) y. \tag{2.2.2}$$

The background flow (2.2.1-2.2.2) results in the equations of motion (1.3.64-1.3.65) with $\kappa_1 = \kappa_2 = 0$, repeated here for convenience:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)\left(\nabla^2 \psi_1' - \frac{\psi_1' - \psi_2'}{2}\right) + \left(\beta + \frac{1}{2}\right)\frac{\partial \psi_1'}{\partial x} + J\left(\psi_1', q_1'\right) = 0, \tag{2.2.3}$$

$$\left(\frac{\partial}{\partial t}\right)\left(\nabla^2 \psi_2' + \frac{\psi_1' - \psi_2'}{2}\right) + \left(\beta - \frac{1}{2}\right)\frac{\partial \psi_2'}{\partial x} + J\left(\psi_2', q_2'\right) = 0. \tag{2.2.4}$$

The boundary conditions for the flow are given by (1.3.67-1.3.68) are

$$\frac{\partial \psi_i}{\partial x} = 0, \qquad \frac{\partial^2 \overline{\psi}_i}{\partial t \partial y} = 0, \qquad \text{on} \qquad y = \pm \frac{W}{2}. \tag{2.2.5}$$

Recall that the flow in this problem is entirely determined by the (β, W) parameter space and the initial conditions, which if sufficiently small are unimportant in the flow evolution.

2.2.2 Conservation relations

The flow within the channel domain $D = \{x \in [0, L], y \in [-W/2, W/2]\}$ is subject to conservation of the integral quantities defined in section (1.3.3). Firstly, Kelvin's impulse is conserved according to (1.3.79)

$$M = \int_D yq_1 + yq_2 d^2\mathbf{x} = M(0) = \frac{1}{6}LW^3\beta. \tag{2.2.6}$$

where, neglecting the infinitesimal noise, the initial Kelvin's impulse $M(0)$ has been calculated using (2.2.2).

The total energy given by (1.3.80) is also conserved

$$\begin{aligned} E &= \frac{1}{2}\int_D |\nabla\psi_1|^2 + |\nabla\psi_2|^2 + \frac{1}{2}\left(\psi_1 - \psi_2\right)^2 d^2\mathbf{x}, \\ = E(0) &= \frac{1}{2}LW + \frac{1}{48}LW^3. \end{aligned} \tag{2.2.7}$$

2

Finally, the upper and lower layer planetary enstrophies are conserved and calculated to be

$$Z_i = \frac{1}{2}\int_D q_i^2 d^2\mathbf{x} = \frac{LW^3}{24}\begin{cases} \left(\beta + \frac{1}{2}\right)^2 & i = 1, \\ \left(\beta - \frac{1}{2}\right)^2 & i = 2. \end{cases} \tag{2.2.8}$$

2.2.3 Description of the physical problem and numerical set-up

The physical problem under consideration is that of an infinite channel of fixed width W. The initial conditions consist of the background flow plus an infinitesimal perturbation so that the evolution of the flow is dominated by the fastest growing linear mode (which can be calculated explicitly from linear theory, see section 2.3). The total wavenumber of the fastest growing mode is denoted a_m and zonal wavenumber is $k_m = \sqrt{a_m^2 - (\pi/W)^2}$. The flow is therefore $2\pi/k_m$ periodic in the x-direction enforced by the periodicity condition (1.3.55).

The restriction to infinitesimal noise initial conditions, allows integration of the Phillips model equations (2.2.3-2.2.4) over the sub-domain $(x, y) \in [-\pi/2k_m, 3\pi/2k_m] \times [-W/2, W/2]$, which is carried out using the numerical scheme described in section 1.4. PV diffusion in each layer ($\nabla^2 q_i$ on the right-hand side of equations 2.2.3-2.2.4) is used to remove the build-up of enstrophy at the grid-scale as discussed in section 1.4. The values of diffusivity ν_q, time step δt, number of Fourier modes and grid points in the y-direction are chosen to run simulations at the three different resolutions described in Table 2.1. The time steps δt are chosen to satisfy the Courant-Friedrichs-Lewy criterion for each simulation and scales approximately with the grid-scale. Note that a higher PV diffusivity ν_q is required at lower inverse criticalities β.

The results of three numerical solutions are shown in Fig. 2.2.1 where the quantity plotted in the upper panel is $|A_1^f(t)|$, with $A_i^f(t)$ a complex wave

Resolution	Fourier modes (x)	Grid points (y)	dt	ν_q
low (LR)	128	64	5.0×10^{-3}	5.0×10^{-5}
medium (MR)	256	128	2.5×10^{-3}	2.5×10^{-5}
high (HR)	512	256	1.0×10^{-3}	1.0×10^{-5}

Table 2.1: The three resolutions of the numerical simulations for the $(\beta, W) = (0.48, 2^{3/4}\pi)$ flow.

amplitude defined to be

$$A_i^f(t) = \frac{4}{LW} \int_D q_i(x, y, t) e^{-ik_m x} \cos\left(\frac{\pi y}{W}\right) d^2\mathbf{x}, \quad i = 1, 2. \tag{2.2.9}$$

Equation (2.2.9) is a measure of the amplitude of the baroclinic wave in each layer of the model. Panel (a) of Fig. 2.2.1 shows the evolution of the baroclinic wave amplitude $|A_1^f(t)|$ for the three numerical resolutions detailed in Table 2.1. All three resolutions (LR, MR and HR) agree at the earliest times and well into the nonlinear stage of the lifecycle ($t \lesssim 300$). After this, the solutions diverge slightly, though more rapidly between the LR and MR evolutions than the MR and HR evolutions. Panel (b) of Fig. 2.2.1 shows the total energy E/LW^3 of the flow, which is a conserved quantity in the absence of any dissipation, $\nu_q = 0$. Energy is conserved during the linear evolution of the flow, whilst during the nonlinear stage the LR simulation dissipates energy most rapidly at a uniform rate. At the highest resolution, energy is conserved very well up to $t = 1000$. Panels (c) and (d) show the upper and lower layer planetary enstrophies (Z_1/LW and Z_2/LW respectively). The upper layer planetary enstrophy Z_1/LW is conserved well throughout the numerical simulation. In the lower layer PV cascades to smaller scales and Z_2/LW is conserved only up to the maximum wave amplitude ($t \approx 250$), but is then dissipated in all three simulations. Quantities like pseudomomentum and pseudoenergy, to be

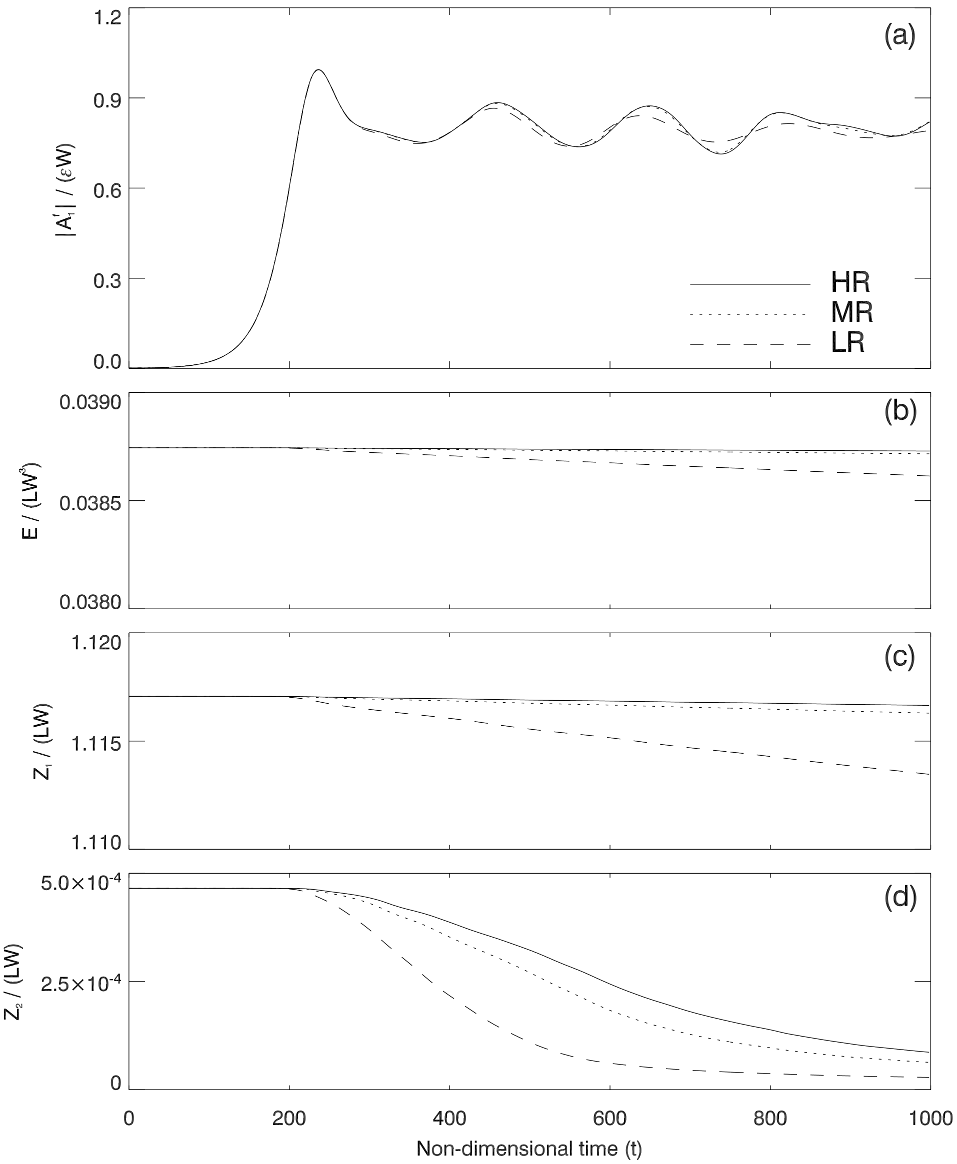

Figure 2.2.1: Panel (a) shows the time evolution of the upper layer wave amplitude $|A_1^f(t)|/\epsilon W$ in a simulation with $(\beta, W) = (0.48, 2^{3/4}\pi)$ for the three numerical resolutions in Table 2.1 as labelled. Panel (b) shows the time evolution of the total energy E/LW^3 and panels (c) and (d) show the evolution of the total enstrophy in the upper (Z_1/LW) and lower (Z_2/LW) layers respectively.

defined below, are conserved well up to the time at which the maximum wave amplitude is reached.

Other convergence tests like the one in Fig. 2.2.1 were carried out across the (β, W) parameter space to verify numerical results. The accuracy of the numerical simulations is also evident through a comparison of results, such as those in Fig. 2.2.1, to the analytical WGP solution.

2.3 Linear Theory

Before proceeding to the WGP weakly nonlinear problem, the linear solution of (2.2.3-2.2.4) is reviewed. The linear problem is relevant for initially small disturbances when nonlinear terms may be neglected and was first solved by Phillips [95].

The equations of motion (2.2.3-2.2.4) are linearized (neglect the Jacobian terms) to give

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\left(\nabla^2\psi_1'-\frac{\psi_1'-\psi_2'}{2}\right)+\left(\beta+\frac{1}{2}\right)\frac{\partial\psi_1'}{\partial x} &= 0, \quad (2.3.10)\\ \left(\frac{\partial}{\partial t}\right)\left(\nabla^2\psi_2'+\frac{\psi_1'-\psi_2'}{2}\right)+\left(\beta-\frac{1}{2}\right)\frac{\partial\psi_2'}{\partial x} &= 0, \quad (2.3.11)\end{aligned}$$

and the perturbation PV is defined in (1.3.61) to be

$$q_i' = \nabla^2\psi_i' + (-1)^i\left(\frac{\psi_1'-\psi_2'}{2}\right). \tag{2.3.12}$$

The solution to equations (2.3.10-2.3.11) with the boundary conditions (2.2.5) support a plane wave solution of the form

$$\begin{pmatrix}\psi_1'\\ \psi_2'\end{pmatrix} = \mathrm{Re}\left\{\begin{pmatrix}1\\ \gamma\end{pmatrix}Ae^{ik(x-ct)}\cos\left(\frac{\pi y}{W}\right)\right\}, \tag{2.3.13}$$

Re denotes the real part, A is a complex valued amplitude and γ is a complex constant describing the phase shift ($arg(\gamma)$) and amplitude ratio ($|\gamma|$) of the waves in each layer. k and $l = n\pi/W$ are the x- and y-wavenumbers respectively with total wavenumber $a = \sqrt{k^2 + l^2}$. The linear solution (2.3.13) has a discretized meridional wavenumber $l = n\pi/W$ due to the finite width of the channel. The restriction to the fastest growing mode means only the gravest mode $n = 1$ need be considered here.

Substitution of (2.3.13) into the linearized equations of motion (2.3.10-2.3.11) and a little algebra yields the dispersion relation for the phase speed c as

$$c^{\pm} = \frac{1}{2} - \frac{(a^2 + \frac{1}{2})}{a^2(a^2+1)}\beta \pm \frac{1}{2a^2(a^2+1)}\left(\beta^2 + a^4(a^4-1)\right)^{\frac{1}{2}}, \quad (2.3.14)$$

and the complex streamfunction ratio

$$\gamma^{\pm} = 2\left(a^2 + \frac{1}{2}\right) + 2\left(\frac{\beta + \frac{1}{2}}{c^{\pm} - 1}\right). \quad (2.3.15)$$

If the imaginary part of the phase speed c_i in (2.3.14) is positive then the system is unstable and the wave will grow (e.g. Vallis [133] p.271-277). The marginal stability curve $\beta_c^I(a)$ defines the point at which a disturbance of wavenumber a becomes unstable $Im(c^+) = 0$. The condition for baroclinic instability is then

$$\beta^2 < (\beta_c^I(a))^2 = a^4\left(1 - a^4\right). \quad (2.3.16)$$

Alternatively, if

$$\beta < \frac{1}{2} \quad (2.3.17)$$

then from (2.3.16) there is a band of unstable wavenumbers

$$\frac{1}{2} - \sqrt{\frac{1}{4} - \beta^2} < a^4 < \frac{1}{2} + \sqrt{\frac{1}{4} - \beta^2}. \quad (2.3.18)$$

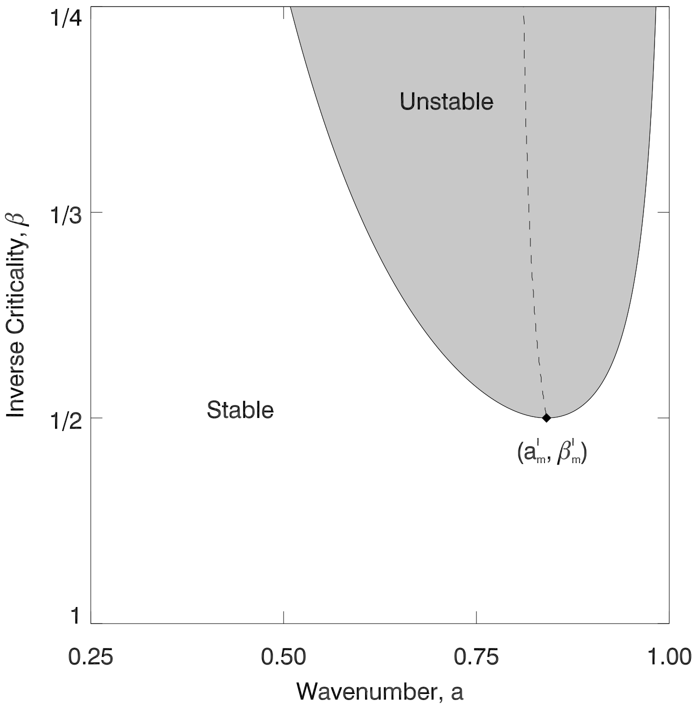

Figure 2.3.1: Curve of marginal stability for a frictionless uniform flow profile in the two-layer model. Wavenumber a is plotted on the horizontal axis and decreasing inverse criticality β on the vertical axis. The shaded area corresponds to the band of unstable wavenumbers and the point of minimum critical shear $(\beta^I_m, a^I_m) = (2^{-1}, 2^{-1/4})$ is marked. The dashed line corresponds to the wavenumbers of the fastest growing modes $a_m(\beta)$ in a channel of width $W = 2^{3/4}\pi$.

The condition (2.3.16) is plotted in Fig. 2.3.1, which clearly shows the band of unstable wavenumbers (2.3.18)

The upper and lower layer background potential vorticities are given by (2.2.2) as $Q_1 = (\beta + \frac{1}{2})y$ and $Q_2 = (\beta - \frac{1}{2})y$ respectively. The upper layer PV gradient is positive for all values of β, whereas the lower PV gradient is negative for $\beta < \frac{1}{2}$ and positive for $\beta > \frac{1}{2}$. The condition for baroclinic instability (2.3.17) is a specific case of the more general Charney-Stern-Pedlosky sufficient condition for instability [18, 90] that the PV gradient (generalized to include boundary conditions) must change sign somewhere within the domain (e.g.

Vallis [133], section 6.4.3). In the case of a uniform flow profile, the condition that the PV gradient reverses is satisfied across the whole channel. The reversal of PV gradients suggests an elegant way to describe baroclinic instability in terms of counter-propagating Rossby waves (CRWs) [14] and this is discussed at length in chapter 3.

The vertical shear in the problem is fixed since the background velocities were set as $U_1 = 1$, $U_2 = 0$ in the description of the model in chapter 1. In the literature [91, 109, 134] the stability condition (2.3.16) is usually written as a condition on the vertical shear $U_1 - U_2$ including an unscaled version of β. This introduces an unnecessary degeneracy of parameters, which are simplified here by the inclusion of the inverse criticality β as defined by (1.2.24). The minimum critical shear (the vertical shear at which the flow first becomes unstable) is equivalent to the largest value of β for which the flow is unstable, which occurs at

$$\left(a^I_m, \beta^I_m\right) = \left(2^{-1/4}, 2^{-1}\right). \tag{2.3.19}$$

The point of minimum critical shear (2.3.19) is marked in Fig.2.3.1. Substitution of (2.3.19) into (2.3.15) gives the amplitude ratio and phase difference of the waves in each layer

$$\gamma_m = \sqrt{2} - 1. \tag{2.3.20}$$

The ratio γ_m is strictly real, indicating that there is no phase shift between the waves in the linear problem and any exchange of momentum between the layers must be associated with wave growth and decay.

It is important to understand what is happening at minimum critical shear.

Substitution of $\beta_m^I = 2^{-1}$ into the phase speed (2.3.14) gives

$$c^{\pm} = \frac{(a^4 - \frac{1}{2})}{2a^2(a^2+1)}(1 \pm 1) \tag{2.3.21}$$

i.e. $c^-(a) = 0$ for all wavenumbers, or alternatively c^- is equal to the lower layer background velocity. At minimum critical shear, the lower branch of the dispersion curve (2.3.21) is stationary so that all modes interacting with the fundamental are resonant and important in the evolution of the flow. In the reference frame of the wave, in the lower layer, the advection of PV by the fundamental wave vanishes and subsequently, at minimum critical shear, the entire lower layer is a critical layer [76].

Pedlosky [92] argued that, as for critical layers in parallel shear flows [8, 20], nonlinearities are important in the lower layer and a consideration of a wide range of harmonics is required, invalidating Pedlosky's single wave theory at minimum critical shear [91] (this was first reported in numerical experiments by Boville [11]). Pedlosky [93] later derived a critical layer model, valid at minimum critical shear consisting of an infinite number of differential equations, which must be solved numerically. WG89 solved the truly frictionless, critical layer problem providing a considerably simpler analytic expression for the evolution of the flow. This solution, introduced earlier as the WGP solution, is reviewed in the next section. The transition between the Pedlosky and WGP weakly nonlinear theories as one moves along the marginal stability curve away from the point of minimum shear was also studied by Gauthier (1990) [36] in terms of a 'de-tuning' of the WGP solution.

2.4 Weakly nonlinear theory

This section reviews the weakly nonlinear theory of WG89 concerning the growth of an initially small wave. An alternative, more direct derivation is followed. Attention is restricted to the infinitesimal noise initial condition [1].

The formulation of the WGP theory is for flows with supercriticality defined by

$$\epsilon = \left(\frac{1}{2} - \beta\right)^{1/2} \qquad (2.4.22)$$

which is formally small ($\epsilon \ll 1$). Substitution of the supercriticality (2.4.22) into (2.2.2) gives the background potential vorticity in each layer as

$$Q_1(y) = (1 - \epsilon^2)y, \qquad Q_2(y) = -\epsilon^2 y. \qquad (2.4.23)$$

Substitution of (2.4.22) and $a_m^I = 2^{-1/4}$ into the phase speed (2.3.14) shows that, at minimum critical shear

$$c^{\pm} = \pm\frac{\sqrt{2}\epsilon}{\sqrt{2}+2}i. \qquad (2.4.24)$$

The wave therefore grows in a reference frame at rest (or propagating with the lower layer velocity $U_2 = 0$) on timescales of order ϵ^{-1}. This suggests that the perturbation streamfunctions ψ_1', ψ_2' should be functions of two time variables, a fast time t and a slow time defined by

$$\tau = \epsilon t. \qquad (2.4.25)$$

Since the lower layer background velocity is equal to the real part of the phase speed given by (2.4.24) (they are both zero), solutions depend only on the

---

[1] WG89 refer to this lifecycle as a ‘subliminal disturbance’.

slow timescale τ. Upon substitution of (2.4.23) and (2.4.25) the equations of motion (2.2.3-2.2.4) become

$$\begin{aligned}
\frac{\partial q_1'}{\partial x} + \frac{\partial \psi_1'}{\partial x} &= -\epsilon\left(\frac{\partial q_1'}{\partial \tau} + J\left(\psi_1', q_1'\right)\right) - \epsilon^2\left(\frac{\partial \psi_1'}{\partial x}\right), && (2.4.26)\\
\frac{\partial q_2'}{\partial \tau} + J\left(\psi_2', q_2'\right) &= \epsilon\left(\frac{\partial \psi_2'}{\partial x}\right). && (2.4.27)
\end{aligned}$$

Power series solutions of (2.4.26-2.4.27) are sought by expanding the perturbation streamfunction and PV in ϵ so that the total streamfunctions and potential vorticities are

$$\begin{aligned}
\psi &= -y\delta_{i1} + \epsilon\psi_i^{(0)} + \epsilon^2\psi_i^{(1)} + \epsilon^3\psi_i^{(2)} + \ldots, && (2.4.28)\\
q_i &= \left(\delta_{i1} - \epsilon^2\right) + \epsilon q_i^{(0)} + \epsilon^2 q_i^{(1)} + \epsilon^3 q_i^{(2)} + \ldots, && (2.4.29)\\
\text{with} \qquad q_i^{(j)} &= \nabla^2\psi_i^{(j)} + (-1)^i\left(\frac{\psi_1^{(j)} - \psi_2^{(j)}}{2}\right), && (2.4.30)
\end{aligned}$$

where $\psi_i^{(j)}$ and $q_i^{(j)}$ are $O(1)$ quantities and, following the method of multiple scales (e.g. Hinch [52]), are allowed to depend on the slow time variable $\tau = \epsilon t$. δ_{i1} is the Kronecker delta.

Substitution of the asymptotic expansion (2.4.28-2.4.29) into equations (2.4.26-2.4.27) gives a series of problems in ϵ. At leading order in ϵ the solution is precisely that of the linear problem for the marginally stable wave (2.3.13) with total wavenumber $a_m^I = 2^{-1/4}$ and $\gamma_m^I = \sqrt{2} - 1$ (the superscript I is now dropped). The branch of the linear solution corresponding to the growing mode is

$$\begin{pmatrix} \psi_1^{(0)} \\ \psi_2^{(0)} \end{pmatrix} = Re\left\{A(\tau)\begin{pmatrix} 1 \\ \gamma_m \end{pmatrix} e^{ik_m x}\cos\left(\frac{\pi y}{W}\right)\right\}. \qquad (2.4.31)$$

From (2.4.24), the real phase speed of the marginally stable wave is zero and the leading order equation is independent of the fast time variable t. Higher order terms are forced by the leading order solution and are therefore also independent of t. Substitution of (2.4.31) into (2.4.30) reveals that the leading order potential vorticities are

$$q_1^{(0)} = -\psi_1^{(0)}, \tag{2.4.32}$$

$$q_2^{(0)} = 0. \tag{2.4.33}$$

The perturbation PV in the lower layer remains an order of magnitude smaller than the growing wave. In the upper layer the PV distribution consists of the uniform background gradient (2.4.23) plus an $O(\epsilon)$ perturbation and the flow remains 'Rossby wave-like'. In the lower layer both the background and perturbation PV fields are $O(\epsilon^2)$ quantities.

The expansions (2.4.28-2.4.29) are substituted into the definitions of Kelvin's impulse (2.2.6) and upper layer enstrophy (2.2.8) to give

$$M = M(0) + \epsilon \int_D y q_1^{(0)} d^2\mathbf{x} + \epsilon^2 \int_D y q_1^{(1)} + y q_2^{(1)} d^2\mathbf{x} + O(\epsilon^3), \tag{2.4.34}$$

$$Z_1 = Z_1(0) + \epsilon \int_D y q_1^{(0)} d^2\mathbf{x} + \epsilon^2 \int_D y q_1^{(1)} + \frac{1}{2}\left(q_1^{(0)}\right)^2 d^2\mathbf{x} + O(\epsilon^3), \tag{2.4.35}$$

where $M(0)$ and $Z_1(0)$ are the initial values of Kelvin's impulse and upper layer enstrophy from (2.2.6) and (2.2.8) respectively. Both Kelvin's impulse and upper layer enstrophy are conserved at all times, $M = M(0)$, $Z_1 = Z_1(0)$, and the remaining terms in (2.4.34-2.4.35) are zero. At $O(\epsilon)$ in both equations, it is seen that

$$\int_D y q_1^{(0)} d^2\mathbf{x} = 0, \tag{2.4.36}$$

which is directly verified upon substitution of (2.4.31). At $O(\epsilon^2)$, the elimination of the integral of $yq_1^{(1)}$ between equations (2.4.34) and (2.4.35) gives

$$\int_D \frac{1}{2}\left(q_1^{(0)}\right)^2 d^2\mathbf{x} = \int_D yq_2^{(1)} d^2\mathbf{x}. \tag{2.4.37}$$

Use of (2.4.32) and substitution of the leading order wave (2.4.31) into the left-hand side of (2.4.37) results in

$$|A(\tau)|^2 = \frac{8}{LW}\int_D yq_2^{(1)} d^2\mathbf{x}, \tag{2.4.38}$$

where the integral on the right-hand side of (2.4.38) measures the change in the Kelvin's impulse in the lower layer. Equation (2.4.38) is precisely the conservation of pseudomomentum in the system. Pseudomomentum invariants are constructed from the linear impulse conservation law $M = M(0)$ (2.2.6) and a Casimir invariant (1.3.83). In the case of pseudomomentum, the Casimir invariant is the enstrophy in the upper layer Z_1.

In the lower layer, the leading order perturbation PV is $q_2^{(1)}$ (since $q_2^{(0)} = 0$) governed by (2.4.26-2.4.27), which upon insertion of (2.4.28-2.4.29) at $O(\epsilon^2)$ gives

$$\left(\frac{\partial}{\partial\tau} - \frac{\partial\psi_2^{(0)}}{\partial y}\frac{\partial}{\partial x} + \frac{\partial\psi_2^{(0)}}{\partial x}\frac{\partial}{\partial y}\right)\left(q_2^{(1)} - y\right) = 0. \tag{2.4.39}$$

Equation (2.4.39) was recognised to be a passive tracer advection equation in WG89 where the structure of the advecting streamfunction $\psi_2^{(0)}$ is explicitly given by the leading order solution (2.4.31). The passive tracer in (2.4.39) is the leading order PV in the lower layer $Q = q_2^{(1)} - y$. Potential vorticity is usually an active tracer as it instantaneously determines the flow via PV inversion (e.g. section 1.2.7), however here Q is a passive tracer because the advecting flow in the lower layer is determined entirely by the leading order

PV in the upper layer $q_1^{(0)}$. Usefully, equation (2.4.39) can be re-cast into the advection of the passive tracer by a *steady* velocity field. Since (2.4.39) is linear in $\psi_2^{(0)}$, the complex amplitude $A(\tau)$ can be chosen to be real ($|A(\tau)|$ is used hereafter). The change of time variables

$$\eta = \gamma_m \int_0^\tau |A(\tilde{\tau})| d\tilde{\tau} \tag{2.4.40}$$

can be made, resulting in

$$\left(\frac{\partial}{\partial \eta} - \frac{\partial \Psi}{\partial y}\frac{\partial}{\partial x} + \frac{\partial \Psi}{\partial x}\frac{\partial}{\partial y} \right) Q = 0, \tag{2.4.41}$$

where

$$\Psi(x, y) = \cos\left(k_m x\right) \cos\left(\frac{\pi y}{W}\right) \tag{2.4.42}$$

For the infinitesimal lifecycle described above, the initial condition is $Q(x, y, 0) = -y$ and WG89's solution (presented in appendix 2.A) of equation (2.4.41) is

$$Q(x, y, \eta) = \\ -\frac{W}{\pi}\sin^{-1}\left(\frac{\sin\left(\frac{\pi y}{W}\right)\mathrm{cn}\left(\frac{k_m \pi}{W}|m\right)\mathrm{dn}\left(\frac{k_m \pi}{W}|m\right) + \sin\left(k_m x\right)\cos^2\left(\frac{\pi y}{W}\right)\mathrm{sn}\left(\frac{k_m \pi}{W}|m\right)}{1 - \sin^2\left(\frac{\pi y}{W}\right)\mathrm{sn}^2\left(\frac{k_m \pi}{W}|m\right)} \right), \tag{2.4.43}$$

where

$$m(x, y) = 1 - \cos^2\left(k_m x\right)\cos^2\left(\frac{\pi y}{W}\right) \tag{2.4.44}$$

and $\mathrm{sn}(\cdot|\cdot)$, $\mathrm{cn}(\cdot|\cdot)$ and $\mathrm{dn}(\cdot|\cdot)$ are Jacobi elliptic functions (e.g. [1]).

The solution (2.4.43) is shown in Fig. 2.4.1 at five snapshots in time $\eta = 0, 0.79, 5.55, 10, 15W/k_m\pi$, displaying the evolution of the lower layer critical layer. Fig. 2.4.1 may be directly compared to the similar plot in Fig. 2 of WG89. In panel (a) the initial meridional PV gradient in the lower layer is

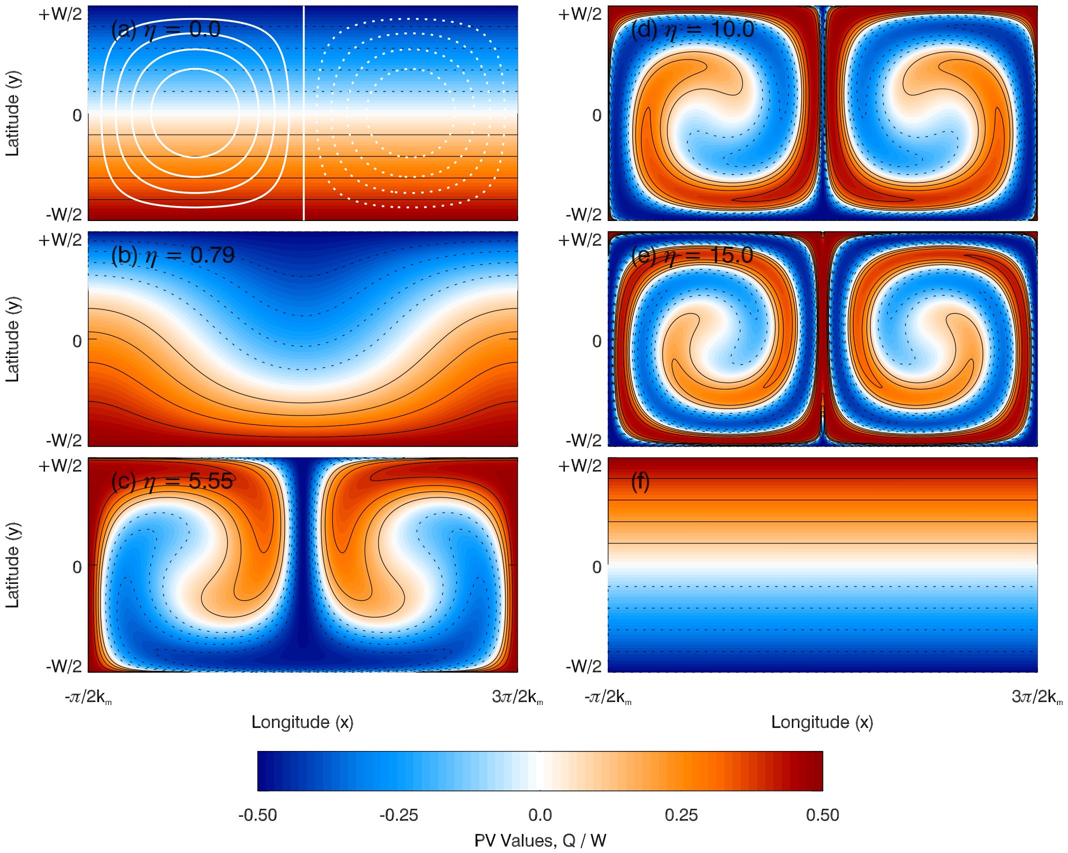

Figure 2.4.1: Snapshots of the total potential vorticity in the lower layer $Q(x, y, \eta)/W$ in the WGP solution, calculated using (2.4.43) at scaled times $\eta = 0, 0.78, 5.55, 10, 15 W/k_m\pi$ (panels a-e). The contour interval is 0.1. In panel (a), the advecting streamfunction $\psi_2^{(0)}$ is contoured in white, with solid lines representing anti-cyclonic (clockwise) circulation and dotted lines representing a cyclonic (anti-clockwise) cell. Its structure, but not its magnitude is invariant in time and hence no fixed contour interval applies. The final panel (f) shows the 'rearranged' PV field used for the WGP pseudomomentum bound (equation 2.5.50).

constant ($=-\epsilon^2$). The disturbance grows in each half of the channel and the circulation cells (white contours in panel a) cause the PV contours to wrap up (panels b-e). As they do so increasingly smaller scales develop, which at long times require greater resolution of quadrature to resolve. There is a corresponding enstrophy cascade to smaller scales. At long times the PV contours continue to wrap so that the PV field is forever transient.

The expression for the lower layer PV (2.4.43) is used to evaluate the integral in (2.4.38) in order to calculate the wave amplitude $|A(\tau)|$. Noting that $q_2^{(1)} = y + Q$, the integral is solved by numerical quadrature using the coordinate transformation discussed in appendix 2.A. Substitution of (2.4.43) into (2.4.38) gives the WGP solution

$$|A(\tau)|^2 = W^2\left(\frac{2}{3} - \frac{16}{\pi^4} I\left(\frac{k_m \pi}{W}\eta\right)\right), \tag{2.4.45}$$

where $I(z)$ denotes the definite integral

$$I(z) = \int_0^1 \int_0^{K(m)} \frac{\sin^{-1}(m^{1/2}\mathrm{sn}(\alpha - z|m))\sin^{-1}(m^{1/2}\mathrm{sn}(\alpha|m))}{(1-m)^{1/2}} d\alpha dm \tag{2.4.46}$$

and $K(m)$ is the complete elliptic integral of the first kind. The integral $I(z)$ has the following properties

(i) $I(0) = \pi^4/24$ and so the unperturbed initial state $\eta = |A(0)| = 0$ is a solution of (2.4.45).

(ii) $I(z) \to 0$ as $z \to \infty$, meaning that the long time solution of (2.4.45) is

$$\lim_{\tau\to\infty} |A| = |A|_{eq} = \left(\frac{2}{3}\right)^{1/2} W, \tag{2.4.47}$$

which corresponds to a completely wrapped up, coarse-grain homogenized, lower layer PV.

(iii) $I(z)$ decreases monotonically to its global minimum at $z_m = 5.55$. At this global minimum, $I(z_m) \approx -2.257$ and the maximum amplitude is therefore

$$|A|_{max} \approx 1.02W. \tag{2.4.48}$$

The point z_m is where Kelvin's impulse and upper layer wave amplitude are simultaneously maximized at time $\eta \approx 5.55W/k_m\pi$ and the lower layer PV field at this time is shown in Fig. 2.4.1 panel (c). This is contrasted to the 'perfect rearrangement' of lower layer PV in Fig. 2.4.1 panel (f) to be discussed below.

The solution (2.4.45) is implicit, since the amplitude $|A(\tau)|$ is related to the rescaled time $\eta(\tau)$ through (2.4.40). The solution (2.4.45) can be re-written using $|A(\tau)| = \gamma_m^{-1} d\eta/dt$ as

$$\tau = \frac{1}{\gamma_m W} \int_{\eta_0}^{\eta} \left(\frac{2}{3} - \frac{16}{\pi^4} I \left(\frac{k_m \pi}{W} \overline{\eta} \right) \right)^{-1/2} d\overline{\eta}. \tag{2.4.49}$$

Calculation of the amplitude must proceed from the initial infinitesimal wave, from which η_0 is determined. Changes in η_0 do not affect the solution provided that it is sufficiently small. Equation (2.4.49) may be inverted numerically to give $\eta(\tau)$ and hence $|A(\tau)|$ explicitly.

2.5 Bounds on wave amplitude and perturbation energy

2.5.1 Held-Shepherd bound on wave amplitude

Related to the WGP solution are upper bounds on the wave amplitude that can be derived using the methods of Arnol'd [4]. The bound relevant to the two-layer model here was calculated by Shepherd [116], but in that work was attributed to I. M. Held and will hence be referred to here as the Held-Shepherd bound.

From the WGP amplitude equation (2.4.38), the wave amplitude is entirely governed by the Kelvin's impulse in the lower layer. The Kelvin's impulse has an upper bound, which is calculated using the rearranged PV field shown in Fig. 2.4.1 panel (f) where $q_2^{(1)} = Q + y = 2y$. Integrating equation (2.4.38)

$$|A(\tau)|^2 = \frac{8}{LW}\int_D y q_2^{(1)} d^2\mathbf{x} \leq \frac{8}{LW}\int_D 2y^2 d^2\mathbf{x} = \frac{4}{3}W^2. \tag{2.5.50}$$

The Held-Shepherd bound is an extension of this idea to the finite criticality case. The details of the derivation of the Held-Shepherd bound are shown in appendix 2.B and the end result is a bound on the upper layer fundamental wave amplitude given by

$$|A_1^f|^2 \leq \frac{1}{6}W^2 \begin{cases} 16\beta\left(\frac{1}{2}-\beta\right) & \frac{1}{6} \leq \beta \leq \frac{1}{2}, \\ 2\left(\beta+\frac{1}{2}\right)^2 & 0 < \beta < \frac{1}{6}. \end{cases} \tag{2.5.51}$$

A_1^f is identified with $-\epsilon A$ in the WGP solution. Writing $\epsilon = \frac{1}{2} - \beta$ and taking the limit $\epsilon \to 0$ ($\beta \to \frac{1}{2}$), equation (2.5.51) recovers the WGP result (2.5.50).

In appendix 2.B a bound on the lower layer wave amplitude is also reviewed and is given by

$$|A_2^f|^2 \leq \frac{1}{3}W^2\left(\frac{1}{2}-\beta\right)^2. \tag{2.5.52}$$

2.5.2 Shepherd's pseudoenergy bound on wave energy

A similar approach is taken to derive and improve upon pseudoenergy bounds by Shepherd [118] where the objective is to find a bound on

$$\begin{aligned} E' + \lambda Z' &= \frac{1}{2}\int_D |\nabla\psi_1'|^2 + |\nabla\psi_2'|^2 + \frac{1}{2}\left(\psi_1'-\psi_2'\right)^2 \\ &\quad + \frac{\lambda}{2}\left((q_1')^2+(q_2')^2\right) d^2\mathbf{x}. \end{aligned} \tag{2.5.53}$$

Equation (2.5.53) is known as the pseudoenergy, the sum of perturbation energy and the weighted perturbation enstrophy. From this a bound can then be found on the perturbation energy

$$E' = \frac{1}{2}\int_D |\nabla\psi_1'|^2 + |\nabla\psi_2'|^2 + \frac{1}{2}\left(\psi_1'-\psi_2'\right)^2 d^2\mathbf{x}. \tag{2.5.54}$$

Shepherd's bound (see equation 8.2 of Shepherd (1993) [118]) in the current notation is given by

$$E' \leq \frac{1}{6}LW^3\begin{cases} \left(\frac{1}{2}-\beta\right)\left(1-\frac{48-2W^2}{W^2}\left(\frac{1}{2}-\beta\right)\right), \\ \frac{1}{8}\left(1+\frac{12}{W^2}\right). \end{cases} \tag{2.5.55}$$

An improved bound, the details of which are found in appendix 2.C, is given by

$$E' \leq \frac{1}{6}LW^3\begin{cases} \left(\frac{1}{2}-\beta\right)\left(1-\frac{2W^2}{W^2+12}\left(\frac{1}{2}-\beta\right)\right), & \frac{1}{4}-\frac{3}{W^2} \leq \beta < \frac{1}{2}, \\ \frac{1}{8}\left(1+\frac{12}{W^2}\right), & 0 < \beta < \frac{1}{4}-\frac{3}{W^2}, \end{cases} \tag{2.5.56}$$

which improves upon the Shepherd pseudoenergy bound (2.5.55).

2.6 Comparison of analytical results to numerical experiments

The numerical model set up in section 2.2.3 is used to perform experiments across the (β, W) parameter space. The first suite of experiments is carried out in a channel of fixed width $W = 2^{3/4}\pi \approx 5.283$. This channel width is significant since the fastest growing mode (calculated in section 2.3) is then isotropic with $k_m = \pi/W = 2^{-1/2}a_m = 2^{-3/4}$ in the limit of vanishing supercriticality $\epsilon \to 0$.

2.6.1 Relevance of the WGP solution at finite criticality

Fig. 2.6.1 compares the evolution over scaled time η of the WGP wave amplitude $|A|/W$, given by (2.4.45), to the results of the fully nonlinear numerical model at inverse criticalities $\beta = 0.495, 0.48, 0.45, 0.42$. The numerical wave amplitudes $|A_1^f|/\epsilon W$ are calculated using equation (2.2.9) and the scaled time for the simulations is calculated from

$$\eta = \gamma_m \int_0^t |A_1^f(t')|dt'. \tag{2.6.57}$$

The fully nonlinear solutions in Fig. 2.6.1, plotted with broken lines, are accurately captured by the WGP solution (solid line) as $\beta \to \frac{1}{2}$ for times $\eta \lesssim 10W/k_m\pi$. The maximum wave amplitude achieved over the lifecycle is therefore accurately predicted by WGP. At later times $\eta \gtrsim 10W/k_m\pi$, the WGP solution and nonlinear simulations diverge as the WGP solution rapidly approaches the constant amplitude $|A|^2 \to 2W^2/3$, whereas the numerical

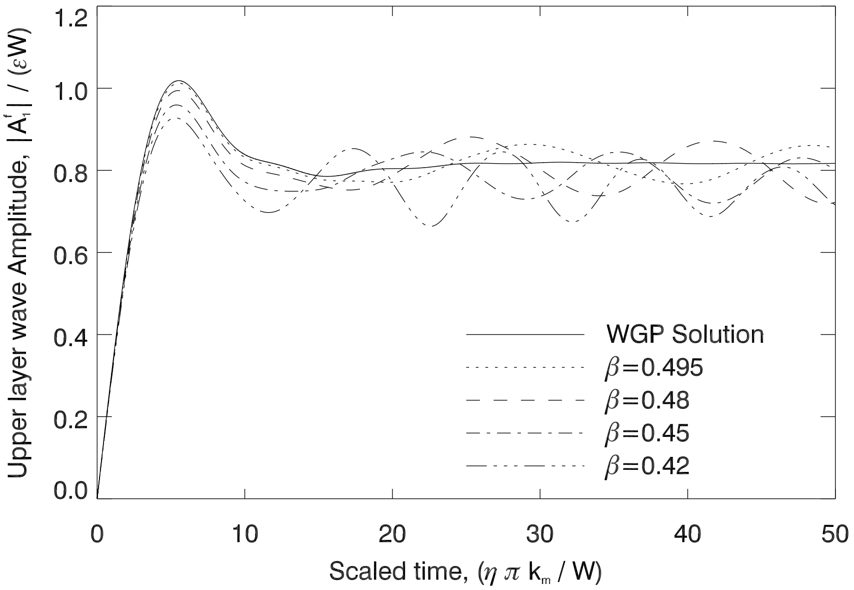

Figure 2.6.1: Evolution in scaled time η of the upper layer wave amplitude $|A_1^f|/\epsilon W$ for numerical simulations with $W = 2^{3/4}\pi$ and $\beta = 0.495, 0.48, 0.45, 0.42$ (broken curves, as labelled), together with the analytical WGP solution (solid curve).

solutions oscillate about this value. It is not clear whether in the absence of numerical diffusion the amplitude of the oscillation decays as $\beta \rightarrow \frac{1}{2}$. There is a decrease in the amplitude of the oscillation observed in Fig. 2.6.1, specifically in the $\beta = 0.495$ curve, which could be due to the much longer integration times. Integration times scale as ϵ^{-1} and therefore the time-integrated effect of artificial diffusion increases as $\epsilon \rightarrow 0$. The period of oscillation and early time behaviour of the solution were shown to be insensitive to the numerical resolution (and ν_q) in Fig. 2.2.1.

Equation (2.4.38) shows that the upper layer wave amplitude is governed by the lower layer PV field and this is therefore the key to understanding the difference between the fully nonlinear wave amplitudes observed in simulations and the WGP solutions seen in Fig. 2.6.1. Fig. 2.6.2 is a plot of the total lower layer PV field $q_2/\epsilon^2 W$ at scaled times $\eta = 5.55W/k_m\pi$ (left column) and $\eta = 10W/k_m\pi$ (right column) for $W = 2^{3/4}\pi$ and $\beta = 0.2, 0.4, 0.495$ compared to the WGP field Q/W calculated directly from (2.4.43). The WGP solution (panels a-b) is very similar to the smallest supercriticality simulation ($\beta = 0.495$ in panels c-d) for both snapshots. There is an asymmetry between the circulation cells in the $\beta = 0.495$ simulation, which are exactly anti-symmetric in the WGP solution. There is greater entrainment of cyclonic PV (red) than anticyclonic PV (blue) by the cyclonic (anti-clockwise) circulation cell in the right-hand side of the channel. The asymmetry is much greater at lower supercriticalities, $\beta = 0.4$ in panels (e-f) and $\beta = 0.2$ in (g-h). For each of these flows a vortex forms within each circulation cell and at $\eta = 10W/k_m\pi$, in panels (f) and (h), the vortices are exactly in phase with the upper layer wave. The upper layer PV contours for each of the flows in Fig. 2.6.2 are shown in Fig. 2.6.3 confirming that the lower layer vortices of the $\beta = 0.2, 0.4$ flows are in phase with the upper layer disturbances (panels d and f of Fig. 2.6.3). In each of these panels, the sign of the vortices is such that the circulation in each cell is enhanced: there is a tendency towards barotropization in fully nonlinear flows similar to that seen in isotropic quasi-geostrophic two-layer turbulence (see Vallis [133], sec. 9.2.3). Barotropization is a tendency of the energy in the system to be transferred to the barotropic mode at scales larger than the deformation radius where it is trapped. The barotropization of the flow is best observed in panels (h) of Fig. 2.6.2 and (f) of Fig. 2.6.3 where signed vortices in each layer are in phase. Panels (a-b) of Fig. 2.6.3 also show the Rossby wave-like behaviour of the upper layer at small supercriticality.

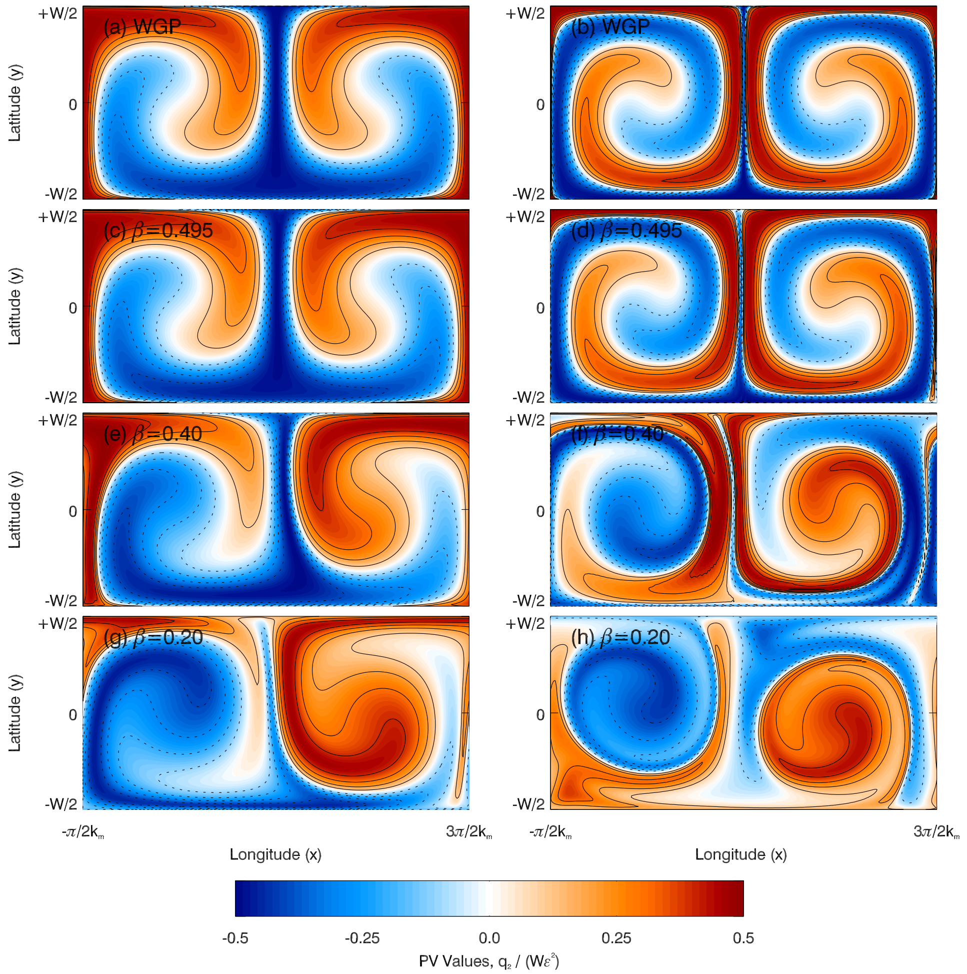

Figure 2.6.2: The (scaled) lower layer PV field $q_2/\epsilon^2 W$ at scaled times $\eta = 5.55W/k_m\pi$ (left panels) and $\eta = 10W/k_m\pi$ (right panels). Panels (a-b) show the WGP solution, panels (c-h) show the results of numerical experiments: panels (c-d) $\beta = 0.495$, panels (e-f) $\beta = 0.4$ and (g-h) $\beta = 0.2$. All simulations have channel width $W = 2^{3/4}\pi$. The contour interval for all panels is 0.1.

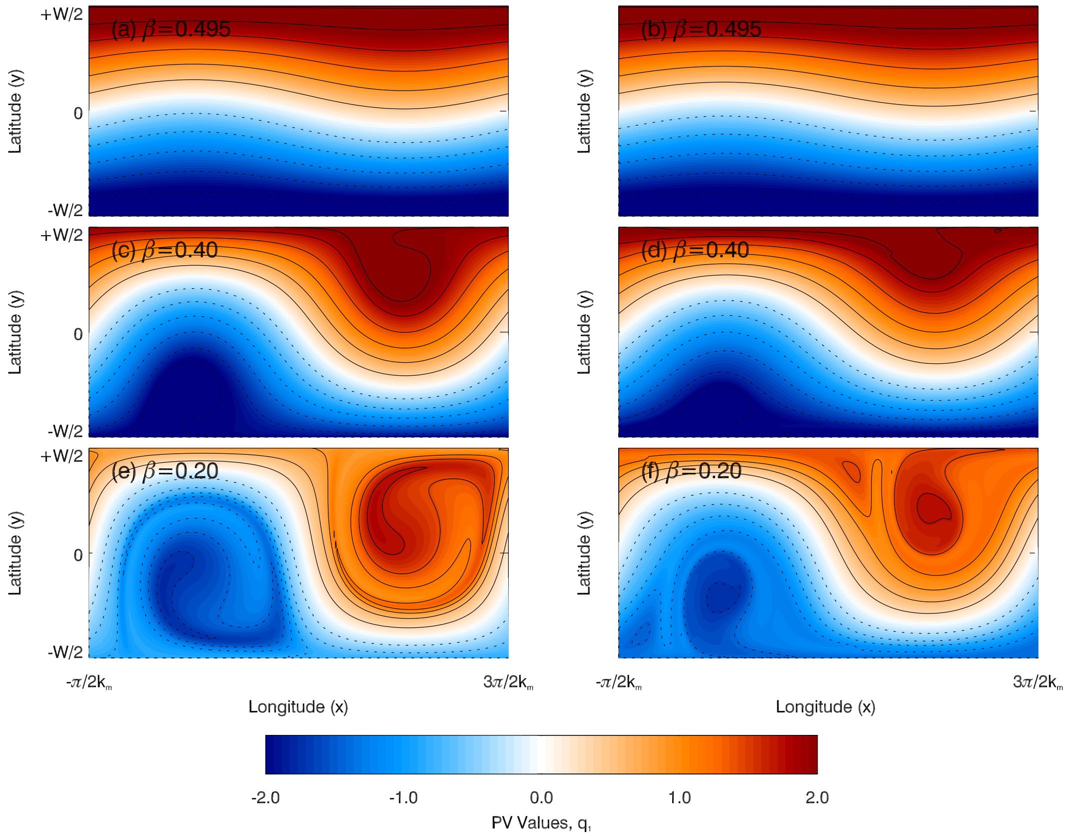

Figure 2.6.3: As for Fig. 2.6.2 panels (c-h), but the upper layer (unscaled) PV fields q_1 are plotted. Note that the PV values are not scaled on supercriticality and the contour levels in all panels are 0.4

The mechanism of equilibration changes as the supercriticality increases, but in all cases takes place via the removal of the lower layer zonal mean PV gradient. In the WGP solution, at low supercriticalities, the lower layer PV is stirred, eventually leading to the (coarse-grain) homogenization of the entire lower layer PV field. At higher supercriticalities, the lower layer PV rolls up into a pair of oppositely signed vortices with one pair per wavelength of the upper layer wave. The transition between these two behaviours is discussed below.

2.6.2 Sensitivity of finite criticality simulations to channel width

Fig. 2.6.4 shows snapshots of the lower layer PV field for the WGP solution and simulations with $\beta = 0.48$ for three widths, $W = 4, 2^{3/4}\pi, 10$. Fig. 2.6.4 shows that there is a significant dependence of the numerical simulations on W. In order to understand these differences, it is useful to consider the different aspect ratios $k_m W/\pi$ of the circulation cells in the three simulations of panels (c-d), (e-f) and (g-h). These channel widths differ since the fastest growing mode emerging from the infinitesimal noise initial conditions has a different structure in each case. For each of $W = 4, 2^{3/4}\pi, 10$, the aspect ratios of the fastest growing normal mode for the $\beta = 0.48$ flow are $k_m W/\pi = 0.757, 1.000, 1.893$ respectively. The circulation cells for $W = 2^{3/4}\pi$ are nearly isotropic (they are exactly isotropic in the limit $\epsilon \to 0$), those for $W = 10$ are elongated in the meridional direction and those for $W = 4$ are elongated zonally. The $W = 2^{3/4}\pi$ simulation is closest to the WGP solu-

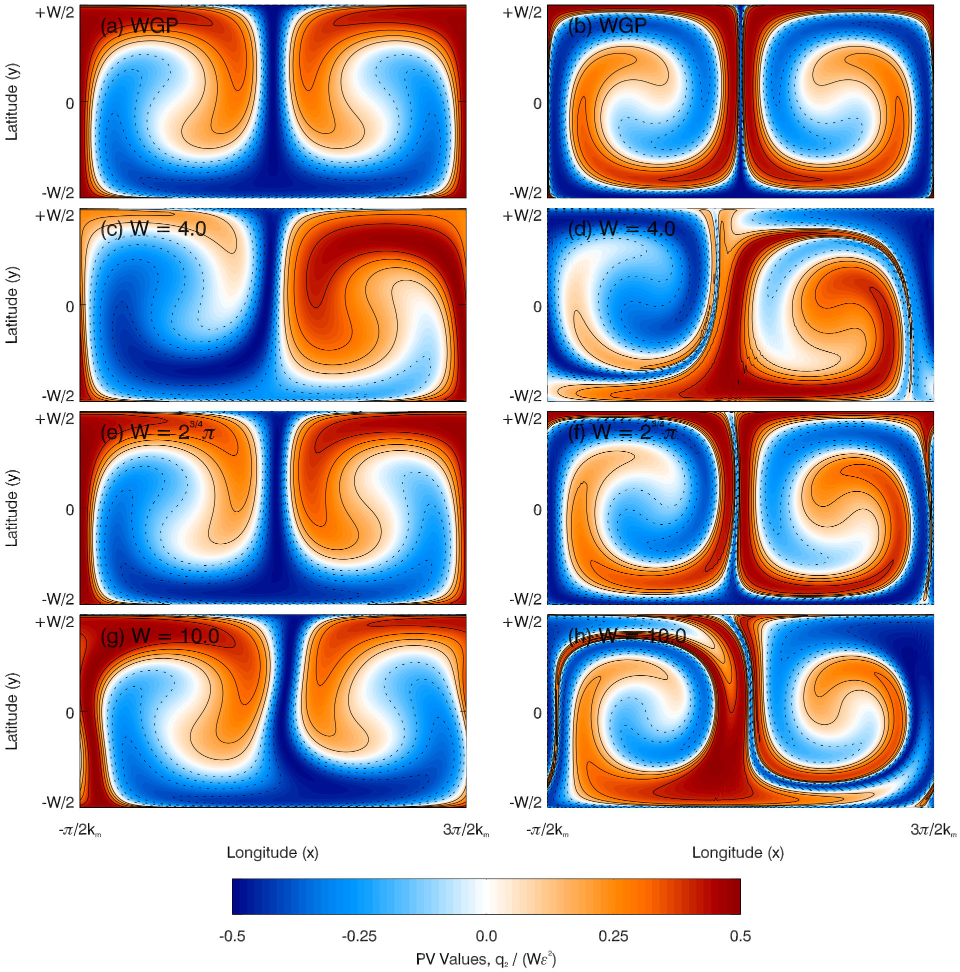

Figure 2.6.4: The (scaled) lower layer PV field $q_2/\epsilon^2 W$ at scaled times $\eta = 5.55W/k_m\pi$ (left panels) and $\eta = 10W/k_m\pi$ (right panels) for inverse criticality $\beta = 0.48$. Panels (a-b) show the WGP solution, panels (c-h) show the results of numerical experiments with: panels (c-d) $W = 4$, panels (e-f) $W = 2^{3/4}\pi$, (g-h) $W = 10$. The contour interval for all panels is 0.1.

tion (compare panels e-f to a-b in Fig. 2.6.4). The $W = 4$ simulation has a greater finite criticality asymmetry between its zonally elongated circulation cells and the vortex roll-up and barotropization resembles that seen for the $W = 2^{3/4}\pi$ simulations at much higher criticality (see panels (c,f) in Fig. 2.6.2). In the $W = 10$ simulation, the asymmetry within each circulation cell is very different. In Fig. 2.6.4 panel (h), the north-south branches are almost antisymmetric in x, but the east-west branches are not quite antisymmetric in y, suggesting that the asymmetry causing the discrepancy with WGP is rotated by $\pi/2$ compared to other simulations. In this case, the asymmetry does not lead to barotropization, but a more complex route to stirring and coarse-grain homogenization of the lower layer PV field.

2.6.3 Oscillation of wave amplitude in numerical simulations

The divergence between numerical simulations and the WGP solution at low supercriticality, observed in Fig. 2.6.1, is now investigated using the following diagnostics. A measure of the difference between the lower layer PV in simulations and the WGP solution is possible using the error measure

$$\mathcal{E}(\eta) = \frac{1}{LW^2}\|\epsilon^{-2}q_2 - Q\|_1, \tag{2.6.58}$$

where the norm $\|\cdot\|$ is the L^1 norm defined as

$$\|\cdot\|_1 = \int_D |\cdot| d^2\mathbf{x}. \tag{2.6.59}$$

The evolution of $\mathcal{E}(\eta)$ is plotted in the left-hand panel of Fig. 2.6.5 for $W = 2^{3/4}\pi$ and $\beta = 0.495, 0.48, 0.44, 0.40$. The error measure $\mathcal{E}(\eta)$ grows faster as

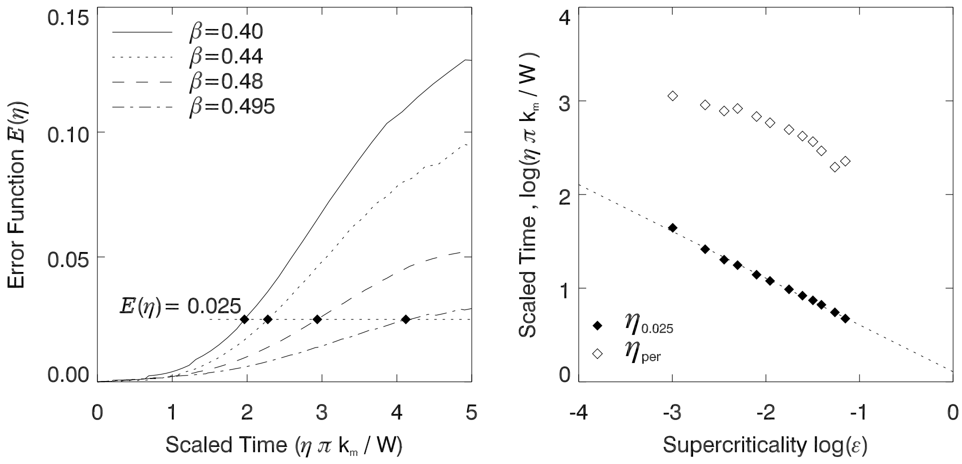

Figure 2.6.5: Left panel shows the evolution of the error measure $\mathcal{E}(\eta)$ for inverse criticalities $\beta = 0.40, 0.44, 0.48, 0.495$. The right panel shows a log-log plot of the growth rate of the error measure, determined from the scaled time $\eta_{0.025}$ defined in the text, and the period of the resulting oscillations η_{per} as a function of the supercriticality ϵ.

the supercriticality increases and at short times is quadratic in η. According to equation (2.6.57) the growth of the error is therefore exponential in time t.

Using the error measure (2.6.58), the time $\eta_{0.025}$ is defined to be the scaled time satisfying $\mathcal{E}(\eta_{0.025} k_m \pi / W) = 0.025$. The right panel of Fig. 2.6.5 shows the dependence of the time $\eta_{0.025}$ on the supercriticality ϵ. The slope of the fit in the log-log plot is almost exactly $-\frac{1}{2}$ and therefore the divergence of the numerical simulation from the WGP solution is on a scaled timescale $\eta \sim \epsilon^{-1/2}$, or in non-dimensional (units L_D/U), physical timescale $t \sim \epsilon^{-3/2}$. The WGP solution develops on timescales $\sim \epsilon^{-1}$ and the theory is formally valid out to times $\tau = \epsilon t = O(1)$. Therefore, the divergence of the numerical simulations from the WGP solution on timescales $O(\epsilon^{-3/2})$ is not inconsistent with the method of multiple scales used to obtain the WGP solution. Also plotted in the right-hand panel of Fig. 2.6.5 are estimates of the period η_{per} of the amplitude oscillations seen in Fig. 2.6.1 at late times. The log-log plot shows that η_{per} also scales approximately with $\epsilon^{-1/2}$ in scaled time or $\epsilon^{-3/2}$ in physical time t.

2.7 Assessment of pseudomomentum and pseudoenergy bounds

In section 2.5 bounds were derived on the wave amplitude (equations 2.5.51 and 2.5.52) and perturbation energy (equation 2.5.55 and 2.5.56) and an obvious question is how closely these are attained by numerical simulations. This is an important question as it concerns the extent to which baroclinic flows are controlled by their dynamical constraints, a topic of great importance to predictive theories such as that investigated in chapter 5.

Fig. 2.7.1 (top panels) shows a plot of the maximum upper layer wave amplitude $|A_1^f|_{max}/W$ achieved by numerical simulations (solid points), the equilibrated amplitudes for those same simulations (unfilled points), the Held-Shepherd amplitude bound (2.5.51) (solid line) and the predictions for WGP (dashed lines), each as a function of inverse criticality β. There are results for three widths $W = 4, 2^{3/4}\pi, 10$, which are plotted with different symbols (as labelled) and the corresponding results for the lower layer with the appropriate bound (2.5.52) are plotted in the lower panels.

At small supercriticalities, $0.4 \lesssim \beta \lesssim 0.5$, Fig. 2.7.1 confirms that the WGP solutions are very successful and become increasingly inaccurate with increasing supercriticality. The maximum wave amplitudes from simulations

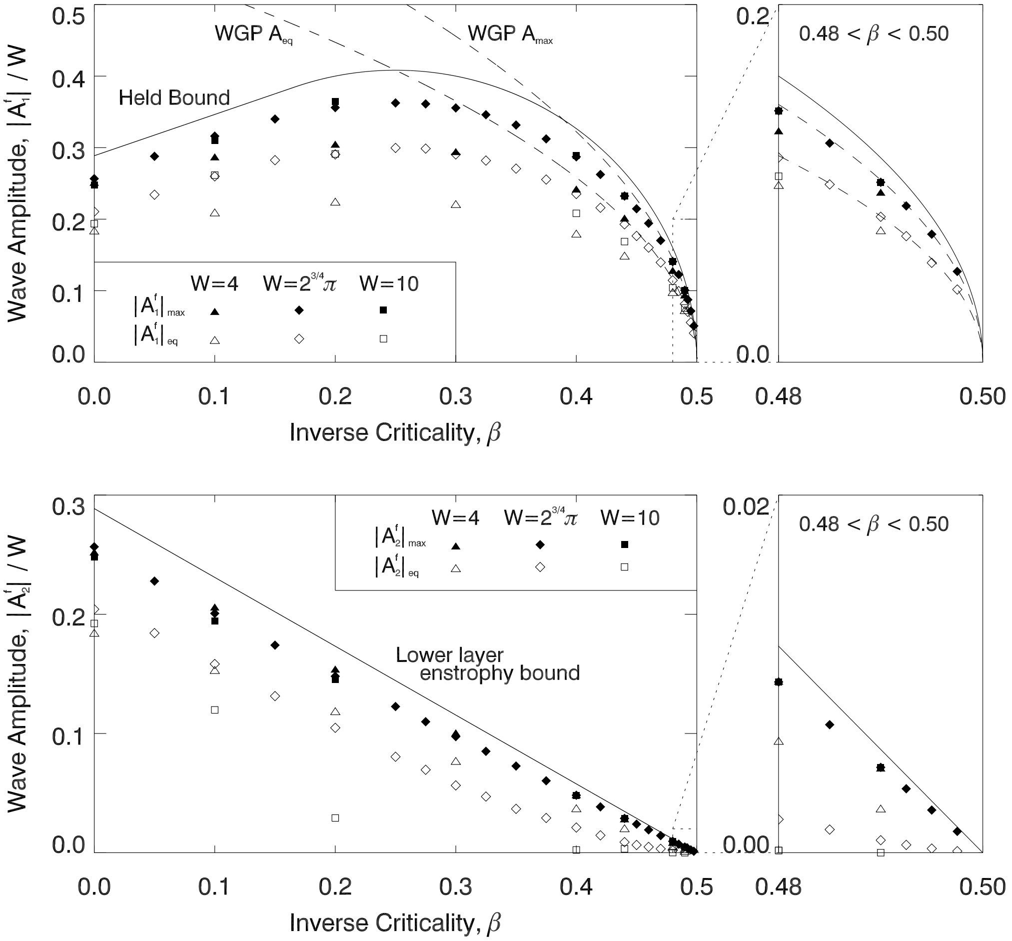

Figure 2.7.1: The maximum ($|A_i^f|_{max}/W$, solid symbols) and equilibrated ($|A_i^f|_{eq}$, unfilled symbols) wave amplitudes in the upper (top panel) and lower (bottom panel) layers from numerical simulations as a function of inverse criticality β. The dashed lines in the top panel show the WGP predictions $|A|_{eq}$ and $|A|_{max}$ defined in section 2.4. The solid lines are the amplitude bounds, (2.5.51) for the upper layer (top panel) and (2.5.52) for the lower layer (bottom panel), following from the Held-Shepherd bound derived in section 2.5.1. There are three symbols for numerical simulations corresponding to different width parameters $W = 4, 2^{3/4}\pi, 10$ as indicated. The right-hand panels are blow-ups of the left-hand panels for the smallest supercriticalities.

follow the Held-Shepherd bounds across the whole range of inverse criticalities and in the upper layer particularly seems to attain a fixed percentage ($\approx 88\%$) of the bound (2.5.51). This trend is less accurate for the narrower channel $W = 4$.

In Fig. 2.7.1 (lower panels), the lower layer wave amplitude $|A_2^f|_{eq}/W$ (unfilled points), allows for the diagnosis of the different equilibration mechanisms described in section 2.6.1. Under coarse-grain homogenization of the lower layer PV field, the lower layer wave amplitude must decay to zero as in the WGP solution. Alternatively, if the PV distribution rolls-up into coherent vortices in the lower layer, the equilibrated lower layer amplitude will be close to the bound (2.5.51) plotted as a solid curve in Fig. 2.7.1 (lower panel). The results confirm that PV homogenization occurs in the wide channel ($W = 10$) simulations and that the roll-up of vortices is efficient in the narrow ($W = 4$) channel, even at relatively low supercriticalities.

Fig. 2.7.2 shows the maximum (solid points) and equilibrated (unfilled points) perturbation energy (E'/LW^3) in the simulations as a function of inverse criticality. The bound (2.5.56) and Shepherd's bound (2.5.55) are also plotted in Fig. 2.7.2 (solid and dotted lines respectively). The wide channel ($W = 10$) numerical simulations in Fig. 2.7.2 almost attain the perturbation energy bound whereas the narrow ($W = 4$) simulations, however, only attain around 50% of the bound. It seems that the pseudoenergy bounds do not appear to contain a great amount of information about the lifecycle dynamics' dependence on W. The perturbation energy bounds are most useful in the widest channel where simulations show that almost all the energy is extracted

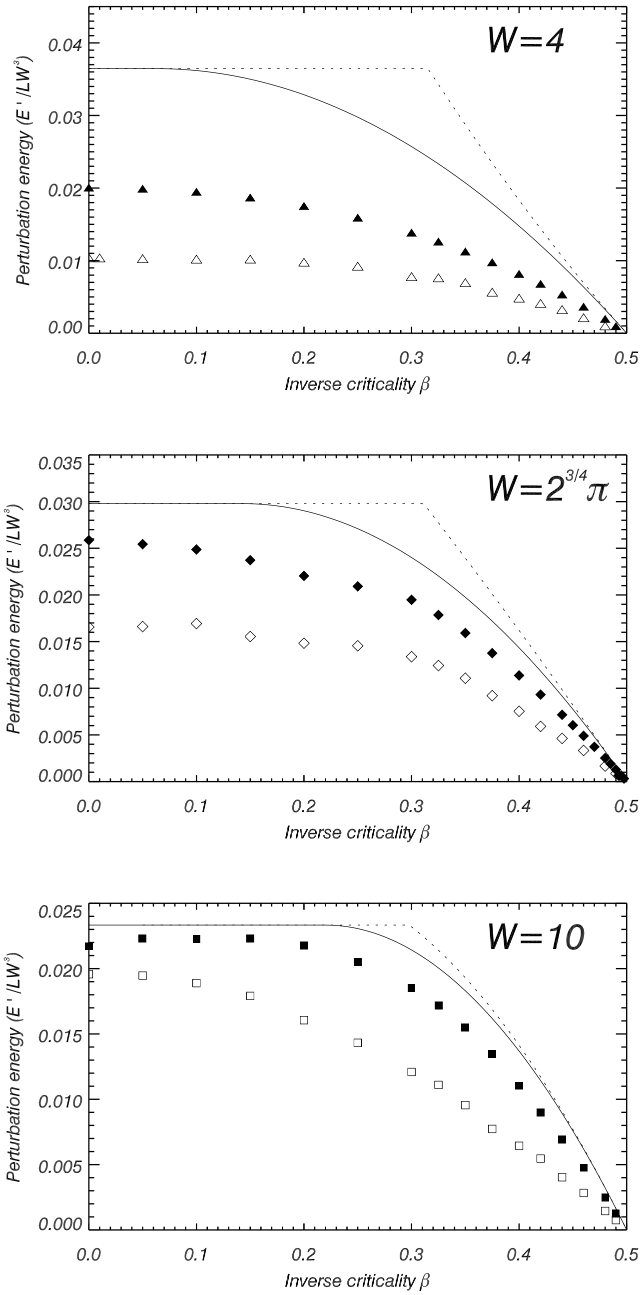

Figure 2.7.2: The maximum (solid points) and equilibrated (unfilled points) perturbation energy E'/LW^3 as a function of inverse criticality β in the simulations with channel widths $W = 4$ (upper panel), $W = 2^{3/4}\pi$ (middle panel) and $W = 10$ (lower panel). Also plotted are the pseudoenergy bounds (2.5.56, dotted curve) and (2.5.55, solid curve).

from the mean flow once the inverse criticality satisfies $\beta \lesssim \frac{1}{4}$.

2.8 Conclusions

This chapter has investigated nonlinear baroclinic equilibration in the Phillips model. The analytical solution of Warn & Gauthier [134] has been reviewed using a shorter derivation than originally put forward. The WGP solution, pseudoenergy and pseudomomentum bounds have been compared to high resolution numerical simulations. The questions put forward in the introduction to this chapter have been addressed as follows:

- At low, finite criticalities $\epsilon \ll 1$, the WGP solution is followed precisely up to a short time into the nonlinear evolution of the lower, critical layer. There was no formation of partial critical layers of the type reported by Gauthier [36] in any of the infinitesimal noise simulations reported here, though, they can be found in experiments where there is strong wavenumber discretization due to the effects of finite channel length. At low supercriticality, the WGP predictions for the maximum wave amplitude are very accurate and, provided the channel width is not too narrow, remain reasonable for $\beta \gtrsim 0.3$. At high supercriticality, the WGP solution is inaccurate (it is no longer formally valid) and contradicts the pseudomomentum bound on the maximum wave amplitude. The WGP solution also gives good predictions for the equilibrated wave amplitude, although in simulations at finite criticality the upper layer amplitude oscillates about the steady value predicted by WGP. The difference in the numerical simulations is the result of an asymmetry in the entrainment of PV into the positive and negative circulation cells in each half of the channel, which is not present in the WGP solution. The resulting oscillation in the upper layer fundamental wave amplitude has a period that scales with $\epsilon^{-3/2}$. There is not any evidence to suggest that the departure of numerical simulations from the analytical solution is the result of a barotropic instability of the kind in the Stewartson-Warn-Warn Rossby-wave critical layer reported by Killworth & McIntyre [60] and Haynes [40, 41].

- At large supercriticalities ($\beta \lesssim 0.3$) the equilibration of the flow is not due to coarse-grain homogenization of the lower layer PV field. Instead, the lower layer zonal mean PV gradient is removed by the formation of a train of alternately signed vortices, which are stable at long times. In the narrower channels (e.g. $W = 4$), the vortex roll-up is observed even at low supercriticalities, whereas for wide channels ($W = 10$), the PV homogenization mechanism occurs for all $\beta \gtrsim 0.15$.

- At low supercriticalities ($\beta \gtrsim 0.4$), the maximum wave amplitude in the upper layer $|A_1^f|_{max}$ is within approximately 12% of the Held-Shepherd pseudomomentum bound (2.5.51). At higher supercriticalities ($\beta < \frac{1}{6}$), the Held-Shepherd bound is an enstrophy bound and the maximum wave amplitude is within $8 - 14\%$ of this bound for the channel with width $W = 2^{3/4}\pi$. In the narrow channel, for which the roll up of vortices occurs across most of the parameter space, the maximum wave amplitudes are considerably less ($16-26\%$) than the bound on wave amplitude (2.5.51).

- The bound on perturbation energy (2.5.56) is an improvement on Shepherd's bound (2.5.55) and was found to be within 20% of numerical simulations with a wide channel ($W = 10$). The bounds were far from being attained in the narrow ($W = 4$) channel.

2.A Solution of the lower layer PV advection equation (2.4.41)

Equation (2.4.41) is

$$\left(\frac{\partial}{\partial\eta} - \frac{\partial\Psi}{\partial y}\frac{\partial}{\partial x} + \frac{\partial\Psi}{\partial x}\frac{\partial}{\partial y}\right) Q = 0,$$

where

$$\Psi = \cos\left(k_m x\right)\cos\left(\frac{\pi y}{W}\right),$$

and the initial condition is

$$Q(x, y, 0) = -y.$$

The homogeneous linear partial differential equation can be solved using the method of characteristics (e.g Riley et al. [104], section 18.6) as

$$Q(x(x_0, y_0, \eta), y(x_0, y_0, \eta), \eta) = -y_0,$$

where

$$\frac{dx}{d\eta} = -\frac{\partial\Psi}{\partial y}(x, y) \quad \text{and} \quad \frac{dy}{d\eta} = \frac{\partial\Psi}{\partial x}(x, y), \quad \text{with} \quad x(0) = x_0, y(0) = y_0. \tag{2.A.60}$$

Physically, a fluid parcel at coordinates (x, y) at (re-scaled) time η was initially (at $\eta = 0$) at coordinates (x_0, y_0), and conserves its initial PV value so that $Q(x, y, \eta) = -y_0$. Equations (2.A.60) are Hamilton's equations and since Ψ does not depend explicitly on the re-scaled time η, it is an invariant (i.e. the Hamiltonian). This recovers the fact that fluid particles in steady flows move along streamlines ($\Psi = \mathit{const.}$). The initial conditions then provide the identity

$$\Psi(x, y) = \cos(k_m x) \cos\left(\frac{\pi y}{W}\right) = \cos(k_m x_0) \cos\left(\frac{\pi y_0}{W}\right) = \Psi_0(x_0, y_0). \quad (2.A.61)$$

The identity (2.A.61) is used to re-write the second equation of (2.A.60) as

$$\frac{dy}{d\eta} = \mp k_m m_0^{1/2} \left(1 - m_0^{-1} \sin^2\left(\frac{\pi y}{W}\right)\right)^{1/2}, \quad y(0) = y_0, \quad (2.A.62)$$

where $m_0 = 1 - \Psi_0^2$ is treated as a constant. Take the negative branch in the subdomain $x \in [-\pi/2k_m, \pi/2k_m]$ and then the positive branch in the subdomain $x \in [\pi/2k_m, 3\pi/2k_m]$ to ensure that the lower layer flow consists of two disjoint circulation cells (with oppositely signed circulation in each cell as illustrated in panel a of Fig. 2.4.1). If focus is restricted to the cell in $(x, y) \in [-\pi/2k_m, \pi/2k_m] \times [-W/2, W/2]$, the full solution can be easily constructed using the symmetry

$$Q\left(x + \frac{\pi}{2k_m}, y, t\right) = Q\left(\frac{\pi}{2k_m} - x, y, t\right).$$

Equation (2.A.62) is integrated using the definition of elliptic integrals (e.g. Abramowitz & Stegun [1]), which gives

$$F\left(\sin^{-1}\left(m_0^{-1/2} \sin\left(\frac{\pi y}{W}\right)\right) | m_0\right) + \frac{k_m \pi}{W} \eta = F\left(\sin^{-1}\left(m_0^{-1/2} \sin\left(\frac{\pi y_0}{W}\right)\right) | m_0\right), \quad (2.A.63)$$

where

$$F(z|m) = \int_0^z \frac{1}{\left(1 - m\sin^2 t\right)^{1/2}} dt$$

is the incomplete elliptic integral of the first kind and (2.A.63) uses the fact that

$$m^{-1/2} F(z|m^{-1}) = F(\sin^{-1}(m^{-1/2} \sin z)|m).$$

The equation (2.A.63) is used to obtain the latitude of a fluid parcel y in terms of its initial position (x_0, y_0). An alternative is to notice that $m = m(\Psi)$ is an invariant, and so substitution of $m = 1 - \Psi^2 = 1 - \cos^2(k_m x)\cos^2(\pi y/W)$ for m_0 in (2.A.63) allows y_0 to be found in terms of the current position (x, y). The initial latitude y_0 provides the solution to equation (2.4.41) through $Q(x, y, \eta) = -y_0$. Then apply the Jacobi elliptic function $\mathrm{sn}(\cdot|m)$ to (2.A.63) and use the standard identities

$$\begin{aligned}
\mathrm{sn}(F(z|m)|m) &= \sin z, \\
\mathrm{cn}(F(z|m)|m) &= \cos z, \\
\mathrm{dn}(F(z|m)|m) &= (1 - m\sin^2 z)^{1/2}, \\
\mathrm{sn}(u + v|m) &= \frac{\mathrm{sn}(u|m)\mathrm{cn}(v|m)\mathrm{dn}(v|m) + \mathrm{sn}(v|m)\mathrm{cn}(u|m)\mathrm{dn}(u|m)}{1 - m\mathrm{sn}^2(u|m)\mathrm{sn}^2(v|m)},
\end{aligned}$$

to recover the solution

$$Q(x, y, \eta) = -\frac{W}{\pi}\sin^{-1}\left(\frac{\sin\left(\frac{\pi y}{W}\right)\mathrm{cn}\left(\frac{k_m \pi}{W}|m\right)\mathrm{dn}\left(\frac{k_m \pi}{W}|m\right) + \sin(k_m x)\cos^2\left(\frac{\pi y}{W}\right)\mathrm{sn}\left(\frac{k_m \pi}{W}|m\right)}{1 - \sin^2\left(\frac{\pi y}{W}\right)\mathrm{sn}^2\left(\frac{k_m \pi}{W}|m\right)}\right).$$

This is equation (2.4.43) in the main text.

The numerical quadrature of the integral (2.4.38) is achieved through the coordinate transformation $(x, y) \to (\alpha, m)$ where $m(x, y)$ is defined by (2.4.44)

and

$$\alpha(x,y) = \frac{k_m\pi}{W}\eta + F\left(\sin^{-1}\left(m^{-1/2}\sin\left(\frac{\pi y}{W}\right)\right)|m\right). \tag{2.A.64}$$

Applying the transformation (2.4.44) and (2.A.64) to (2.4.43) gives

$$Q = -\frac{W}{\pi}\sin^{-1}\left(m^{1/2}\mathrm{sn}\left(\alpha|m\right)\right).$$

Taking sn(·) of (2.A.64) and rearranging gives

$$y = \frac{W}{\pi}\sin^{-1}\left(m^{1/2}\mathrm{sn}\left(\alpha - \frac{k_m\pi}{W}\eta|m\right)\right).$$

Together with the Jacobian

$$\frac{\partial x}{\partial \alpha}\frac{\partial y}{\partial m} - \frac{\partial x}{\partial m}\frac{\partial y}{\partial \alpha} = \frac{W}{2k_m\pi}\left(1-m\right)^{-1/2},$$

the integral in (2.4.38) can be written as

$$|A(\tau)|^2 = W^2\left(\frac{2}{3} - \frac{16}{\pi^4}I\left(\frac{k_m\pi}{W}\eta\right)\right),$$

which is equation (2.4.45) and $I(z)$ is given in the main text by (2.4.46). The limit $K(m) = F(\pi/2|m)$ corresponds to integration over a quarter of a circulation cell (defined above), with the integral over the full cell being four times greater than this due to symmetries in the x and y directions.

2.B A bound on perturbation enstrophy

A bound on the perturbation enstrophy specific to the two-layer model will be derived following the work of Shepherd [116]. The bound is derived by exploiting a Liapunov stability theorem for subcritical flows, itself a generalization of the Charney-Stern theorem to finite amplitude disturbances. In fact, in the two-layer model, if the initial disturbance is infinitesimal then an even tighter bound on the perturbation enstrophy and wave amplitude may be found.

Start with the two-layer quasi-geostrophic equations (2.2.3-2.2.4) with the boundary conditions (2.2.5). The initial potential vorticity is given by (2.2.2) plus an infinitesimal non-zonal perturbation. As has been discussed, if $\beta < \frac{1}{2}$ then the flow is unstable to baroclinic instability. Consider the conservation of energy E and momentum M and the existence of conserved Casimir quantities $\mathcal{C}$ which may be written as

$$\begin{aligned} M &= \int_{\mathcal{D}} y\left(Q_1 + Q_2\right) d^2\mathbf{x}, && (2.B.65)\\ E &= \int_{\mathcal{D}} \frac{1}{2}\left(\left|\nabla\Psi_1\right|^2 + \left|\nabla\Psi_2\right|^2\right) + \frac{1}{4}\left(\Psi_1 - \Psi_2\right)^2 d^2\mathbf{x}, && (2.B.66)\\ \mathcal{C} &= \int_{\mathcal{D}} C_1[q_1] + C_2[q_2] d^2\mathbf{x}, && (2.B.67) \end{aligned}$$

where the integral is over the channel domain $L \times W$. Equation (2.B.67) holds for any choice of functions C_1 and C_2. A particularly useful choice is the layer weighted enstrophy

$$Z(\lambda) = \frac{1}{2}\int_{\mathcal{D}} q_1^2 + \lambda q_2^2 d^2\mathbf{x}.$$

The bound described in Shepherd [116] is attributed to Isaac Held and will be described here as the Held-Shepherd bound. It is a bound of the perturbation enstrophy

$$Z' = \frac{1}{2}\int_{\mathcal{D}} (q_1')^2 + (q_2')^2 d^2\mathbf{x} \tag{2.B.68}$$

where $q_i' = Q_i - \overline{Q}_i$ and it is constructed from the conserved quantities $Z(\lambda)$

and M, which from (2.2.6) and (2.2.8) are

$$\begin{aligned} Z(\lambda) &= \frac{1}{24}LW^3\left(\left(\beta+\frac{1}{2}\right)^2+\lambda\left(\beta-\frac{1}{2}\right)^2\right), \\ M &= \frac{1}{6}LW^3\beta. \end{aligned}$$

Now note that

$$\begin{aligned} Z' &\leq \tfrac{1}{2}\int_{\mathcal{D}}(q_1')^2+\lambda(q_2')^2d^2\mathbf{x} && \text{for } \lambda \geq 1, \\ &\leq \tfrac{1}{2}\int_{\mathcal{D}}(Q_1-\mu y)^2+\lambda(Q_2-\tfrac{\mu}{\lambda}y)^2d^2\mathbf{x} && \text{for any } \mu, \\ &= Z(\lambda)-\mu M+\tfrac{1}{24}LW^3\tfrac{\mu^2(\lambda+1)}{\lambda}, \\ &= \tfrac{1}{24}LW^3\left(\left(\beta+\tfrac{1}{2}\right)^2+\lambda\left(\beta-\tfrac{1}{2}\right)^2-4\mu\beta+\tfrac{\mu^2(\lambda+1)}{\lambda}\right). \end{aligned} \tag{2.B.69}$$

This bound holds for $\lambda \in [1,\infty)$ and $\mu \in (-\infty,\infty)$. The tightest possible bound is at $(\lambda,\mu)=(\lambda_c,\mu_c)$ where (λ_c,μ_c) is a critical point of the function

$$f(\lambda,\mu)=\left(\beta+\frac{1}{2}\right)^2+\lambda\left(\beta-\frac{1}{2}\right)^2-4\mu\beta+\frac{\mu^2(\lambda+1)}{\lambda}.$$

Solving this it is straightforward to show that for $\beta > \frac{1}{4}$ the relevant critical point is

$$(\lambda_c,\mu_c)=\left(\frac{3\beta-\frac{1}{2}}{\frac{1}{2}-\beta},3\beta-\frac{1}{2}\right),$$

and for $\beta \geq \frac{1}{4}$ the tightest bound is at

$$(\lambda_c,\mu_c)=(1,\beta).$$

Therefore, the bounds are

$$Z' \leq \frac{1}{48}LW^3\begin{cases} 16\beta\left(\frac{1}{2}-\beta\right) & \frac{1}{4}\leq\beta\leq\frac{1}{2}, \\ 1 & \beta>\frac{1}{4}. \end{cases} \tag{2.B.70}$$

This is Held's result (substitute $U=1$, $F=\frac{1}{2}$ into equation (5.6) of [116] with $W=1$ and divide by L to get zonal means here).

More useful in this work is a bound on the upper layer perturbation enstrophy from which it is possible to get a bound on the fundamental wave amplitude in the upper layer.

$$Z_1' = \frac{1}{2}\int_{\mathcal{D}} (q_1')^2 d^2\mathbf{x}.$$

This may be derived in much the same way noting that in (2.B.69), λ may now take any value in the range $[0,\infty)$ for the first inequality to be true when Z' is replaced with Z_1'. The tightest bound for $\beta < \frac{1}{6}$ then comes for $(\lambda,\mu) = (0,0)$ giving the result

$$Z_1' \leq \frac{1}{48} LW^3 \begin{cases} 16\beta\left(\frac{1}{2}-\beta\right) & \frac{1}{6} \leq \beta \leq \frac{1}{2}, \\ 2\left(\beta+\frac{1}{2}\right)^2 & \beta < \frac{1}{6}. \end{cases} \tag{2.B.71}$$

The Held-Shepherd bound on enstrophy can be re-written as a bound on the amplitude of the fundamental in the upper layer using Bessel's inequality for Fourier series. The upper layer enstrophy can be written as the sum of the Fourier coefficients of the potential vorticity q_1' and since we are dealing with a plane wave solution (2.3.13), the Fourier coefficients are simply the amplitudes of the upper layer waves at each wavenumber. Writing A_{kl} as the amplitude of a wave with wavenumber (k,l) and calculating the Fourier coefficients gives

$$|A_1^f|^2 = |A_{11}|^2 \leq \sum_{k,l} |A_{kl}|^2 \leq \frac{8}{LW} Z_1' \tag{2.B.72}$$

Equation (2.B.71) may then be substituted into (2.B.72) to get

$$|A_1^f|^2 \leq \frac{W^2}{6} \begin{cases} 16\beta\left(\frac{1}{2}-\beta\right) & \frac{1}{6} \leq \beta \leq \frac{1}{2}, \\ \left(\beta+\frac{1}{2}\right)^2 & \beta < \frac{1}{6}. \end{cases} \tag{2.B.73}$$

This is equation (2.5.51) in the main text. It is also possible to form a bound on the lower layer wave amplitude $|A_2^f|$ from the lower layer enstrophy:

$$Z_2' \leq Z_2 = \frac{1}{24} L W^3 \left(\frac{1}{2} - \beta \right)^2 , \tag{2.B.74}$$

and so

$$|A_2^f|^2 \leq \frac{8 Z_2'}{LW} \leq \frac{1}{3} W^2 \left(\frac{1}{2} - \beta \right)^2 . \tag{2.B.75}$$

This is equation (2.5.52) in the main text.

2.C Pseudoenergy bounds on wave energy

The pseudoenergy bound follows the work of Shepherd (1993) [118] and aims to find a bound on the total energy of the perturbation, given by

$$E' = \frac{1}{2} \int_D |\nabla \psi_1'|^2 + |\nabla \psi_2'|^2 + \frac{1}{2} \left(\psi_1' - \psi_2' \right) d^2\mathbf{x}.$$

A 'basic state' streamfunction, $\Psi_i(y)$ for $i = 1, 2$, is introduced and from the symmetry in the problem is assumed to be an odd function of y satisfying the boundary conditions

$$\Psi_i = \mp \alpha_i W \quad \text{on} \quad y = \pm \frac{W}{2},$$

where the constants α_i are as yet undetermined. Each constant α_i corresponds to the cross-channel mean of the basic state $U_i = -\Psi_{iy}$ in layer i.

The basic streamfunction Ψ_i is used to form the following inequality

$$\begin{aligned}
E' &= \frac{1}{2}\int_D |\nabla\psi_1'|^2 + |\nabla^2\psi_2'|^2 + \frac{1}{2}\left(\psi_1' - \psi_2'\right) d^2\mathbf{x}, \\
&\leq \frac{1}{2}\int_D |\nabla(\psi_1 - \Psi_1)|^2 + |\nabla(\psi_2 - \Psi_2)|^2 + \frac{1}{2}\left((\psi_1 - \psi_2) - (\Psi_1 - \Psi_2)\right) d^2\mathbf{x}, \\
&= E - \int_D \nabla\psi_1 \cdot \nabla\Psi_1 + \nabla\psi_2 \cdot \nabla\Psi_2 + \frac{1}{2}\left(\psi_1 - \psi_2\right)\left(\Psi_1 - \Psi_2\right) d^2\mathbf{x} \\
&\quad + \frac{1}{2}\int_D \left(\Psi_{1y}\right)^2 + \left(\Psi_{2y}\right)^2 + \frac{1}{2}\left(\Psi_1 - \Psi_2\right)^2 d^2\mathbf{x}, \\
&= E + \int_D \left(\Psi_1 q_1 + \Psi_2 q_2\right) d^2\mathbf{x} - \alpha_1 LW \\
&\quad + \frac{1}{2}\int_D \left(\Psi_{1y}\right)^2 + \left(\Psi_{2y}\right)^2 + \frac{1}{2}\left(\Psi_1 - \Psi_2\right)^2 - 2\beta y\left(\Psi_1 + \Psi_2\right) d^2\mathbf{x}.
\end{aligned} \tag{2.C.76}$$

The first term in the integrand of (2.C.76) is the only one that is not easily bounded. A similar approach is used to find a bound on $\nu Z'$ where $\nu > 0$ is a constant and Z' is the perturbation enstrophy.

$$\begin{aligned}
\nu Z' &= \frac{1}{2}\nu \int_D (q_1')^2 + (q_2')^2 d^2\mathbf{x}, \\
&\leq \frac{1}{2}\nu \int_D (q_1')^2 + \lambda(q_2')^2 d^2\mathbf{x} \qquad\qquad \text{for} \quad \lambda \geq 1, \\
&\leq \frac{1}{2}\nu \int_D \left(q_1 - \frac{1}{\nu}\left(\Psi_1 + \mu y\right)\right)^2 + \lambda(q_2 - \frac{1}{\lambda\nu}\left(\Psi_2 + \mu y\right))^2 d^2\mathbf{x}, \\
&= \nu Z_1 + \nu\lambda Z_2 - \mu M - \int_D \left(\Psi_1 q_1 + \Psi_2 q_2\right) d^2\mathbf{x} \\
&\quad + \frac{1}{2\nu}\int_D \left(\Psi_1 + \mu y\right)^2 + \frac{1}{\lambda}\left(\Psi_2 + \mu y\right)^2 d^2\mathbf{x}.
\end{aligned} \tag{2.C.77}$$

Adding the inequalities (2.C.76) and (2.C.77) eliminates the integral that is 'difficult-to-bound' so that

$$\begin{aligned}
E' + \nu Z' \leq\ & E + \nu Z_1 + \nu\lambda Z_2 - \mu M - \alpha_1 LW \\
& + \frac{1}{2}\int_D \left(\Psi_{1y}\right)^2 + \left(\Psi_{1y}\right)^2 + \frac{1}{2}\left(\Psi_1 - \Psi_2\right)^2 \\
& \qquad - 2\beta y\left(\Psi_1 + \Psi_2\right) + \frac{1}{\nu}\left(\Psi_1 + \mu y\right)^2 + \frac{1}{\lambda\nu}\left(\Psi_2 + \mu y\right)^2 d^2\mathbf{x}
\end{aligned} \tag{2.C.78}$$

where M, E, Z_1 and Z_2 are given by their initial values, as in appendix 2.B.

The bound derived by Shepherd is found by considering a uniform basic state, or

$$\Psi_1 = -\alpha_1 y, \qquad \Psi_2 = -\alpha_2 y,$$

where the integral in (2.C.78) is calculated to be

$$\begin{aligned} E' + \nu Z' \;\le\; & \frac{1}{2} LW\left((1-\alpha_1)^2 + \alpha_2^2\right) + \frac{1}{48}\left(1 + 2\nu\left(\beta + \frac{1}{2}\right)^2 + 2\lambda\nu\left(\beta - \frac{1}{2}\right)^2\right. \\ & \left. + 4\beta\left(\alpha_1 + \alpha_2 - 2\mu\right) + 2\left(\alpha_1 - \alpha_2\right)^2 + \frac{4}{\nu}\left(\alpha_1 - \mu\right)^2 + \frac{4}{\lambda\nu}\left(\alpha_2 - \mu\right)^2\right). \end{aligned} \tag{2.C.79}$$

It is possible to minimize the bound with respect to the parameters $(\alpha_1, \alpha_2, \mu, \nu, \lambda)$ over the domain $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times (0, \infty) \times [1, \infty)$. Equation (2.C.79) is quadratic in α_1, α_2 and μ and so the optimal bound with respect to these parameters is found by solving a linear system, which occurs at the critical values

$$\begin{aligned} \alpha_{1,2c} &= \frac{1}{2} \pm \frac{\nu\left(6 - \beta W^2 + \lambda(6 + \beta W^2)\right)}{2W^2 + \nu(1+\lambda)(12 + W^2)}, \\ \mu_c &= \frac{2W^2 + 4\nu^2\lambda\beta(12 + W^2) + \nu((2\beta + 1)W^2 + \lambda(24 + W^2 + 2\beta W^2))}{2(2W^2 + \nu(1+\lambda)(12 + W^2))}. \end{aligned} \tag{2.C.80}$$

Substitution of the critical values (2.C.80) into the inequality (2.C.79) yields

$$E' + \nu Z' \le \frac{LW^3}{24} \frac{\left(1 + \nu\left(\beta + \frac{1}{2}\right) - \nu\lambda\left(\beta - \frac{1}{2}\right)\right)^2 (12 + W^2)}{2W^2 + \nu(1+\lambda)(12 + W^2)}. \tag{2.C.81}$$

Differentiation of (2.C.81) with respect to λ and ν shows that the critical point with respect to these variables lies on the curve

$$\lambda_c(\nu) = \frac{24 + (8\beta - 2)W^2 + \nu(6\beta - 1)(W^2 + 12)}{2\nu\left(\frac{1}{2} - \beta\right)(W^2 + 12)},$$

which in the inequality (2.C.81) gives

$$E' + \nu Z' \leq \frac{LW^3}{6}\left(\frac{1}{2} - \beta\right)\left(1 - \frac{2\left(\frac{1}{2} - \beta\right)W^2}{W^2 + 12} + 2\beta\nu\right). \tag{2.C.82}$$

Recall the constraint $\lambda \geq 1$, which means that the critical point only provides the optimal bound on E' when $\nu \to 0$ on condition that

$$\beta \geq \frac{1}{4} - \frac{3}{W^2}.$$

Otherwise the lower bound occurs at $(\lambda, \mu) = (1, 0)$, which gives

$$E' \leq \frac{LW}{48}\left(W^2 + 12\right). \tag{2.C.83}$$

Combining the results (2.C.82) and (2.C.83) gives the bound on E'

$$E' \leq \frac{LW^3}{6}\begin{cases} \left(\frac{1}{2} - \beta\right)\left(1 - \frac{2\left(\frac{1}{2} - \beta\right)W^2}{W^2+12}\right) & \frac{1}{4} - \frac{3}{W^2} \leq \beta < \frac{1}{2}, \\ \frac{1}{8}\left(1 + \frac{12}{W^2}\right) & 0 < \beta < \frac{1}{4} - \frac{3}{W^2}. \end{cases} \tag{2.C.84}$$

This is equation (2.5.56) in the main text. The bound (2.C.84) may be compared with the corresponding bound (8.2) in Shepherd (1993) [118], which is

$$E' \leq \frac{LW^3}{6}\begin{cases} \left(\frac{1}{2} - \beta\right)\left(1 - \frac{48-2W^2}{W^2}\left(\frac{1}{2} - \beta\right)\right) & \frac{1}{4} - \frac{3}{W^2} \leq \beta < \frac{1}{2} \\ \frac{1}{8}\left(1 + \frac{12}{W^2}\right) & 0 < \beta < \frac{1}{4} - \frac{3}{W^2}. \end{cases} \tag{2.C.85}$$

This is equation (2.5.55) in the main text. The bound (2.C.84) is an improvement upon (2.C.85).

Chapter 3

Mechanisms of instability and dissipative destabilization

3.1 Introduction

In chapter 2 baroclinic instability in the absence of Ekman friction was studied in the context of Phillips' two-layer model [95, 96]. The relevant linear theory was described and a curve of marginal stability $\beta_c^I(a)$ (equation 2.3.16) was established. It was stated that the condition for instability corresponds to a reversal of the PV gradient in the lower layer allowing the instability mechanism to be described in terms of counter-propagating Rossby waves (CRWs). This 'PV thinking' approach was first put forward for a two-layer model by Bretherton [14]. Above the stability threshold (2.3.17), $\beta < \beta_m^I = \frac{1}{2}$, the background PV gradient in the lower layer is negative, whilst in the upper layer it is positive. Therefore, relative to the mean flow, Rossby waves propagate westwards in the upper layer and eastwards in the lower layer. The Rossby waves can enter into a configuration where the velocity field induced by one acts to amplify the other. Below the stability threshold $\beta > \frac{1}{2}$, the PV gradient in each layer is positive, Rossby waves propagate westwards in both layers and the flow is stable (recall that this corresponds precisely to the Charney-Stern-Pedlosky condition for instability [18] in the Phillips model, see e.g. Vallis [133] p.275-277). Bretherton's description of interacting Rossby waves in regions of oppositely signed PV gradients is also relevant to more general models of geophysical flows as discussed by Hoskins et al. [55].

A natural next step is to investigate the effects of Ekman friction at the top and bottom boundaries of the channel. The linear problem in the Phillips model with Ekman friction was first studied by Holopainen [53]. Holopainen discovered that the presence of Ekman friction provides a new instability mechanism, which will hereafter be referred to as the 'Holopainen mechanism'. Waves that are otherwise stable when no Ekman friction is present can be 'dissipatively destabilized' by the Holopainen mechanism. The result is an order one displacement of the marginal stability curve in parameter space, which importantly, is robust in the limit of vanishing Ekman friction. Swaters [130] offered a modal, kinematic wave description of this somewhat counter-intuitive effect of Ekman friction and remarks that the 'physical reason for the Ekman-induced destabilization of inviscidly stable baroclinic quasi-geostrophic flow has yet to be given'.

More generally, it is known that the dissipative destabilization mechanism is not simply a peculiarity of Phillips' model and has been described in a dynamical systems framework by Krechetnikov and Marsden [62, 63]. The specific case of the Holopainen mechanism is discussed as part of wider set of dynamical systems using the underlying Hamiltonian structure of the evolution equations for various models (note that dissipative destabilization is referred to here as the Holopainen instability mechanism when specifically referring to Phillips' model). In a more realistic model of a jet-like flow, Lee [67] diagnoses dissipative destabilization by surface Ekman layers as providing an eddy energy source for baroclinic waves. Dissipative destabilization has also been reported in the context of the Eady model [24] by Weng [135] and Weng and Barcilon [136] following earlier studies of Ekman friction effects in heated, rotating annulus experiments by Barcilon [6]. In the Eady model, dissipative destabilization can only occur in the presence of sloping boundaries, which are analogous to the β-effect in the Phillips model (known as the topographic β effect). In other models, the dissipative destabilizing effects are more complicated [71, 88], but an understanding of the dissipative destabilization mechanism is necessary for understanding the stability properties of a wider context of baroclinic flows and therefore Phillips' model is a sensible starting point for such investigations.

Can the Holopainen mechanism be understood using the 'PV thinking' framework that uncovered Bretherton's CRW mechanism for baroclinic instability? Clearly the same mechanism cannot be at work since it turns out that there is not necessarily a reversal of the PV gradient when the flow is unstable to Holopainen instability. A useful framework with which to investigate the linear CRW mechanism was introduced by Bishop [9, 10], and later extended by Davies & Bishop [19] and Heifetz et al. [42, 43, 44] for a range of geophysical flows. Using this approach, the PV disturbances associated with the Rossby wave in each layer are written as perturbations about the background flow proportional to $e^{i(kx+\theta(t))}\cos(ly)$, i.e. modal solutions with a time dependent phase, which reveal a fruitful perspective on the linear dynamics of the interacting Rossby waves. Modal solutions will be applied here to the Holopainen mechanism to highlight the similarities and differences with the CRW mechanism. This novel application of Bishop's PV oriented approach will then be compared to the description of the Holopainen mechanism put forward by Swaters [130]. The dissipative destabilization in the Eady model will also be studied using the PV approach and shown to be exactly analogous to that of Phillips' model.

The plan for this chapter is as follows. Section 3.2 will discuss linear theory and dissipative destabilization by Ekman friction in Phillips' model. Section 3.3 will describe the method of Bishop using frictionless baroclinic instability in the specific case of Phillips' model. This method will be extended to consider the Holopainen mechanism, including a comparison to Swaters' approach. The analogous mechanism of dissipative destabilization in the Eady model is presented in section 3.4 before conclusions are drawn in section 3.5.

3.2 Linear theory and dissipative destabilization by Ekman friction

The equations of motion (1.3.64-1.3.65) for the Phillips model including Ekman friction, with a uniform background velocity in the upper layer and zero velocity in the lower layer, were derived in section1.3.1 and are repeated here for convenience

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\left(\nabla^2\psi_1'-\frac{\psi_1'-\psi_2'}{2}\right)+\left(\beta+\frac{1}{2}\right)\frac{\partial\psi_1'}{\partial x}+\kappa\nabla^2\psi_1' \quad = \quad -J\left(\psi_1',q_1'\right), \tag{3.2.1}$$

$$\left(\frac{\partial}{\partial t}\right)\left(\nabla^2\psi_2'+\frac{\psi_1'-\psi_2'}{2}\right)+\left(\beta-\frac{1}{2}\right)\frac{\partial\psi_2'}{\partial x}+\kappa\nabla^2\psi_2' \quad = \quad -J\left(\psi_2',q_2'\right), \tag{3.2.2}$$

where the Jacobian is defined as $J(\psi,\phi)=\psi_x\phi_y-\psi_y\phi_x$ and the boundary conditions are

$$\begin{aligned}\frac{\partial^2\overline{\psi'}_i}{\partial t\partial y}+\kappa\frac{\partial\overline{\psi'}_i}{\partial y} &= 0, &&\text{on} \quad y=\pm\frac{W}{2},\\ \frac{\partial\psi_i'}{\partial x} &= 0, &&\text{on} \quad y=\pm\frac{W}{2},\\ \psi_i'(x,y,t) &= \psi_i'(x+L,y,t), && \end{aligned} \tag{3.2.3}$$

for $i=1,2$ with $\overline{(\ldots)}$ denoting a zonal mean.

3.2.1 Linear stability

Following Holopainen [53] (see also Romea [109]) and in parallel to section 2.3, the linear stability of (3.2.1-3.2.2) is investigated by looking for normal mode solutions of the form

$$\begin{pmatrix}\psi_1'\\ \psi_2'\end{pmatrix}=Re\ \tilde{\epsilon}\begin{pmatrix}1\\ \gamma\end{pmatrix}e^{ik(x-ct)}\cos ly, \tag{3.2.4}$$

where k and l $(=\pi/W$ in a channel of width W) are the zonal and meridional wavenumbers, $\tilde{\epsilon}$ is the initial wave amplitude and γ is the complex-valued phase difference between the streamfunctions in the upper and lower layer

($a = \sqrt{k^2 + l^2}$ is again the total wavenumber). The ansatz (3.2.4) is substituted into the equations of motion (3.2.1-3.2.2) and neglecting terms of $O(\tilde{\epsilon}^2)$ (arising from the Jacobian terms on the right-hand side) leads to the dispersion relation for the complex phase speed.

$$\begin{aligned} c^{\pm} &= \frac{1}{2} - \frac{\left(a^2 + \frac{1}{2}\right)}{a^2(a^2+1)}\beta - i\kappa\frac{a^2 + \frac{1}{2}}{k(a^2+1)} \\ &\quad \pm \frac{1}{2a^2(a^2+1)}\left(\beta^2 + a^4(a^4-1) - \frac{a^4\kappa^2}{k^2} + 2i\frac{a^2\kappa}{k}\beta\right)^{\frac{1}{2}}. \end{aligned} \tag{3.2.5}$$

The solution also specifies that the complex streamfunction ratio is

$$\gamma^{\pm} = 2\left(a^2 + \frac{1}{2}\right) + 2\left(\frac{\beta + \frac{1}{2} + i\frac{a^2\kappa}{k}}{c^{\pm} - 1}\right). \tag{3.2.6}$$

When Ekman friction is included in the problem, $\gamma^{\pm}$ is complex valued at every wavenumber and hence there is a leading order phase shift between the waves in each layer. This is in contrast to the WGP problem of chapter 2 for which the neutral waves are in phase. This distinction is important for the derivation of a weakly nonlinear theory and is the subject of chapter 4.

Instability occurs when the imaginary part of the phase speed is positive, which may only occur for the positive branch of the dispersion relation c^+, i.e. $\mathrm{Im}\,(c^+ > 0)$. The marginal stability curve $\beta_c^R(a,\kappa)$ is the threshold between stability and instability and is found by setting $\mathrm{Im}\,(c^{\pm}) = 0$ in (3.2.5). A little algebra leads to the condition for instability.

$$\beta < \beta_c^R(a,\kappa) = \left(a^2 + \frac{1}{2}\right)\left(a^2\left(1 - a^2\right) - \frac{4a^4}{k^2}\kappa^2\right)^{1/2}. \tag{3.2.7}$$

The marginal stability curve (3.2.7) in the limit of vanishing Ekman friction ($\kappa \to 0$) is given by

$$\beta_c^R(a,0) = \left(a^2 + \frac{1}{2}\right)\left(a^2\left(1 - a^2\right)\right)^{1/2}. \tag{3.2.8}$$

3.2.2 The Holopainen instability

Recall that the criterion for instability in the frictionless problem is given by the marginal stability curve (2.3.16) as

$$\beta < \beta_c^I(a) = a^2\left(1 - a^4\right)^{1/2}. \tag{3.2.9}$$

A comparison of (3.2.9) to the marginal stability curve in the limit of vanishing Ekman friction (3.2.8) shows that $\beta_c^R(a,0) > \beta_c^I(a)$ and therefore an infinitesimal amount of Ekman friction moves the marginal stability curve by an order one amount in the (β, a) parameter space. Waves with total wavenumber a satisfying $\beta_c^I(a) < \beta < \beta_c^R(a,0)$ are dissipatively destabilized by the introduction of Ekman friction (discussed in e.g. Romea [109], Krechetnikov & Marsden [63], Swaters [130]).

The marginal stability curves (3.2.7), (3.2.8) and (3.2.9) are plotted in Fig. 3.2.1 for $\kappa = 0$ (solid line), $\kappa \to 0$ (dashed line) and $\kappa = 0.2$ (dotted line). The shaded region of Fig. 3.2.1, bounded above by $\beta_c^I(a)$ and below by $\beta_c^R(a,0)$, shows the region of (β, a) parameter space occupied by waves which are dissipatively destabilized by the Holopainen mechanism.

$\beta_c^R(a,0)$ differs from the frictionless value $\beta_c(a)$ by a factor of $(a^2+\frac{1}{2})/a^2(a^2+1)$ as shown in Fig. 2 of Romea [109] and reproduced in Fig. 3.2.2. It is evident in Fig. 3.2.2 that Ekman friction has a stronger effect on long waves, which may be argued on scaling grounds directly from the equations of motion (3.2.1-3.2.2). Advection scales with the gradient of the perturbation vorticity whereas friction is proportional to vorticity itself. Therefore in longer waves, where the vorticity gradients are smaller, viscous effects are of greater rela-

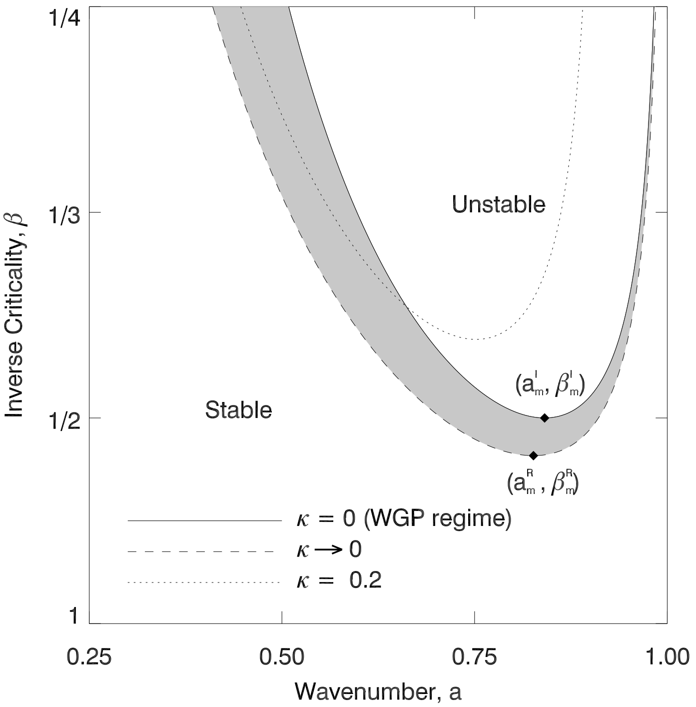

Figure 3.2.1: Curves of marginal stability (3.2.7-3.2.9) with total wavenumber a on the horizontal axis and decreasing inverse criticality β on the vertical axis. Three curves are shown: (i) zero Ekman friction ($\kappa = 0$, solid line), (ii) the limit of vanishing friction ($\kappa \to 0$, dashed line), and (iii) large Ekman friction ($\kappa = 0.2$, dotted). The shaded area corresponds to the region where waves are dissipatively destabilized by the Holopainen mechanism. Also labelled are the points of minimum critical shear on the frictionless $(\beta^I_m, a^I_m) = (2^{-1}, 2^{-1/4})$ and $\kappa \to 0$ (β^R_m, a^R_m) marginal stability curves.

tive importance. The dependence is also noticeable in the dispersion relation (3.2.5), where the Ekman number κ always appears with a factor k^{-1}, i.e. the effect of Ekman friction is greatest for low values of the x-wavenumber k.

3.2.3 Discontinuity in the limit of vanishing friction

The details of the discontinuity between the $\kappa = 0$ and $\kappa \to 0$ marginal stability curves is perhaps best explained by examining how the growth rate of a wave, of fixed wavenumber a, varies with inverse criticality β when Ekman

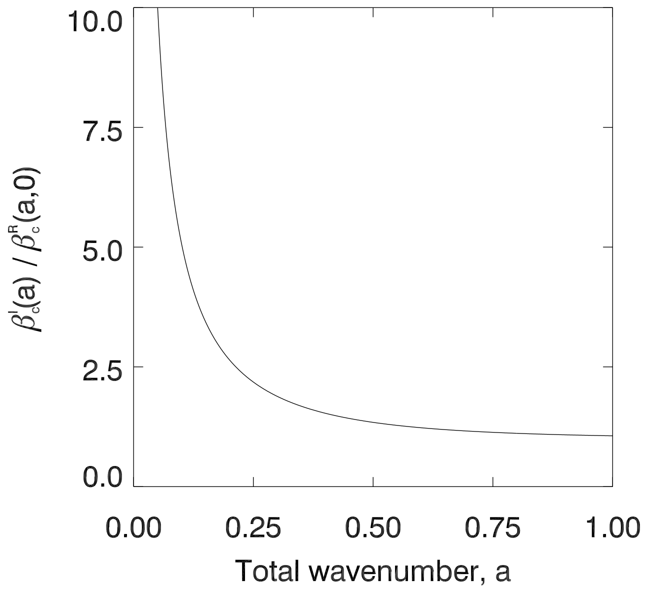

Figure 3.2.2: The ratio $\beta_c^R(a,0)/\beta_c^I(a)$ as a function of total wavenumber a.

friction is present and when it is not. Growth rates $\mathrm{Im}(kc^+)$ for a wave with wavenumber $a = 2^{-1/4}$ are calculated from (3.2.5) as a function of β and plotted in Fig. 3.2.3 for no Ekman friction ($\kappa = 0$, solid line) and for weak Ekman friction ($\kappa = 0.005$, dashed line). The $\kappa = 0$ curve shows the normal mode solutions discussed in section 2.3 with a bifurcation at the frictionless minimum critical shear ($\beta_m^I = \frac{1}{2}$). For stable flows ($\beta > \frac{1}{2}$) the two normal modes have distinct real phase speeds and zero growth rate. For unstable flows ($\beta < \frac{1}{2}$) the phase speeds are a complex conjugate pair with one growing and one decaying mode. These normal mode solutions correspond to a pair of coupled Rossby waves as described by Bretherton [14]. At marginal criticality $\beta = \frac{1}{2}$, the two normal modes coalesce as the PV gradient in the lower layer vanishes and the system can only support a single Rossby wave in the upper layer.

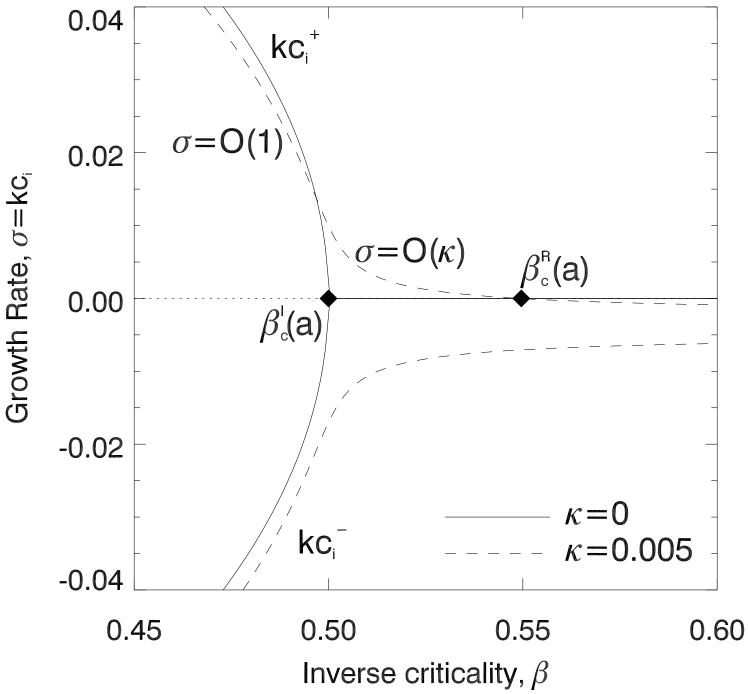

Figure 3.2.3: Bifurcation diagram of the growth rates of the fastest growing wavenumber against inverse criticality β for frictionless flows ($\kappa = 0$, solid line) and flows with small Ekman friction ($\kappa = 0.005$, dashed line).

The addition of weak Ekman friction to the system makes a complex-valued perturbation to the phase speeds, changing the growth rates by an amount linear in κ. The perturbation is different for each normal mode (notice the change in the complex factor γ in 3.2.4) and hence the bifurcation structure at $\beta = \frac{1}{2}$ is completely changed by any non-zero value of κ, however small. The phase speed is always complex and for $\beta_c^I(a) < \beta < \beta_c^R(a, 0)$, the dissipatively destabilized wave has a small, but positive growth rate of $O(\kappa)$. As the frictionless threshold is passed at $\beta < \frac{1}{2}$, the growth rate becomes $O(1)$ because the (frictionless) baroclinic instability mechanism is present, though growth rates are reduced due to the dissipative effect of Ekman friction.

3.2.4 Kinematic wave description of the Holopainen instability mechanism

A modal description of the Holopainen instability has been given by Swaters [130] who describes the mechanism in terms of wave modes and their relationship to the onset of classical (frictionless) baroclinic instability (as described in sections 2.3 and 3.2.2). This work will now be reviewed so that a comparison can be made to a PV-based description of the Holopainen mechanism to be described in section 3.3.

Swaters' approach is to recast the linearized equations of motion into two types of wave operator, a dynamic part, representing Rossby wave evolution and a kinetic wave operator (defined below) due to Ekman friction. The result is a hyperbolic partial differential equation whose characteristics are determined by the dynamic Rossby wave operator. Subsequently, for the problem to be well-posed, the phase speed associated with the kinematic part of the wave must lie within the phase speeds determined by the dynamic part of the wave operator.

The linearized equations of motion (3.2.1-3.2.2) with a plane wave solution (3.2.4) may be re-written (equations (12-14) of Swaters with $F = \frac{1}{2}$) into the single equation

$$\left(\frac{\partial}{\partial t} + c_- \frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c_+ \frac{\partial}{\partial x}\right)\psi_i' = -\frac{2\kappa(a^2+\frac{1}{2})}{(a^2+1)}\left(\frac{\partial}{\partial t} + c_0 \frac{\partial}{\partial x}\right)\psi_i', \quad (3.2.10)$$

where

$$c_\pm = \frac{a^2(a^2+1) - 2(a^2+\frac{1}{2})\beta \pm \sqrt{\beta^2 - a^4(1-a^4) + 4a^6(a^2+1)\kappa^2/k^2}}{2a^2(a^2+1)}, \quad (3.2.11)$$

and

$$c_0 = \frac{1}{2} - \frac{\beta}{a^2 + \frac{1}{2}}, \tag{3.2.12}$$

for $i = 1, 2$. Note that $c_\pm$ are the phase speeds of the Doppler-shifted baroclinic and barotropic Rossby wave modes and contain the neutral phase shift due to the presence of dissipation (the κ^2 term inside the square root). The form of equations (3.2.10-3.2.12) is useful as a direct comparison may be made to the analysis of kinematic waves presented by Lighthill & Whitham [70, 137].

The limit $\kappa \to 0$ in (3.2.11) yields exactly the Rossby wave solutions for the frictionless problem (i.e. equation 3.2.5 with $\kappa = 0$) as only the left-hand side of (3.2.10) remains. The second order differential operator on the left-hand side of (3.2.10) is the dynamic part of the wave operator, described by Whitham [137] as a wave in which particle accelerations occur against a background restoring gradient. In the case of Rossby waves, patches of anomalous PV accelerate against the restoring effect of the background PV gradient.

In contrast, taking the low frequency or low wavenumber limit leaves only the first order, right-hand side of the operator (3.2.11), which has phase speed c_0 and corresponds to a kinematic wave. A kinematic wave is described by Lighthill & Whitham [70, 137] as a travelling wave solution to a first order conservation law.

If there is no Ekman friction then only the dynamic wave part of the operator exists (right-hand side of 3.2.10 vanishes) and there are two Rossby waves propagating on the background PV gradients. Thus the dynamic operator in (3.2.11) is the case of baroclinic instability where instability depends on the propagation of Rossby waves in opposite directions (CRWs) and is possible if

$\beta < \frac{1}{2}$. When Ekman friction is non-zero ($\kappa \neq 0$), equation (3.2.10) is a hyperbolic partial differential equation whose characteristics are entirely determined by the dynamic wave, second-order part of the operator [1]. Swaters argues that for the problem to be well-posed, the first order kinematic wave part of the solution must be consistent with these characteristics, i.e. the condition for stability is that the phase speed of the kinematic wave lies inside the interval spanned by the Doppler-shifted barotropic and baroclinic Rossby waves

$$c_- \leq c_0 \leq c_+. \tag{3.2.13}$$

This condition is completely different to the case of baroclinic instability, which from Fig. 3.2.3 becomes stable when the two normal modes coalesce. The instability threshold with Ekman friction (3.2.7) may be recovered by substitution of (3.2.11) and (3.2.12) into (3.2.13). The condition (3.2.13) is described by Swaters as the 'onset of disorder' [130] (see also Baines [5]), a consequence of the low frequency-wavenumber kinematic waves being inconsistent with the propagation properties of the whole dynamical system.

3.3 Instability of the two-layer model in terms of potential vorticity

The Bretherton [14] description of counter-propagating Rossby waves (CRWs) or PV-thinking gives a qualitative description of the mechanism of baroclinic

---

[1] see Whitham [137] chapter 10 for a discussion of wave hierarchies. In this case the 'higher order' part of the hyperbolic partial differential operator determines the properties for wave propagation

instability in the two-layer model and also the Eady model [15]. Hoskins et al. [55] discuss the conservation of PV and interacting Rossby waves on isentropic surfaces in the atmosphere. The qualitative PV description was given a quantitative form by Bishop [9, 10] and applied to numerous geophysical instabilities in later works (e.g. Davies & Bishop [19], Heifetz et al. [42, 43, 44]). Here Bishop's quantitative approach is reviewed before being extended to gain an insight into the Holopainen instability mechanism.

3.3.1 A PV-thinking approach to instability

The upper and lower layer PV perturbations are written in the form

$$\begin{aligned} q_1'(x,y,t) &= \mathrm{Re}\left\{Q(t)e^{i(kx+\theta_1(t))}\cos(ly)\right\}, \\ q_2'(x,y,t) &= \mathrm{Re}\left\{R(t)e^{i(kx+\theta_2(t))}\cos(ly)\right\}, \end{aligned} \tag{3.3.14}$$

where $Q(t)$ and $R(t)$ are real valued wave amplitudes of the upper and lower waves respectively and $\theta_1(t)$ and $\theta_2(t)$ are the corresponding phases. Re denotes the real part. The perturbation streamfunction in each layer is written in terms of these potential vorticities as

$$\psi_1' = -\frac{\left(a^2+\frac{1}{2}\right)q_1' + \frac{1}{2}q_2'}{a^2(a^2+1)}, \qquad \psi_2' = -\frac{\frac{1}{2}q_1' + \left(a^2+\frac{1}{2}\right)q_2'}{a^2(a^2+1)}, \tag{3.3.15}$$

where a is the total wavenumber ($a^2 = k^2 + l^2$). The ansatz (3.3.14-3.3.15) is substituted into the linearized equations of motion (3.2.1-3.2.2 with the Jacobians on the right-hand side neglected) and upon taking real and imaginary parts this gives

$$\dot{Q} = \frac{k}{2a^2(a^2+1)}\left(\left(\frac{1}{2}+\beta\right)R\sin\theta - \frac{2\kappa a^2}{k}\left[\left(a^2+\frac{1}{2}\right)Q+\frac{1}{2}R\cos\theta\right]\right), \tag{3.3.16}$$

$$\dot{R} = \frac{k}{2a^2(a^2+1)}\left(\left(\frac{1}{2}-\beta\right)Q\sin\theta - \frac{2\kappa a^2}{k}\left[\left(a^2+\frac{1}{2}\right)R+\frac{1}{2}Q\cos\theta\right]\right), \tag{3.3.17}$$

$$\begin{aligned}\dot{\theta} = \frac{k}{2a^2(a^2+1)}&\left(1-2a^4+\beta\left(\frac{R}{Q}-\frac{Q}{R}\right)\cos\theta\right. \\ &\left.+\left(\frac{R}{Q}+\frac{Q}{R}\right)\left(\frac{1}{2}\cos\theta+\frac{\kappa}{k}\sin\theta\right)\right),\end{aligned} \tag{3.3.18}$$

where $\theta = \theta_1 - \theta_2$ is the phase difference between the upper and lower layer disturbances.

Equations (3.3.16-3.3.18) are analogous to those put forward by Bishop [10] and give a quantitative form to the qualitative arguments motivated by the PV-thinking approach of Bretherton. The Rayleigh problem [100] of shear instability was considered by Heifetz et al. [42] and further work by Heifetz et al. [43] described how the CRW viewpoint could be extended for more complicated flows, applying it to the Charney model of baroclinic instability [44].

The approach using upper- and lower-level PV disturbances also allows the study of structures undergoing non-modal growth, i.e. disturbances which amplify at a rate faster than that of the fastest growing normal mode. This kind of non-modal growth has been observed in numerical experiments [122, 120]. The optimum growth rate is dependent on the initial phase configuration of the PV perturbations [32] and the optimum growth structure has been investigated by Farrell [33] and O'Brien [86]. Despite this important line of work, this chapter will focus on the transition to normal mode growth with little discussion of non-normal, transient growth rates.

In equations (3.3.16-3.3.17) there are two different effects caused by Ekman friction, which can be seen by considering terms proportional to κ, inside the square brackets:

(i) The first terms inside the square brackets in (3.3.16-3.3.17) are proportional to $-\kappa Q$ and $-\kappa R$ respectively and represent the direct, dissipative action of Ekman friction on the disturbances. These terms reduce the growth rate of the disturbance in each layer in proportion the Ekman number κ.

(ii) The second, possibly destabilizing effect of friction is seen in the second terms inside the square brackets of (3.3.16-3.3.17). These terms are proportional to $-\kappa R \cos\theta$ and $-\kappa Q \cos\theta$ respectively. If the phase difference between the disturbances in each layer satisfies $\pi/2 < |\theta| < \pi$ then these terms will be positive and the disturbances grow in proportion to the amplitude of the disturbance in the other layer. Conversely, if $0 < |\theta| < \pi/2$, then the frictional effect in each layer inhibits the growth of the wave in the opposite layer. Physically, the drag acting in one layer induces a velocity in the opposite layer causing the wave amplitude to amplify or attenuate depending on the phase difference θ.

In this section, the focus is on the destabilizing effect of friction. To remove the dissipative effect of Ekman friction in equations (3.3.16-3.3.18) the time and wave amplitudes are re-scaled as

$$\begin{aligned} Q^* &= Qe^{-2\left(a^2+\frac{1}{2}\right)\kappa^* t^*}, \\ R^* &= Re^{-2\left(a^2+\frac{1}{2}\right)\kappa^* t^*}, \\ t^* &= \frac{kt}{2a^2(a^2+1)}, \end{aligned} \tag{3.3.19}$$

where $\kappa^* = a^2\kappa/k$. The channel width enters the problem via the x-wavenumber k and a wider channel reduces the growth rate of the disturbance for a fixed wavenumber a (see time scaling in in 3.3.19). The x-wavenumber also appears in the re-scaled Ekman number κ^* so that for a PV disturbance with a given total wavenumber a, the effects of Ekman friction are greater on a disturbance with a smaller zonal wavenumber (greater x-wavelength).

Under the rescaling (3.3.19) and writing the supercriticality as $\epsilon^* = \frac{1}{2} - \beta$, [2] equations (3.3.16-3.3.18) become

$$\dot{Q}^* = R^*\left((1-\epsilon^*)\sin\theta - \kappa^*\cos\theta\right), \tag{3.3.20}$$

$$\dot{R}^* = Q^*\left(\epsilon^*\sin\theta - \kappa^*\cos\theta\right), \tag{3.3.21}$$

$$\dot{\theta} = 2\alpha + \left((1-\epsilon^*)\frac{R^*}{Q^*} + \epsilon^*\frac{Q^*}{R^*}\right)\cos\theta + \kappa^*\left(\frac{Q^*}{R^*} + \frac{R^*}{Q*}\right)\sin\theta, \tag{3.3.22}$$

where

$$\alpha = \frac{1}{2} - a^4, \tag{3.3.23}$$

Dots now denote differentiation with respect to t^*. Equations (3.3.20-3.3.22) are a simplification of the dynamical system (3.3.16-3.3.18) in order to emphasise the dependence on just three parameters ϵ^*, α and κ^*.

---

[2] Note that this is *not* the supercriticality of chapter 2 where $\epsilon = \sqrt{\frac{1}{2} - \beta}$.

The PV description (3.3.20-3.3.22) will be used next to develop a novel understanding of the Holopainen instability mechanism, which can be directly contrasted to the CRW mechanism of baroclinic instability. It is therefore necessary to first review the CRW mechanism of baroclinic instability in the two-layer model.

3.3.2 Baroclinic instability in terms of counter - propagating Rossby waves (CRWs)

Setting $\kappa^* = 0$ in (3.3.20-3.3.22) and further re-scaling of of the wave amplitudes and times

$$\tilde{Q} = \frac{Q^*}{\sqrt{|\epsilon^*|}} \quad \tilde{R} = \frac{R^*}{\sqrt{1-\epsilon^*}} \quad \tilde{t} = \sqrt{|\epsilon^*|(1-\epsilon^*)}t, \tag{3.3.24}$$

gives

$$\begin{aligned} \dot{\tilde{Q}} &= \tilde{R}\sin\theta, & (3.3.25) \\ \dot{\tilde{R}} &= \tilde{Q}\mathrm{sgn}(\epsilon^*)\sin\theta, & (3.3.26) \\ \dot{\theta} &= 2\tilde{\alpha} + \left(\frac{\tilde{Q}}{\tilde{R}} + \frac{\tilde{R}}{\tilde{Q}}\right)\cos\theta, & (3.3.27) \end{aligned}$$

where $\tilde{\alpha} = \alpha/\sqrt{|\epsilon^*|(1-\epsilon^*)}$ and differentiation is with respect to $\tilde{t}$ [3]. The parameters $(\epsilon^*, \tilde{\alpha})$ completely determine the dynamical system (3.3.25-3.3.27), which are identical to (7a-d) in Davies & Bishop [19] (describing the Eady problem) and (14a-d) in Heifetz et al [42] (describing the Rayleigh problem), since they are describing the same qualitative scenario of counter-propagating

---

[3] Note that equations (3.3.25-3.3.27) are singular at $\epsilon^* = 0$, in which case, the relevant dynamical system is (3.3.20-3.3.22) with $\epsilon^* = \kappa^* = 0$.

Rossby waves. Note that the parameter ϵ^* influences the qualitative behaviour of the system (3.3.25-3.3.27) only through its sign, i.e. whether the system is stable ($\epsilon^* < 0$) or unstable ($\epsilon^* > 0$). The behaviour of the system of ODEs (3.3.25-3.3.27) is identical to those in Heifetz et al. [42, 43, 44] and Davies & Bishop [19], the basic features of which are now discussed.

The strength of the coupling between the upper and lower Rossby waves is governed by the parameter $\tilde{\alpha}$ in equation (3.3.27), which itself depends on the supercriticality ϵ^* and the total wavenumber a. A growing normal mode in an unstable flow ($\epsilon^* > 0$) exists when the waves can enter into a phase-locked configuration: $\dot{\theta} = 0$ with $\theta \in (0, \pi)$ and $\dot{\tilde{Q}} = \dot{\tilde{R}} = \tilde{Q}\sin\theta = \tilde{R}\sin\theta$. According to (3.3.27) a phase-locked solution $\dot{\theta} = 0$ exists if $-1 < \tilde{\alpha} < 1$ which upon rearrangement and substitution of $\epsilon^* = \frac{1}{2} - \beta$ yields precisely the marginal stability curve (2.3.16)

$$\begin{aligned} |\tilde{\alpha}| = \frac{\left|\frac{1}{2} - a^4\right|}{\sqrt{|\epsilon^*|(1-\epsilon^*)}} = \frac{\left|\frac{1}{2} - a^4\right|}{\sqrt{\frac{1}{4} - \beta^2}} \quad &< \quad 1 \\ \Rightarrow \quad \beta^2 \quad &< \quad a^4(1 - a^4). \end{aligned} \tag{3.3.28}$$

The inequality $\tilde{\alpha} < -1$ corresponds to the short-wave cut-off and $\tilde{\alpha} > 1$ is the long-wave cut-off, to be discussed below.

The dynamical system (3.3.25-3.3.27) can be integrated numerically subject to initial conditions $(\tilde{Q}_0, \tilde{R}_0, \theta_0)$. The initial conditions $(\tilde{Q}_0, \tilde{R}_0, \theta_0) = (10^{-3}, 10^{-3}, 0)$ have been used to calculate the wave evolutions at the parameter settings $(\epsilon^*, \tilde{\alpha})$ given in Table 3.1. For these systems, the evolution of the phase difference between the upper and lower layer PV disturbances is plotted in Fig. 3.3.1.

Solution	ϵ^*	$\tilde{\alpha}$	Description
I	0.01	0	Unstable
II	-0.01	0	Stable, below marginal stability curve
III	0.01	> 1	Stable, below low wavenumber cut-off
IV	0.01	< -1	Stable, above high wavenumber cut-off

Table 3.1: Details of inverse criticality β and total wavenumber a used for solutions I-IV to (3.3.25-3.3.27).

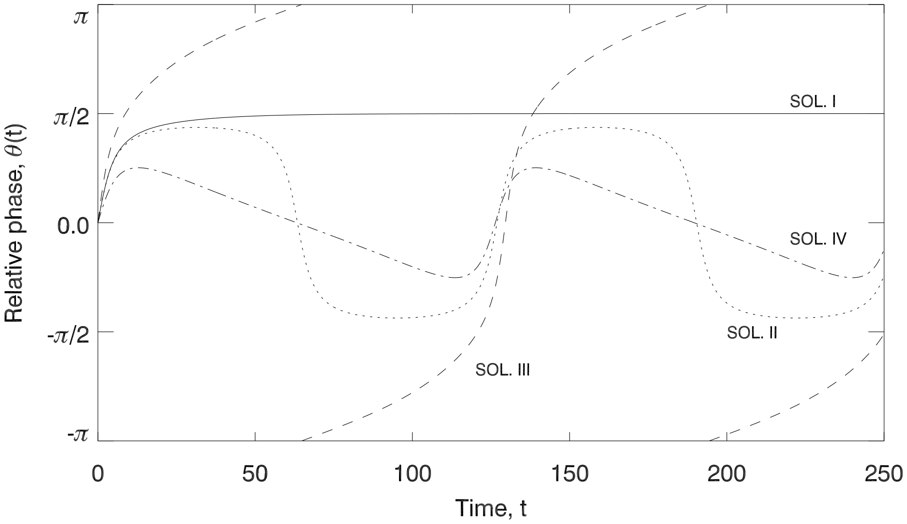

Figure 3.3.1: Evolution of the phase difference θ of PV disturbances calculated from (3.3.25-3.3.27). The parameters $(\epsilon^*, \tilde{\alpha})$ for each solution I-IV are given in Table 3.1 with initial conditions $(\tilde{Q}_0, \tilde{R}_0, \theta_0) = (10^{-3}, 10^{-3}, 0)$.

Solution I (solid line) is a flow that satisfies the frictionless condition for instability $-1 < \tilde{\alpha} < 1$. The phase difference between the two layers rapidly converges to a constant value, $\theta = \pi/2$ (by $t = 80$) and remains there for all time. This demonstrates phase-locking as the two PV disturbances quickly reach, and remain in, a constant phase configuration relative to each other where $\dot{\tilde{Q}} = \dot{\tilde{R}} > 0$ and the two waves grow exponentially.

The transition of the system from the initial conditions to a phase-locked configuration is explained by the tendency of the phase difference $\dot{\theta}$, which according to (3.3.27) is proportional to $\cos\theta$. If $0 < |\theta| < \frac{\pi}{2}$ then the upper layer wave is less than a quarter of a wavelength out of phase with the lower layer wave, then $\dot{\theta} > 0$ and the phase difference increases. In this case, the velocity field induced by the Rossby wave in the upper layer enhances the propagation of the Rossby wave in the lower layer and the velocity field induced by the lower layer Rossby wave enhances the propagation of the upper layer wave. If $\pi/2 < |\theta| < \pi$ then the waves in each layer are between a quarter and half wavelength out of phase, $\dot{\theta} < 0$ and each Rossby wave hinders the propagation of the wave in the opposite layer. As the phase difference modulates, the growth of the waves is governed by (3.3.25-3.3.26). If $\theta \in (0, \pi)$ then both $\tilde{Q}$ and $\tilde{R}$ are growing in proportion to the magnitude of the wave in the opposite layer and the smaller magnitude wave grows more rapidly.

Solutions II-IV in Fig. 3.3.1 each show stable configurations of equations (3.3.25-3.3.27). Solution II (dotted line, $\tilde{\alpha} = 0$) is for a flow above the frictionless stability threshold $\epsilon^* < 0$ where the phase difference never converges, but oscillates about $\theta = 0$. In this case the wave amplitudes vary according to $\dot{\tilde{Q}}/\tilde{Q} = -\dot{\tilde{R}}/\tilde{R} = \sin\theta$, which means that when the wave in one layer is growing, the opposite wave is decaying according to the sign of $\sin\theta$. As the phase difference oscillates about $\theta = 0$ the waves in each layer grow and decay alternately and there is no sustained growth of the wave amplitudes.

Solutions III (dashed line) and IV (dot-dashed lines) lie within the frictionless threshold $\epsilon^* > 0$, but are outside the marginal stability curve $|\tilde{\alpha}| > 1$. Solution III ($\tilde{\alpha} > 1$) is below the low-wavenumber cut-off, which means that $\dot{\theta} > 0$ at all times and the phase difference is always increasing. At low wavenumbers (relatively large wavelengths), the restoring effect of PV displacements against the background PV gradient dominates over advection (described by Bretherton [14]) and the Rossby waves are strongly coupled.

Solution IV shows a flow above the high wavenumber cut-off ($\tilde{\alpha} < -1$) where the phase difference oscillates about zero (similar to solution II). At high wavenumbers (short wavelengths), advection dominates over the amplifying effects of PV perturbations against the background PV gradient and the two Rossby waves are uncoupled. The upper layer Rossby wave propagates westwards, but with a phase speed that is not great enough to overcome the background (eastward) advection. The lower layer wave propagates eastwards on a negative PV gradient (in the absence of background flow). Therefore, both PV disturbances are moving eastwards (in a stationary frame of reference) resulting in an oscillation of the phase difference θ and wave amplitudes $\tilde{Q}(t)$, $\tilde{R}(t)$.

The discussion now returns to the unstable flow of solution I. The phase-locked configuration $\dot{\theta} = 0$, $\dot{\tilde{Q}} = \dot{\tilde{R}}$ of (3.3.25-3.3.27) is written explicitly as

$$\tilde{R} = \tilde{Q}, \quad \dot{\tilde{Q}} = \dot{\tilde{R}} = \tilde{Q} \sin\theta, \quad \text{where} \quad \theta = \arcos\left(-\tilde{\alpha}\right). \tag{3.3.29}$$

Fig. 3.3.2 shows the total PV field in each layer in a growing, phase-locked solution with the parameters $\epsilon^* = 0.1$, $\tilde{\alpha} = 0$. Fig. 3.3.2 highlights the mechanism of CRWs in a growing normal mode similar to the figures of Bretherton (1966), for the two-layer model (Fig. 2 [14]), Heifetz (2000) for the case of barotropic instability (Fig. 2,4 [42]) and Heifetz et al. (2004) constructed for a general flow (Fig. 1 [43]).

In Fig. 3.3.2, the solid, blue lines show the PV perturbations in each layer

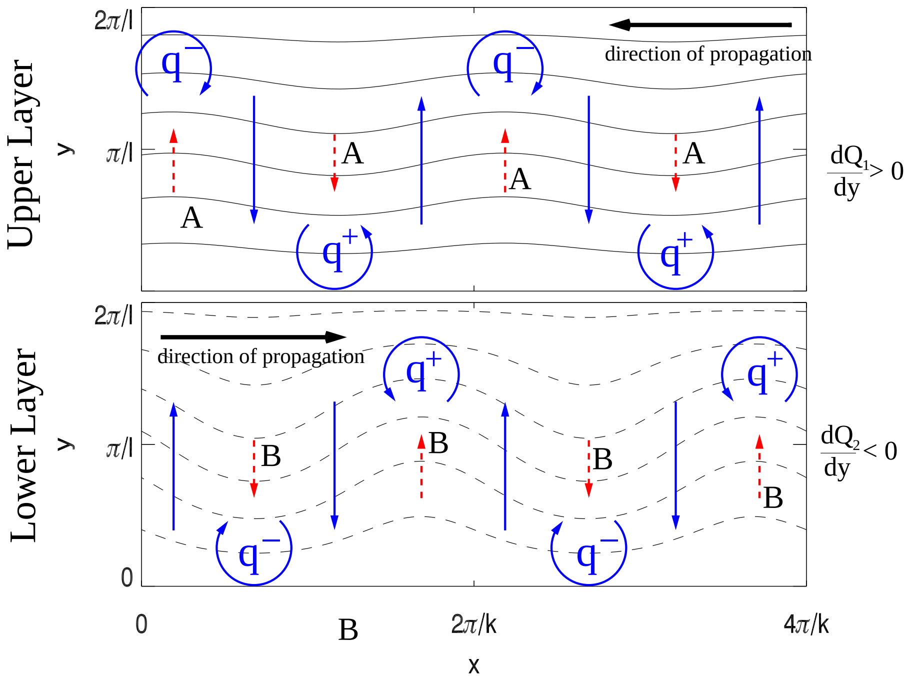

Figure 3.3.2: Contour plot of total PV in each layer with $\epsilon^* = 0.1$ and $\tilde{\alpha} = 0$. In this configuration, the waves are phase-locked and growing. Positive and negative PV disturbances are marked with q^+ and q^- respectively. The vertical blue arrows show the velocity of fluid patches and the dashed red lines represent the corresponding velocity induced in the opposite layer. The arrows marked with A show the velocity due to the lower layer advecting PV to reinforce the anomalies in the upper layer. The corresponding effect in the lower layer is marked B.

and the red, dashed lines show the projection of this motion onto the opposite layer. The Rossby wave in the upper layer is propagating westward on a positive PV gradient (eastward zonal flow). The PV perturbation induced by the westward-propagating upper layer wave acts to advect positive and negative PV, into the crests and troughs of the lower Rossby wave respectively (arrows labelled B in Fig. 3.3.2). In the lower layer, there is an eastward propagating Rossby wave on a negative PV gradient. The PV perturbation in the upper layer induced by the lower layer velocity field similarly acts to displace positive and negative PV in the same sense as the upper layer Rossby wave (arrows

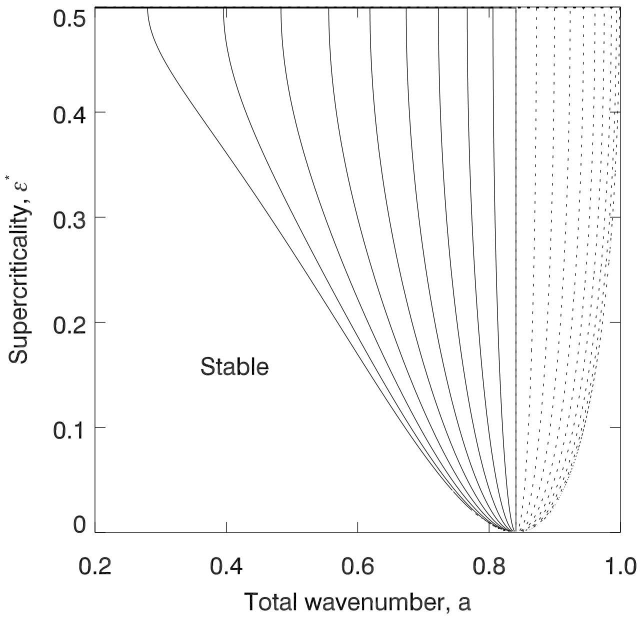

Figure 3.3.3: Contour plot of phase difference θ of the normal mode solutions $\dot{\theta} = 0$, $\dot{\tilde{Q}} = \dot{\tilde{R}}$ to equations (3.3.25-3.3.27) over the range of unstable wavenumbers (alternatively, variation of $\tilde{\alpha}$). All growing configurations show a westward tilt of the disturbance with height: solid contours show phase differences $\theta \in \left(\frac{\pi}{2}, \pi\right)$ and dotted contours show phase differences $\theta \in \left(0, \frac{\pi}{2}\right)$. The contour interval is $\pi/20$.

labelled A in Fig. 3.3.2). Therefore in the phase locked configuration with $\theta \in (0, \pi)$ there is a mutual reinforcement of the Rossby waves in each layer and the disturbances grow exponentially.

The configuration in Fig. 3.3.2 shows that the upper layer wave is to the west of the lower wave ($\theta > 0$). This westward phase tilt with height occurs for the whole range (ϵ^*, a) of unstable waves (it is more convenient here to use total wavenumber a rather than $\tilde{\alpha}$ as the second parameter). Fig. 3.3.3 shows how the phase-locked configuration θ of the growing normal mode varies with total wavenumber (or $\tilde{\alpha}$) and supercriticality. Fig. 3.3.3 shows that the equilibrium phase difference θ is a trade-off between the competing effects of the background advection due to $\tilde{\alpha}$ and wave coupling (the $\cos\theta$ term in 3.3.27). At $a = 2^{-1/4}$ ($\tilde{\alpha} = 0$) all supercriticalities phase-lock with phase difference $\theta = \frac{\pi}{2}$, where the upper wave is exactly quarter of a wavelength to the west of the lower layer wave, the configuration plotted in the mechanism schematic Fig. 3.3.2.

As the wavenumber increases (α increases) there is a decrease in the westward phase tilt of the phase-locked configuration. Fig. 3.3.3 shows that at larger wavenumbers where $-1 < \tilde{\alpha} < 0$, the westward tilt of the phase locked configuration is such that $\theta \in \left(0, \frac{\pi}{2}\right)$ (dotted contours) and that the phase locked configuration for smaller wavenumbers $0 < \tilde{\alpha} < 1$ is tilted further westwards $\theta \in \left(\frac{\pi}{2}, \pi\right)$ (solid lines).

3.3.3 Holopainen instability in terms of PV disturbances

The Holopainen instability mechanism will now be discussed using the approach introduced in section 3.3.1. The most transparent case where dissipative destabilization occurs is for $\epsilon^* = 0$, since there is no PV gradient in the lower layer ($Q_2(y) = 0$). Setting $\epsilon^* = 0$ in equations (3.3.20-3.3.22) and dropping the stars from Q^* and R^* gives

$$\begin{aligned} \dot{Q} &= R\left(\sin\theta - \kappa^*\cos\theta\right), && (3.3.30)\\ \dot{R} &= -Q\kappa^*\cos\theta, && (3.3.31)\\ \dot{\theta} &= 2\alpha + \frac{R}{Q}\cos\theta + \left(\frac{Q}{R} + \frac{R}{Q}\right)\kappa^*\sin\theta. && (3.3.32) \end{aligned}$$

The form of the solution (3.3.14) corresponds to a vortex train, coupled to the westward propagating Rossby wave in the upper layer. A further simplification to equation (3.3.32), which does not alter the behaviour of the dynamical system, can be made by setting $\alpha = 0$, i.e. choosing the wavenumber $a = 2^{-1/4}$ which corresponds to minimum critical shear in the frictionless problem.

The first term in equation (3.3.30) is the effect of the lower layer PV disturbance on the upper layer Rossby wave as discussed above for the $\kappa^* = 0$ case: If $0 < \theta < \pi$, the disturbance induced in the upper layer by the lower layer PV field advects PV in the same sense as the upper layer Rossby wave. Unlike the CRW scenario however, there is no corresponding term in equation (3.3.31), but only the dissipative destabilization term $-Q\kappa^* \cos\theta$, which causes the lower layer disturbance to grow if $\pi/2 < |\theta| < \pi$.

The effect of Ekman friction on the modulation of the phase difference θ is governed by the $\kappa^* \sin\theta$ term in equation (3.3.32). If $\kappa^* \gtrsim (R^2/Q^2)\text{cot}\theta$, Ekman friction effects dominate the direct wave-wave interaction and alters the phase tendency near $\theta = \pm\pi/2$, which is sufficient to allow the phase difference to reach a locked configuration. This is in contrast to solution II of section 3.3.2 where in the absence of friction at the same supercriticality the phase oscillates about $\theta = 0$ (Fig. 3.3.1).

Equations (3.3.30-3.3.32) can be solved exactly for $\alpha = 0$ to find the phase-locked solution $\dot{\theta} = 0$, $\dot{Q} = \dot{R}$,

$$Q = \frac{\left(1 + (\kappa^*)^2\right)^{1/4}}{(\kappa^*)^{1/2}} R \qquad \theta = \arctan\left(\kappa^* - \left(1 + (\kappa^*)^2\right)^{1/2}\right). \qquad (3.3.33)$$

If $\kappa^* \ll 1$, then $Q \approx (\kappa^*)^{-1/2} R$ and $\theta \approx \arctan(-1) = 3\pi/4$ for a growing pair of disturbances. Fig. 3.3.4 shows the potential vorticity distribution in each layer for $(\epsilon^*, \alpha, \kappa^*) = (0, 0, 0.01)$ in a phase-locked configuration with arrows indicating the mechanism by which the disturbance grows. The red arrows

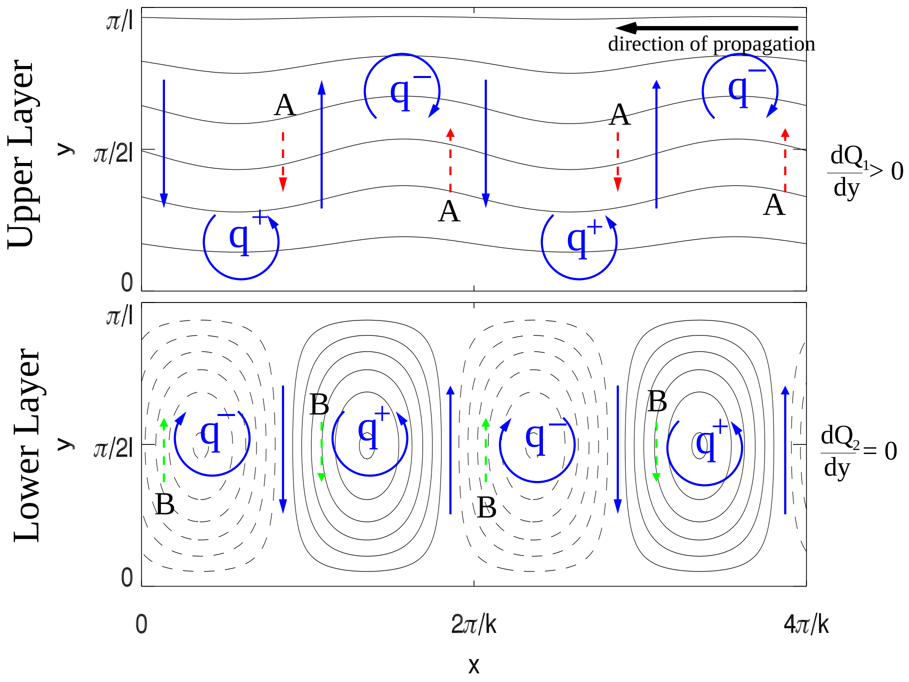

Figure 3.3.4: Contour plot of total PV in each layer calculated from (3.3.30-3.3.32) with $\epsilon^* = 0$, $\alpha = 0$ and $\kappa^* = 0.01$. The upper layer Rossby wave propagates to the left on a positive background PV gradient. In the lower layer, the background PV gradient is zero and does not support Rossby wave propagation and the PV disturbance is a train of vortices. The red arrows (marked A) show the projection of the lower layer flow onto the upper layer Rossby wave as in Fig. 3.3.2. The green arrows (marked B) show the effect of friction on the upper layer wave in reinforcing the vortices in the lower layer.

(A) show the sense of the flow induced by the lower layer vortex train within the upper layer. The induced flow acts to perturb positive PV southwards and negative PV northwards in such a way as to amplify the upper layer Rossby wave (as in the CRW mechanism). The key idea in the Holopainen instability mechanism is that Ekman friction acts on the flow induced by the upper layer wave within the lower layer, which spins up a train of vortices. The green arrows (B) show the flow induced by the upper layer wave disturbances within the lower layer. If the PV disturbances are close to being in anti-phase (with $\theta < \pi$ as in Fig. 3.3.4) then this flow reinforces the circulation of the lower layer vortex train and the PV disturbances in both layers grow.

Note that here is the same action of friction on the lower layer vortex train, which induces a flow in the upper layer. However, since this is an $O(\kappa^*)$ effect, it is dominated by the $O(1)$ effect of the CRW-type amplification (marked A). The amplifying effect of friction is the dominant mechanism in the lower layer since for $\epsilon^* = 0$ there is no baroclinic instability mechanism.

3.3.4 Phase portrait description of instability mechanisms

A useful way to see the complete range of wave development in equations (3.3.20-3.3.22) is to use a phase diagram (similar to the Bishop [9]). These phase diagrams are of use in describing non-modal disturbances, but are used here to clearly contrast the mechanisms behind the classic baroclinic and Holopainen instabilities.

The ratio of the wave amplitudes Q/R and the phase difference θ completely describe the structure of the PV disturbance. Therefore one can introduce a new variable $\xi(t) = \arctan\left(Q(t)/R(t)\right)$, satisfying $0 \leq \xi \leq \pi/2$. The values $\xi = \pi/2$ and $\xi = 0$ then correspond to the lower and upper layer disturbances respectively having zero amplitude. Recasting equations (3.3.20-3.3.22) in terms of ξ and θ gives

$$\begin{aligned} \dot{\xi} &= \left(\frac{1}{2} - \frac{1}{2}\cos 2\xi - \epsilon^*\right)\sin\theta + \kappa^*\cos 2\xi\cos\theta && (3.3.34) \\ \dot{\theta} &= 2\alpha + \frac{2}{\sin 2\xi}\left(\left(\frac{1}{2} + \left(\frac{1}{2} - \epsilon^*\right)\cos 2\xi\right)\cos\theta + \kappa^*\sin\theta\right) && (3.3.35) \end{aligned}$$

The functions $\xi(t)$ and $\theta(t)$ then uniquely describe the evolution of the disturbance structure given initial conditions $\xi(0)$ and $\theta(0)$. It is possible to plot phase diagrams with contours parallel to $(\dot{\xi}, \dot{\theta})$ on the (ξ, θ) plane. The phase diagrams display the entire evolution of the flow from the initial conditions to an equilibrium point or cycle.

Fig. 3.3.5 shows four such phase portraits. Panel (a) shows an unstable flow with no Ekman friction ($\epsilon^* = 0.02$,$\kappa^* = 0$), panel (b) a stable flow with no Ekman friction ($\epsilon^* = -0.02, \kappa^* = 0$), panel (c) shows a baroclinically unstable flow with Ekman friction ($\epsilon^* = 0.02, \kappa^* = 0.01$) and panel (d) shows a flow in the Holopainen instability region ($\epsilon^* = -0.02, \kappa^* = 0.01$). Panels (a) and (b) are based on Fig. 3 (a) and (c) respectively of Heifetz et al. [42] showing unstable and stable configurations of the Rayleigh model of shear instability.

Panel (a) of Fig. 3.3.5 shows frictionless baroclinic instability, with $\epsilon^* = 0.02, \alpha = 0$, where the wave is unstable and according to equations (3.3.25-3.3.26), both wave amplitudes are growing when $0 < \theta < \pi$ and decaying when $-\pi < \theta \in 0$, i.e. on the right-hand side of panel (a) the waves are growing and on the left-hand side the waves are decaying. There are two stationary points corresponding to normal modes, one of which is growing (marked g, the stable node) and the other decaying (marked d, the unstable node). The phase speeds of these normal modes calculated from linear theory are complex conjugates (see equation 2.3.14). For $\alpha = 0$, when $0 < |\theta| < \pi/2$, the phase difference is increasing, whilst for $\pi/2 < |\theta| < \pi$ the phase difference is decreasing. All trajectories are towards $\theta = \frac{\pi}{2}$ as the phase modulates to the locked configuration described in section 3.3.2 (e.g. solution I above).

A comparison of panel (c) to panel (a) shows that the addition of a small

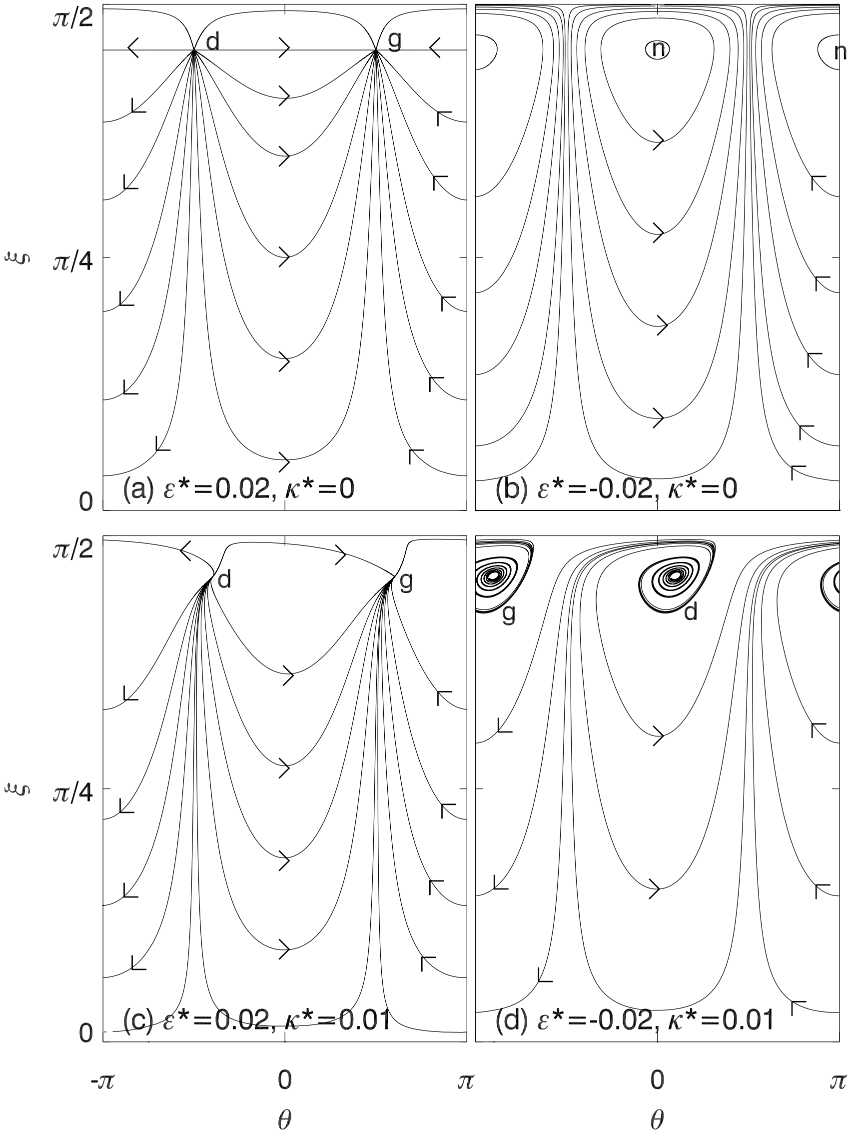

Figure 3.3.5: Phase portraits for four points in the parameter space (ϵ^*, κ^*) calculated using equations (3.3.34-3.3.35). Panel (a) $\epsilon^* = 0.02$, $\kappa^* = 0$: in the WGP regime, panel (b), $\epsilon^* = -0.02$, $\kappa^* = 0$ lying in the the baroclinically stable region, panel (c) $\epsilon^* = 0.02$, $\kappa^* = 0.01$, a baroclinically unstable flow with Ekman friction, panel (d) $\epsilon^* = -0.02$, $\kappa^* = 0.01$, a dissipatively destabilized flow. All four panels are plotted with $\alpha = 0$.

amount of Ekman friction ($\kappa^* = 0.01$ in this case) has little effect on an unstable baroclinic flow. The phase portrait in (c) is modified slightly, but is structurally identical to (a) with one growing and one decaying normal mode, marked g and d as before.

The stable flow in panel (b), $(\epsilon^*, \kappa^*) = (-0.02, 0)$, has very different behaviour to the unstable flow of panel (a). In this case, where $\epsilon^* < 0$, equations (3.3.25-3.3.26) show that Q is growing and R is decaying for $0 < \theta < \pi$ whilst for $-\pi < \theta < 0$, R is growing and Q decaying (this is the situation in solution II above). The phase speeds of the normal modes are both real indicating two neutral modes represented on the phase portrait (b) as centres and labelled n. All trajectories orbit one of these centres and as they do the ratio ξ and hence the amplitudes of the upper and lower layer waves oscillate between growth and decay in the two halves of the phase portrait.

Panel (d) of Fig. 3.3.5 shows how the stable wave structure in (b) is altered by the introduction of Ekman friction. When κ^* becomes non-zero, a bifurcation occurs as the two stationary points, previously centres, become an unstable and stable node labelled g and d respectively, representing the growing and decaying normal mode solutions. The local values of $\dot{\theta}$ and $\dot{\xi}$ show that the transition to the growing normal mode g takes longer than in panel (a) as the locking mechanism is $O(\kappa^*)$. In this case, the phase difference oscillates about the locked value and the amplitudes $Q(t)$ and $R(t)$ grow and decay alternately. Eventually, the solution reaches the normal mode solution in a phase configuration similar to that of Fig. 3.3.4. The phase difference of the growing mode is close to $\theta = \pi$, so that the lower layer vortex train has almost the opposite phase to the upper layer Rossby wave. The phase portrait of panel (d) demonstrates the earlier remark that friction induced effects dominate phase modulation for $\kappa^* \gtrsim (R^2/Q^2)\cot\theta$, i.e. close to $\xi = \frac{\pi}{2}$ and near $\theta = \frac{\pi}{2}$ and $\theta = -\frac{\pi}{2}$. In these cases $\dot{\theta} > 0$ and $\dot{\theta} < 0$ respectively creating stable and unstable nodes in place of centres.

The bifurcation occurring between panels (b) and (d) of Fig. 3.3.5 as κ^* becomes non-zero is exactly the change in the dynamical system suggested by Swaters [130]. Swaters' kinematic wave picture is a useful modal interpretation of the Holopainen instability and is complemented by the Bishop approach of this section. The Bishop approach offers a better physical description of the mechanism at work in terms of PV thinking, as summarised in Fig. 3.3.4. It is especially useful since it is easily contrasted with the CRW picture of classic baroclinic instability in Fig. 3.3.2. A succinct description of the transition to the normal mode and non-modal growth is provided by the phase portraits in Fig. 3.3.5.

3.4 Dissipative destabilization in the Eady model

The Holopainen instability mechanism in the two-layer model is a specific example of dissipative destabilization, which takes place in a variety of models as described by Krechetnikov and Marsden [62, 63]. An Eady-like model [24] is another flow where dissipative destabilization can occur.

The Eady model is a continuously stratified model in a three-dimensional channel where the effects of differential rotation are ignored so that flow is on the f-plane ($\beta = 0$, Vallis [133] p.66). Eady [24] considered flow between flat, smooth upper and lower boundaries so that no Ekman friction or topographic forcing (due to sloping boundaries) is present. The first study of Ekman friction in the Eady model was made by Barcilon [6] who considered symmetric (i.e. equal) Ekman layers at the top and bottom of the channel. It was found that although Ekman pumping into the interior releases energy for instability, there is no frictional instability equivalent to the Holopainen instability. Hide [50] later extended the Eady model to include sloping boundaries, but studied the effects of Ekman friction or sloping topography independently. Weng [135] included both Ekman friction and topography and discovered that a small amount of asymmetric Ekman friction (i.e. with Ekman layers of different thickness at the top and bottom of the channel) destabilizes baroclinic waves in a similar way to the Holopainen mechanism. Weng & Barcilon [136] later studied the effects of symmetric and asymmetric Ekman friction on the sloping boundaries in the Eady model and found that a small amount of symmetric Ekman friction dissipatively destabilizes the flow in a channel with sloping boundaries.

It should be emphasized that oppositely sloping boundaries are necessary for dissipative destabilization of the Eady model. The sloping boundaries create a topographic β-effect similar to the β-effect caused by differential rotation in the two layer model. Annulus experiments by Mason [77] confirm that if the boundaries slope with equal but opposite angles, then an effect similar to β is created. In the absence of sloping boundaries, there is no dissipative destabilization of the Eady model. This is equivalent to the absence of the β-effect in the two-layer model where, on the f-plane, dissipative destabilization does not occur [91].

The PV-based approach of section 3.3.1 has been applied to the Eady model with no Ekman friction by Davies & Bishop [19]. In the Eady model, the background PV in the interior of the channel is zero and instead of two counter-propagating Rossby waves, there are counter-propagating edge waves on the upper and lower vertical boundaries, which interact with each other. Bretherton [15] first describes baroclinic instability of the Eady model in terms of these two interacting edge waves.

It is possible to adapt the Davies & Bishop [19] description of Eady edge wave propagation to the dissipative destabilization of the Eady model described by Weng and Weng & Barcilon [135, 136] in order to make a direct comparison to the Holopainen mechanism in the two-layer model.

3.4.1 Description of the Eady model

The Eady model ([24] or Vallis [133] p.265) consists of an incompressible fluid that is uniformly stratified (buoyancy frequency $N^2 = const.$). The differential rotation caused by the sphericity of the Earth is ignored so motion is on the f-plane ($\beta = 0$). The non-dimensionalized equations of motion here follow those in Weng [135] and Weng & Barcilon [136]. The basic state is a uniform vertical shear, which has streamfunction and velocity profiles

$$\Psi(y, z) = -\Lambda yz \quad \Leftrightarrow \quad U(z) = \Lambda z, \tag{3.4.36}$$

where Λ is the (non-dimensionalized) constant vertical shear so that the channel has depth 1 and a zonal velocity of Λ at $z = 1$. The basic state potential vorticity, calculated from (3.4.36) is given by

$$Q = \frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + S\frac{\partial^2 \Psi}{\partial z^2} = 0, \tag{3.4.37}$$

and the background buoyancy is

$$B = \frac{\partial \Psi}{\partial z} = -\Lambda y. \tag{3.4.38}$$

S is a stratification parameter, which measures the importance of the background buoyancy. For a full discussion of the non-dimensionalized parameters, see appendix 3.A. The physical fields are written in terms of the background and a small amplitude, three-dimensional disturbance in the geostrophic streamfunction ψ', potential vorticity q' and buoyancy b' about the basic flow

$$\begin{aligned} \psi &= -\Lambda yz + \psi', \\ q &= 0 + q', \\ b &= -\Lambda y + b'. \end{aligned} \tag{3.4.39}$$

The perturbation potential vorticity and buoyancy are defined in terms of the geostrophic streamfunction as

$$\begin{aligned} q' &= \frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2} + S\frac{\partial^2 \psi'}{\partial z^2}, \\ b' &= \frac{\partial \psi'}{\partial z}. \end{aligned} \tag{3.4.40}$$

The flow is contained within a three-dimensional channel bounded in the zonal direction by $x \in [0, 2\pi/k]$ and in the meridional direction by $y \in [0, \pi/l]$. A $y - z$ cross-section of the three-dimensional model is shown in Fig. 3.4.1. At

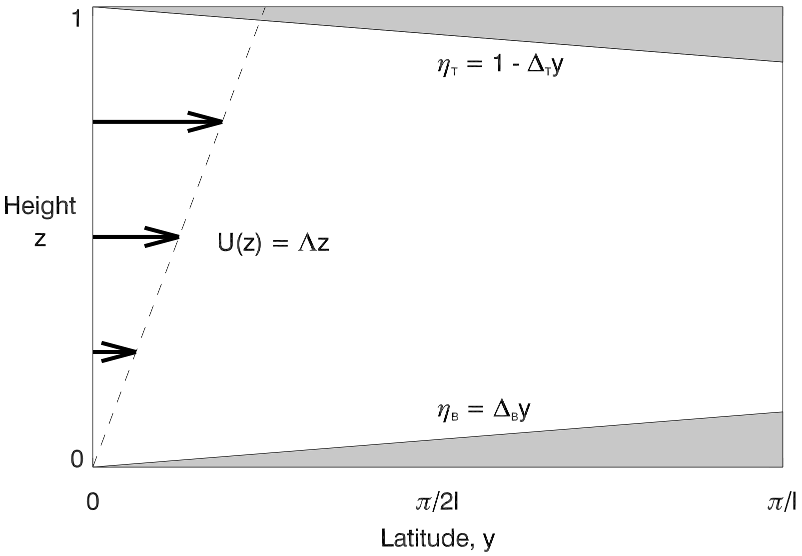

Figure 3.4.1: A schematic of the $y - z$ cross-section in the Eady model with sloping boundaries at the top and bottom of the channel.

the top and bottom of the channel are oppositely sloping boundaries, η_T and η_B respectively, given by

$$\eta_T = 1 - \Delta_T y, \tag{3.4.41}$$

$$\eta_B = \Delta_B y, \tag{3.4.42}$$

where Δ_T and Δ_B are the gradients of the slopes on each vertical boundary. Small amplitude disturbances (3.4.39) are governed by the linearized equation for conservation of quasi-geostrophic potential vorticity (e.g. Vallis [133], p.218,266)

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right) q' = 0, \quad z \in [0, 1]. \tag{3.4.43}$$

If the initial PV in the interior is zero, $q'(x, y, z, 0) = 0$, then PV conservation (3.4.43) implies that $q'(x, y, z, t) = 0$ at all subsequent times. In the Eady model the PV in the interior of the channel is therefore zero. In order to solve equation (3.4.43) vertical boundary conditions are required at the top and bottom of the channel, which are given by (Vallis [133] p.219) the linearized buoyancy equations

$$\left(\frac{\partial}{\partial t}+\Lambda\frac{\partial}{\partial x}\right)b'+\frac{dB}{dy}\frac{\partial\psi'}{\partial x}+\frac{S}{Ro}w'=0, \qquad b'=\frac{\partial\psi'}{\partial z}, \quad \text{on} \quad z=0,1, \tag{3.4.44}$$

where w' is the vertical velocity and Ro is the Rossby number (1.2.7).

There is a contribution to the vertical velocity at each boundary due to Ekman pumping into the interior (as described in section 1.2.8) and the sloping surface given by (3.4.41-3.4.42):

$$\begin{aligned}\frac{1}{Ro}w' &= -\kappa_T\nabla^2\psi'+\mathbf{u}\cdot\nabla\eta_B,\\ &= -\kappa_T\nabla^2\psi'-\Delta_T\frac{\partial\psi'}{\partial x}, \quad \text{on} \quad z=1,\end{aligned} \tag{3.4.45}$$

and

$$\frac{1}{Ro}w'=\kappa_B\nabla^2\psi'+\Delta_B\frac{\partial\psi'}{\partial x}, \quad \text{on} \quad z=0, \tag{3.4.46}$$

where κ_T and κ_B are the Ekman parameters at the top and bottom of the channel proportional to the thickness of the Ekman layer at each boundary. These are explicitly defined in appendix 3.A. In equations (3.4.45-3.4.46), the leading order vertical velocity is $O(Ro)$ as in the two-layer model, which accounts for the Ro^{-1} factors of w'.

For flat, smooth boundaries at $z=0,1$, $w'=0$ in each of (3.4.45) and (3.4.46) and inserting this into (3.4.44) the upper and lower buoyancy equations are

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)b'_T-\Lambda\frac{\partial\psi'}{\partial x} = 0, \tag{3.4.47}$$

$$\left(\frac{\partial}{\partial t}\right)b'_B-\Lambda\frac{\partial\psi'}{\partial x} = 0, \tag{3.4.48}$$

where buoyancy at the upper and lower boundaries are written as b'_T and b'_B respectively. Equations (3.4.47-3.4.48) governing Eady edge wave interaction are equation (2) of Davies & Bishop [19].

Alternatively, with Ekman pumping on each of the sloping boundaries η_T and η_B, substitution of (3.4.45) and (3.4.46) into the buoyancy equation (3.4.44) gives (e.g. equation 1 of Weng [135] or equation 1 of Weng & Barcilon [136])

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\Lambda\frac{\partial}{\partial x}\right)b'_T-(\Lambda+S\Delta_T)\frac{\partial\psi'}{\partial x}-\kappa_T\nabla^2\psi' &= 0, &(3.4.49)\\ \left(\frac{\partial}{\partial t}\right)b'_B-(\Lambda-S\Delta_B)\frac{\partial\psi'}{\partial x}+\kappa_B\nabla^2\psi' &= 0. &(3.4.50)\end{aligned}$$

3.4.2 Eady edge wave description of instability

Edge waves are written in terms of the buoyancy field at the vertical channel boundaries as

$$\begin{aligned}b'_T &= Re\left\{T(t)e^{i(kx+\theta_T(t))}\sin(ly)\sinh(\mu(1-z))\right\}, &(3.4.51)\\ b'_B &= Re\left\{B(t)e^{i(kx+\theta_B(t))}\sin(ly)\sinh(\mu z)\right\}, &(3.4.52)\end{aligned}$$

where (T,θ_T) are the time-dependent amplitude and phase of the upper Eady wave at $z=1$ and (B,θ_B) are the amplitude and phase of the lower wave at $z=0$. The parameter $\mu=\sqrt{S}a$ is the horizontal wavenumber projected onto the vertical coordinate with the Rossby radius of deformation (see appendix 3.A).

From the definition $b'=\partial\psi'/\partial z$, the streamfunction field at the upper and lower boundaries are

$$\psi'_T = \frac{1}{\mu\sinh\mu} Re\left\{\left(T(t)e^{i\theta_T(t)}\cosh\mu + B(t)e^{i\theta_B(t)}\right)e^{ikx}\sin(ly)\right\}, \tag{3.4.53}$$

$$\psi'_B = \frac{1}{\mu\sinh\mu} Re\left\{\left(T(t)e^{i\theta_T(t)} + B(t)e^{i\theta_B(t)}\cosh\mu\right)e^{ikx}\sin(ly)\right\}. \tag{3.4.54}$$

The edge wave description (3.4.51-3.4.54) is inserted into the linearized equations of motion (3.4.49-3.4.50). For simplicity, the slope at the top and bottom of the channel are set to be equal, $\Delta_B = \Delta_T = \Delta$, as is the Ekman friction in each case, $\kappa_T = \kappa_B = \kappa$. Upon taking real and imaginary parts this gives

$$\dot{T} = \frac{k}{\mu\sinh\mu}\left(-(\Lambda + S\Delta)B\sin\theta - \frac{a^2\kappa}{k}(T\cosh\mu - B\cos\theta)\right), \tag{3.4.55}$$

$$\dot{B} = \frac{k}{\mu\sinh\mu}\left(-(\Lambda - S\Delta)T\sin\theta - \frac{a^2\kappa}{k}(B\cosh\mu - T\cos\theta)\right), \tag{3.4.56}$$

$$\begin{aligned}\dot{\theta} = \frac{k}{\mu\sinh\mu}\Bigg(&-\Lambda\mu\sinh\mu + 2\Lambda\cosh\mu + S\Delta\left(\frac{T}{B} - \frac{B}{T}\right)\cos\theta \\ &+ \left(\frac{T}{B} + \frac{B}{T}\right)\left(-\Lambda\cos\theta - \frac{a^2\kappa}{k}\sin\theta\right)\Bigg).\end{aligned} \tag{3.4.57}$$

where $\theta = \theta_T - \theta_B$ is the phase difference between the upper and lower edge waves. Equations (3.4.55-3.4.57) are (7a-d) in Davies & Bishop if $\Delta = \kappa = 0$.

Equations (3.4.55-3.4.57) may be re-scaled using

$$T_E = Te^{-(\kappa_E\cosh\mu)t_E} \quad B_E = e^{-(\kappa_E\cosh\mu)t_E} \quad t_E = \frac{2\Lambda kt}{\mu\sinh\mu}, \tag{3.4.58}$$

where

$$\kappa_E = \frac{a^2\kappa}{2\Lambda k}, \tag{3.4.59}$$

and writing

$$\begin{aligned}\alpha_E &= \frac{1}{2}\left(\cosh\mu - \frac{1}{2}\mu\sinh\mu\right),\\ \delta_E &= \frac{1}{2} - \frac{S\Delta}{2\Lambda},\\ \theta_E &= \theta + \pi,\end{aligned} \tag{3.4.60}$$

gives

$$\dot{T}_E = B_E\left((1-\delta_E)\sin\theta_E - \kappa_E\cos\theta_E\right), \tag{3.4.61}$$

$$\dot{B}_E = T_E\left(\delta_E\sin\theta_E - \kappa_E\cos\theta_E\right), \tag{3.4.62}$$

$$\dot{\theta}_E = 2\alpha_E - \left((1-\delta_E)\frac{B_E}{T_E} + \delta_E\frac{T_E}{B_E}\right)\cos\theta_E + \left(\frac{T_E}{B_E} + \frac{B_E}{T_E}\right)\sin\theta_E, \tag{3.4.63}$$

where differentiation is with respect to t_E. Equations (3.4.61-3.4.63) are of exactly the same form as (3.3.20-3.3.22) with

$$(T_E, B_E, \theta_E, t_E, \delta_E, \alpha_E, \kappa_E) = (Q^*, R^*, \theta, t^*, \epsilon^*, \alpha, \kappa^*). \tag{3.4.64}$$

The process of dissipative destabilization in the two-layer and Eady models are therefore governed by an identical set of ODEs.

The re-scaled phase difference (3.4.60) shows that the Eady and two-layer dynamical systems are π out of phase. This is because the Eady equations are written in terms of Eady edge waves. When written as PV δ-functions a positive buoyancy b' on the boundary corresponds to a negative PV disturbance and a negative buoyancy to positive PV. Therefore, if (3.4.55-3.4.57) are considered from a PV perspective, the phases θ must be re-written in terms of the PV phase difference θ_E.

3.5 Conclusions

The focus of this section has been the dissipative destabilization of waves, which are stable in the frictionless problem, but become unstable when Ekman friction is included in each layer. The mechanism by which dissipative destabilization occurs in the two-layer model has been labelled the Holopainen instability mechanism.

- The idea of baroclinic instability in terms of counter-propagating Rossby waves and 'PV thinking' first introduced by Bretherton [14] was explained and reviewed for the two-layer model using a quantitative approach attributed to Bishop [9, 10]. A dynamical system is derived that fully describes the phase modulation and wave amplification of the interacting PV disturbances.

- A novel approach to describe the Holopainen instability mechanism was introduced in terms of the interaction between an upper layer Rossby wave and lower layer vortex train. The PV approach reveals that the Holopainen mechanism consists of a train of vortices in the lower layer phase-locking and mutually amplifying a Rossby wave in the upper layer. The key difference with the CRW-type mechanism is that the lower layer vortices are spun up by the effect of drag on motions induced by the upper layer Rossby wave. This amplification cannot occur without Ekman friction in the problem.

- The use of phase portraits allows the full range of wave behaviours to be explored, specifically showing the change of behaviour in the interacting PV disturbances when non-zero Ekman friction is included. The instability mechanism of baroclinically unstable flows is not qualitatively changed by the presence of small Ekman friction. The Holopainen instability is governed by a dynamical system distinct from baroclinic instability, which is emphasised by the respective phase portraits.

- The Holopainen instability mechanism has been described in terms of kinematic waves by Swaters [130], which was reviewed. This approach relates dynamic Rossby waves to a kinematic wave, which appears due to non-zero Ekman friction. The condition for instability is then a condition on the phase speeds of these two different waves. The PV thinking approach presented here is arguably more successful in that it reveals the mechanism by which disturbances interact to reach a normal mode configuration and also isolates the process by which friction amplifies the lower layer disturbance.
- It was also demonstrated that the dissipative destabilization of the Eady model is equivalent to the two-layer Holopainen instability as it is governed by an identical dynamical system under appropriate scaling.

3.A Non-dimensionalization of the Eady model

The physical fields introduced in section 3.4.1 are non-dimensionalized with horizontal length scales L, the height of the channel H and horizontal velocity U. This results in non-dimensional parameters as follows

- Dimensionless vertical shear Λ so that $U = 0$ at $z = 0$ and $U = \Lambda$ at $z = 1$.

- Stratification parameter

$$S = \frac{N^2H^2}{f_0^2L^2} = \frac{Ro^2}{Fr^2} = \frac{L_D^2}{L^2}, \tag{3.A.65}$$

  where N is the Brunt-Väisäla frequency determining the background stratification and L_D is the Rossby radius of deformation defined for continuous stratification as $L_D = NH/f_0$.

- Gradient of the slope at the top and bottom of the channel,

$$\Delta_{T,B} = \frac{1}{Ro}\frac{h_{T,B}}{H}, \tag{3.A.66}$$

  where $h_T/H = (1/H)d\eta_T/dy$ and $h_B/H = (1/H)d\eta_B/dy$ for slopes η_T and η_B respectively. The slopes are assumed shallow so that $h_{T,B}/H = O(Ro)$ and $\Delta_{T,B}$ are $O(1)$ quantities.

- $\mu = S^{1/2}a = S^{1/2}(k^2+l^2)^{1/2}$ is the projection of the horizontal wavenumber a onto the vertical coordinate with associated length scale $S = L_D/L$ (Vallis [133] p.218-219).

- Ekman friction is

$$r = \frac{E_k^{1/2}S}{2Ro}, \tag{3.A.67}$$

  defined in the same way as κ for the two-layer model.

Chapter 4

Nonlinear baroclinic equilibration in the presence of Ekman friction

4.1 Introduction

Chapter 2 considered the finite amplitude equilibration of Phillips' two layer model in the absence of Ekman friction. An important feature of this flow at minimum critical shear is that the whole lower layer is a critical layer. A weakly nonlinear theory first derived by Warn & Gauthier [134] (WGP solution) was reviewed and compared to numerical simulations. Chapter 3 introduced Ekman friction in each layer and, following Holopainen [53], described waves that are linearly stable when no Ekman friction is present, but are dissipatively destabilized by the introduction of a small amount of Ekman friction. This dissipative destabilization of the two-layer model was introduced as the Holopainen instability and is associated with an order one movement of the curve of marginal stability.

An alternative weakly nonlinear theory for dissipatively destabilized waves has been given by Romea [109], who made a multiple scales expansion about the marginal stability curve of Holopainen's analysis. In Romea's solution, there is no formation of a critical layer, but rather a Landau-type amplitude equation. It was suggested by Boville [11] that Romea's theory is only applicable below the frictionless marginal stability curve (for dissipatively destabilized waves) though this has not been tested in the literature. There is no obvious reason why Romea's theory should not be accurate for large Ekman numbers, when baroclinic instability is modified by Ekman friction as discussed in chapter 3.

The two weakly nonlinear theories of Warn & Gauthier and Romea will be investigated. The predictions of each weakly nonlinear theory will be tested against two-layer baroclinic lifecycles, initialized with an unstable zonal flow plus near-infinitesimal noise. Following Romea, the QG Ekman number is taken to be equal in each layer, although the relaxation of this restriction does not greatly change the theory [27]. The baroclinic lifecycles including Ekman friction, modelled by the numerical experiments, are arguably more relevant than those of chapter 2 to rotating annulus experiments and oceanic channels forced by wind-stress.

A suite of numerical simulations is performed across the (β, κ) parameter space in order to answer the following questions:

(i) When Ekman friction is present ($\kappa \neq 0$), are the WGP predictions at all relevant? How are the dynamics of nonlinear equilibration changed by dissipative destabilization?

(ii) Flows which satisfy the frictionless criterion for instability $\beta < \frac{1}{2}$ are driven by baroclinic instability rather than the Holopainen mechanism. Are predictions made by Romea's theory relevant to these flows? What happens in the nonlinear stages of the Holopainen regime, particularly in the limit of vanishing Ekman friction $\kappa \to 0$?

(iii) What are the implications for more general hypotheses of nonlinear equilibration? This question is important for 'baroclinic adjustment' theories, which some authors describe as a process during which regions of unstable flow with a negative latitudinal PV gradient are homogenized by PV mixing, leaving the flow stable to further instability (by the Charney-Stern-Pedlosky criterion described in section 2.3). Is the frictionless concept of baroclinic adjustment ever relevant for flows with $\kappa \neq 0$?

The plan of this chapter is as follows. Section 4.2 briefly describes the physical problem and numerical set-up, specifically how it differs from that of chapter 2. Section 4.3 reviews the weakly nonlinear theory of Romea, which builds on the linear theory introduced in chapter 3. The results of numerical simulations are discussed in section 4.4 and conclusions are presented in section 4.5.

4.2 Physical problem and numerical set-up

4.2.1 The physical problem

The physical problem addressed in this chapter is the same infinite channel of section 2.2 with the exception that Ekman layers are present at the upper and lower boundaries. As described in section 1.2.8 the equations of motion are (1.2.54), repeated here for convenience

$$\begin{aligned}\left(\frac{\partial}{\partial t}-\frac{\partial\psi_i}{\partial y}\frac{\partial}{\partial x}+\frac{\partial\psi_i}{\partial x}\frac{\partial}{\partial y}\right)q_i &= -\kappa\nabla^2\psi_i, \qquad (4.2.1)\\ q_i &= \beta y+\nabla^2\psi_i+(-1)^i\left(\frac{\psi_1-\psi_2}{2}\right).\end{aligned}$$

The no-normal flow and Phillips boundary conditions, (1.3.56) and (1.3.59) respectively together with periodicity (1.3.55) in the re-circulating channel are

$$\begin{aligned}\frac{\partial\psi_i}{\partial x} &= 0 \quad \text{on} \quad y=\pm\frac{1}{2}W, \qquad (4.2.2)\\ \frac{\partial^2\overline{\psi}_i}{\partial t\partial y}+\kappa\left(\frac{\partial\overline{\psi}_i}{\partial y}+\delta_{1i}\right) &= 0 \quad \text{on} \quad y=\pm\frac{1}{2}W, \qquad (4.2.3)\\ \psi_i(x,y) &= \psi_i(x+L,y), \qquad (4.2.4)\end{aligned}$$

for $i=1,2$, where δ_{1i} is the Kronecker delta. The four parameters in (4.2.1-4.2.4) have been previously defined as the inverse criticality β, the QG Ekman number κ, the non-dimensional channel width W and non-dimensional channel length L.

As in section 2.2.3, the initial conditions consist of the basic flow plus an infinitesimal perturbation so that the subsequent evolution is dominated by the fastest growing normal mode with zonal wavenumber k_m. The channel width parameter is fixed here to be $W = 2^{3/4}\pi$ (which in frictionless flow means that the fastest growing mode is isotropic at minimum critical shear, $k_m = \pi/W$). In this chapter, the dependence of the baroclinic lifecycles on the width parameter W is not investigated. All the simulations of this chapter are therefore for nearly isotropic waves, which may be partly justified on physical grounds, since approximately isotropic waves emerge spontaneously in most annulus experiments, atmospheric and oceanic flows (e.g. James [58], section 8.2).

4.2.2 Numerical implementation

The quasi-geostrophic equations (4.2.1) are integrated using exactly the same method as section 2.2.3 with a PV diffusion $\nu_q \nabla^2 q_i$ on the right-hand side of (4.2.1) to maintain numerical stability (as discussed in section 1.4). The spatial discretizations and corresponding diffusivities ν_q and time-steps δt are those in Table 2.1 for the smallest Ekman numbers, although less PV diffusion is required when the Ekman number κ is larger and is chosen to be as small as possible. It was shown in section 2.2.3 (see Fig. 2.2.1 and surrounding discussion) that the results of numerical simulations do not depend on the value of the PV diffusivity ν_q.

A sweep of the (β, κ) parameter space is performed with the channel width fixed at $W = 2^{3/4}\pi$ throughout, and the fastest growing wavenumber a_m is calculated separately for each simulation using the linear theory of section 3.2. In total, over 330 experiments are performed within the parameter space $\beta \in [0, 0.55]$, $\kappa \in [0, 0.2]$ for which the flow is unstable.

4.3 Romea's weakly nonlinear theory

4.3.1 Linear theory

The linear theory for the physical problem with Ekman friction in each layer was reviewed in section 3.2. The key point discussed in that section is that, due to the dissipative destabilization mechanism, the marginal stability curve when Ekman friction is absent ($\kappa = 0$, equation 2.3.16) and when Ekman friction is vanishingly small ($\kappa \to 0$, equation 3.2.8) are different, as demonstrated in Fig. 3.2.1.

Section 3.2 showed that the plane wave solution (3.2.4) in the equations of motion (4.2.1) leads to the dispersion relation (3.2.5) and corresponding complex streamfunction ratio (3.2.6), all of which are repeated here for convenience as they will be used in the derivation of Romea's weakly nonlinear theory below,

$$\begin{aligned}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} &= -\begin{pmatrix}y\\ 0\end{pmatrix} + Re\ \tilde{\epsilon}\begin{pmatrix}1\\ \gamma\end{pmatrix} e^{ik(x-ct)}\cos ly, &(4.3.5)\\ c^{\pm} &= \frac{1}{2} - \frac{\left(a^2+\frac{1}{2}\right)}{a^2(a^2+1)}\beta - i\kappa\frac{a^2+\frac{1}{2}}{k(a^2+1)} &(4.3.6)\\ &\quad \pm\frac{1}{2a^2(a^2+1)}\left(\beta^2 + a^4(a^4-1) - \frac{a^4\kappa^2}{k^2} + 2i\frac{a^2\kappa}{k}\beta\right)^{\frac{1}{2}}, &\\ \gamma^{\pm} &= 2\left(a^2+\frac{1}{2}\right) + 2\left(\frac{\beta+\frac{1}{2}+i\frac{a^2\kappa}{k}}{c^{\pm}-1}\right). &(4.3.7)\end{aligned}$$

The marginal stability curve (3.2.7) then corresponds to the parts of the parameter space where $\operatorname{Im} c^+ = 0$, the transition between regions that are stable and those regions where instability occurs ($\operatorname{Im} c^+ > 0$). The marginal stability curve (3.2.7) is

$$\beta_c^R(a,\kappa) = \left(a^2 + \frac{1}{2}\right)\left(a^2\left(1-a^2\right) - \frac{4a^4}{k^2}\kappa^2\right)^{1/2}. \tag{4.3.8}$$

As in the WGP solution, the minimum critical shear β_m^R is defined to be the maximum value of β for which instability occurs. The minimum critical shear for the $\kappa \to 0$ marginal stability curve is marked on Fig. 3.2.1. The total wavenumber at minimum critical shear $\beta = \beta_m^R$ will be labelled a_m, and the corresponding zonal wavenumber is $k_m = (a_m^2 - \pi^2/W^2)^{1/2}$. Recall in chapter 2 that for frictionless flow ($\kappa = 0$), $\beta_m^I = \frac{1}{2}$ and $a_m = 2^{-1/4}$. When $\kappa \neq 0$, β_m^R and a_m are determined by finding

$$\beta_m^R(\kappa) = \sup_{a>\pi/W}\left\{\beta_c^R(a,\kappa)\right\}. \tag{4.3.9}$$

The superscript R will hereafter be dropped from β_m^R.

4.3.2 Weakly nonlinear theory

This section reviews the work of Romea [109] and updates his analysis using a more modern mathematical approach. Romea's theory is formulated for flows close to marginal stability $\beta = \beta_m$ for waves with the marginally stable total wavenumber a_m as determined by (4.3.9). In fact, unlike the WGP solution, the Romea analysis is applicable to any point on the marginal stability curve (4.3.8).

To make a perturbation expansion a new small parameter, the 'Ekman friction supercriticality' ϵ_* is introduced as

$$\epsilon_* = \sqrt{\beta_m - \beta}. \tag{4.3.10}$$

Note that the Ekman friction supercriticality ϵ_* is distinct from the supercriticality ϵ in the WGP problem of chapter 2, because in general $\beta_m \neq \frac{1}{2}$, i.e. the frictionless value. The natural slow timescale with Ekman friction is $T = \epsilon_*^2 t$, which has a slower (quadratic) dependence on supercriticality compared to the linear dependence on ϵ in the WGP problem.

Substitution of $\frac{\partial}{\partial t} \rightarrow \frac{\partial}{\partial t} + \epsilon_*^2 \frac{\partial}{\partial T}$ into the equations of motion (4.2.1) and writing the streamfunctions and potential vorticities in terms of the basic states plus a perturbation (see equations 1.3.64-1.3.65) gives

$$\left(\frac{\partial}{\partial t} + \epsilon_*^2 \frac{\partial}{\partial T} + \frac{\partial}{\partial x}\right) q_1' + \left(\beta_m + \frac{1}{2} - \epsilon_*^2\right) \frac{\partial \psi_1'}{\partial x} + J\left(\psi_1', q_1'\right) = -\kappa \nabla^2 \psi_1', \tag{4.3.11}$$

$$\left(\frac{\partial}{\partial t} + \epsilon_*^2 \frac{\partial}{\partial T}\right) q_2' + \left(\beta_m - \frac{1}{2} - \epsilon_*^2\right) \frac{\partial \psi_2'}{\partial x} + J\left(\psi_2', q_2'\right) = -\kappa \nabla^2 \psi_2'. \tag{4.3.12}$$

As in the WGP problem (equations 2.4.28-2.4.29) the streamfunctions ψ_i and potential vorticities q_i are expanded in ascending powers of ϵ_*

$$\psi_i = -y\delta_{1i} + \epsilon_* \psi_i^{(0)} + \epsilon_*^2 \psi_i^{(1)} + \epsilon_*^3 \psi_i^{(2)} + \dots, \tag{4.3.13}$$

$$q_i = \left(\beta_m - \epsilon_*^2 + (-1)^{i+1}\frac{1}{2}\right) + \epsilon_* q_i^{(0)} + \epsilon_*^2 q_i^{(1)} + \epsilon_*^3 q_i^{(2)} + \dots, \tag{4.3.14}$$

$$\text{where} \qquad q_i^{(j)} = \nabla^2 \psi_i^{(j)} + (-1)^i \left(\frac{\psi_1^{(j)} - \psi_2^{(j)}}{2}\right).$$

Substitution of (4.3.13-4.3.14) into (4.3.11-4.3.12) yields a sequence of equations at successive orders of ϵ_*, which may be solved to yield an amplitude equation for the upper layer wave.

Collecting terms of $O(\epsilon_*)$ yields equations satisfied by the marginal wave

$$\mathcal{L}_R \mathbf{\Psi^{(0)}} = 0 \tag{4.3.15}$$

where the matrix differential operator $\mathcal{L}_R$ is given by

$$\mathcal{L}_R \equiv \begin{pmatrix} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)\left(\nabla^2 - \frac{1}{2}\right) + \left(\beta_m + \frac{1}{2}\right)\frac{\partial}{\partial x} + \kappa \nabla^2 & \frac{1}{2}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) \\ \frac{1}{2}\frac{\partial}{\partial t} & \left(\frac{\partial}{\partial t}\right)\left(\nabla^2 - \frac{1}{2}\right) + \left(\beta_m - \frac{1}{2}\right)\frac{\partial}{\partial x} + \kappa \nabla^2 \end{pmatrix} \tag{4.3.16}$$

acting on the vector streamfunction

$$\mathbf{\Psi^{(0)}} = \begin{pmatrix} \psi_1^{(0)} \\ \psi_2^{(0)} \end{pmatrix}. \tag{4.3.17}$$

The solution to equation (4.3.15) is the linear plane wave solution (4.3.5) at minimum critical shear given by

$$\mathbf{\Psi^{(0)}} = \begin{pmatrix} \psi_1^{(0)} \\ \psi_2^{(0)} \end{pmatrix} = \mathrm{Re} \begin{pmatrix} 1 \\ \gamma_m \end{pmatrix} A\left(T\right) e^{ik_m(x - c_m t)} \cos\left(\frac{\pi y}{W}\right), \tag{4.3.18}$$

with corresponding potential vorticities

$$\begin{pmatrix} q_1^{(0)} \\ q_2^{(0)} \end{pmatrix} = \mathrm{Re} \begin{pmatrix} -\left(a_m^2 + \frac{1}{2}\right) + \frac{1}{2}\gamma_m \\ -\left(a_m^2 + \frac{1}{2}\right)\gamma_m + \frac{1}{2} \end{pmatrix} A\left(T\right) e^{ik_m(x - c_m t)} \cos\left(\frac{\pi y}{W}\right). \tag{4.3.19}$$

In the WGP problem the leading order potential vorticity in the lower layer is zero, whereas in the Romea problem, both leading order potential vorticities (4.3.19) are non-zero and the subsequent nonlinear behaviour is determined by the interaction of both layers.

$A(T)$ is the complex amplitude of the wave, γ_m is the complex amplitude ratio and c_m is the real phase speed. c_m and γ_m are calculated from (4.3.6) and (4.3.7) respectively as

$$c_m = \frac{1}{2} - \frac{\beta_m}{a_m^2 + \frac{1}{2}}, \tag{4.3.20}$$

and

$$\gamma_m = 2\left(a_m^2 + \frac{1}{2}\right) - 2\left(\frac{\beta_m + \frac{1}{2} + \frac{ia_m^2\kappa}{k}}{\frac{1}{2} + \beta_m/(a_m^2 + \frac{1}{2})}\right). \tag{4.3.21}$$

The non-zero imaginary part of γ_m means that there is a leading order phase shift between the upper and lower layer due to the Ekman friction. The growth of the dissipatively destabilized wave is due to the release of more energy by the Ekman friction induced phase shift than is dissipated by Ekman friction. In the WGP problem the amplitude coefficient is strictly real at leading order (viz. equation 2.3.20) as the phase shift is proportional to dA/dT, not A, which fundamentally alters the asymptotic expansion. It is for this reason that the slow timescale T in the Romea problem is $O(\epsilon_*^2)$ rather than $O(\epsilon)$ as in the WGP problem. The leading order phase shift also means that there is a non-zero PV flux in each layer, which is necessary for momentum transport between the layers. The momentum transport is proportional to $|A|^2$ and it is reduced by the restoring effect of momentum transport within the Ekman layers. The WGP problem is entirely different as the only momentum transport is that associated with wave growth and decay.

At $O(\epsilon_*^2)$ the expansion (4.3.11-4.3.12) reveals

$$\mathcal{L}_R\mathbf{\Psi^{(1)}} = \mathbf{f^{(1)}}, \tag{4.3.22}$$

with the forcing term

$$\mathbf{f}^{\mathbf{(1)}} = \begin{pmatrix} -J\left(\psi_1^{(0)}, q_1^{(0)}\right) \\ -J\left(\psi_2^{(0)}, q_2^{(0)}\right) \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{\pi k_m \gamma_i |A|^2}{4W} \sin\left(\frac{2\pi y}{W}\right), \tag{4.3.23}$$

where $\gamma_i = \mathrm{Im}\{\gamma_m\}$. The Jacobians in (4.3.23) are directly evaluated using the $O(\epsilon_*)$ solution (4.3.18) and since γ_i is proportional to κ, they are non-zero when $\kappa \neq 0$. The Romea problem is therefore invalid if $\kappa = 0$, yet remains well-defined in the limit $\kappa \to 0$.

The right-hand side of equation (4.3.23) is dependent on only y and T and so particular integrals of the form $\psi_1^{(1)}(y,T)$ and $\psi_2^{(1)}(y,T)$ are sought. See appendix 4.A for the full derivation of the Romea solution, which at $O(\epsilon_*^2)$ is

$$\mathbf{\Psi}^{\mathbf{(1)}} = \begin{pmatrix} \psi_1^{(1)} \\ \psi_2^{(1)} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{a_m^2\left(a_m^2 + \frac{1}{2}\right)W}{4\pi\left(2\beta_m + a_m^2 + \frac{1}{2}\right)} |A|^2 \left(\sin\left(\frac{2\pi y}{W}\right) + \frac{2\pi}{W} y\right). \tag{4.3.24}$$

Equation (4.3.24) is the $O(\epsilon_*^2)$ correction to the mean flow, which results from the competition between eddy fluxes and Ekman friction, each of which are linear in κ and act to force the mean flow. The streamfunction $\mathbf{\Psi}^{\mathbf{(1)}}$, represents the balance between these two effects and has only a weak dependence on κ due to β_m, a_m and k_m through equation (4.3.8).

At the next order $O(\epsilon_*^3)$, equations (4.3.11-4.3.12) take the form

$$\mathcal{L}_R \mathbf{\Psi}^{\mathbf{(2)}} = \mathbf{f}^{(2)}, \tag{4.3.25}$$

where

$$\mathbf{f}^{(2)} = \begin{pmatrix} -\frac{\partial q_1^{(0)}}{\partial T} + \frac{\partial \psi_1^{(0)}}{\partial x} - J(\psi_1^{(0)}, q_1^{(1)}) - J(\psi_1^{(1)}, q_1^{(0)}) \\ -\frac{\partial q_2^{(0)}}{\partial T} + \frac{\partial \psi_2^{(0)}}{\partial x} - J(\psi_2^{(0)}, q_2^{(1)}) - J(\psi_2^{(1)}, q_2^{(0)}) \end{pmatrix}. \tag{4.3.26}$$

Equation (4.3.25) does not need to be solved explicitly to derive the amplitude evolution equation for $A(T)$. The amplitude evolution equation instead derives from a solvability condition arising because the operator $\mathcal{L}_R$ has a non-trivial kernel. If the forcing $\mathbf{f}^{(2)}$ were to project onto the kernel of $\mathcal{L}_R$, then the response of $\mathbf{\Psi^{(2)}}$ would include resonantly excited, or so-called secular terms that grow linearly in time. If secular terms were to be present, then the expansion of the equations of motion (4.2.1) would become disordered on timescales $t \sim \epsilon_*^{-2}$ (or $T \sim O(1)$). For the expansion (4.3.13-4.3.14) to remain uniformly valid on the long timescale, the solvability condition must be enforced to remove secular terms from the solution.

An elegant method[1] to derive the solvability condition is to consider the adjoint of the operator $\mathcal{L}_R$. The adjoint is defined with reference to an inner product acting on a pair of streamfunctions $\mathbf{\Phi} = (\phi_1, \phi_2)^\dagger$ and $\mathbf{\Psi} = (\psi_1, \psi_2)^\dagger$, as follows

$$\langle \mathbf{\Phi}, \mathbf{\Psi} \rangle = \frac{1}{\overline{T}} \int_0^{\overline{T}} \int_D \phi_1\psi_1 + \phi_2\psi_2 d^2\mathbf{x}dt \tag{4.3.27}$$

where integration in space is over the domain of the channel, and the time integral corresponds to averaging over an interval of length $\overline{T}$ where $O(1) \ll \overline{T} \ll O(\epsilon_*^{-2})$. The adjoint operator $\mathcal{L}_R^\dagger$ is then defined by

$$\langle \mathbf{\Phi}, \mathcal{L}_R \mathbf{\Psi} \rangle = \langle \mathcal{L}_R^\dagger \mathbf{\Phi}, \mathbf{\Psi} \rangle, \tag{4.3.28}$$

---

[1] The solvability condition is equivalent to the procedure in Romea's original paper [109] where the secular terms in $\psi_1^{(2)}$ and $\psi_2^{(2)}$ are removed to preserve the ordering of magnitudes in the perturbation expansion. However, proceeding with the use of a solvability condition as in Warn & Gauthier [134] is a slightly more elegant explanation.

and using integration by parts twice, is given by

$$\mathcal{L}_R^\dagger \equiv \begin{pmatrix} \begin{array}{c} -\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\left(\nabla^2-\frac{1}{2}\right) \\ -\left(\beta_m+\frac{1}{2}\right)\frac{\partial}{\partial x}+\kappa\nabla^2 \end{array} & -\frac{1}{2}\frac{\partial}{\partial t} \\ -\frac{1}{2}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right) & \begin{array}{c} -\left(\frac{\partial}{\partial t}\right)\left(\nabla^2-\frac{1}{2}\right) \\ -\left(\beta_m-\frac{1}{2}\right)\frac{\partial}{\partial x}+\kappa\nabla^2 \end{array} \end{pmatrix}. \tag{4.3.29}$$

Now consider any pair of functions $\mathbf{\Phi}_0 = (\phi_{0,1}, \phi_{0,2})^\dagger$ belonging to the kernel of $\mathcal{L}_R^\dagger$ so that $\mathcal{L}_R^\dagger\mathbf{\Phi}_0 = 0$. Taking the inner product of equation (4.3.25) with $\mathbf{\Phi}_0$ gives

$$\langle \mathbf{\Phi}_0, \mathcal{L}_R\mathbf{\Psi}^{(2)}\rangle = \langle \mathcal{L}_R^\dagger\mathbf{\Phi}_0, \mathbf{\Psi}^{(2)}\rangle = \langle \mathbf{\Phi}_0, \mathbf{f}^{(2)}\rangle. \tag{4.3.30}$$

The solvability condition is therefore simply

$$\langle \mathbf{\Phi}_0, \mathbf{f}^{(2)}\rangle = 0. \tag{4.3.31}$$

A bit of work shows that the relevant solution of $\mathcal{L}_R^\dagger\mathbf{\Phi}_0 = 0$ is

$$\mathbf{\Psi}_0 = \mathrm{Re}\begin{pmatrix} -c_m \\ (1-c_m)\gamma_m^* \end{pmatrix} e^{ik_m(x-c_mt)}\cos\left(\frac{\pi y}{W}\right) \tag{4.3.32}$$

where γ_m^* is the complex conjugate of (4.3.21). After a little manipulation, evaluation of the solvability condition (4.3.31, see appendix 4.A), gives

$$A_T = (\sigma_r + i\sigma_i)k_mA - (\mu_r + i\mu_i)k_m|A|^2A, \tag{4.3.33}$$

where

$$\begin{aligned} \sigma &= \sigma_r + i\sigma_i \\ &= \left(\frac{\kappa}{k_m}\right)\frac{4a_m^2\left(a_m^2+\frac{1}{2}\right)\left(\beta_m - 4a_m^2\left(a_m^2+\frac{1}{2}\right)^2(\kappa/k_m)\,i\right)}{\beta_m^2 + 16a_m^4\left(a_m^2+\frac{1}{2}\right)^4(\kappa/k_m)^2}, \end{aligned} \tag{4.3.34}$$

and

$$\begin{aligned}\mu &= \mu_r + i\mu_i \\ &= \frac{a_m^2\left(a_m^2+\frac{1}{2}\right)^2\left(3a_m^2\left(1-a_m^2\right)+4\left(a_m^2-\frac{1}{2}\right)\pi^2/W^2\right)}{2\left(2\beta_m+a_m^2+\frac{1}{2}\right)\left(\beta_m^2+16a_m^4\left(a_m^2+\frac{1}{2}\right)^4\left(\kappa/k_m\right)^2\right)} \\ &\qquad \times\left(4a_m^2\left(a_m^2+\frac{1}{2}\right)^2\frac{\kappa}{k_m}+i\beta_m\right). \qquad (4.3.35)\end{aligned}$$

The expressions (4.3.34) and (4.3.35) are much simplified from the corresponding expression in Romea ([109], see his A1 and A2) and make clear the parameter dependence of the solution.

Calculation of ($A^*\times$(4.3.33)$+A\times$(4.3.33)*) results in the equation governing the evolution of the real amplitude

$$\frac{d|A|^2}{dT} = 2k_m|A|^2\left(\sigma_r-\mu_r|A|^2\right). \qquad (4.3.36)$$

Equation (4.3.36) is a Landau-type equation [64], common for dynamical systems near marginal stability (e.g. Drazin & Reid [22], p.370-379). It governs a variety of fluid dynamical systems starting with Stuart [125] to describe Poiseuille flow at relatively low Reynolds number. In such flows, linear instability and nonlinear equilibration requires $\sigma_r, \mu_r > 0$, which in the present context corresponds to supercritical flow. The linear growth rate in time (which from equation 4.3.36 is $2k_m\sigma_r\epsilon_*^2$) can be obtained directly from the dispersion relation (4.3.6) by a Taylor expansion about $\beta=\beta_m$. The stabilizing effect of the mean flow correction (4.3.24) on the fundamental wave is governed by the coefficient μ_r, which involves a direct calculation from β_m, W, a_m, k_m and κ.

The Romea solution for the amplitude evolution is found upon integration of (4.3.36) to be

$$|A(T)|^2 = \frac{|A(0)|^2 e^{2k_m\sigma_r T}}{1 + (\mu_r/\sigma_r)\,|A(0)|^2\left(e^{2k_m\sigma_r T} - 1\right)}. \tag{4.3.37}$$

A wave is unstable if $\sigma_r > 0$ and the wave amplitude is a monotonically increasing function of the slow timescale T, which at long times approaches the asymptotic limit

$$\begin{aligned}
\epsilon_*|A|_{eq} &= \lim_{T\to\infty} \epsilon_*|A(T)|, \\
&= (\beta_m - \beta)^{1/2}\left(\frac{\sigma_r}{\mu_r}\right)^{1/2}, \\
&= (\beta_m - \beta)^{1/2}\left(\frac{2\beta_m\left(2\beta_m + a_m^2 + \frac{1}{2}\right)W^2}{a_m^2\left(a_m^2 + \frac{1}{2}\right)^3\left(3a_m^2\left(1 - a_m^2\right)W^2 + 4\left(a_m^2 - \frac{1}{2}\right)\pi^2\right)}\right)^{1/2}.
\end{aligned} \tag{4.3.38}$$

The asymptotic limit (4.3.38) reveals the parameter dependency of the equilibrated wave amplitude. There is no explicit dependence on the Ekman number κ, which only influences the equilibrated wave amplitude via changes in the inverse criticality β_m and total wavenumber a_m at marginal stability.

Explicit expressions for (4.3.38) in the limit of vanishing Ekman friction $\kappa \to 0$ may be found by calculating explicit expressions for the inverse criticality and total wavenumber at marginal stability from (4.3.8) and (4.3.9):

$$\beta_m = \frac{\sqrt[4]{3}\left(3 + \sqrt{3}\right)}{8\sqrt{2}}, \qquad a_m = \frac{1}{2}\left(1 + \sqrt{3}\right)^{1/2} \quad \text{as} \quad \kappa \to 0. \tag{4.3.39}$$

Substitution of these values into (4.3.38) gives

$$\lim_{\kappa\to 0} \epsilon_*|A|_{eq} = (\beta - \beta_m)^{1/2}\left(\frac{32\sqrt[4]{3}\left(\sqrt{2} + \sqrt[4]{3}\right)W^2}{8\left(3 + \sqrt{3}\right)\pi^2 + 9\left(2 + \sqrt{3}\right)W^2}\right), \tag{4.3.40}$$

the equilibrated wave amplitude in the limit of vanishing Ekman friction

A direct comparison may be made between the result (4.3.40) and the corresponding result for the WGP theory given by (2.4.47), noting that the supercriticalities in these two expressions are different ($\beta_m \neq \frac{1}{2}$ in 4.3.40). The most significant difference between the two formulations is that Romea predicts an equilibrated wave amplitude that is independent of W in the wide channel limit $W \rightarrow \infty$, whereas WGP predicts an amplitude proportional to W. For frictionless instability, it is required that $W > 2^{1/4}\pi$ and therefore, the coefficient multiplying the supercriticality ϵ_* is always smaller than that in WGP. The equilibrated amplitudes (4.3.38) and (4.3.40) will be compared to numerical simulations in the next section.

4.4 Behaviour at finite supercriticality

In this section the Romea analytical solution (4.3.37) will be compared directly to numerical simulations. The characteristic features of the flow will be discussed and contrasted to the altogether different evolution of the WGP problem where no Ekman friction is present. An exploration of the (β, κ) parameter space will be made and an intermediate region of behaviour is discovered, aspects of which are described by both the WGP and Romea weakly nonlinear theories.

4.4.1 Properties of the flow near marginal stability

The first task is to ascertain where in the parameter space the Romea analytic solution (4.3.37) is relevant. The sweep of the (β, κ) parameter space was described in section 4.2.2 and from this space six numerical simulations in particular, labelled EXPTS I-VIa and detailed in Table 4.1, will be used to describe a variety of flow behaviours.

It is possible to estimate the type of flows that the parameters in Table 4.1 might correspond to. For example, for a flow with Rossby number $Ro = 0.1$ (typical of the ocean or annulus experiments), a QG Ekman number $\kappa = 0.04$ (as in EXPT I) corresponds to an Ekman number $E_k^{1/2} \approx \kappa Ro = 4 \times 10^{-3}$ or $E_k \approx 10^{-5}$. For a fluid such as water (with kinematic viscosity of the order $\nu = 10^{-6}$), in an annulus of depth $H = 0.1$, this would correspond to a rotation rate of $\Omega \approx 10\text{rad s}^{-1}$, which is around twice the rotation rates typically investigated in annulus experiments and so within the realm of of feasible experiment design[2]. A QG Ekman number of $\kappa = 0.1$ on the other hand would correspond to a rotation rate of $\Omega \approx 1\text{rad s}^{-1}$, which is more achievable. This cursory examination suggests that the small Ekman numbers of the dissipatively destabilized flows in particular might be difficult to access in a laboratory setting (and be outside observed flows in the atmosphere or ocean). The question remains as to whether the weakly nonlinear regime can be reproduced in a laboratoryor observed in real geophysical flows and the issues associated with this are discussed in chapter 6.

EXPTS I and II lie near the curve of marginal stability and therefore Romea's solution might be expected to be accurate. EXPT I, $(\beta, \kappa) = (0.53, 0.04)$ is dissipatively destabilized, which means that the sole destabilizing mechanism is the Holopainen instability mechanism (as described in chapter 3). If there

---

[2] from Peter Read, direct communication

Simulation	Inverse Criticality β	QG Ekman number κ	Description
EXPT I	0.53	0.04	Dissipatively destabilized. Romea is accurate.
EXPT II	0.31	0.16	Strong Ekman friction. Romea is accurate.
EXPT III	0.50	0.04	Dissipatively destabilized. Illustrates Holopainen mechanism.
EXPT IV	0.495	0	Weak baroclinic instability. WGP is accurate.
EXPT V	0.48	0.0025	Weak baroclinic instability and weak Ekman friction. WGP predicts early times, Romea predicts equilibrium.
EXPT VIa	0.47	0.005	As EXPT V, but at greater supercriticality.
EXPT VIb	0.47	PV damping $r = 0.005$	Weak baroclinic instability with PV damping. Forced WGP is accurate.

Table 4.1: Summary of selected simulations. All experiments have channel width $W = 2^{3/4}\pi$.

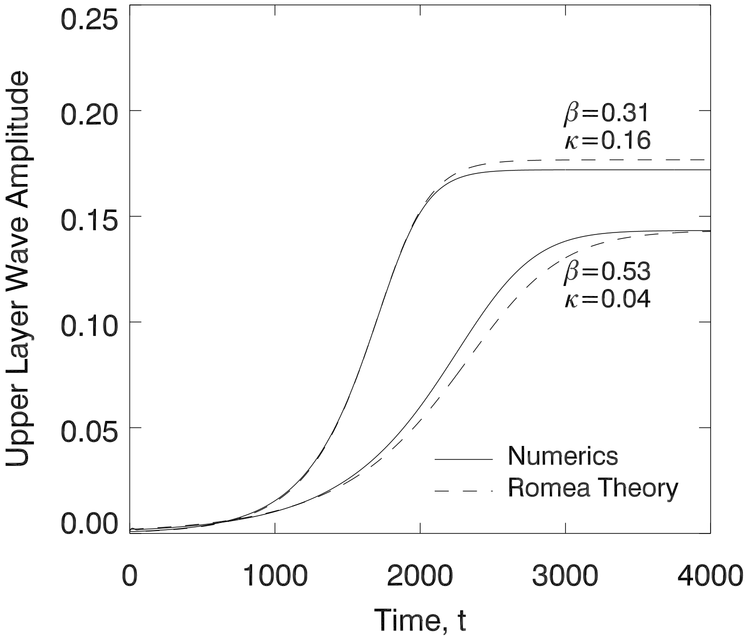

Figure 4.4.1: Evolution of the fundamental wave amplitudes $|A_1^s|$ for EXPT I $(\beta, \kappa) = (0.53, 0.04)$ and EXPT II $(0.31, 0.16)$. Numerical results are solid lines and analytical solutions calculated with Romea's nonlinear theory $\epsilon_*|A|$ are dashed lines.

were no Ekman friction ($\kappa = 0$), then EXPT I would be stable to baroclinic instability. In contrast, EXPT II with $(\beta, \kappa) = (0.31, 0.16)$ is close to marginal stability because of strong Ekman friction and in the absence of Ekman friction would be strongly unstable to baroclinic instability.

The time evolution of the upper layer wave amplitudes for EXPTS I and II are shown in Fig. 4.4.1. The numerical simulations are plotted as solid curves, and the corresponding results from Romea's equation (4.3.37) as dashed curves. The complex amplitude of the waves in numerical simulations A_i^s are calculated from the upper layer streamfunction using

$$A_i^s(t) = \frac{4}{LW} \int_D \psi_i(x, y, t) e^{-ik_m x} \cos\left(\frac{\pi y}{W}\right) d^2\mathbf{x}, \quad i = 1, 2. \tag{4.4.41}$$

The wave amplitudes are then given by $|A_i^s|$. Note that the wave amplitudes

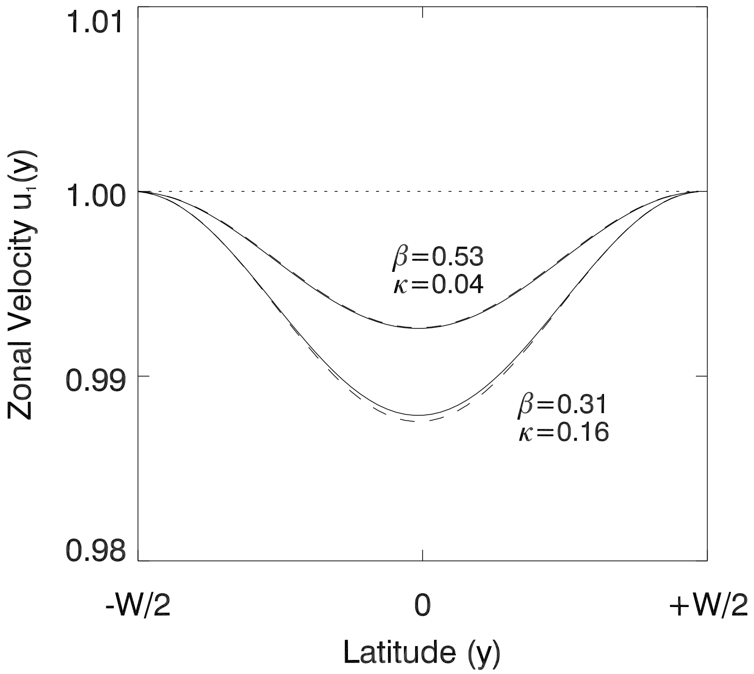

Figure 4.4.2: The zonal velocity of the equilibrated state in the upper layer for the two cases (i) $\beta = 0.53, \kappa = 0.04$ and (ii) $\beta = 0.31, \kappa = 0.19$. In each case, the numerical simulations are solid lines and the analytic corrected mean flow as dashed lines.

in chapter 2 were defined by (2.2.9) using the PV field rather than the streamfunction in order to exploit the results derived from pseudomomentum bounds in section 2.5.1. The amplitude $|A_1^s(t)|$ can be directly compared to $\epsilon_*|A(T)|$ in Romea's theory. From Fig. 4.4.1, the Romea theory is clearly accurate for both EXPT I and II where the maximum difference between the wave amplitudes in either simulation is 12% (during EXPT I). The equilibrated amplitude is predicted by Romea's theory (equation 4.3.38) to within 3% in both experiments. Fig. 4.4.2 shows the upper layer zonal mean winds $\overline{u}_1 = -\overline{\psi}_{1y}$ (solid curves) for EXPTS I and II at late times in the wave evolution when the flow is steady. The zonal mean correction predicted by Romea's theory in equation (4.3.24), $-\epsilon_*^2\psi_{1y}^{(1)}$, is also plotted in Fig. 4.4.2 (dashed lines). The mean flow correction is well-captured by Romea's theory for each of EXPTS I and II. The upper layer zonal velocity decreases across the channel with the largest adjustment at the channel centre (the lower layer zonal flow adjustment is not plotted, since according to (4.3.24) it is simply oppositely signed to the upper layer adjustment).

The discussion now turns to observations of the lower layer PV field, the development of which is plotted for EXPTS III and IV in Fig. 4.4.3. EXPT III is a point in the parameter space $(\beta, \kappa) = (0.5, 0.04)$ that illustrates a 'clean' example of Romea-type equilibration: the inverse criticality $\beta = 0.5$ eliminates the background PV gradient in the lower layer and the vortex train behaviour is clear (this was also the value of β chosen to describe the Holopainen instability mechanism in Fig.3.3.4). In contrast EXPT IV, with $(\beta, \kappa) = (0.495, 0)$, shows a frictionless WGP-type equilibration (from chapter 2). As emphasised in section 2.6.1, the development of the lower layer PV field is key to understanding the equilibration mechanism for EXPTS III and IV.

The dynamical behaviour of EXPT III is qualitatively similar to EXPT I with waves that are dissipatively destabilized and grow due to the Holopainen instability, although the flow is slightly more unstable in EXPT III. The Ekman friction acts to relax the system back to the initial zonal flow throughout the evolution. Therefore, since $q_2 = 0$ initially, the existence of PV in the lower layer must ultimately be due to the action of Ekman friction on the velocity field induced by the PV distribution in the upper layer.

Initially, in Fig. 4.4.3 panel (a) the linear growing mode is present at small amplitude. The disturbance is amplified during the linear growth stage

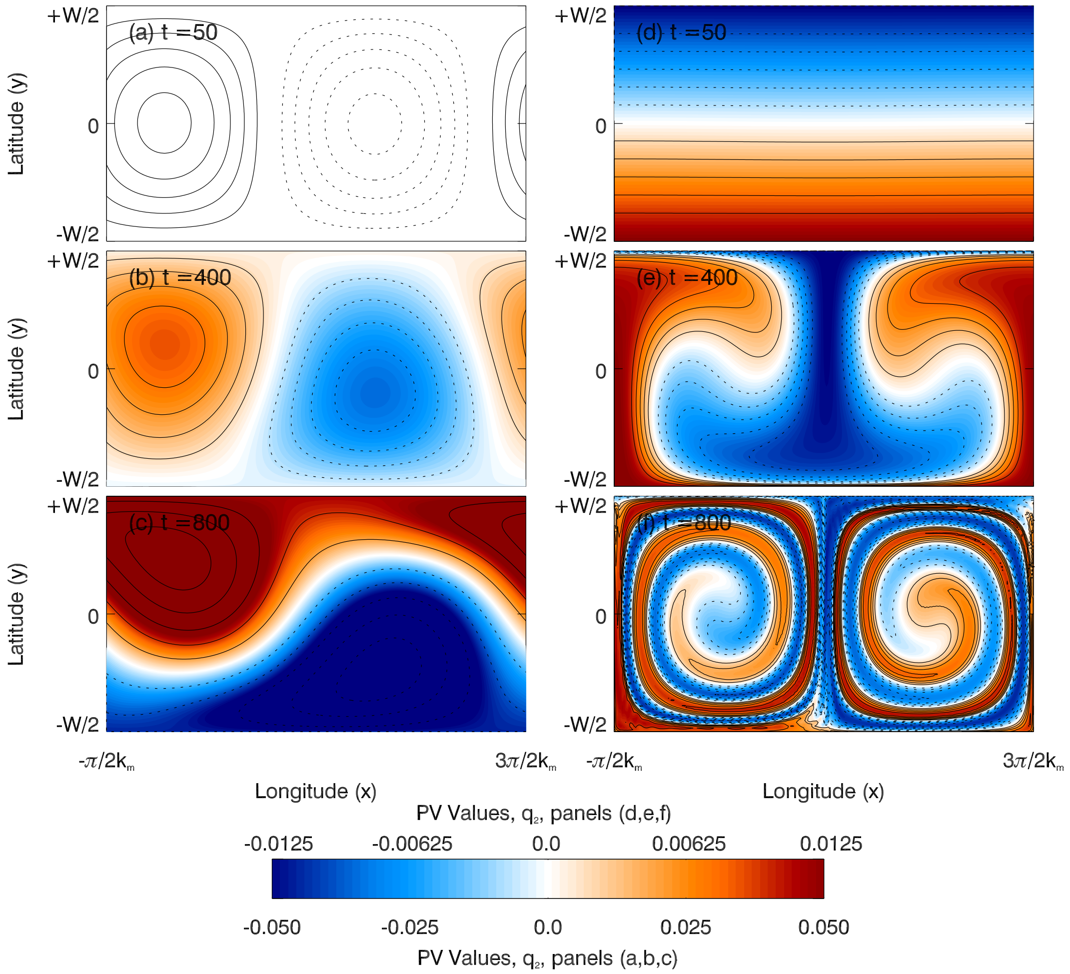

Figure 4.4.3: Panels (a-c) are snapshots of the lower layer PV field at three times ($t = 0, 400, 800$) for EXPT III $(\beta, \kappa) = (0.5, 0.04)$. Blue/dashed lines correspond to negative PV values and red/solid lines correspond to positive PV values. The contour intervals in each of the panels are (a) 1.5×10^{-3}, (b) 7.5×10^{-3}, (c) 2.0×10^{-2}. Panels (d-f) are the same snapshots in time as (a-c), but for EXPT IV $(\beta, \kappa) = (0.495, 0)$. Contour intervals for panels (d-f) are 8.0×10^{-3}.

approximately up to time $t = 400$ as seen in panel (b) and the PV distribution remains very close to the plane wave solution (4.3.5) taking the form of a pair of oppositely signed vortices in the lower layer of the channel. The vortices in the lower layer grow exponentially due to the Holopainen instability mechanism as described in chapter 3. The vortex train in the lower layer eventually reaches a finite amplitude and a significant mean flow feedback ensues. The changes to the mean flow act to advect both the upper layer Rossby wave and the lower layer vortex train, which disrupts the phase locking between the layers and inhibits growth. The Holopainen mechanism of instability is equilibrated and the flow approaches a final, steady state. The steady state in the lower layer, shown in Fig. 4.4.3 panel (c), consists of finite amplitude vortices and a weak positive PV gradient associated with the feedback of the adjusted mean flow.

The behaviour of the lower layer PV field in EXPT III is in sharp contrast to that of EXPT IV which has parameters $(\beta, \kappa) = (0.495, 0)$ and is plotted in Fig. 4.4.3 panels (d-f) at the same snapshots in time as for EXPT III. The behaviour of the lower layer PV field for such flows was discussed in section 2.6.1. Recall that in the absence of Ekman friction, potential vorticity is conserved. EXPT IV is weakly unstable to baroclinic instability and at early times (panel d) there is a uniform negative PV gradient in the lower layer. In panels (e-f) the PV gradient is removed by stirring throughout the lower layer (which is a critical layer), which eventually leads to coarse-grain homogenization. The elimination of the PV gradient occurs throughout the parameter space investigated in chapter 2, although at lower criticalities occurs due to a roll-up of vortices in the lower layer, rather than a stirring of PV as predicted by the WGP solution (a full discussion of this point was given in section 2.6.1).

EXPTS I-III lie in a subset of the (β, κ) parameter space that can be described as 'Romea-like'. This is contrasted to behaviour such as that observed in small supercriticality simulations in chapter 2 or the flow evolution in EXPT IV here, which may be referred to as 'WGP-like'. Both Romea [109] and Boville [11] suggested that the Romea-like region in the (β, κ) parameter space might be restricted to the region where disturbances are dissipatively destabilized with $\frac{1}{2} < \beta < \beta_m$. However, EXPT II shows that the Romea-like regime extends far into the region of frictionless instability $\beta < \frac{1}{2}$, provided that Ekman friction is sufficiently strong.

A useful way to distinguish between WGP- and Romea-like evolutions is by consideration of the upper layer wave amplitude. In the WGP lifecycle the wave amplitude reaches a maximum value approximately 25% greater than the equilibrated value. In contrast, in Romea-like evolutions the upper layer wave amplitude monotonically increases with time up to the maximum equilibrated value as seen in Romea's solution (4.3.37) and EXPTS I and II. The full extent of the WGP- and Romea-like regions, diagnosed in this way, across the (β, κ) parameter space will be determined in the next section.

4.4.2 Exploration of the (β, κ) parameter space

In this section the full results of numerical simulations across the (β, κ) parameter space are reported. Fig. 4.4.4 panels (a-b) contrast Romea's prediction for the equilibrated wave amplitude $\epsilon_* |A|_{eq}$ from equation (4.3.38) with the equilibrated wave amplitudes attained during numerical simulations, calculated using (4.4.41). The thick, solid curve in all three panels of Fig. 4.4.4 is the inverse criticality $\beta_m = \beta_m(\kappa)$ associated with marginal stability.

The first region of parameter space in Fig. 4.4.4 is the Romea-like regime lying close to the marginal stability curve, where the supercriticality $\epsilon_* = (\beta_m - \beta)^{1/2}$ is small. The agreement between Romea's theory and simulations within this region (where EXPTS I-III are located) is very good. As expected, the maximum wave amplitudes, shown in Fig. 4.4.4 panel (c), are close to the equilibrated amplitudes (in panel b) within this region.

The second region of parameter space observed in Fig. 4.4.4 panel (b) is a 'boundary layer' at $\kappa = 0$ within the range $\beta \in [0, 0.5)$. There is a discontinuity in $|A_1^s|_{eq}$ at the boundary layer emphasising the difference between the equilibrium wave amplitude in the limit $\kappa \to 0$ and in the frictionless problem ($\kappa = 0$), a point which will be discussed further below.

The third region of the (β, κ) parameter space in Fig. 4.4.4 is located far from both the marginal stability curve and from $\kappa = 0$ and will be referred to as the 'strongly nonlinear' region. In this region, the amplitude of the equilibrated wave amplitude in simulations $|A_1^s|_{eq}$ is overestimated by a factor of two or more, with $|A_1^s|_{max} > |A_1^s|_{eq}$ throughout. The dependence of $|A_1^s|_{eq}$ on (β, κ) throughout a large part of this region is relatively weak.

Fig. 4.4.5 examines the accuracy of Romea's theory as the supercriticality is increased from marginal stability and the system changes from the Romea-like region of parameter space to the strongly nonlinear region. Equilibrated amplitudes $|A_1^s|_{eq}$ from numerical simulations are plotted as a function of in-

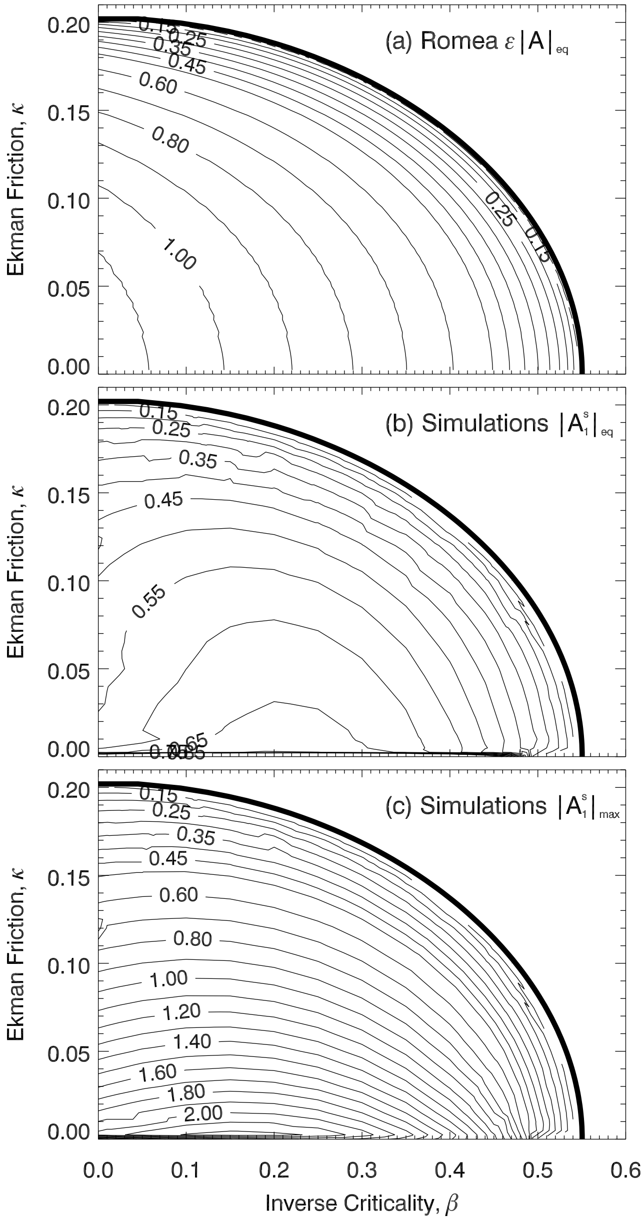

Figure 4.4.4: Panel (a) shows the equilibrated wave amplitude predicted by Romea's theory, $\epsilon_*|A|_{eq}$ from equation (4.3.38), as a function of inverse criticality β and quasi-geostrophic Ekman number κ. Panel (b) shows the equilibrated amplitude attained in numerical simulations, $|A_1^s|_{eq}$ calculated using (4.4.41). Panel (c) shows the maximum wave amplitude attained during the simulations. Panels (b) and (c) have been drawn using results from over 330 numerical experiments as described in section 4.2.2.

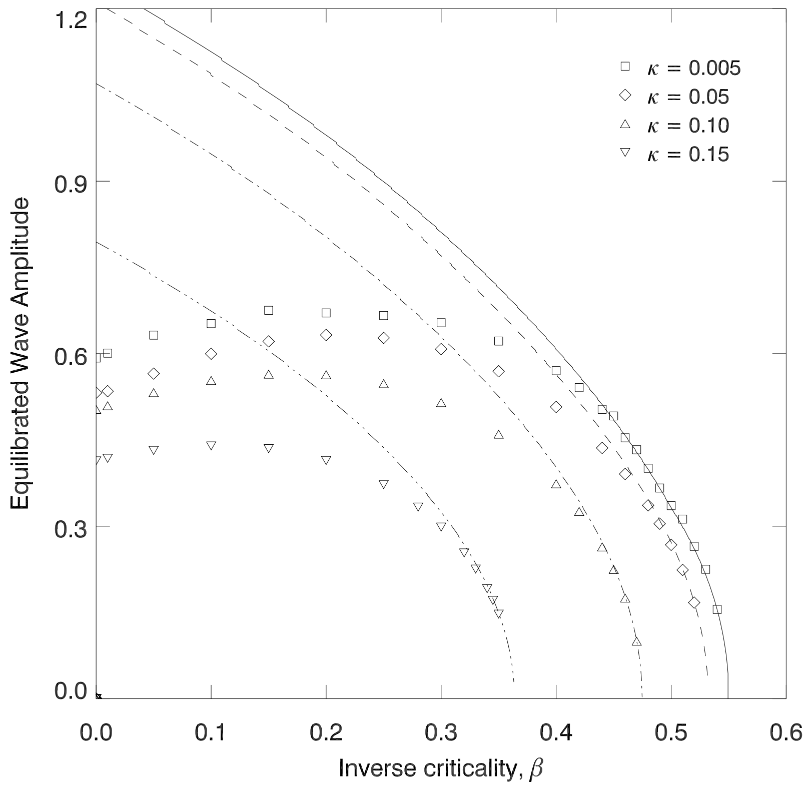

Figure 4.4.5: The equilibrated amplitude as a function of inverse criticality β for quasi-geostrophic Ekman numbers $\kappa = 0.02, 0.05, 0.1, 0.15$. The numerical simulations $|A_1^s|_{eq}$ are points (as labelled) while the curves show predictions by Romea's theory $\epsilon_*|A|_{eq}$ from equation (4.3.38). Numerical experiments in this figure correspond to those in Fig.4.4.4 where the channel has a fixed width $W = 2^{3/4}\pi$.

verse criticality β together with the corresponding predictions from Romea, for $\kappa = 0.02, 0.05, 0.1, 0.15$. The equilibrated amplitudes remain close to Romea's predictions for $\beta_m - \beta \lesssim 0.06$ (or $\epsilon_* \lesssim 0.25$) for each value of κ. As supercriticality ϵ_* increases, the equilibrated amplitudes attain a maximum value at $\beta_m - \beta \approx 0.25$ (or $\epsilon_* \approx 0.5$), before decreasing.

It is also possible to carry out numerical experiments and choose the channel dimensions such that the fastest growing mode is isotropic (i.e. set $W = \pi/k_m$ for given the parameters (β, κ). The equilibrated amplitudes of

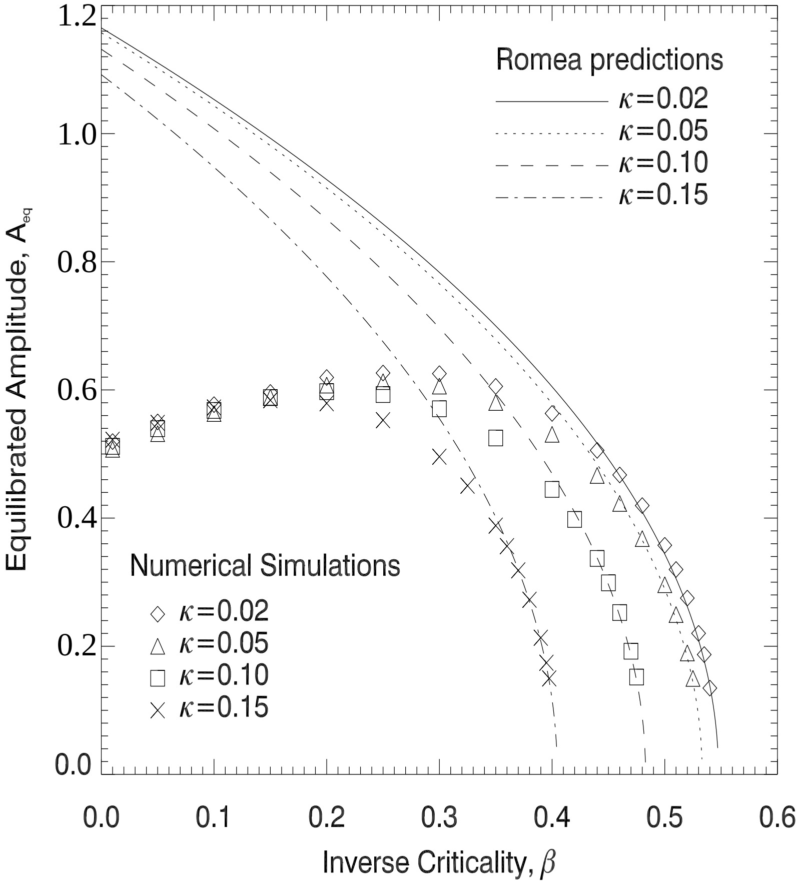

Figure 4.4.6: As for Fig. 4.4.5, except that amplitudes plotted are for the fastest growing isotropic modes. This is achieved by setting the width parameter $W = \pi/k_m$.

numerical experiments with isotropic disturbances are compared to predictions made using Romea's weakly nonlinear theory in Fig. 4.4.6. Equilibrated amplitudes for $\kappa = 0.02, 0.05, 0.10, 0.15$ are plotted as a function of the inverse criticality β. The agreement between Romea's solution and numerical experiments is very good and follows a similar pattern to Fig.4.4.5. As supercriticality increases, the equilibrated amplitude attains a maximum around $\beta_m - \beta = \epsilon_* \approx 0.5$, before decreasing as β approaches zero. A striking contrast to Fig. 4.4.5 is that for the isotropic disturbances of Fig. 4.4.6, as β

approaches zero the equilibrated amplitudes are almost independent of κ.

A fourth region of (β, κ) parameter space in Fig. 4.4.4 approximately occupies $0.4 \lesssim \beta < 0.5$ and $0 < \kappa \lesssim 0.05$ and will be referred to as the 'mixed behaviour' regime. In this region of parameter space, Romea accurately predicts the equilibrated amplitude (see Fig. 4.4.4 panels a-b), while the maximum amplitude (shown in Fig. 4.4.4 panel c) is much larger. The evolution of the disturbance is therefore not Romea-like, since Romea's solution (4.3.37) predicts that the equilibrated and maximum wave amplitudes are identical. The 'mixed behaviour' region of parameter space is the subject of the next section.

4.4.3 Evolution in the 'mixed behaviour' region of parameter space

In section 4.4.2 a region of the (β, κ) parameter space, which approximately occupies $0.4 \lesssim \beta < 0.5$ and $0 < \kappa \lesssim 0.05$, was identified and termed the 'mixed behaviour' region. Within this region, the flow is unstable to baroclinic instability (since $\beta < \frac{1}{2}$) and the Ekman friction is small ($\kappa \to 0$). Romea's theory makes an accurate prediction of the equilibrated wave amplitude $|A_1^s|_{eq}$, but the flow is not Romea-like since the maximum wave amplitude attained $|A_1^s|_{max}$ is significantly larger than the equilibrated wave amplitude. The mixed behaviour region of parameter space is a region where the WGP theory, formally valid for $\kappa = 0$, might be expected to be relevant. Most flows in the mixed behaviour region have the property that the timescale $T_R \sim \kappa^{-1}\epsilon_*^{-2}$ for the development of the Romea solution is much longer than the timescale for the frictionless baroclinic lifecycle $T_I \sim \epsilon^{-1}$. Note the different supercriticalities: ϵ defined in the WGP solution through (2.4.22) and ϵ_* given by (4.3.10) for the problem with Ekman friction. The appropriate question is then whether flows in the mixed behaviour region where $T_I \ll T_R$ should be expected to show features of the frictionless solution at short times.

EXPT V described in Table 4.1 as $(\beta, \kappa) = (0.48, 0.0025)$ is an example of a flow in the mixed behaviour region. Fig. 4.4.7 panel (a) shows the evolution of the amplitude of the upper layer fundamental wave during EXPT V (solid line). The dashed line in the same panel shows the evolution of the WGP solution given by equation (2.4.45) for $\beta = 0.48$ and the dotted line is the Romea analytical solution (4.3.37). The initial growth rate and maximum amplitude of the wave are captured reasonably well by WGP theory up to $t \approx 400$ where the wave begins to oscillate about the WGP equilibrated amplitude in the manner described in section 2.6.3. The wave amplitude then decays on the much longer timescale T_R towards an equilibrated amplitude, which is predicted well by Romea's theory (4.3.38). Similar behaviour is observed in panel (b) of Fig. 4.4.7, where the solid line is a plot of the evolution of the upper layer wave amplitude for EXPT VIa, $(\beta, \kappa) = (0.47, 0.005)$. The contrast with EXPT VIb, plotted as a dashed line, where PV damping replaces Ekman friction[3], is discussed further below.

Fig. 4.4.8 shows snapshots of the lower layer PV field during EXPT V, which may be contrasted to the Romea-like behaviour observed in EXPT III

---

[3] When PV damping replaces Ekman friction, $-rq_i$ replaces $-\kappa\nabla^2\psi_i$ on the right-hand side of the first equation in (4.2.1)

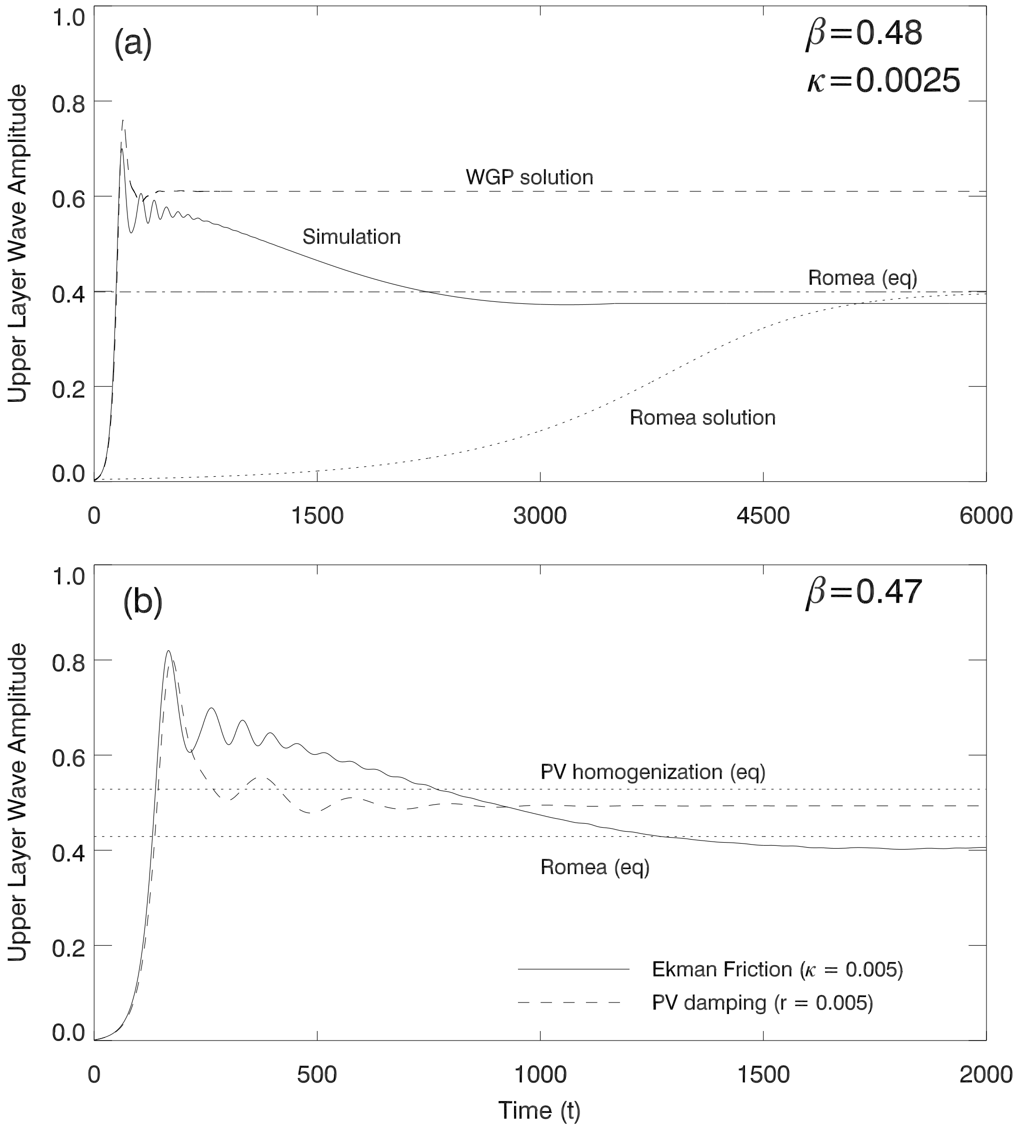

Figure 4.4.7: Panel (a) plots the evolution in time of the upper layer wave amplitude $|A_1^s|$ during EXPT V with $(\beta, \kappa) = (0.48, 0.0025)$ as a solid curve. The dashed curve shows the corresponding WGP solution for $\kappa = 0$ in equation (2.4.45) and the dotted curve shows the Romea solution of equation (4.3.37). Also plotted (dotted-dashed line) is the asymptotic value for Romea's solution given by (4.3.38). Panel (b) is as for panel (a), but for EXPTS VIa, $(\beta, \kappa) = (0.47, 0.005)$ - solid curve, and VIb, $(\beta, r) = (0.47, 0.005)$ - PV damping, dashed curve. The dotted curves show the equilibria predicted by Romea for EXPT VIa and WGP for EXPT VIb.

(Fig. 4.4.3 panels a-c) and the WGP-like behaviour seen in EXPT IV (Fig. 4.4.3 panels d-f). The initial lower layer PV distribution in Fig.4.4.8 panel (a) has a negative latitudinal PV gradient, required for baroclinic instability. As the disturbance grows, in Fig. 4.4.8 panels (b-c), the evolution of the PV field is very close to the critical layer behaviour observed in the corresponding frictionless flow and predicted by the WGP solution. At these early times, the flow in EXPT V is essentially undergoing a frictionless baroclinic lifecycle as Ekman friction is not strong enough to act on the baroclinic lifecycle timescale ($T_I \ll T_R$). There are slight asymmetries apparent between the two circulation cells as negative PV becomes concentrated in the cyclonic (anti-clockwise) cell and positive PV in the anti-cyclonic cell. This is in precisely the opposite sense as the finite amplitude effect reported for frictionless flows in chapter 2, where positive PV becomes concentrated in the cyclonic cell. By time $t = 225 - 275$, panels (d-e) of Fig. 4.4.8, coherent Romea-like vortices have formed in the lower layer due to the cumulative effect of Ekman friction on longer timescales. These vortices continue to grow via the Holopainen mechanism, described by Fig. 4.4.3 and discussed in section 4.4.1. By time $t = 2000$, panel (f) of Fig. 4.4.8 shows that a significant positive PV gradient has emerged due to the induced mean flow, which acts to stabilize the flow.

EXPT V undergoes a near-frictionless lifecycle on the timescale T_I where the initial negative PV gradient in the lower layer is homogenized by stirring. The flow undergoes an adjustment on a much longer timescale T_R where forced vortices due to the Ekman friction are observed in the lower layer. There is subsequently the generation of a positive PV gradient in the lower layer

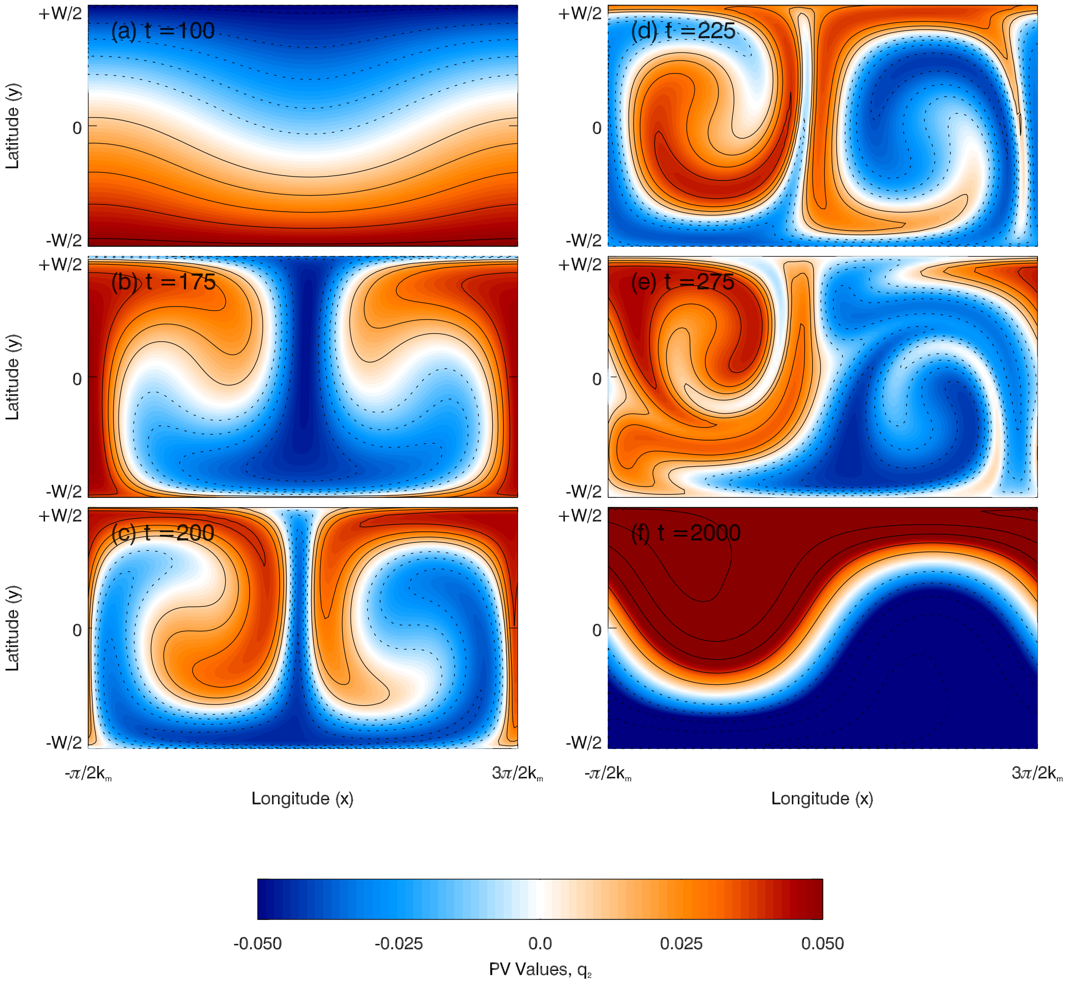

Figure 4.4.8: Snapshots of the lower layer PV field for EXPT V $(\beta, \kappa) = (0.48, 0.0025)$ at times $t = 100, 175, 200, 225, 275$ and 2000. Blue (and dashed contours) corresponds to negative PV values and red (and solid contours) corresponds to positive PV values. Contour intervals in panels (a-e) are 0.01 and in panel (f) are 0.02.

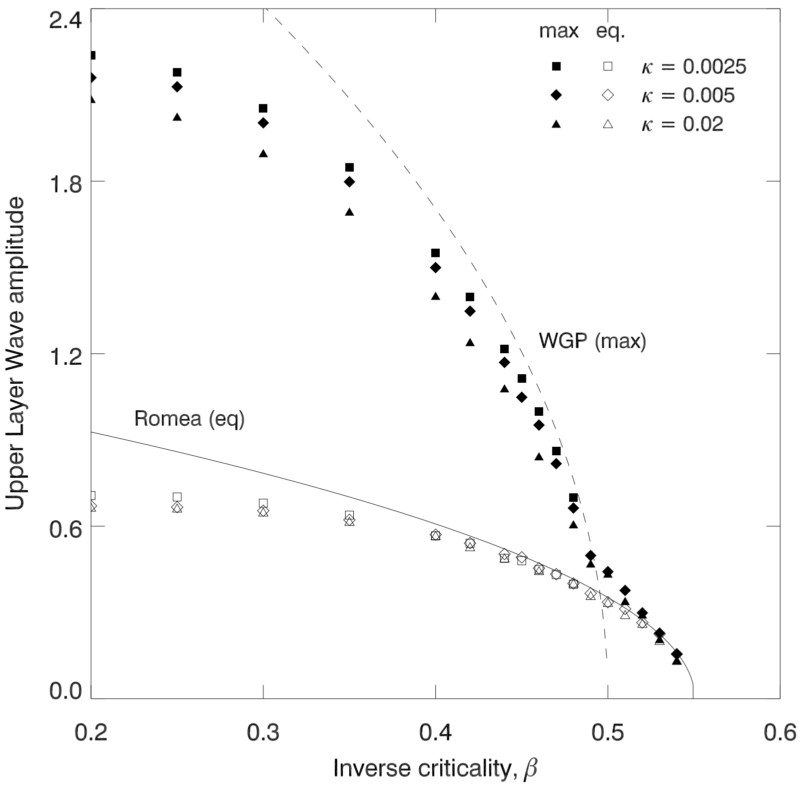

Figure 4.4.9: The equilibrated amplitudes $|A_1^s|_{eq}$ (unfilled points) and maximum amplitudes $|A_1^s|_{max}$ (solid points) as a function of inverse criticality β in numerical simulations with $\kappa = 0.0025, 0.005, 0.02$, as labelled. The dashed curve shows the WGP prediction (2.4.48) for the maximum amplitude and the solid curve, the Romea predictions for the equilibrated amplitude (4.3.40) in the $\kappa \to 0$ limit.

associated with the nonlinear correction to the mean flow and an equilibrated state determined by the Ekman friction is eventually reached.

The flow evolution seen for EXPT V is representative of the wider 'mixed behaviour' region of parameter space. Fig. 4.4.9 plots the maximum and equilibrated wave amplitudes, $|A_1^s|_{max}$ (unfilled points) and $|A_1^s|_{eq}$ (solid points) respectively, for simulations with QG Ekman numbers $\kappa = 0.0025, 0.005, 0.02$ and inverse criticalities $\beta \in [0.2, 0.55]$. The WGP prediction (2.4.48) for the maximum wave amplitude (solid curve) and the Romea prediction (4.3.40) for the equilibrated amplitude (dashed curve) are both plotted. For $\beta > \frac{1}{2}$ the maximum and equilibrated amplitude are approximately equal (as for EXPT I) and are reasonably well predicted by Romea's theory, especially as $\kappa \to 0$. At $\beta = \frac{1}{2}$, there is a pivot point: the equilibrated amplitude continues to be well-predicted by Romea's theory whereas the maximum amplitude is now much better predicted by the WGP theory.

4.5 Conclusions

This chapter has examined the relevance of the Romea and WGP weakly nonlinear theories, each formally valid under different assumptions. Three questions were posed in the introduction to this chapter, which can now be answered:

(i) The WGP theory captures the early time evolution of baroclinic lifecycles provided that they (a) satisfy the frictionless instability criterion $\beta < \frac{1}{2}$ with a supercriticality $\epsilon = (\frac{1}{2} - \beta)^{1/2}$ that is not too large and (b) Ekman friction is relatively weak, $\kappa \lesssim 0.05$. For these lifecycles, the wave equilibrates on a timescale $T_I \sim \epsilon^{-1}$ via the formation and roll-up of a nonlinear critical layer in the lower layer. On this timescale, the maximum wave amplitude is attained and is well-predicted by WGP. Beyond this timescale, as seen in EXPT V of section 4.4.3, the upper layer wave amplitude briefly oscillates about the WGP prediction for the equilibrated amplitude reported in chapter 2. On the longer timescale $T_R \sim \kappa^{-1}\epsilon_*^{-2}$ (where ϵ_* is the supercriticality for the flow in the presence of Ekman friction, given by equation 4.3.10), the wave amplitude decays from the WGP predicted value. As this occurs, vortices magnify in the lower layer and destroy the early time critical layer evolution. The late-time behaviour is then similar to the equilibration of dissipatively destabilized waves, such as EXPT III.

(ii) Romea's theory accurately predicts the nonlinear evolution of lifecycles in which waves are dissipatively destabilized by the Holopainen mechanism, but are otherwise stable to classic baroclinic instability, $\frac{1}{2} < \beta < \beta_m$ and $\kappa \lesssim 0.05$. EXPT I in section 4.4.1 is an example of such behaviour. For stronger Ekman friction, $0.05 \lesssim \kappa \lesssim 0.2$ where the flow is baroclinically unstable $\beta_m < \frac{1}{2}$, as in EXPT II, Romea's theory is found to be accurate close to marginal stability $\epsilon_* = (\beta_m - \beta)^{1/2} \ll 1$. For flows subject to baroclinic instability at small Ekman numbers ($\beta < \frac{1}{2}$, $\kappa \ll 1$), Romea's theory had not been thought relevant (see e.g. Boville [11]), nevertheless numerical experiments show that Romea's predictions for the equilibrated amplitude and zonal flow, established on the timescale T_R, are accurate. The timescale T_R is longer than the WGP timescale T_I throughout much of the parameter space and so there is no contradiction between the short time behaviour predicted by WGP and the late-time, equilibrium behaviour is well-captured by Romea.

(iii) The frictionless baroclinic adjustment hypothesis (e.g. section 12.6.2 of Vallis [133]) is that, in general, flows adjust through PV homogenization to a state that is stable under the Charney-Stern-Pedlosky criterion whereby PV gradients have a single sign over the whole domain). The final, equilibrated state of the WGP solution is an example of just such an 'adjusted' state as pointed out by Shepherd [118], since the lower layer PV is completely homogenized at late times in the WGP solution. However, the presence of Ekman friction causes the equilibrium states to completely differ from the adjusted states. An example of this is in panel (f) of Fig. 4.4.8 where the equilibrated PV field of EXPT V has a positive PV gradient in the lower layer, an indication of an 'over-adjustment'. The nonlinear baroclinic equilibration in the presence of Ekman friction is therefore a clear example of a flow where baroclinic adjustment fails.

These conclusions raise questions about flows for which the 'baroclinic adjustment' argument is valid in the prediction of equilibrated states. The systems where the adjustment hypothesis is valid are those where weakly forced-dissipative experiments equilibrate at a state near the equilibrated state of the unforced problem. In the current chapter, the two-layer β-channel *in the special case of PV damping* (e.g. Pedlosky [93], Warn & Gauthier [134]) is just such a flow, e.g. the equilibrated state of EXPT VIb in Fig. 4.4.7 converges to a state which is close to that predicted by the forced-dissipative formulation of the WGP theory. The forced dissipative prediction of WGP is $|A|_{eq} = \epsilon W/\sqrt{3}$, valid in the limit of vanishing PV damping $r \to 0$. The equilibrium state of EXPT VIb is an 'adjusted' state since the lower layer PV is homogenized, like the unforced baroclinic lifecycles discussed in chapter 2.

It is the presence of Ekman friction in baroclinic lifecycles that prevents the success of adjustment hypotheses since ultimately the flow must be rendered stable to dissipative destabilization. This chapter demonstrates a concrete example of equilibration that occurs via a mechanism distinct from a baroclinic adjustment hypothesis. Romea's theory has been shown to be accurate in predicting the equilibrated state of flows across a wide region of the parameter space.

4.A Derivation of Romea's evolution equations

The working in this appendix follows Romea [109], with some improvements, and is carried out as an expansion about minimum critical shear of the marginal stability curve (3.2.7), which occurs at inverse criticality β_m and total wavenumber $a_m = \sqrt{k_m^2 + \pi^2/W^2}$.

The equations of motion (4.3.11-4.3.12) at $O(\epsilon_*^2)$ are given by (4.3.22), which is

$$\mathcal{L}_R \mathbf{\Psi^{(1)}} = \mathbf{f^{(1)}}, \tag{4.A.42}$$

or written in full

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) q_1^{(1)} + \left(\beta_m + \frac{1}{2}\right) \frac{\partial \psi_1^{(1)}}{\partial x} + \kappa \nabla^2 \psi_1^{(1)} = -J\left(\psi_1^{(0)}, q_1^{(0)}\right), \tag{4.A.43}$$

$$\frac{\partial q_2^{(1)}}{\partial t} + \left(\beta_m - \frac{1}{2}\right) \frac{\partial \psi_2^{(1)}}{\partial x} + \kappa \nabla^2 \psi_2^{(1)} = -J\left(\psi_2^{(0)}, q_2^{(0)}\right). \tag{4.A.44}$$

Unlike the WGP expansion, the right-hand side of (4.A.43) and (4.A.44) are non-zero. This is due to the presence of Ekman friction, which means the ratio of the wave amplitude γ_m given by (4.3.21) is not strictly real. The right-hand side of (4.A.43-4.A.44) is explicitly calculated using the linear solution (4.3.18), which gives

$$\mathbf{f}^{(1)} = \begin{pmatrix} -J\left(\psi_1^{(0)}, q_1^{(0)}\right) \\ -J\left(\psi_2^{(0)}, q_2^{(0)}\right) \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{k\pi\gamma_i |A(T)|^2}{4W} \sin\left(\frac{2\pi y}{W}\right) \qquad (4.A.45)$$

where γ_i is the imaginary part of γ_m. The forcing $\mathbf{f}^{(1)}$ satisfies the solvability condition to be discussed below and all that remains is to find the particular integral and complementary functions for $\mathbf{\Psi^{(1)}} = (\psi_1^{(1)}, \psi_2^{(1)})^T$, which via use of the Phillips boundary condition (1.3.59) gives

$$\mathbf{\Psi^{(1)}} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \frac{k_m W \gamma_i |A|^2}{16\pi\kappa} \left(\sin\left(\frac{2\pi y}{W}\right) + \frac{2\pi}{W} y\right). \qquad (4.A.46)$$

The streamfunction is defined up to a constant and so the choice $\mathbf{\Psi^{(1)}} = \mathbf{0}$ at $y = 0$ is made. Alternatively, substitution of γ_i into (4.A.46) gives

$$\mathbf{\Psi^{(1)}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{a_m^2 \left(a_m^2 + \frac{1}{2}\right) W}{4\pi \left(2\beta_m + a_m^2 + \frac{1}{2}\right)} |A|^2 \left(\sin\left(\frac{2\pi y}{W}\right)) + \frac{2\pi}{W} y\right). \qquad (4.A.47)$$

This is equation (4.3.24) in the main text.

Proceeding to the next order in ϵ_*, the equations of motion (4.3.11-4.3.12) are

$$\mathcal{L}_R \mathbf{\Psi^{(2)}} = \mathbf{f^{(2)}}, \qquad (4.A.48)$$

or written in full

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)q_1^{(2)} &+ \left(\beta_m+\frac{1}{2}\right)\frac{\partial \psi_1^{(2)}}{\partial x}+\kappa\nabla^2\psi_1^{(2)} \\ &= -\frac{\partial q_1^{(0)}}{\partial T}+\frac{\partial \psi_1^{(0)}}{\partial x}-J\left(\psi_1^{(0)},q_1^{(1)}\right)-J\left(\psi_1^{(1)},q_1^{(0)}\right),\end{aligned} \tag{4.A.49}$$

$$\begin{aligned}\frac{\partial q_2^{(2)}}{\partial t} &+ \left(\beta_m-\frac{1}{2}\right)\frac{\partial \psi_2^{(2)}}{\partial x}+\kappa\nabla^2\psi_2^{(2)} \\ &= -\frac{\partial q_2^{(0)}}{\partial T}+\frac{\partial \psi_2^{(0)}}{\partial x}-J\left(\psi_2^{(0)},q_2^{(1)}\right)-J\left(\psi_2^{(1)},q_2^{(0)}\right).\end{aligned} \tag{4.A.50}$$

Each of the terms on the right-hand side of equations (4.A.49-4.A.50) may be calculated explicitly as $\mathbf{f}^{(2)}=(f_1^{(2)},f_2^{(2)})^T$ where

$$\begin{aligned}f_1^{(2)} &= -\left(a^2+\frac{1}{2}-\frac{\gamma_m}{2}\right)\frac{dA}{dT}+ik_mA \\ &-\left(ik_mAe^{ik_m(x-c_mt)}\cos\left(\frac{\pi y}{W}\right)\right)B|A|^2\left(-\frac{2\pi}{W}\left(\frac{4\pi^2}{W^2}+1\right)\cos\left(\frac{2\pi y}{W}\right)-\frac{2\pi}{W}\right) \\ &+\left(\left(-\left(a_m^2+\frac{1}{2}\right)+\frac{\gamma_m}{2}\right)ik_me^{ik_m(x-c_mt)}\cos\left(\frac{\pi y}{W}\right)\right) \\ &\times B|A|^2\left(\frac{2\pi}{W}\cos\left(\frac{2\pi y}{W}\right)+\frac{2\pi}{W}\right),\end{aligned} \tag{4.A.51}$$

and

$$\begin{aligned}f_2^{(2)} &= -\left(\left(a^2+\frac{1}{2}\right)\gamma_m-\frac{1}{2}\right)\frac{dA}{dT}+ik_m\gamma_mA \\ &-\left(ik_m\gamma_mAe^{ik_m(x-c_mt)}\cos\left(\frac{\pi y}{W}\right)\right)B|A|^2\left(\frac{2\pi}{W}\left(\frac{4\pi^2}{W^2}+1\right)\cos\left(\frac{2\pi y}{W}\right)+\frac{2\pi}{W}\right) \\ &+\left(\left(-\left(a_m^2+\frac{1}{2}\right)\gamma_m+\frac{1}{2}\right)ik_me^{ik_m(x-c_mt)}\cos\left(\frac{\pi y}{W}\right)\right) \\ &\times B|A|^2\left(-\frac{2\pi}{W}\cos\left(\frac{2\pi y}{W}\right)-\frac{2\pi}{W}\right),\end{aligned} \tag{4.A.52}$$

with

$$B=\frac{a_m^2(a_m^2+\frac{1}{2})W}{4\pi(2\beta_m+a_m^2+\frac{1}{2})}=-\frac{\gamma_ik_mW}{16\pi\kappa}.$$

As discussed in the main text, the solvability condition (4.3.31) is given by

$$\langle \mathbf{\Phi}_0, \mathbf{f}^{(2)} \rangle = \mathbf{0}, \tag{4.A.53}$$

where $\mathbf{\Phi}_0$ lies in the kernel of $\mathcal{L}^\dagger$ and so satisfies $\mathcal{L}^\dagger \mathbf{\Phi}_0 = \mathbf{0}$. The solution that results in a non-vanishing inner product is

$$\mathbf{\Phi}_0 = \begin{pmatrix} -c_m \\ \\ (1-c_m)\gamma_m^* \end{pmatrix} e^{ik_m(x-c_m t)} \cos\left(\frac{\pi y}{W}\right), \tag{4.A.54}$$

where * refers to the complex conjugate. The solvability condition is therefore (4.A.51), (4.A.52) and (4.A.54) substituted into (4.A.53), which upon integration gives

$$\begin{aligned} 0 \;=\; & \left(c_m \left(-a_m^2 - \frac{1}{2} + \frac{\gamma_m}{2} \right) - (1-c_m)\,\gamma_m \left(-a_m^2 \gamma_m - \frac{\gamma_m}{2} + \frac{1}{2} \right) \right) \frac{dA}{dT} \\ & + i k_m \left(-c_m + (1-c_m)\,\gamma_m^2 \right) A \\ & + \frac{2\pi i k_m B}{W} A|A|^2 \left(-c_m \left(\frac{3}{4} \left(-a_m^2 - \frac{1}{2} + \gamma_m \right) + \frac{\pi^2}{W^2} + \frac{3}{4} \right) \right. \\ & \left. + (1-c_m)\,\gamma_m \left(-\frac{3}{4} \left(-a_m^2 \gamma_m - \frac{\gamma_m}{2} + \frac{1}{2} \right) - \frac{\gamma_m \pi^2}{W^2} - \frac{3\gamma_m}{4} \right) \right) \end{aligned} \tag{4.A.55}$$

Alternatively, writing

$$p_1 = (1-c_m)\,\gamma_m = a_m^2 - \frac{1}{2} - \frac{2a_m^2 \kappa}{k_m} i, \tag{4.A.56}$$

and

$$p_2 = c_m = \frac{1}{2} - \frac{\beta_m}{\left(a_m^2 + \frac{1}{2}\right)}, \tag{4.A.57}$$

allows equation (4.A.55) to be re-written as

$$\frac{dA}{dT} = (\sigma_r + i\sigma_i)\, k_m A - (\mu_r + i\mu_i)\, k_m |A|^2 A, \tag{4.A.58}$$

where the coefficients σ and μ are given by

$$\sigma = \sigma_r + i\sigma_i = \frac{ik_m\left(p_2 - \gamma_m p_1\right)}{\left(a_m^2\gamma_m + \frac{\gamma_m}{2} - \frac{1}{2}\right)p_1 - \left(a_m^2 + \frac{1}{2} - \frac{\gamma_m}{2}\right)p_2}, \tag{4.A.59}$$

and

$$\mu = \mu_r + i\mu_i = \left(\frac{3ik\gamma_i}{16\kappa}\right)\frac{\left(\left(a_m^2 - \frac{1}{2}\right)\gamma_m - \frac{1}{2} - \frac{4\pi^2}{3W^2}\gamma_m\right)p_1 - \left(\frac{1}{2} - a_m^2 + \frac{\gamma_m}{2} + \frac{4\pi^2}{3W^2}\right)p_2}{\left(a_m^2\gamma_m + \frac{\gamma_m}{2} - \frac{1}{2}\right)p_1 - \left(a_m^2 + \frac{1}{2} - \frac{\gamma_m}{2}\right)p_2}. \tag{4.A.60}$$

Equations (4.A.56-4.A.60) are the evolution equations for the wave amplitude as given in Romea, albeit, derived via a slightly modernised approach. It is possible to simplify the coefficients σ and μ by substitution of (4.A.56-4.A.57) into (4.A.59-4.A.60) and use of the expression for the marginal stability curve β_m^2 (equation 3.2.7). After tedious but straight-forward algebra this results in

$$\begin{aligned}
\sigma &= \frac{\frac{4ia_m^2\kappa/k_m}{\frac{1}{2}+\beta_m/\left(a_m^2+\frac{1}{2}\right)}\left(\frac{2ia_m^2\kappa}{k_m} + i\left(a_m^2 - \frac{1}{2}\right)\right)}{\frac{-1}{\frac{1}{2}+\beta_m/\left(a_m^2+\frac{1}{2}\right)}\left(\frac{2ia_m^2\kappa}{k_m} + i\left(a_m^2 - \frac{1}{2}\right)\right)\left(4\frac{a_m^2\kappa}{k_m}\left(a_m^2 + \frac{1}{2}\right) - i\beta_m/\left(a_m^2 + \frac{1}{2}\right)\right)} \\
&= \left(\frac{\kappa}{k_m}\right)\frac{4a_m^2\left(a_m^2 + \frac{1}{2}\right)\left(\beta_m - 4\left(a_m^2 + \frac{1}{2}\right)^2\left(\kappa/k_m\right)i\right)}{\beta_m^2 + 16a_m^4\left(a_m^2 + \frac{1}{2}\right)^4\left(\kappa/k_m\right)^2},
\end{aligned} \tag{4.A.61}$$

and

$$\begin{aligned}
\mu &= \frac{\frac{3ia_m^2}{8\left(\frac{1}{2}+\beta_m/\left(a_m^2+\frac{1}{2}\right)\right)}\left(\frac{2a_m^2\kappa}{k_m} + i\left(a_m^2 - \frac{1}{2}\right)\right)\left(a_m^2 - a_m^4 + \frac{4\pi^2}{W^2}\left(a_m^2 - \frac{1}{2}\right)\right)}{\frac{-1}{\frac{1}{2}+\beta_m/\left(a_m^2+\frac{1}{2}\right)}\left(\frac{2a_m^2\kappa}{k_m} + i\left(a_m^2 - \frac{1}{2}\right)\right)\left(\frac{4a_m^2\kappa}{k_m}\left(a_m^2 + \frac{1}{2}\right) - i\beta_m/\left(a_m^2 + \frac{1}{2}\right)\right)} \\
&= \frac{a_m^2\left(a_m^2 + \frac{1}{2}\right)^2\left(3a_m^2\left(1 - a_m^2\right) + 4\left(a_m^2 - \frac{1}{2}\right)\pi^2/W^2\right)}{2\left(2\beta_m + a_m^2 + \frac{1}{2}\right)\left(\beta_m^2 + 16a_m^4\left(a_m^2 + \frac{1}{2}\right)^4\left(\kappa/k_m\right)^2\right)} \\
&\qquad \times\left(4a_m^2\left(a_m^2 + \frac{1}{2}\right)\frac{\kappa}{k_m} + i\beta_m\right).
\end{aligned} \tag{4.A.62}$$

Equations (4.A.58), (4.A.61) and (4.A.62) are stated in the main text as (4.3.33), (4.3.34) and (4.3.35) respectively.

Chapter 5

Isolated jets and Equilibration by Potential Vorticity Homogenization

5.1 Introduction

Chapters 2, 3 and 4 were concerned with the development of baroclinic instability of a single wave mode in the two-layer model with a uniform vertical shear. This simple problem was shown to be amenable to the development of weakly nonlinear theories, which were studied in detail in chapters 2 and 4. The current chapter replaces the uniform flow profile of the upper layer with a jet to provide a simple model of a jet stream in the terrestrial atmosphere, which is of interest for a number of reasons discussed in section 1.3.2.

The formation of jets is a common feature of geostrophic turbulence in the presence of a sufficiently strong gradient of potential vorticity [101, 102]. Jet streams are ubiquitous in the oceans and atmosphere of the Earth [79, 103] and are also observed in the atmospheres of the gas giants [139, 115].

Full baroclinic lifecycles of realistic tropospheric jets have been the subject of numerical studies, for example of the primitive equations by Simmons and Hoskins [119, 121] and Thorncroft et al.[132]. These studies reveal an upgradient momentum flux at the jet's core leading to a strengthening and sharpening of the jet. During the nonlinear stages as the jet reaches an equilibrated state, the eddy momentum flux is dominant over the eddy heat flux and so the resulting behaviour was termed a 'baroclinic-growth, barotropic decay' lifecycle.

This study will focus on the equilibration of a single jet in the two-layer model. Previous studies of this kind have focused on the properties of the most unstable linear mode [34, 85] , which have established that if critical levels of this mode appear to the flanks of the jet in the upper layer, then the jet's structure is asymmetric in the vertical. PV mixing is observed to the flanks of the jet in the upper layer and across the centre of the jet in the lower layer. The increase of the jet strength and associated upgradient momentum flux is intrinsically linked to PV mixing in the elegant description of jet formation given by Dritschel & McIntyre [23] and 'Rossby-wave elasticity' [80]. In regions where the latitudinal PV gradient is strong the flow is wave-like, whereas in turbulent regions, the PV gradient (or Rossby elasticity) is weakened creating a positive feedback whereby the region is susceptible to further weakening. The result is that the formation of jets is a consequence of horizontal mixing of PV, which results in regions of strongly mixed PV adjacent to regions where the PV gradient is sharp. By the PV invertibility principle, this entirely determines the balanced flow. Strong eastward jets are co-located with strong PV gradients and broader westward jets are in regions where PV is almost homogeneous. The equilibrated PV pattern is reached by downgradient mixing of PV, associated with the upgradient momentum flux responsible for jet sharpening and maintenance.

This chapter builds on work by Esler (2008) [29], E08 hereafter, which developed a theory to predict the final, near-steady, equilibrated state of a jet following baroclinic instability and nonlinear equilibration. E08's theory, which relies on no explicit calculation of the linear dynamics, is named *Equilibration via Potential Vorticity Homogenization* (EPVH hereafter) and uses the global constraints of energy and momentum conservation combined with the assumption that available potential energy is minimized over the course of the lifecycle. The latter assumption enforces a requirement that the instability proceeds to completion and that the final state is stable to further disturbances. A final assumption is that PV is completely homogenized within de-lineated regions, a consequence of 'Rossby wave elasticity', and is motivated by numerical simulations of a variety of geophysical flows [101, 102, 23].

Well de-lineated PV mixing regions are observed in simulations to the flanks of the jet in the upper layer and across the centre of the jet in the lower layer. EPVH makes predictions for the widths of these de-lineated mixing regions. The theory is successful in predicting the structure of the equilibrated jet for a range of symmetric initial jet profiles [29]. Significantly, it was shown to be more successful than alternative maximum-entropy theories [82, 105, 106] subject to the same global constraints.

The EPVH assumption that available potential energy is minimized over the course of the lifecycle is necessary to close the set of nonlinear equations derived from dynamical constraints. A natural question is, could other variational principles be equally successful? It is possible to close the EPVH nonlinear equations with alternative assumptions, two of which are the subject of the first part of the chapter. The first alternative is that the area of the the mixing regions over both layers of the channel is maximized. This assumption may be argued to be a consequence of Rossby-wave elasticity [80] and may also be related to theories of maximum mixing entropy from statistical mechanics [82, 105]. The second alternative assumption is that the zonal mean enstrophy of the system is minimized over the lifecycle and and is motivated by studies of two-dimensional turbulence in which some invariant quantities are dissipated on much shorter time scales than others [16, 69]. In this case, there is a cascade of enstrophy to small scales where it is dissipated on time scales much shorter than the energy.

To assess these alternative variational principles, the second part of the chapter introduces a linear, barotropic shear to the otherwise symmetric background flow. This is an important question since baroclinic flows are sensitive to horizontal structure, which predictive theories such as baroclinic adjustment are unable to capture [85]. The presence of a barotropic component to the flow in baroclinic lifecycles has been explored with the focus on the analysis of the linear stability problem [57, 83]. Observations of such flows show that barotropic shear reduces the growth rate of the fastest growing normal mode. One idea that offers an explanation of the reduction in growth rate is the 'barotropic governor' mechanism first put forward by James and Gray [59]. If no barotropic shear is present, the horizontal structure is almost optimally configured to release available energy. When the shear is present, James [57] argued, the meridional structure of the fastest growing normal mode is modified and subsequently the rate at which available potential energy can be released is reduced.

In section 5.2 the physical problem, equations of motion and numerical model are reviewed. Section 5.3 discusses the properties of typical baroclinic lifecycles in the two-layer model, with and without a linear, barotropic shear. The details of EPVH are then discussed in section 5.4 where the theory is extended to include asymmetric flows and is also recast to form separate sets of equations for the two alternative hypotheses suggested above. The linear solution of the problem is also discussed. In section 5.5, the predictions of EPVH are directly compared to numerical simulations for symmetric initial flow profiles. In section 5.6 the predictions of EPVH in the presence of a background, barotropic shear are presented and compared to numerical simulations. Finally, conclusions are drawn in section 5.7.

5.2 Introduction to the physical problem

5.2.1 The two-layer model with a linear, background, meridional shear

The equations of motion for an upper layer jet in the two-layer model with zero zonal flow on the channel walls are given by (1.3.72-1.3.77). The latter part of this chapter will investigate the effect of a barotropic linear meridional shear in each layer super-imposed on top of the 'sech-squared' profile described by (1.3.69). The non-dimensional rate of shear is given by Γ and replacing (1.3.69), the background velocities are therefore given by

$$\begin{aligned} U_1(y) &= -\frac{\partial \Psi_1}{\partial y} = \text{sech}^2\left(\frac{y}{\sigma}\right) + \Gamma y, \\ U_2(y) &= -\frac{\partial \Psi_2}{\partial y} = \Gamma y. \end{aligned} \tag{5.2.1}$$

The background PV fields $Q_1(y)$ and $Q_2(y)$ given by (1.3.70) are altered by the addition of a constant $-\Gamma$ term in the presence of a meridional shear:

$$\begin{aligned} Q_1(y) &= \beta y + \left(\frac{2}{\sigma}\text{sech}^2\left(\frac{y}{\sigma}\right) + \frac{\sigma}{2}\right)\tanh\left(\frac{y}{\sigma}\right) - \Gamma, \\ Q_2(y) &= \beta y - \frac{\sigma}{2}\tanh\left(\frac{y}{\sigma}\right) - \Gamma. \end{aligned} \tag{5.2.2}$$

Importantly, the background PV gradients remain unchanged by the addition of the barotropic shear and the condition for baroclinic instability remains the same as an unsheared flow ($\beta < \frac{1}{2}$).

The boundary conditions on $y = \pm W/2$ are altered by the background shear such that (1.3.76-1.3.77) are replaced by

$$\frac{\partial \psi_i}{\partial x} = 0 \quad \text{on} \quad y = \pm\frac{W}{2}, \tag{5.2.3}$$

Resolution	Fourier modes (x)	Grid points (y)	dt	ν_4
Low (LR)	256	225	5.0×10^{-3}	2.5×10^{-3}
Medium (MR)	512	450	2.5×10^{-3}	7.5×10^{-4}
High (HR)	1024	900	1.0×10^{-3}	2.0×10^{-4}

Table 5.1: Parameter values used in numerical simulations.

and

$$\frac{\partial \overline{\psi}_i}{\partial y} = \mp \frac{\Gamma W}{2} \quad \text{on} \quad y = \pm \frac{W}{2}, \tag{5.2.4}$$

for $i = 1, 2$.

5.2.2 Outline of Numerical Simulations

The numerical model described in section 1.4 is employed to obtain solutions of the equations of motion (1.3.72-1.3.73) with hyperdiffusion proportional to $\nabla^4\psi_i$ as in E08. The flow is governed by the PV equations

$$\begin{aligned} \frac{Dq_i}{Dt} &= \nu_4 \nabla^4 \psi_i, \\ q_i &= \nabla^2 \psi_i + (-1)^i \left(\frac{\psi_1 - \psi_2}{2} \right), \qquad \text{for} \qquad i = 1, 2, \end{aligned} \tag{5.2.5}$$

together with boundary conditions (5.2.3) and (5.2.4). The details of the three resolutions used to integrate (5.2.5) are shown in Table 5.1. In each case, the grid resolution in the x- and y-directions, the time step δt and the artificial hyperdiffusion ν_4 are chosen so that the numerical solutions converge with increasing resolution. The convergence is directly demonstrated for this problem in sections 5.3.1 and 5.5 using the three resolutions in Table 5.1.

In this chapter, the channel length and width are set to be $L = 20\pi$ and

$W = 7\pi$ Rossby radii respectively[1], which minimizes the effect of the sidewalls on the simulations and are also representative of the extra-tropical troposphere.

All numerical simulations are initialized using a PV perturbation to the upper layer of the form

$$\tilde{q}_1(x,y) = \epsilon \left(x - \frac{1}{2}L\right) \exp\left(\frac{-((x-L/2)^2 + y^2)}{R^2}\right), \tag{5.2.6}$$

with $\epsilon = 0.04$ and $R = 2$. This disturbance is local allowing the fastest growing modes to emerge naturally.

5.2.3 Physical Constraints

The flow in the two-layer Phillips model is subject to the dynamical constraints of section 1.3.3, which must be realized if a theory to predict the equilibrated state of a given flow is to be successful.

Kelvin's impulse M is conserved between the two layers as given by (1.3.79) and is always equal to the initial Kelvin's impulse M_0,

$$M = M_0 \quad = 4L\sigma\tanh\left(\tfrac{W}{\sigma}\right). \tag{5.2.7}$$

The total energy E is also conserved for an inviscid flow as in equation (1.3.80). In the presence of hyperdiffusion, the energy satisfies

$$\begin{aligned} \frac{dE}{dt} &= \frac{1}{2}\frac{d}{dt}\int_D |\nabla\psi_1|^2 + |\nabla\psi_2|^2 + \frac{1}{2}(\psi_1 - \psi_2)^2 d^2\mathbf{x}, \\ &= -\nu_4 \int_D |\nabla^2\psi_1|^2 + |\nabla^2\psi_2|^2 d^2\mathbf{x}. \end{aligned} \tag{5.2.8}$$

The last term of (5.2.8) due to hyperdiffusion is strictly less than zero and ensures that over the entire evolution of the flow the energy satisfies $E \leq E_0$

---

[1] The channel width employed here is wider than that used in E08 where $W = 5\pi$.

where E_0 is the initial energy given by

$$E_0 = \frac{1}{2}L\int_{-W/2}^{W/2} |\nabla\Psi_1|^2 + |\nabla\Psi_2|^2 + \frac{1}{2}\left(\Psi_1 - \Psi_2\right)^2 dy. \tag{5.2.9}$$

The third set of constraints is due to the fluid parcel-wise conservation of PV (5.2.5) and as discussed in section 1.3.3, if the flow is inviscid $\nu_4 = 0$, the functional

$$\mathcal{C}[q_1, q_2] = \int_D C_1(q_1) + C(q_2)d^2\mathbf{x}, \tag{5.2.10}$$

is conserved for arbitrary functions C_1 and C_2. One way to ensure that (5.2.10) is satisfied is to ensure that the evolving PV field q_i belongs to the set of 'conservative rearrangements' of the initial PV field $Q_i(y)$. This property is exploited and will be explained in the context of the EPVH hypothesis in section 5.4. In the presence of hyperdiffusion, as in the numerical simulations of this chapter, PV is not conserved exactly and the final PV field is therefore not a precise rearrangement of the initial PV field. However, equation (5.2.10) still places restrictions on the final PV field, which can be represented as a parcel-wise rearrangement of PV followed by the application of a smoothing function consistent with the presence of hyperdiffusion.

5.2.4 Description of the parameter space

The three parameters in this problem are inverse criticality β, jet half-width σ and barotropic shear Γ. E08 considered the problem for $\Gamma = 0$ and performed a suite of numerical simulations in the region of parameter space $\beta \in [0.2, 0.325]$, $\sigma \in [1.5, 3.0]$. The choice of parameter space ensures that the instability is purely baroclinic, that is, no barotropic instability takes place as may be demonstrated through consideration of the background potential vorticity gradients in each layer (1.3.71)

$$\frac{dQ_1}{dy} = \beta + \left(\frac{\sigma^2-8}{2\sigma^2}\right)\operatorname{sech}^2\left(\frac{y}{\sigma}\right) + \frac{6}{\sigma^2}\operatorname{sech}^4\left(\frac{y}{\sigma}\right), \tag{5.2.11}$$

$$\frac{dQ_2}{dy} = \beta - \frac{1}{2}\operatorname{sech}^2\left(\frac{y}{\sigma}\right). \tag{5.2.12}$$

It should be emphasised that the addition of the background shear in (5.2.1) leaves the potential vorticity gradients dQ_i/dy unchanged from the $\Gamma = 0$ flow.

The first consideration is to ensure that a given flow is unstable to baroclinic instability. The Charney-Stern-Pedlosky criterion [94, 133] for baroclinic instability is satisfied by ensuring that a reversal of PV gradient occurs in the lower layer. The PV gradient in the lower layer (5.2.11) is strictly positive if $\beta > \frac{1}{2}$ and so for the flow to be unstable to baroclinic instability, it is required that

$$\beta < \frac{1}{2}. \tag{5.2.13}$$

This condition is precisely the stability condition for the uniform, frictionless flow studied in chapter 2. Unlike the uniform flow problem of chapter 2, the criterion (5.2.13) is not satisfied across the whole channel, but within a central region whose width depends upon the parameters (β, σ).

The second consideration is to ensure that there is no possibility of barotropic instability, which would change the character of the instability and may occur if the PV gradient of the upper layer PV field $Q_1(y)$ or the layer-wise mean PV field $Q_1(y)+Q_2(y)$ changes sign (Rayleigh's inflection point criterion, e.g. Vallis [133], p.258). Using (5.2.11) the upper layer PV gradient does not change sign if

$$\beta > \frac{2}{3\sigma^2}\left(1 - \frac{\sigma^2}{8}\right)^2, \tag{5.2.14}$$

and the layer-wise mean PV field $Q_1(y) + Q_2(y)$ is given by

$$Q_1(y) + Q_2(y) = 2\beta y + \frac{2}{\sigma}\text{sech}^2\left(\frac{y}{\sigma}\right)\tanh\left(\frac{y}{\sigma}\right), \tag{5.2.15}$$

the gradient of which does not change sign if

$$\beta > \frac{1}{3\sigma^2}. \tag{5.2.16}$$

The parameter space (β, σ) is chosen such that (5.2.14) and (5.2.16) are satisfied, which precludes the possibility of an instability with a mixed barotropic-baroclinic character.

The three conditions (5.2.13), (5.2.14) and (5.2.16) are plotted in Fig. 5.2.1 within the (β, σ) parameter space. The points in Fig. 5.2.1 correspond to a suite of numerical simulations with $\beta \in [0.2, 0.325]$ and $\sigma \in [1.5, 3.0]$ (as used in E08). The restriction of $\sigma \leq 3.0$ is to ensure that the jet width σ is less than the Rhines scale [101], otherwise there is the possibility of multiple jet formation [89][2]. For a wider jet, the instability develops further from the centre of the channel and so $\sigma \leq 3.0$ also avoids any interference from the channel sidewalls. The final restriction on the extent of the parameter space $\beta \geq 0.2$ is due to the limitations of E08's formulation of EPVH theory, which will be discussed in section 5.4. Flows with $\beta \leq 0.2$ were considered in a second paper by Esler [28], the so-called 'leaky-barrier' problem. At these values of

---

[2] The Rhines scale is defined to be $L_R = \sqrt{V/\beta}$ where V is the turbulent eddy velocity. If it is assumed that $V \sim U$ (the initial velocity), then $L_R \ll \sigma$ corresponds to the condition $\sigma \gg \beta^{-1/2}$

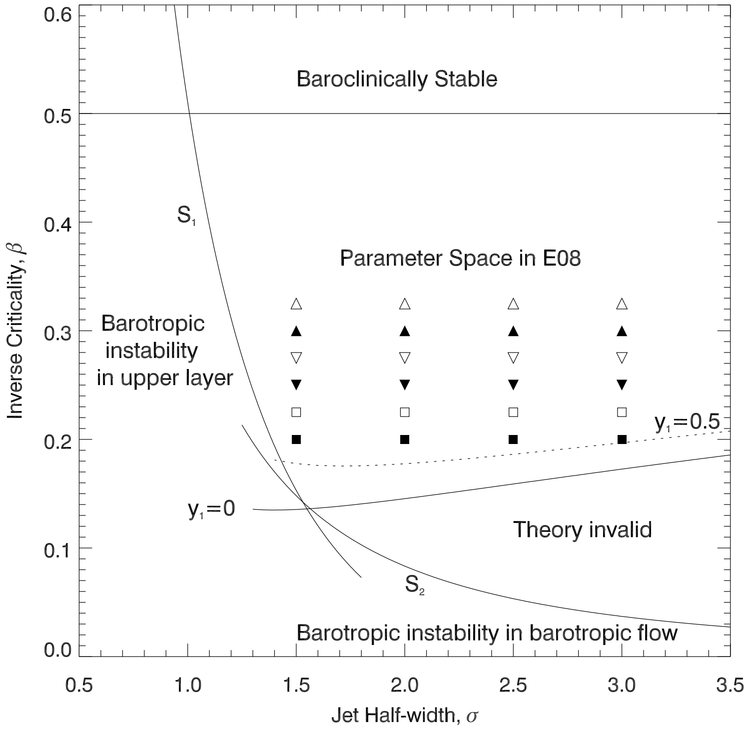

Figure 5.2.1: Regime diagram corresponding to Fig. 2 of E08. The solid line $\beta = 0.5$ marks the boundary between flows that are stable and unstable to baroclinic instability as given by (5.2.13). S_1 and S_2 mark areas where the instability may have a mixed baroclinic-barotropic character as given by (5.2.14) and (5.2.16) respectively. The points mark the parameters at which numerical simulations are carried out to test EPVH. Also plotted are the limits where the EPVH theory might be expected to be invalid, $y_1 = 0$ and $y_1 = \frac{1}{2}$, which will be discussed in section 5.4.7.

the inverse criticality the mixing barrier at the jet's core breaks down allowing the mixing of fluid in the upper layer across the centre of the channel.

5.3 Characteristics of Typical Flow Evolution

5.3.1 Evolution and energetics of a typical flow of a symmetric jet ($\Gamma = 0$)

Fig. 5.3.1 shows snapshots of the total PV field in each layer at times $t = 100$ and $t = 400$ for a flow with parameters $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$ at high numerical resolution (HR in Table 5.1). The initial perturbation (5.2.6) excites all wavenumbers between 4 and 20 and the initial wave growth follows the fastest growing normal mode of the linear dynamics (discussed further in section 5.4.6). Following the initial linear evolution, the waves become saturated first in the lower layer where PV contours wrap up in the centre of the channel as observed at $t = 100$ (panel b). This is similar to the wrapping up of PV contours in chapter 2 (see e.g. Fig. 2.6.2), but in this case it is limited to a finite region in the centre of the channel rather than across the whole lower layer. In the upper layer PV stirring occurs on each flank of the jet as observed at $t = 100$ (panel a). Also in panel (a), the jet core is seen to support large amplitude waves in the upper layer, which as the flow subsequently develops start to decay. Panels (c) and (d) of Fig. 5.3.1 show that by $t = 400$ the waves have almost decayed completely and PV mixing has left regions of weak turbulence in the centre of the lower layer and to the flanks of the jet in the upper layer. Outside of these weakly turbulent areas, the flow is laminar and disturbed only slightly by small amplitude waves.

At the jet's core in the upper layer, in panels (a) and (c), a central region has formed where the PV gradient is very strong. This region acts as a mixing

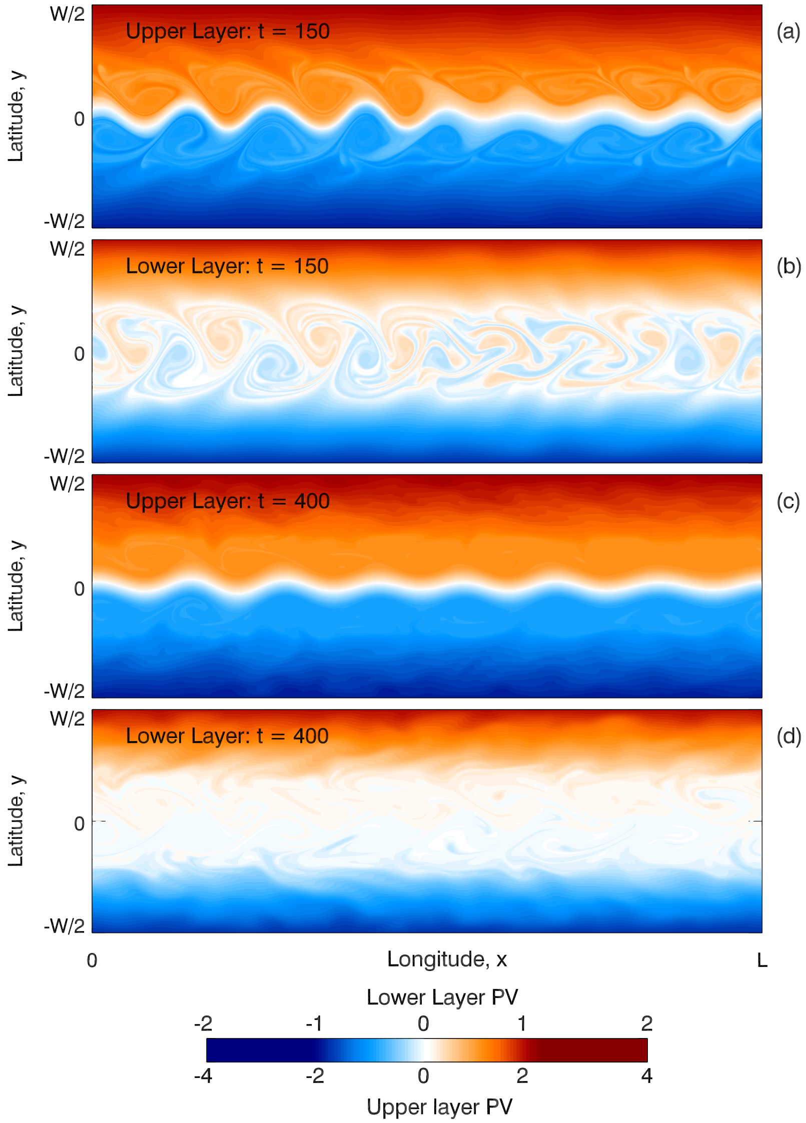

Figure 5.3.1: Snapshots of the PV distribution in the upper and lower layer for the flow with $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$. Positive PV values are blue and negative PV values are red.

barrier and prevents eddy-driven transport across it so that no mixing of PV can occur across the centre of the upper layer. The central mixing barrier and associated PV mixing at the flanks of the jet in the upper layer is consistent with ‘Rossby-wave elasticity’. The magnitude of the Rossby (quasi-) elasticity depends on the latitudinal PV gradient, itself fundamental to Rossby wave propagation. In regions of strong PV gradients (or strong Rossby elasticity), such as at the upper layer jet core, Rossby wave propagation dominates and the behaviour is wave-like. In regions where the PV gradient is weakened, such as to the flanks of the jet, there is a positive feedback since the Rossby elasticity is weakened here, facilitating further mixing of PV. In these regions (panel c), turbulent mixing leads to eventual PV homogenization. Rossby wave elasticity is also intricately related to PV inversion, from which the the velocity profile of the flow may be deduced. Strong eastward jets develop in regions of sharp PV gradients at the centre of the channel and weaker westward jets develop in regions of homogenized PV to the flanks. This will be seen in plots of the zonal wind profiles in section 5.5 (see Figs. 5.5.3 and 5.5.4).

Fig. 5.3.2 shows the time evolution of the energetics of the flow $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$ for three numerical resolutions (specified in Table 5.1). Both linear growth and nonlinear decay are almost entirely inviscid processes since total energy (TE) is very nearly conserved. The total eddy energy (EKE+EPE) is extracted from the available zonal mean potential energy (ZPE) during the linear growth stage ($t \lesssim 100$) and is returned to the zonal flow as kinetic energy (ZKE) through inviscid turbulent processes ($t \gtrsim 100$). The atmospheric literature describes this as a ‘baroclinic-growth barotropic-decay’ cycle [119,

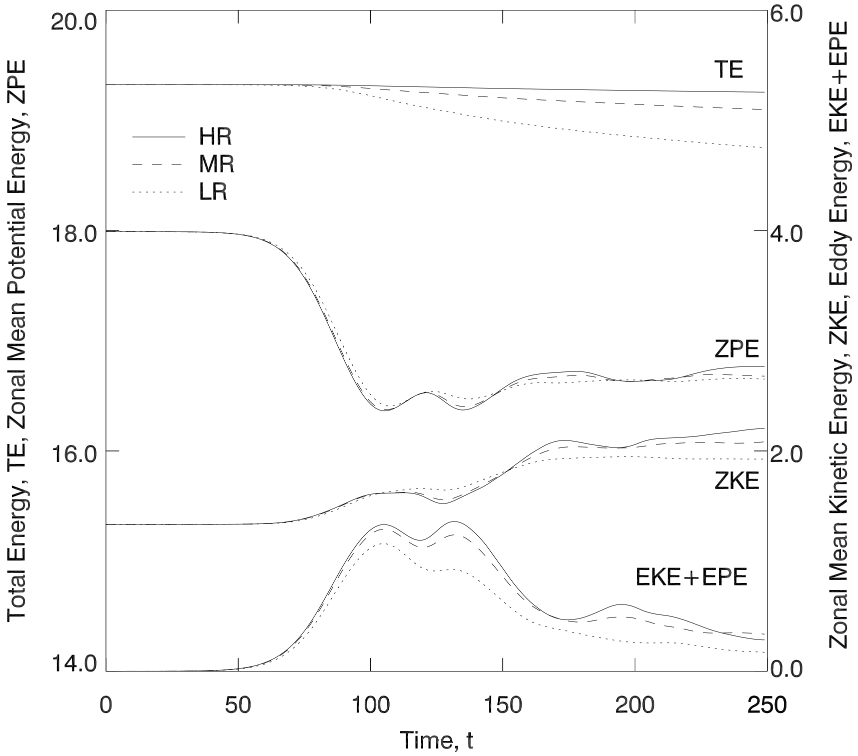

Figure 5.3.2: Evolution of energetics for a jet with parameters $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$. The four curves are total eddy energy (EKE+EPE), zonal kinetic energy (ZKE), zonal potential energy (ZPE) and total energy (TE) over the course of the baroclinic-growth barotropic-decay cycle. The energetics of the three different resolutions stated in Table 5.1 are each plotted as labelled.

132]. At long times ($t \gtrsim 200$), the rate of increase of zonal kinetic energy slows resulting in the PV distribution observed in panels (c) and (d) of Fig. 5.3.1. At this time, the jet has ceased to be forced by eddies and has entered a near-steady state. It is this near-zonal, equilibrated state that EPVH theory aims to predict.

Fig. 5.3.2 compares the energetics of three numerical simulations for the three resolutions in Table 5.1. All resolutions are qualitatively similar over the entire evolution of the flow. Fig. 5.3.2 also shows that at the highest resolution, the total energy (TE) is conserved very well whereas for the LR and MR simulations the total energy decreases very slightly over the course of the lifecycle due to the greater hyperdiffusion in these simulations. At all resolutions, plots of PV contours such as Fig. 5.3.1 reveal that the large scale structures of the flow are the same. There is little difference to report between the MR and HR resolutions.

5.3.2 Evolution and energetics of a typical flow with a background, barotropic shear $\Gamma \neq 0$

The discussion now turns to the evolution of the flow when the background, barotropic shear is non-zero ($\Gamma \neq 0$) focusing on the differences with symmetric jet profiles. The time evolution of an asymmetric flow follows a similar baroclinic-growth, barotropic-decay cycle to that of the symmetric flow plotted in Fig. 5.3.2. Fig. 5.3.3 shows the evolution of the eddy energy (EKE+EPE) over the course of the lifecycles for flows with $\Gamma = 0.00, 0.04, 0.08, 0.12$ and $(\beta, \sigma) = (0.25, 2)$. Fig. 5.3.3 shows that as the barotropic shear Γ increases, the initial linear wave growth is slower and saturates at a lower energy. The nonlinear evolutions of the flows are also affected as the eddy energy takes longer to decay for stronger shears. All flows have reached an equilibrated state by $t = 400$. This is confirmed by plots of the PV distribution such as that in Fig. 5.3.1. The observation that the initial wave growth is inhibited by a background barotropic shear is consistent with the 'barotropic governor' mechanism proposed by James & Gray [59] who argue that barotropic structure in the flow attenuates the baroclinic conversion process: in the absence of a barotropic shear, '...the phase tilts of the most unstable wave are nearly

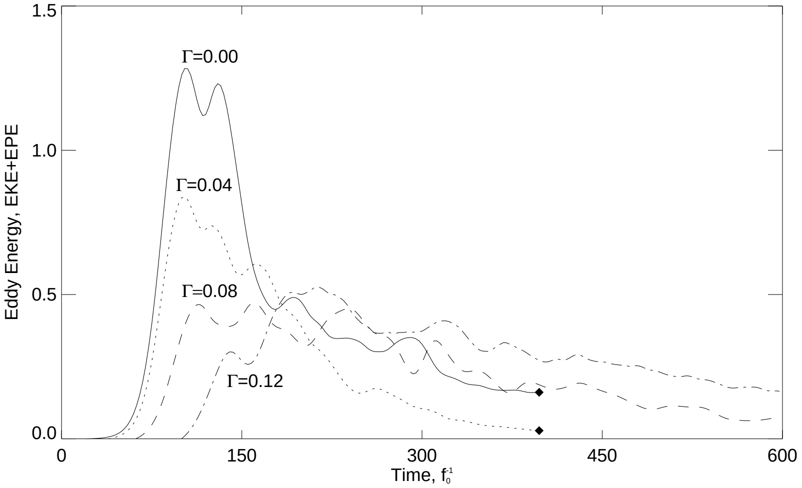

Figure 5.3.3: Evolution of the wave energy (eddy kinetic and potential energy) over the course of a lifecycle for $\Gamma = 0.00, 0.04, 0.08, 0.12$ with $(\beta, \sigma) = (0.25, 2.0)$.

optimally configured for baroclinic energy conversion, any rearrangement of this structure is likely to lead to slower growth rates' [57].

Fig. 5.3.4 shows a comparison of the PV distribution in each layer at two snapshots in time $t = 100$ and $t = 400$ for the asymmetric flows with $(\beta, \sigma) = (0.25, 2)$ and $\Gamma = 0, 0.04, 0.10$. The top panel of Fig. 5.3.4 is a reproduction of panel (a) in Fig. 5.3.1, with a box showing where a close up of the PV distribution has been taken in panels (b-d). The left-hand panels (b,i-iv) show close-ups of the evolution of the PV distribution for the symmetric jet ($\Gamma = 0$). The middle column of panels (c,i-iv) show the evolution of the PV field for $\Gamma = 0.04$ and the right-hand column of panels (d,i-iv) shows the evolution of the PV field for $\Gamma = 0.12$.

The central column of panels, for a flow with a relatively 'weak' shear ($\Gamma = 0.04$) are very similar to the symmetric case in the left-hand column of panels. The role up of PV contours in each layer proceeds in the same way as

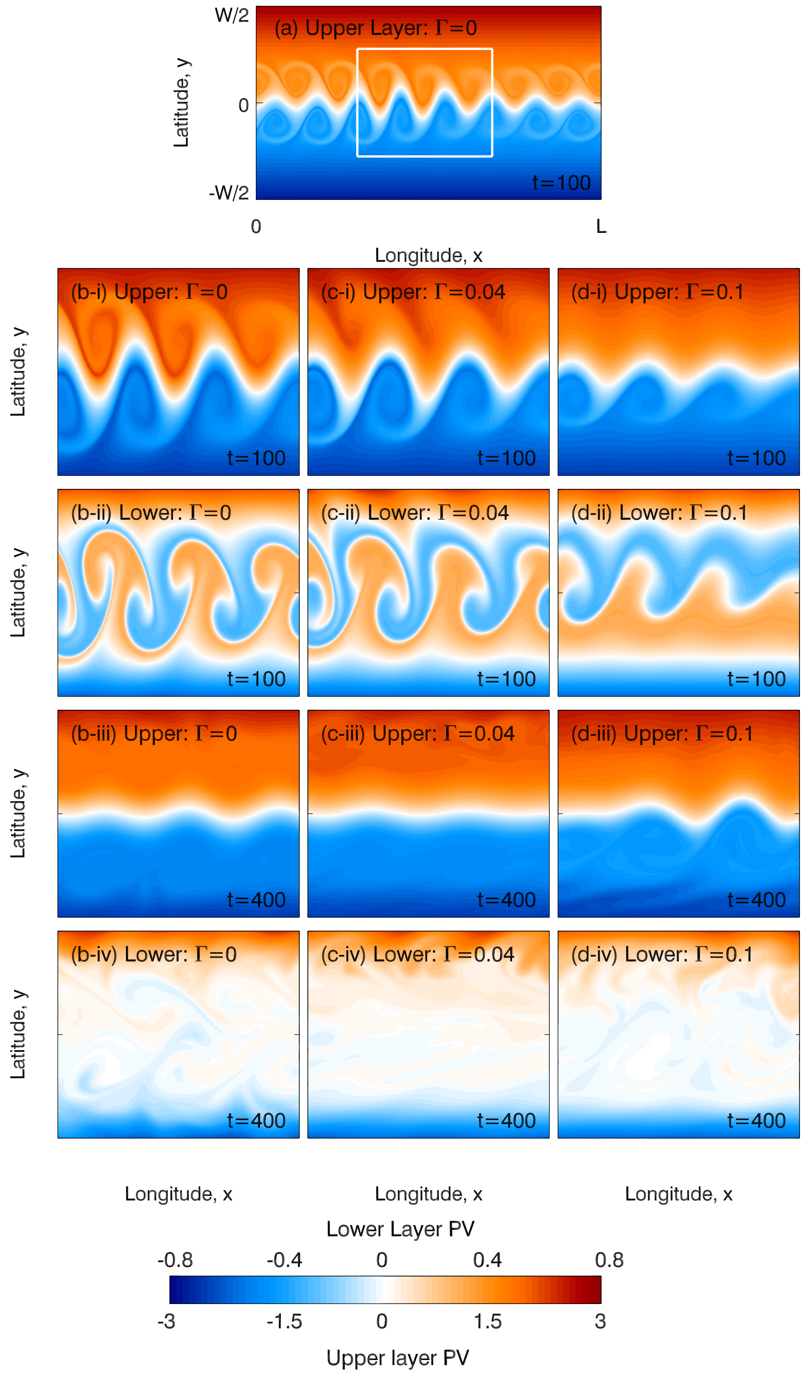

Figure 5.3.4: Plots of PV distributions in the upper and lower layers for an initial jet with $(\beta, \sigma) = (0.25, 2)$ and $\Gamma = 0, 0.04, 0.01$. Panel (a) shows the whole channel with the square showing the subdomain from which panels (b-d) are taken. Panels (b,i-iv) show the upper and lower layer PV distributions at two snapshots in time for the symmetric jet $\Gamma = 0$, as labelled. Panels (c,i-iv) and (d,i-iv) show the corresponding snapshots for $\Gamma = 0.04$ and $\Gamma = 0.10$ respectively.

was described in section 5.3.1 with PV mixing occurring on both flanks of the upper layer jet, however the mixing region on the southern flank is appreciably wider than its northern counterpart. The right-hand column of panels for a 'strong shear' ($\Gamma = 0.1$) which shows a qualitatively different pattern of mixing from the $\Gamma = 0$ and $\Gamma = 0.04$ flows. In panel (d-i) of Fig. 5.3.4, no wrapping up of PV contours is seen on the northern flank of the jet in the upper layer. Correspondingly, in the lower layer (panel d-ii), the pattern of the wrapping up of PV contours is clear to the southern edge of the mixing region, whereas the contours do not fully wrap on the northern edge. Ultimately, for the $\Gamma = 0.1$ flow, there is no wave breaking on the northern flank of the upper layer jet and no region of mixed PV (panel d-iii).

Fig. 5.3.5 shows the zonal mean PV profiles in each layer of the equilibrated PV distributions in panels (b,iii-iv), (c,iii-iv) and (d,iii-iv). Panels (a,c) of Fig. 5.3.5 show that when $\Gamma = 0, 0.04$, i.e. a 'weak shear', there are clearly two regions of the upper layer where the zonal mean PV is homogenized, whereas in panel (e) for the $\Gamma = 0.10$ flow ('strong shear') there is only one. Panels (d-f) show that the lower layer homogenization of the mean PV profiles varies little between $\Gamma = 0, 0.04, 0.10$. The simulations of Figs. 5.3.4 and 5.3.5 span two distinct qualitative regimes of PV mixing for the $(\beta, \sigma) = (0.25, 2.0)$ jets:

(a) For 'weak' shears, turbulent mixing occurs both to the north and south flank of the central jet. As the shear is increased the mixing region on the northern flank ($y > 0$) disappears. Consideration of further numerical simulations shows that the 'weak' shear regime approximately lies in the parameter range $\Gamma \lesssim 0.05$.

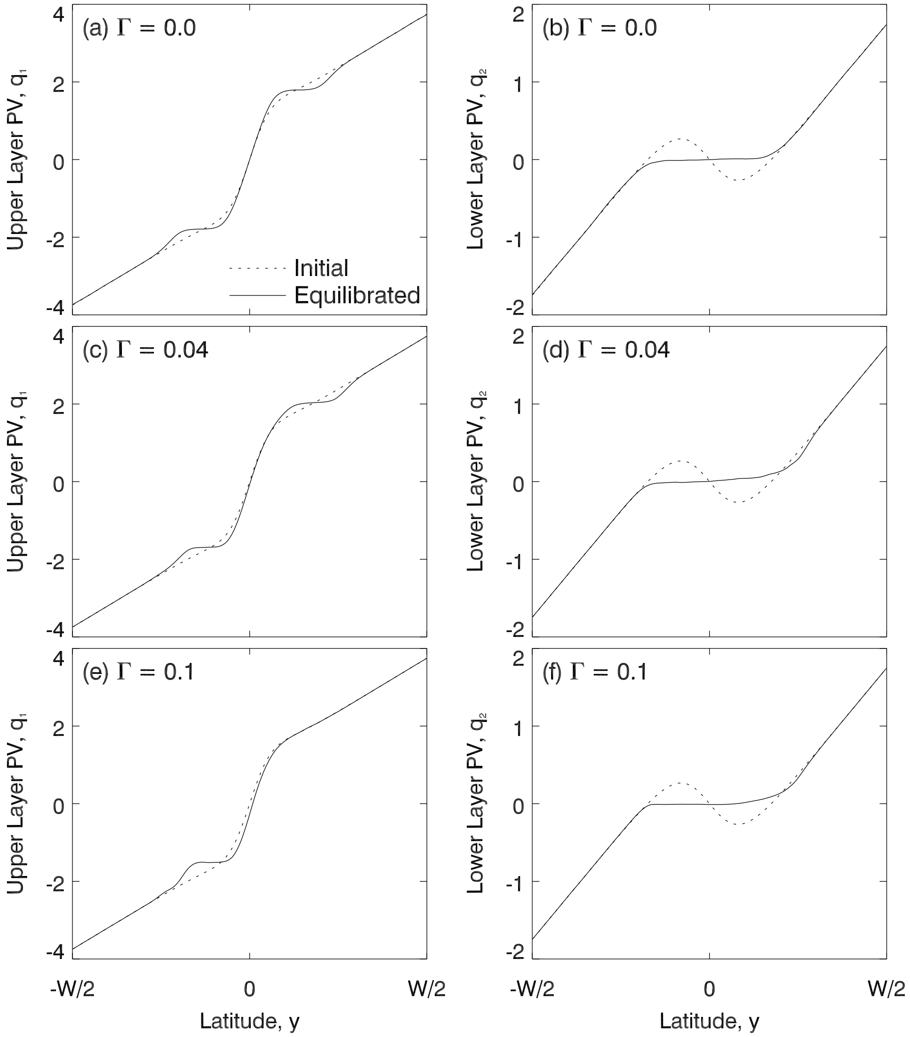

Figure 5.3.5: Zonal mean PV profiles of the equilibrated flow for the $(\beta, \sigma) = (0.25, 2)$ jet at $\Gamma = 0, 0.04, 0.10$ corresponding to the bottom six panels of Fig. 5.3.4. Panels (a,c,e) are for the upper layer and (b,d,f) are for the lower layer.

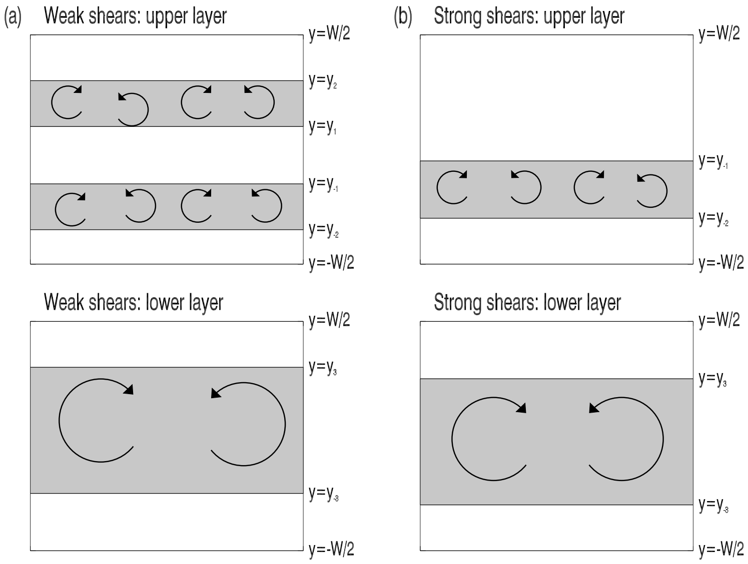

Figure 5.3.6: Schematic showing regions where turbulent mixing occurs in each layer for (a) weak shears and (b) strong shears.

(b) For ‘strong’ shears ($\Gamma \gtrsim 0.05$) there is only one mixing region in the upper layer to the southern flank of the jet.

For both regimes, the mixing region in the lower layer extends across the centre of the channel and is similar in size for all values of Γ.

The two mixing patterns described for ‘weak’ and ‘strong’ shears as set out in (a) and (b) are shown schematically in Fig. 5.3.6. The turbulent mixing in these regions leads to homogenization of PV, which is described as occurring in well de-lineated regions. For weak shears, in the upper layer, mixing occurs in the regions, $y \in [y_1, y_2]$ and $y \in [y_{-2}, y_{-1}]$ with $-W/2 < y_{-2} < y_{-1} < 0 < y_1 < y_2 < W/2$. This description of the mixing regions gives a natural definition of the central mixing barrier, which lies in the range $y \in [y_{-1}, y_1]$. In the lower layer there is just one region where homogenization of PV occurs extending across the centre of the channel $y \in [y_{-3}, y_3]$ with

$-W/2 < y_{-3} < 0 < y_3 < W/2$. For strong shears, mixing only occurs in the upper layer in $y \in [y_{-2}, y_{-1}]$.

The pattern of PV mixing for weak and strong shears is similar for other initial jet profiles (with $(\beta, \sigma) \neq (0.25, 2)$), though the change from one to two upper layer mixing regions occurs at different values of Γ. The existence of clearly de-lineated regions where PV is homogenization occurs will be exploited in the next section where the EPVH analytical theory is described.

5.4 Introduction of EPVH and Linear Instability theory

5.4.1 General Principle of the Hypothesis

Within the E08 parameter space shown in Fig. 5.2.1 the evolution of the energetics and PV distribution of the flow are qualitatively similar to those described by Figs. 5.3.1 and 5.3.2 (see Figs. 3 and 4 of E08). Using the assumption that the baroclinic lifecycle is complete once all available potential energy has been released, E08 forms a hypothesis for baroclinic equilibration:

EPVH-A hypothesis[3]

> The active eddy field forces the mean flow towards a final, near-steady, equilibrated state that has the minimum possible available potential energy subject to the constraints of conservation of total zonal momentum and total zonal energy, and the additional kinematic constraint that all mixing processes result in potential vorticity homogenization within well-delineated regions.

---

[3] verbatim from E08 [29]

The EPVH hypothesis then makes the following assumptions:

(a) A negligible amount of total energy is lost to dissipation.

(b) All but a negligible proportion of total energy is returned to the zonal mean, which is equivalent to saying that the baroclinic-growth barotropic-decay cycle proceeds to completion.

(c) All mixing that occurs during the equilibration results in complete homogenization of PV within distinct regions of the channel that are bounded in latitude.

(d-A) The flow acts to minimize the total available potential energy subject to constraints interpreted via (a-c).

Figs. 5.3.1, 5.3.2, 5.3.4 and 5.3.5 support assumptions (a-c) in the case $(\beta, \sigma) = (0.25, 2)$. Plots of PV contours from simulations elsewhere in the (β, σ) parameter space confirm similar results. Assumption (d-A) closes the system and, simply stated, remarks that the baroclinic instability proceeds until the available potential energy in the system is exhausted subject to the dynamical constraints on the flow.

5.4.2 Formulation of EPVH for an asymmetric flow

E08 only considered initial jets that are symmetric about the centre of the channel $y = 0$ and the theory took this into account by requiring PV mixing regions in both layers to be symmetric about the centre of the channel, i.e. $y_{-i} = -y_i$ for $i = 1, 2, 3$ in Fig. 5.3.6 panel (a). This symmetric theory showed good agreement with the numerical simulations in terms of predicting the equilibrated state of the flow throughout the parameter space. The theory was tested most thoroughly in the parameter space $\beta \in [0.2, 0.325]$, $\sigma \in [1.5, 3]$, but appears to work within the entire parameter region (see Fig. 5.2.1) apart from where it predicts its own breakdown.

The formulation of EPVH theory as introduced in E08 is now adapted to allow the possibility of PV mixing which is not symmetric about the centre of the channel. The mixing latitudes are written as a vector $\mathbf{y} = (y_1, y_2, y_3, y_{-1}, y_{-2}, y_{-3})^T$. If complete homogenization of PV occurs in the regions illustrated in the schematic of Fig. 5.3.6 panel (a), then the equilibrated PV distribution may be written as

$$\begin{aligned} q_1(y) &= \begin{cases} Q_1(y), & y_2 < y < \frac{1}{2}W, \\ \overline{Q}_1^{(1,2)}, & y_1 < y < y_2, \\ Q_1(y), & y_{-1} < y < y_1, \\ \overline{Q}_1^{(-2,-1)}, & y_{-2} < y < y_{-1}, \\ Q_1(y), & -\frac{1}{2}W < y < y_{-2}, \end{cases} \\ q_2(y) &= \begin{cases} Q_2(y), & y_3 < y < \frac{1}{2}W, \\ \overline{Q}_2^{(-3,3)}, & y_{-3} < y < y_3, \\ Q_2(y), & -\frac{1}{2}W < y < y_{-3}, \end{cases} \end{aligned} \tag{5.4.17}$$

where the background PV in each layer $Q_1(y)$ and $Q_2(y)$ are given by (5.2.2) and the mean quantities of the form $\overline{Q}_i^{(j,k)}$ are given by

$$\overline{Q}_i^{(j,k)} = \frac{1}{y_k - y_j} \int_{y_j}^{y_k} Q_i(y) dy. \tag{5.4.18}$$

The EPVH theory makes a prediction $\mathbf{Y}^T = (Y_1, Y_2, Y_3, Y_{-1}, Y_{-2}, Y_{-3})$ for the latitudes that bound the mixing regions given the parameters of the initial jet (β, σ, Γ) by minimizing the available potential energy V, given by equation (1.3.82) as

$$V = \frac{1}{4} \int_D (\psi_1 - \psi_2)^2 d^2\mathbf{x}. \tag{5.4.19}$$

Under assumption (b) of the EPVH-A hypothesis, $V \approx \overline{V}$ at the end of the simulation, hence this is equivalent to minimizing the zonal mean potential energy $\overline{V}$. If the latitudes bounding the mixing regions $\mathbf{y}$ are allowed to vary then the minimum $\mathbf{Y}$ occurs at a stationary point of

$$F(\mathbf{y}|Q_i(y)) = \overline{V}(\mathbf{y}|Q_i(y)) + \lambda\overline{E}(\mathbf{y}|Q_i(y)) + \mu M(\mathbf{y}|Q_i(y)), \tag{5.4.20}$$

where λ and μ are Lagrange multipliers. The system of equations to be solved in $(\mathbf{y}, \lambda, \mu)$ is then

$$\frac{\partial F}{\partial y_i} = 0, \qquad \overline{E}(\mathbf{Y}) = E_0, \qquad M(\mathbf{Y}) = M_0, \qquad i = \pm 1, \pm 2, \pm 3. \tag{5.4.21}$$

Recall that the conserved quantities M_0 and E_0 are given explicitly by (5.2.7) and (5.2.9) respectively.

The partial derivatives of F may be calculated explicitly and are given in appendix 5.A.1. Equations (5.4.21) consist of 8 unknown parameters within 8 equations and can be solved numerically using Broyden's method (e.g. Press et al. [99]). To ensure that a solution $\mathbf{Y}$ is a minimum of available potential energy, it is checked that the potential $\overline{V}$ increases if the solution is perturbed slightly subject to momentum and energy conservation:

$$\overline{V}(\mathbf{Y} + \delta\mathbf{Y}) > \overline{V}(\mathbf{Y}), \tag{5.4.22}$$

for perturbations $\delta\mathbf{Y} = (\delta Y_1, \delta Y_2, \delta Y_3, \delta Y_{-1}, \delta Y_{-2}, \delta Y_{-3},)^T$ where $\delta Y_i \ll 1$ and subject to the constraints

$$\sum_i \delta Y_i \frac{\partial M}{\partial y_i}\bigg|_{\mathbf{y}=\mathbf{Y}} = \sum_i \delta Y_i \frac{\partial \overline{E}}{\partial y_i}\bigg|_{\mathbf{y}=\mathbf{Y}} = 0, \tag{5.4.23}$$

for $i = \pm 1, \pm 2, \pm 3$. A prediction by the EPVH theory for the equilibrated flow is given by the PV field (5.4.17) where the mixing latitudes are the solution to (5.4.21), $(y_1, y_2, y_3, y_{-1}, y_{-2}, y_{-3})^T = (Y_1, Y_2, Y_3, Y_{-1}, Y_{-2}, Y_{-3})^T$.

To find the geostrophic streamfunction from which physical quantities such as the zonal wind may be calculated, the potential vorticity is inverted. The inversion of the mixed zonal PV field (5.4.17) requires use of the boundary conditions (5.2.3) and (5.2.4), which may be recast in a form such that the original boundary conditions of E08 can be used (see appendix 5.A.2).

5.4.3 EPVH for 'strong' shears

For larger values of Γ there is only one mixing region in the upper layer as described by the mixing schematic of Fig. 5.3.6 panel (b). It is possible to solve the minimization problem (5.4.21) with only one region of PV homogenization in the upper layer by setting $y_{-1} = y_{-2}$ and removing any dependence of the mixed PV distribution (5.4.17) on the latitudes y_1, y_2. The minimization problem then consists of 6 equations in 6 unknowns,

$$\frac{\partial F}{\partial y_i} = 0, \qquad \overline{E}(\mathbf{Y}) = E_0, \qquad M(\mathbf{Y}) = M_0, \qquad i = 1, 2, \pm 3. \tag{5.4.24}$$

The new problem minimizes the available potential energy subject to the same constraints of energy and momentum conservation, but with only one mixing region in the upper layer and one mixing region across the centre of the jet in the lower layer.

There is the possibility of multiple equilibrated solutions to (5.4.21) and (5.4.24) for a given initial jet (β, σ, Γ). Therefore a comparison of the potential energies $\overline{V}$ of these states is required to find the global minimum, which under the assumption (d-A) of EPVH-A is the relevant prediction. This will be discussed in section 5.6.

5.4.4 Maximization of mixing area, EPVH-B

Assumption (d-A) of the EPVH-A hypothesis (section 5.4.2) is that the baroclinic lifecycle is complete when the available potential energy is exhausted. Therefore the equilibrated state occurs at the minimum of available potential energy subject to the dynamical constraints of the system. A simple alternative for a variational principle is to maximize the area in which mixing takes place, a heuristic approach, but one which can be related to 'Rossby-wave elasticity' and maximum-entropy ideas in statistical mechanics. Essentially, it is an alternative way to close the set of equations (5.4.20-5.4.21) to provide a comparison to EPVH-A.

Rossby-wave elasticity [80] is a concept by which regions of strong PV gradients are associated with Rossby wave-like activity and regions of weak PV gradients are associated with turbulent mixing [23]. This is due to the positive feedback through which regions of weakened Rossby elasticity (or PV gradient) are susceptible to further weakening and in adjacent regions, the Rossby elasticity is strengthened. It may be argued then that the baroclinic lifecycle is complete when regions of mixed PV and weak Rossby elasticity have reached the greatest extent allowed by the surrounding regions of stronger elasticity. The maximization of mixing regions over the two layers of the model is a statement that the Rossby elasticity over the whole domain has been weakened as far as possible given the dynamical constraints of the flow. This is an alternative statement to the minimization of available potential energy that the baroclinic lifecycle is complete.

An alternative justification comes from ideas in statistical mechanics [82, 105, 106] where a mixing entropy is defined and maximized to determine the equilibrated state of the flow. The mean PV profile of the equilibrated flow or 'macro-state' is considered to be a coarse-grain average of an underlying 'micro-state', itself an arrangement of equal sized, small-scale vortices (this follows from parcel-wise conservation of PV). A measure of mixing entropy is then introduced, which is essentially a count of the number of possible micro-state arrangements for a given macro-state. By the Ergodic hypothesis, all realizable micro-states have similar probabilities and maximizing the mixing entropy corresponds to the most likely macro-state [75]. E08 explored such a theory for the unstable jet problem and demonstrated that EPVH gave superior predictions of equilibrated flows.

One way to improve maximum entropy is to restrict regions in which mixing may take place on the basis of empirical observations. Under the EPVH hypothesis, PV is rearranged in a prescribed way (i.e. it is homogenized within de-lineated regions of the channel following observations of numerical simulations). The maximum mixing entropy within this restricted mixed profile corresponds to the PV profile which has the largest area where PV has been mixed subject to energy and momentum conservation.

The following alternative hypothesis replaces assumption (d-A) of section 5.4.2 with (d-B) The flow acts to maximize the total area of mixing regions where complete homogenization of PV takes place, subject to the physical constraints in the system interpreted via (a-c).

This hypothesis will be referred to as EPVH-B and the minimization of potential energy hypothesis as EPVH-A. The total width of the mixing zones is given by

$$W(y_i) = (y_2 - y_1) + (y_{-1} - y_{-2}) + (y_3 - y_{-3}), \tag{5.4.25}$$

and so the function that must be maximized is

$$G(\mathbf{y}|Q_i(y)) = W + \lambda \overline{E} + \mu M, \tag{5.4.26}$$

The turning points of G are given by 8 equations in 8 unknowns

$$\frac{\partial G}{\partial y_i} = 0 \quad i = \pm 1, \pm 2, \pm 3 \qquad E(\mathbf{Y}) = E_0 \quad M(\mathbf{Y}) = M_0, \tag{5.4.27}$$

where, as before M_0 and E_0 are given by (5.2.7) and (5.2.9) respectively. The derivatives in (5.4.27) are written in terms of the mixed potential vorticity (5.4.17) in appendix 5.A.3 and solved using Broyden's method. EPVH-B can also be formulated for the 'strong shear' case in the same way as EPVH-A, described in section 5.4.3.

5.4.5 Minimization of potential enstrophy, EPVH-C

In addition to the two hypotheses for EPVH suggested in sections 5.4.2 and 5.4.4, a third hypothesis is now considered where the potential enstrophy in the system is minimized. The idea that the potential enstrophy is minimized is an alternative statement that the baroclinic lifecycle is complete and is based on an argument due to Bretherton & Haidvogel [16] and Leith [69].

This argument is known as the 'selective decay hypothesis' whereby some global invariants, called dissipated, undergo a turbulent cascade to small scales and are removed by even a small dissipation, whilst other invariants, called rugged, do not decay in a turbulent cascade. In this case the enstrophy is the dissipated invariant and the energy is the rugged invariant. The 'selective-decay principle' has been applied to two-dimensional turbulence over topography [16], magnetohydrodynamic turbulence [78], two-dimensional vortices by Leith [69] and barotropic waves in a channel [143].

In two-dimensional turbulent flows, enstrophy cascades to smaller scales and energy, which is also conserved, cascades to and remains at large scales (e.g. Vallis [133], section 8.3). The cascade of enstrophy continues until viscosity, however small, dissipates the enstrophy at a suitably small scale. Therefore in practice, there is a reduction in the total enstrophy with time. The rate of decay is related to the strain field, which is governed by the larger energy containing scales of motion and is independent of the type of small-scale dissipation. The energy invariant suffers no such decay, or at the very least decays on time scales much longer than the enstrophy. The consequence of this is that the flow relaxes to state, which has the minimum enstrophy for a given energy (and Kelvin's impulse) on time scales much faster than the dissipation of energy.

The assumption that the equilibrated state is independent of x means that minimizing enstrophy Z is equivalent to minimizing the zonal enstrophy $\overline{Z}$.

The hypothesis that the potential enstrophy in the equilibrated flow is minimal is implemented in a new hypothesis, labelled EPVH-C, by replacing assumption (d-A) with

(d-C) The flow acts to minimize the total enstrophy of the flow subject to the physical constraints in the system interpreted via (a-c).

Replacing the available potential energy $\overline{V}$ in (5.4.20) with the zonal enstrophy $\overline{Z}$, the function that must be minimized is

$$H(\mathbf{y}|Q_i(y)) = \overline{Z} + \lambda\overline{E} + \mu M. \tag{5.4.28}$$

The turning points of H are given by the following 8 equations in 8 unknowns.

$$\frac{\partial H}{\partial y_i} = 0 \quad i = \pm 1, \pm 2, \pm 3 \qquad \overline{E}(\mathbf{Y}) = E_0 \quad M(\mathbf{Y}) = M_0 \tag{5.4.29}$$

The nonlinear equations (5.4.29) are expressed in terms of the mixed potential vorticity 5.4.17 in appendix 5.A.4 and solved as before using Broyden's method. It is checked that the total enstrophy $\overline{Z}$ increases if the solution is perturbed slightly subject to momentum and energy conservation as in equations (5.4.22-5.4.23). Again, EPVH-C can be formulated for the 'strong shear' case in the same way as EPVH-A, described in section 5.4.3.

The results of sections 5.4.2, 5.4.4 and 5.4.5 have provided three different hypotheses for a theory to predict the equilibrated state of an initially unstable jet. Before looking at the predictions made by each of these, the linear theory of the flow is reviewed.

5.4.6 Linear Theory

The linear theory describing the initial growth of a small disturbance follows the same procedure as sections 2.3 and 3.2 where a dispersion relation was found analytically for uniform flow profiles. For the isolated jet of this chapter, the linear problem must be solved numerically.

Writing the streamfunction in terms of the zonal flow plus a small perturbation about it $\psi_i = \Psi_i(y) + \psi_i'(x, y, t)$, the linearized equations of motion from (1.3.72-1.3.73) are

$$\begin{aligned} \frac{\partial q_i'}{\partial t} + U_i \frac{\partial q_i'}{\partial y} + \frac{dQ_i}{dy}\frac{\partial \psi_i'}{\partial x} &= 0, \\ \frac{\partial \psi_i'}{\partial y} &= 0 \quad \text{on} \quad y = \pm\frac{W}{2}, \end{aligned} \tag{5.4.30}$$

for $i = 1, 2$. $U_i(y) = -\partial\Psi_i/\partial y$ is the background flow and $Q_i(y)$ is the background potential vorticity given by (5.2.1) and (5.2.2) respectively. The perturbation potential vorticity q_i' is given by

$$q_i' = \nabla^2\psi_i' + (-1)^i\left(\frac{\psi_1' - \psi_2'}{2}\right). \tag{5.4.31}$$

Look for normal mode solutions of the form

$$\psi_i' = \hat{\psi}_i(y)e^{ik(x-ct)}, \qquad i = 1, 2, \tag{5.4.32}$$

where $c = c_r + ic_i$ is the complex wave speed and the growth rate of the disturbance is given by kc_i. Insertion of (5.4.32) into (5.4.30) and (5.4.31) leads to

$$\begin{aligned} (U_i - c)\, q_i' + \frac{dQ_i}{dy}\hat{\psi}_i + (-1)^i\,(U_2 - U_1)\,\hat{\psi}_i &= 0, \\ \frac{\partial^2\hat{\psi}_i}{\partial y^2} - k^2\hat{\psi}_i + (-1)^i\frac{\hat{\psi}_1 - \hat{\psi}_2}{2} &= q_i', \end{aligned} \tag{5.4.33}$$

for $i = 1, 2$. Alternatively, equation (5.4.33) may be written as the eigenvalue problem (e.g. Swanson & Pierrehumbert [128], Esler & Haynes [30])

$$\mathbf{L}\boldsymbol{\psi} = c\mathbf{M}\boldsymbol{\psi}, \tag{5.4.34}$$

where $\boldsymbol{\psi} = (\hat{\psi}_1, \hat{\psi}_2)^T$ with

$$\mathbf{L} = \begin{pmatrix} U_1(y)\left(\frac{d^2}{dy^2} - k^2 - \frac{1}{2}\right) + \frac{dQ_1}{dy} & \frac{1}{2} \\ \frac{1}{2} & U_2(y)\left(\frac{d^2}{dy^2} - k^2 - \frac{1}{2}\right) + \frac{dQ_2}{dy} \end{pmatrix}, \tag{5.4.35}$$

and

$$\mathbf{M} = \begin{pmatrix} \left(\frac{d^2}{dy^2} - k^2 - \frac{1}{2}\right) + \frac{dQ_1}{dy} & \frac{1}{2} \\ \frac{1}{2} & \left(\frac{d^2}{dy^2} - k^2 - \frac{1}{2}\right) + \frac{dQ_2}{dy} \end{pmatrix}. \tag{5.4.36}$$

Eigenfunction solutions $\boldsymbol{\psi}$ then exist if the wavenumber k and frequency $\omega = kc$ satisfy some dispersion relation $D(\omega, k) = 0$. To find the numerical solution of (5.4.34), the flow is discretized on a grid and the y derivatives in the matrix operators $\mathbf{L}$ and $\mathbf{M}$ are approximated by finite differences. The matrix $\mathbf{M}$ is then inverted numerically and the eigenvalues and eigenvectors of $\mathbf{M}^{-1}\mathbf{L}$ are found using a linear algebra package (the QR algorithm, e.g. Press et al. [99]). The eigenvalues c are used to find the growth rate kc_i, from which the fastest growing mode is established. The streamfunction $\boldsymbol{\psi}$ may be found through substitution of the eigenvalue c of the fastest growing mode into equation (5.4.34) and use of inverse iterations to converge to a solution $\boldsymbol{\psi}(y)$. Insert $\boldsymbol{\psi}(y)$ into the linear mode (5.4.32) and take real parts to calculate the disturbance streamfunction.

A snapshot of the total streamfunction[4] $(\Psi_i + \tilde{\epsilon}\psi'_i + c_r y)$ of the fastest

---

[4] Here for simplicity, $\psi'_i \rightarrow \tilde{\epsilon}\psi'_i$, where $\tilde{\epsilon}$ is the streamfunction amplitude and ψ'_i is the

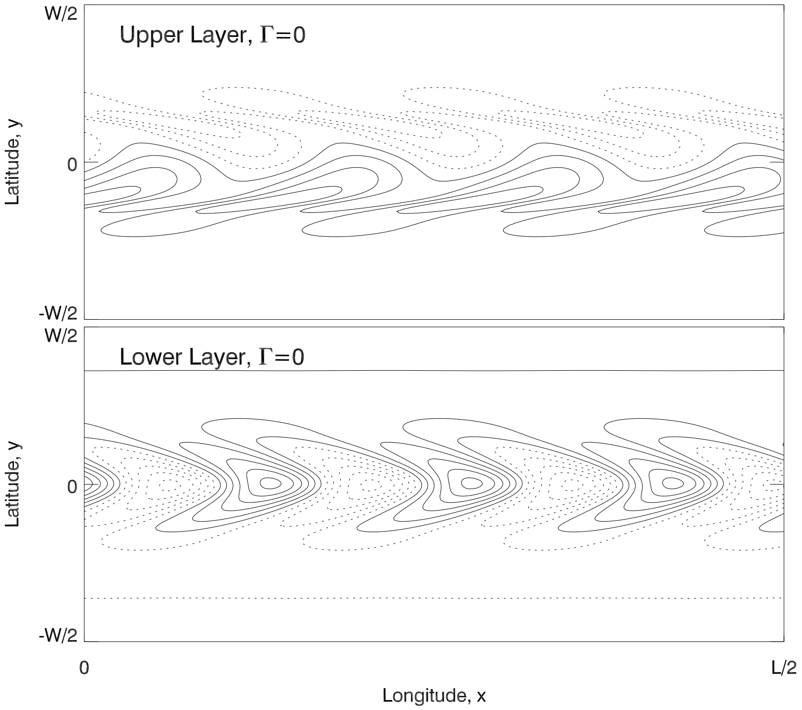

Figure 5.4.1: Total streamfunction $\psi_i = \Psi_i + \tilde{\epsilon}\psi_i' + c_r y$ in the frame of the wave, of the fastest growing linear mode on a jet with $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$. The amplitude of the disturbance is chosen to be $\tilde{\epsilon} = 12.5$. The contour interval is 0.75 in the upper layer and 0.4 in the lower layer. Half of the channel width is plotted in the x-direction for a better resolution of the streamfunction contours.

growing mode on the $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$ jet is shown in Fig. 5.4.1. The total streamfunction pattern in Fig. 5.4.1 is a useful way to describe the generation of Rossby waves and the associated convergent momentum flux (e.g. Vallis [133], section 12.1). Across the centre of the lower layer there are closed contours, which are associated with stirring of the fluid and generation of Rossby waves. The Rossby waves propagate away from the central region to the north and south where they break and dissipate. Associated with the propagation of Rossby waves is a transport of momentum into the central source region. The 'bow' shape of the eddies (closed streamlines) in the lower

---

normalized structure of the normal mode.

layer (Fig. 5.4.1 bottom panel) demonstrates the convergence of momentum. The northern part of the 'bow' is tilted south-east to north-west orientation and so the horizontal eddy velocities u' and v' are directed such that the zonal mean eddy momentum flux $\overline{u'v'} < 0$. The zonal eddy-momentum flux is positive in the southern part of the 'bow' where the tilt is south-west to north-east, resulting in a convergence of momentum into the centre of the channel and an acceleration of the jet. In the upper layer (upper panel of Fig. 5.4.1), there are two sets of closed streamlines, one on the northern flank of the jet where the eddy-momentum flux is negative and one on the southern flank where the eddy-momentum flux is positive (due to the tilt of the eddies in the same way as the lower layer 'bow' contours).

The Rossby waves excited by the stirring of PV in the centre of the lower layer propagate north and south until they break (as in the numerical simulations of Fig. 5.3.1), which occurs near the latitudes where linear theory becomes invalid. Linear theory breaks down at critical lines y_c, which are defined to occur where the real wave speed c_r is equal to the initial velocity in the upper layer, i.e. $U_1(y_c) - c_r = 0$, which means that terms on the leading diagonal of the linear operator $\mathbf{L} - c\mathbf{M}$ in equation (5.4.34) vanish. Nakamura [85] commented that the location of the critical lines of the fastest growing modes, i.e. where linear theory formally breaks down, approximately coincide with the mixing regions in the upper layer [144]. The latitude of the critical line can be calculated numerically for a given flow (β, σ, Γ), and will be compared to the mixing latitudes predicted by EPVH in sections 5.4.7 and 5.6.

Fig. 5.4.2. shows a contour plot of the total streamfunction for the fastest

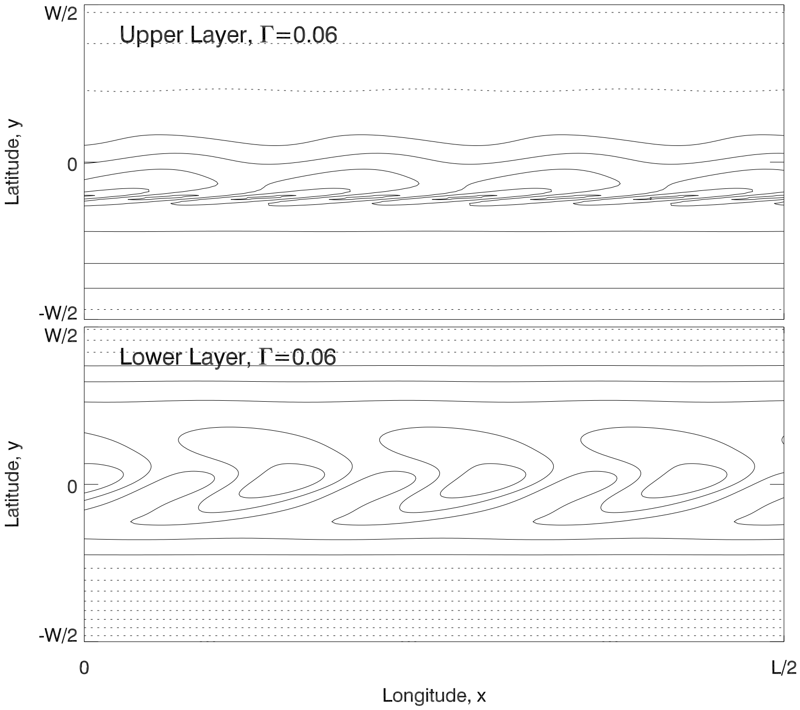

Figure 5.4.2: As for Fig. 5.4.1, but the fastest growing linear mode is for a jet with a background shear $(\beta, \sigma, \Gamma) = (0.25, 2, 0.06)$. The disturbance has amplitude $\tilde{\epsilon} = 5$. The contour intervals in the upper layer is 1 and in the lower layer is 0.4.

growing normal mode on a jet with parameters $(\beta, \sigma, \Gamma) = (0.25, 2, 0.06)$. In Fig. 5.4.2, closed streamlines only occur to the southern flank of the jet in the upper layer. This corresponds to the turbulent mixing occurring on this flank, which is observed in the PV snapshots of Fig. 5.3.4 panel (d-iii) and Fig. 5.3.5 panel (e) (the 'strong shear' case). In the context of eddy-transport, there is an upgradient momentum flux only on this flank of the jet. In the lower layer, the bow-shaped, closed streamlines are tilted by the background shear. The streamfunction patterns in Figs. 5.4.1 and 5.4.2 show that the linear background shear diminishes the regions in which closed streamlines and turbulent mixing occur in the channel. The linear growth rate is correspondingly re-

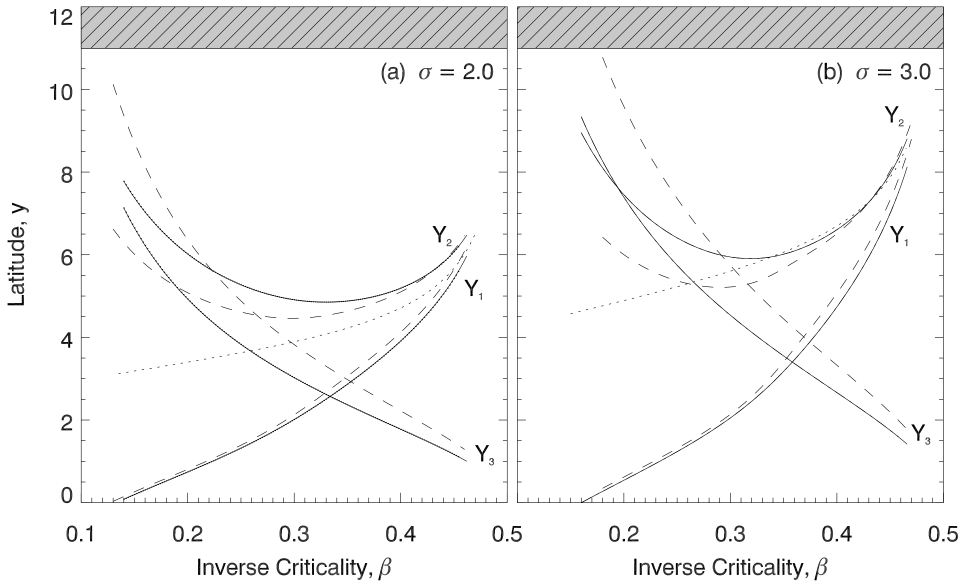

Figure 5.4.3: Solutions $(Y_1, Y_2, Y_3)^T$ of the symmetric $(Y_{-i} = -Y_i)$ nonlinear systems of equations (5.4.21), (5.4.27) and (5.4.29) for EPVH-A (solid lines), EPVH-B (dashed lines) and EPVH-C (solid lines-identical to EPVH-C) against inverse criticality β. Panels (a) and (b) show solutions for initial jet widths $\sigma = 2.0$ and $\sigma = 3.0$ respectively. $(Y_1, Y_2, Y_3)^T$ bound the regions where PV mixing is expected to take place. The dotted line shows the latitude y_c of the upper layer critical line of the fastest growing linear mode as calculated in section 5.4.6.

duced for the asymmetric jet (observed in Fig. 5.3.3) as the transport of eddy momentum is inhibited. The reduction of the linear growth rate and alteration of the structure of the linear modes are consistent with the barotropic governor arguments of James [59, 57]

5.4.7 EVPH predictions for symmetric jets

The results of the analytical EPVH theory are now reviewed for symmetric jets. The latitude predictions $(y_1, y_2, y_3)^T = (Y_1, Y_2, Y_3)^T$ for each of EPVH-A, EPVH-B and EPVH-C are shown in Fig. 5.4.3. EPVH-A and EPVH-C make identical predictions for the latitudes $(Y_1, Y_2, Y_3)^T$, i.e. the hypotheses of minimization of available potential energy and enstrophy are equivalent. For the rest of this section, the results from EPVH-A and EPVH-C will be referred to as EPVH-A/C.

The qualitative description of predicted mixing zones is the same for EPVH-A/C and EPVH-B. If $\beta > 0.5$, the flow is stable and there are no solutions to the EPVH nonlinear equations, i.e. EPVH predicts that no PV mixing takes place in either layer and the initial jet remains unchanged when the initial flow is stable. This is because no latitudinal PV mixing can occur in a stable flow without the total momentum changing [116, 142]. As β increases towards 0.5, the Charney-Stern-Pedlosky condition for instability, the mixing zones in both layers shrink to zero with $Y_3 \to 0$ and $Y_1 \to Y_2$. The PV mixing zone in the lower layer is at its widest across the centre of the channel for low β. The critical lines y_c defined in section 5.4.6 and plotted as dotted lines in Fig. 5.4.3 lie within, or close to, the upper layer mixing regions $Y_1 < y < Y_2$ predicted by all three EPVH hypotheses. This is consistent with the idea that upper layer mixing occurs in the vicinity of the critical lines of the fastest growing linear mode [85], as discussed in section 5.4.6.

An attractive feature of all three EPVH hypotheses is that they predict their own breakdown, which may occur for two reasons. Firstly, if $Y_2 \to W/2$ or $Y_3 \to W/2$ in the upper or lower layer respectively, a mixing region extends to the channel walls which must influence the equilibrated state of the flow. If $Y_2, Y_3 \to W/2$, then this coincides with the breakdown of assumption (b) as a significant proportion of the energy remains in the wave at the end of the lifecycle. Exactly this type of behaviour was observed in the WGP flows of chapter 2 where the equilibrated state is forever transient and the wave continues to wrap up at long times (see Figs. 2.4.1 and 2.6.2). Significantly, it is the presence of the sidewalls that limits the growth of the wave and any potential energy that is released from the initial state remains as eddy energy (recall that the equilibrated eddy energy of the the WGP regime is dependent on the channel width W, e.g. Fig. 2.7.2). The sidewalls may therefore be expected to influence flows with a more uniform initial profile [34]. The second way in which EPVH predicts its own breakdown is when $Y_1 \rightarrow 0$. If $Y_1 = 0$, then the central mixing barrier in the upper layer has vanished, which violates assumption (c) of the EPVH hypotheses that mixing occurs in well-defined regions of the channel. The limit $Y_1 = 0$ for EPVH-A is plotted on Fig. 5.2.1. As the width of the mixing barrier decreases towards this limit, it may be expected to leak so that some mixing of PV may occur across $y = 0$. The line $Y_1 = \frac{1}{2}$ is plotted as a dotted line in Fig. 5.2.1 indicating where mixing might be expected to occur across the central barrier. The so-called ‘leaky barrier’ is the subject of a second Esler paper [28] and is encountered in the region of parameter space $\beta \lesssim 0.2$.

There are some differences between predictions made by the EPVH-A/C and EPVH-B hypotheses. At higher inverse criticalities ($\beta \rightarrow 0.5$), the solutions $(Y_1, Y_2, Y_3)^T$ for EPVH-A/C and EPVH-B are almost identical. At lower inverse criticalities, the EPVH-B theory predicts a much wider mixing region in the lower layer than EPVH-A/C and slightly smaller mixing regions in the upper layer. There is very little difference to report for the width of the central mixing barrier Y_1 in the upper layer over the entire range of β. For this channel width ($W = 7\pi$) and $\beta \lesssim 0.15$ (for the $\sigma = 2$ jet) or $\beta \lesssim 0.2$ (for the $\sigma = 3$ jet), the predicted lower layer mixing region for EPVH-B extends almost all the way across the channel width, whereas the mixing region for EPVH-A/C does not. A comparison of Fig. 5.4.3 to Fig. 5 of E08 (where a narrower channel is used, $W = 5\pi$) shows that there is little difference in the EPVH-A solutions $(Y_1, Y_2, Y_3)^T$ for the two different channel widths.

5.5 Direct comparison of EPVH predictions with numerical simulations of symmetric flows

E08 showed that EPVH-A successfully predicts the shape and strength of the equilibrated jet across a range of the (β, σ) parameter space in the case of an initially symmetric flow. The profiles of the zonal mean velocity profiles $\overline{u}_1$, $\overline{u}_2$ and zonal mean PV profiles $\overline{q}_1$, $\overline{q}_2$ were directly compared to numerical simulations and were shown to be very successful. The aim of this section is to make a comparison of the two alternative hypotheses EPVH-B and EPVH-C to see whether the successful predictions of E08 are the result of the particular form of the minimization problem (5.4.21) or more generally due to the correct treatment of the global constraints.

To facilitate comparison with the analytical theory, it is convenient here to introduce a diagnostic to define regions where PV has been homogenized in the equilibrated state of numerical simulations. By considering the zonal

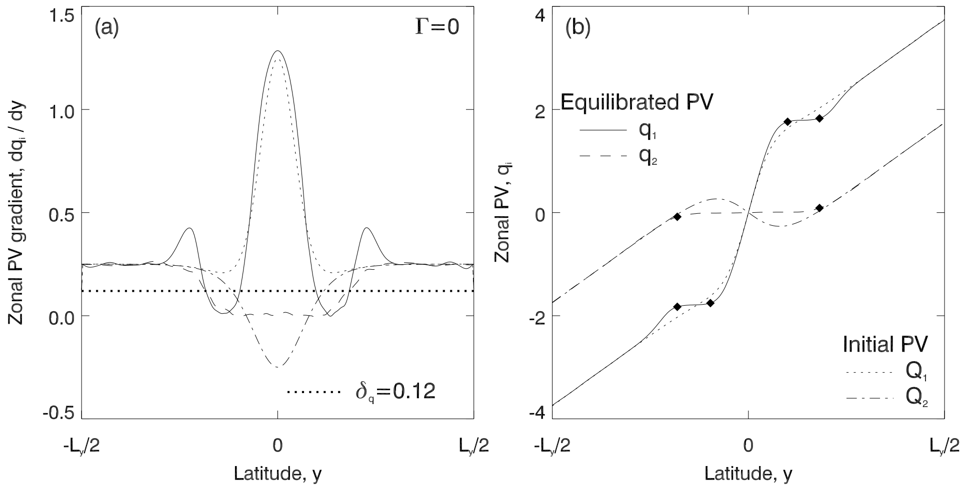

Figure 5.5.1: Panel (a) shows the latitudinal gradient of the zonal mean PV for a flow with parameters $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$. The curves are for the upper and lower layer equilibrated state (solid and dashed lines respectively) and the initial PV gradients in the upper and lower layer (dotted and dotted-dashed lines respectively). Also marked is the diagnostic value $\delta_q = 0.12$ used to define regions of PV homogenization in the equilibrated state. In panel (b) are the corresponding zonal PV profiles, with solid points marking the edges of PV mixing regions diagnosed using the definition (5.5.37).

mean PV in each layer $\overline{q}_1(y)$ and $\overline{q}_2(y)$, a latitude is said to lie within a mixed region if the gradient of the zonal mean PV is less than some threshold value

$$\left|\frac{d\overline{q}_i}{dy}\right| < \delta_q \quad \text{for} \quad i = 1, 2. \tag{5.5.37}$$

The diagnostic (5.5.37) defines two homogenized regions in the upper layer $[m_1, m_2]$, $[m_{-2}, m_{-1}]$ and one in the lower layer $[m_{-3}, m_3]$.

The specific value of δ_q is chosen to be $\delta_q = 0.12$, the relevance of which is shown in Fig. 5.5.1. Panel (a) of Fig. 5.5.1 shows the latitudinal gradient of zonal PV of the equilibrated and initial flows for the parameters $(\beta, \sigma, \Gamma) = (0.25, 2, 0)$. Wherever the PV gradient is less than $\delta_q = 0.12$, then according to the diagnostic (5.5.37), that particular latitude is defined as lying within a homogenized region. The regions defined in this way are marked by solid points in panel (b) of Fig. 5.5.1, which demonstrates that the diagnostic is a reasonable (though crude) estimate of where homogenization of PV occurs in numerical simulations.

Fig. 5.5.2 shows the available potential energy V, width W of homogenized regions and maximum jet strength in each layer $\max\{\overline{u}_i\}$ of the equilibrated states predicted by the EPVH-A/C and EPVH-B hypotheses for jets with half widths $\sigma = 2, 3$. The results from numerical simulations are also plotted as solid points. Panels (a) and (b) show the available potential energy calculated from (1.3.82) compared to the initial available potential energy in the flow (dotted line). More potential energy is released where the flow is furthest from the stability threshold $\beta = 0.5$. The predictions by all three EPVH hypotheses are identical at the highest inverse criticalities ($\beta \to 0.5$) as anticipated from the similarity of the latitude solutions in Fig. 5.4.3. When the flow is close to the stability threshold very little PV is mixed and subsequently the change in the jet is very small. The EPVH-A/C hypotheses capture the potential energy of the numerical simulations more accurately than EPVH-B.

The similarity of the hypotheses' predictions at high inverse criticalities is also evident in panels (c-d) of Fig. 5.5.2, which show the total width of the mixing regions in the channel calculated from (5.4.25). Fig. 5.4.3 shows that the EPVH-B solutions predict a wider mixing region in the lower layer and slightly narrower regions in the upper layer than EPVH-A/C. The prediction by EPVH-B of wider mixing regions is associated with less potential energy released during the lifecycle and therefore a weaker eastward jet acceleration in each layer. Both hypotheses over-estimate the total width of PV homoge-

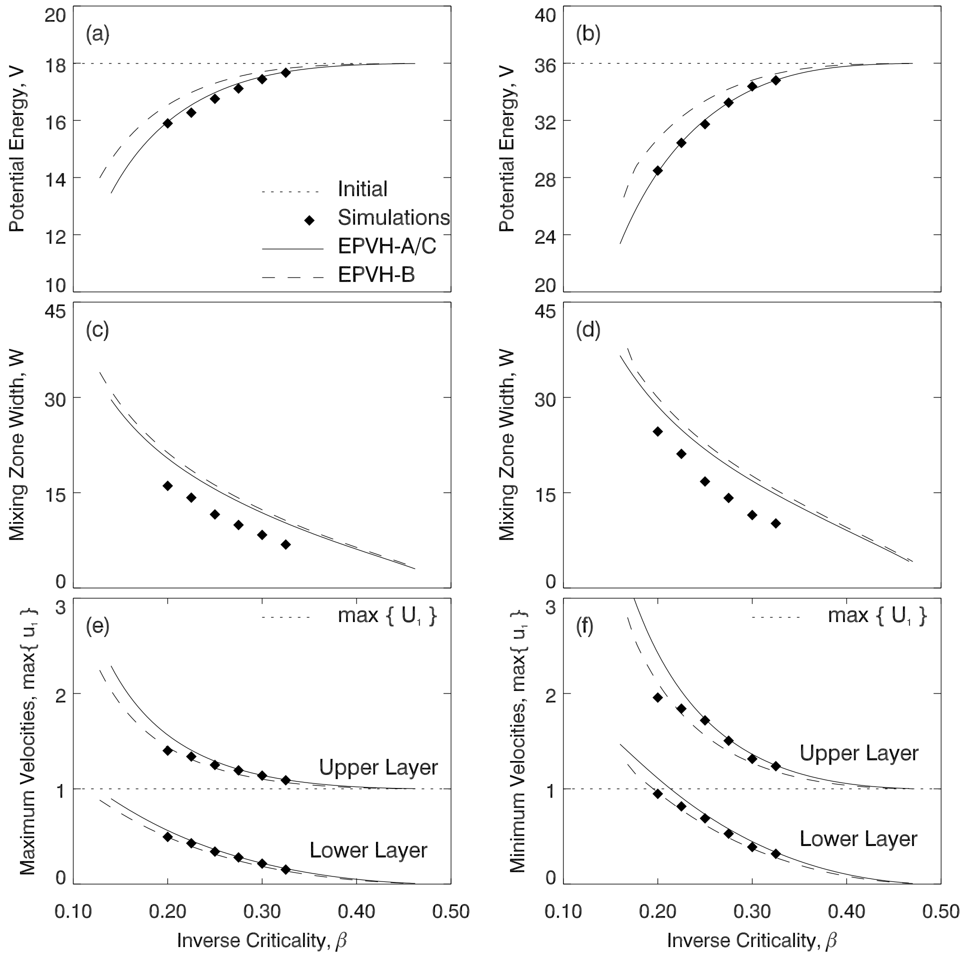

Figure 5.5.2: Available potential energy V, total mixing zone width W and maximum velocities in each layer $\max\{\overline{u}_1\}$ and $\max\{\overline{u}_2\}$ for equilibrated jets. Predictions using EPVH-A/C are solid lines and using EPVH-B are dashed lines. The relevant quantities calculated from numerical simulations are solid points (the widths in panels c-d are calculated using the diagnostic 5.5.37). The left-hand panels (a,c,e) are for initial jets with half width $\sigma = 2$ and the right-hand panels (b,d,f) are for initial jets with half-width $\sigma = 3$.

nization, however the decrease in the total mixing width as inverse criticality increases is qualitatively captured. Panels (e-f) of Fig. 5.5.2 show that the strengths of the equilibrated jets in each layer is bounded below by EPVH-B and above by the EPVH-A/C hypothesis. The variation in the equilibrated jet strength within the (β, σ) parameter space and the shape of individual jet profiles is discussed at greater length below.

Fig. 5.5.3 shows profiles of the zonal mean wind and PV of the equilibrated jet (at time $t = 400$) in each layer for $(\beta, \sigma) = (0.25, 2)$. Panel (a) of Fig. 5.5.3 shows the results of numerical simulations at the three resolutions described in Table 5.1. Panel (a) establishes that the numerical experiments are independent of the artificial diffusivity ν_4 and that the zonal mean velocity converges with decreasing ν_4. There is little difference between the MR and HR resolutions. The LR jet is the weakest, which is consistent with a greater loss of energy due to higher dissipation over the course of the lifecycle as shown in Fig. 5.3.2.

Panels (b-f) of Fig. 5.5.3 show a direct comparison of the predictions by each of the EPVH-A/C and EPVH-B hypotheses with high resolution (HR) numerical simulations. Panel (b) shows the equilibrated zonal wind profile of the HR simulation (solid lines) and zonal wind predictions for EPVH-A/C (dashed lines) and EPVH-B (dashed-dotted lines). The initial jet profiles U_1 and U_2 are also plotted (dotted line). The EPVH-A/C predictions capture both the structure and magnitude of the final jet in each layer. There is not agreement in the weaker negative (westward) winds to the flanks of the main jet in the upper layer. In these regions, EPVH-A/C over-predicts the

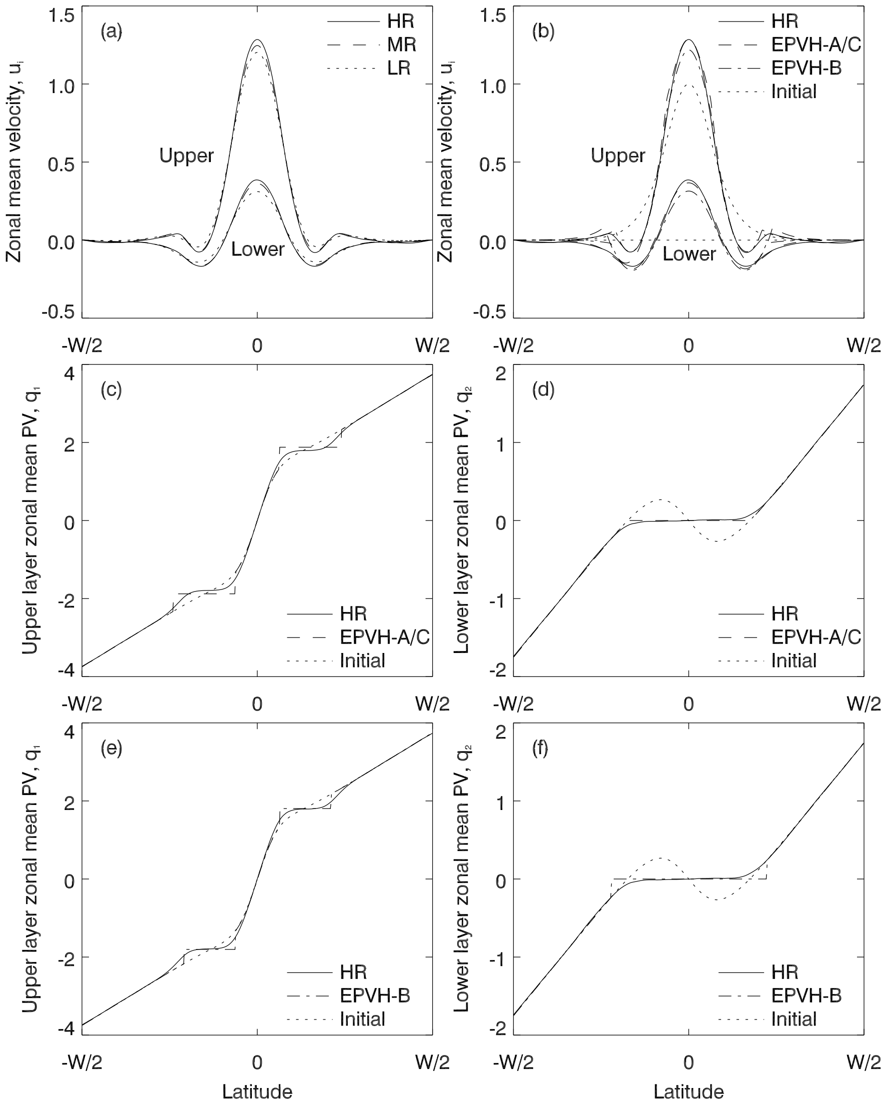

Figure 5.5.3: Zonal velocity and potential vorticity profiles of the equilibrated jet. Numerical simulations are compared to predictions made by EPVH-A/C and EPVH-B. Panel (a) shows the equilibrated zonal wind of the upper and lower layer for the three resolutions described in Table 5.1. Panel (b) shows a comparison of the HR simulation zonal wind to predictions made with EPVH-A/C (dashed line) and EPVH-B (dashed-dotted line). Panels (c) and (d) show a comparison of the zonal PV field predicted by EPVH-A/C to numerical simulations in the upper and lower layer respectively. Panels (e) and (f) are the same as (c) and (d) but show predictions using the EPVH-B hypothesis.

strength of the westward flow. The success of the predictions may be explained by consideration of the corresponding zonal PV profiles $\overline{q}_1$ and $\overline{q}_2$ in panels (c) and (d). EPVH-A/C captures the regions of homogenized PV very well, especially in the lower layer (panel d) and slightly over-predicts the width of the homogenized regions in the upper layer.

Panel (b) shows that the equilibrated zonal wind is also well-captured by the EPVH-B hypothesis. EPVH-B under-predicts the strength of the main jet in each layer (Fig. 5.5.3, panels e-f) and is inferior to EPVH-A/C in this respect. However, EPVH-B is more accurate in predicting the strength and shape of the weaker westward jets in the upper layer. The success of EPVH-B may also be explained by looking at the zonal mean PV in each layer in panels (e) and (f). This shows that EPVH-B over-predicts the size of the homogenized region in the lower layer, but is more accurate than EPVH-A/C at predicting the upper layer mixing regions. It is perhaps the more accurate description of the upper layer mixing regions, which leads to the success of EPVH-B in capturing the weaker westward jets.

How do the EPVH hypotheses perform in the rest of the parameter space? Fig. 5.5.4 shows the equilibrated zonal winds for two initial jets with $(\beta, \sigma) = (0.3, 3)$ (panel a) and $(\beta, \sigma) = (0.2, 2)$ (panel b). For $(\beta, \sigma) = (0.3, 3)$, the initial jet is broader than the $(\beta, \sigma) = (0.25, 2)$ jet and the instability is relatively weak. The mixing regions in the upper layer are further from the jet core and the resulting equilibrated jet is wide with weak westward jets on each flank. The shape of the jet is well-captured by both hypotheses, again EPVH-A/C over-predicts and EPVH-B under-predicts the strength of the jet in each layer

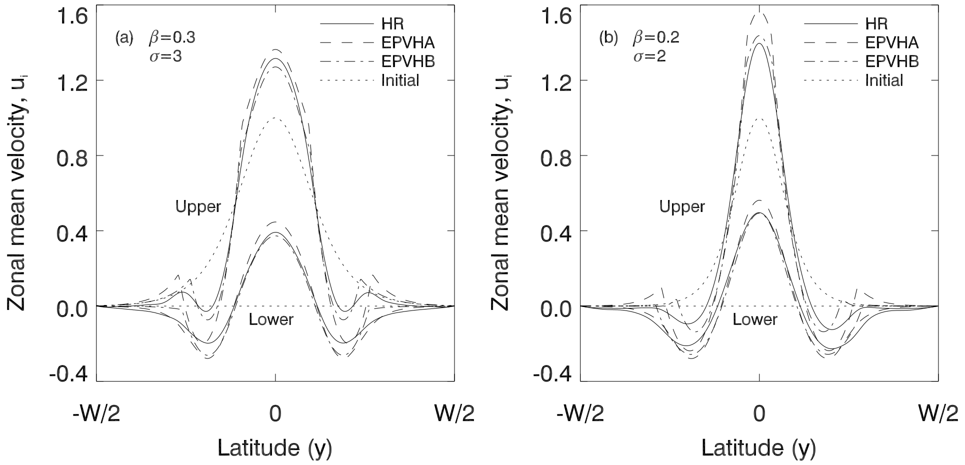

Figure 5.5.4: Zonal wind profiles $\overline{u}_1$ and $\overline{u}_2$ of the equilibrated jets with initial profiles (a) $\beta = 0.3$, $\sigma = 3$ and (b) $\beta = 0.2$, $\sigma = 2$. Numerical simulations (solid lines) are plotted together with predictions made by the EPVH-A/C (dashed lines) and EPVH-B hypotheses (dash-dotted lines) and the initial jet profiles (dotted lines). Both the upper and lower layer profiles are illustrated as labelled.

and EPVH-A/C is less accurate in capturing the westward flows. The $(\beta, \sigma) = (0.2, 2)$ jet is initially thin and equilibrates in a triangular shape. EPVH-A/C greatly over-predicts the magnitude of the jets and this is attributed to the effect of PV mixing across the jet core as $Y_1 \lesssim 1$. EPVH-B is slightly more successful at capturing the shape of the jet at these low inverse criticalites.

A further evaluation of the success of each hypothesis is made by comparing the maximum jet strengths in the upper and lower layers ($\max\{\overline{u}_1\}$ and $\max\{\overline{u}_2\}$ respectively). Fig. 5.5.5 compares the predictions of these jet strengths by each hypothesis to the results of numerical experiments. Panel (a) shows that the EPVH-A/C hypotheses generally over-predicts the upper layer jet magnitude and EPVH-B under-predicts the jet strength throughout the parameter space. Both hypotheses are effective at capturing the lower layer jet magnitude. The worst results for EPVH-A/C are for the $\beta = 0.2$ and

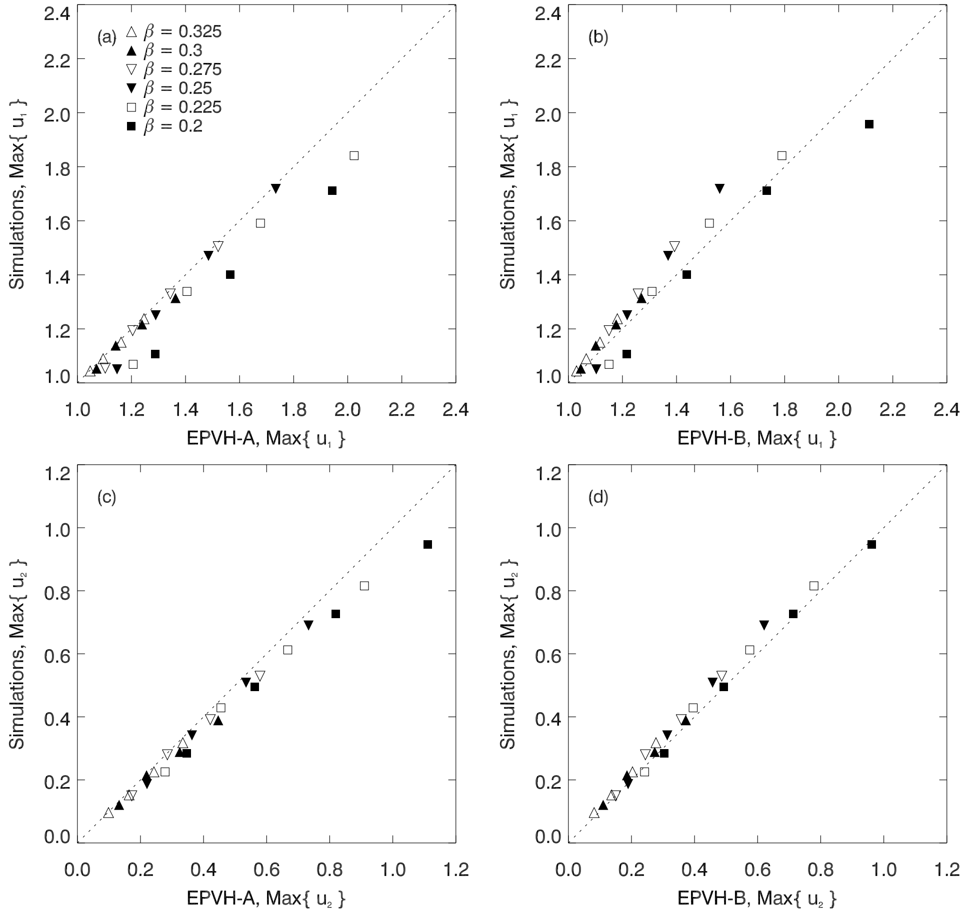

Figure 5.5.5: A comparison of maximum jet strength in each layer $\max\{\overline{u}_1\}$ and $\max\{\overline{u}_2\}$. In each panel, the jet strength taken from numerical simulations is on the vertical axis and predicted jet strengths are on the horizontal axis Panels (a) and (b) show results in the upper layer for EPVH-A/C and EPVH-B respectively. Panels (c) and (d) compare the corresponding results for the lower layer.

$\beta = 0.225$ jets. It is for these flows that the upper layer mixing barrier, lying within $-Y_1 < y < Y_1$, becomes too thin (smaller than the deformation radius $Y_1 \lesssim 1$) to maintain its integrity. Therefore a small amount of mixing across the barrier and the propagation of small-amplitude waves prevents the theory being accurate. These jets are also poorly captured by EPVH-B, although the tendency of this hypothesis to under-predict the jet strength means that the predictions are closer to the simulations than EPVH-A/C.

Figs. 5.5.2 and 5.5.5 provides evidence that both hypotheses reasonably capture the jet strength over a wide part of the parameter space. Fig. 5.5.2 in particular indicates that EPVH-A/C is more accurate at capturing properties of the flow such as available potential energy and also the jet strengths in each layer. In the majority of flows the jet strength is bounded above by EPVH-A/C and below by EPVH-B. The westward flow is poorly captured, though some of this error may be explained by the discontinuity in the predicted PV profiles given by (5.4.17), which introduces a sharp PV gradient at the edges of the mixing regions. This is not present in the smooth profiles of the numerical simulations, e.g. panels (c-f) of Fig. 5.5.3. Upon inversion, the discontinuity in the PV necessarily introduces errors in the predicted zonal wind profile.

The fact that any of the hypotheses introduced, EPVH-A, EPVH-B or EPVH-C, provide reasonably accurate predictions of the jet structure over the majority of the parameter space indicates that it is the correct treatment of global constraints in the problem and not the choice of variational principle that is most important in determining the equilibrated state. This will be further tested in the next sections by considering flows in the presence of a background linear shear on top of the jet profile.

5.6 Assessment of EPVH for asymmetric jets

In this section the predictions for the equilibrated flow of an asymmetric jet ($\Gamma \neq 0$) are presented for each of the hypotheses EPVH-A, EPVH-B and EPVH-C. The predictions made by the EPVH hypotheses are also compared to numerical simulations for flows in the parameter space $\Gamma \in [0, 0.15]$ with $(\beta, \sigma) = (0.25, 2)$. Fig. 5.6.1 shows the latitude solutions $(Y_1, Y_2, Y_3, Y_{-1}, Y_{-2}, Y_{-3})^T$ predicted by the three EPVH hypotheses and estimates of the latitudes delineating the mixing zones in numerical simulations (using the diagnostic 5.5.37). Although the results presented here are only for the $(\beta, \sigma) = (0.25, 2)$ jet, a similar qualitative behaviour is seen for flows in the rest of the (β, σ, Γ) parameter space. The points plotted in Fig. 5.6.1 show the latitudes m_i defined by (5.5.37) for each numerical simulation. The points are separated into the two types of mixing as set out in Fig. 5.3.6. Solid points are for simulations with a 'weak shear', where homogenization of PV occurs on both flanks of the jet in the upper layer. Unfilled points are for simulations with a 'strong shear' where homogenization of PV occurs only on the southern flank of the upper layer jet. As Γ increases over the parameter range, both the southern and northern mixing regions migrate northwards before the northern region disappears. The lower layer mixing region maintains an approximately constant width.

The latitude solutions $(Y_1, Y_2, Y_3, Y_{-1}, Y_{-2}, Y_{-3})^T$ plotted in Fig. 5.6.1 are for each of the hypotheses EPVH-A (equation 5.4.21, dashed and dotted lines),

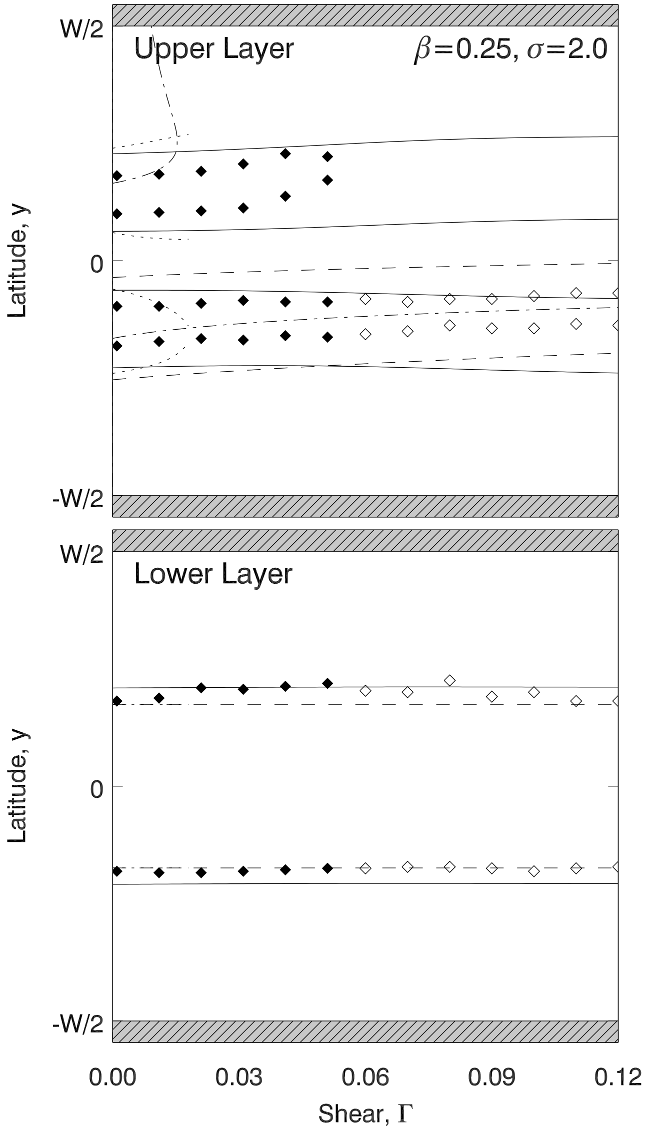

Figure 5.6.1: Latitude solutions $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_{-1}, Y_{-2}, Y_{-3})^T$ predicted by the three EPVH hypotheses for $\Gamma \in [0, 0.15]$ and $(\beta, \sigma) = (0.25, 2)$. Predictions by EPVH-A are plotted for two upper layer mixing regions (dotted lines) and one upper layer mixing region (dashed lines, i.e. set $Y_{-1} = Y_{-2}$, see section 5.4.3). Predictions by EPVH-B are plotted as solid lines and predictions by EPVH-C coincide with the one mixing region solution of EPVH-A (dashed lines). The dashed-dotted line shows the latitude y_c of the upper layer critical line of the fastest growing linear mode as calculated in section 5.4.6.

EPVH-B (equation 5.4.27, solidd line) and EPVH-C (5.4.29, dashed line). In each case, predictions can be found with either two mixing regions in the upper layer (as in sections 5.4.2, 5.4.4 and 5.4.5), or by setting $Y_1 = Y_2$, with one mixing region in the upper layer as in section 5.4.3.

(A) In Fig. 5.6.1, two solutions are plotted for EPVH-A, one with two mixing regions in the upper layer (dotted lines) and another with a single mixing region in the upper layer (dashed line). The two mixing layer regime predicts solutions in the range $\Gamma \in [0, 0.018]$, but as the shear approaches the threshold value $\Gamma \lesssim 0.018$, the southern mixing zone in the upper layer $[Y_{-2}, Y_{-1}]$ decreases in size and then disappears. The single mixing zone solution predicts one mixing region to the southern flank of the jet over the whole parameter range. The solution that is relevant, according to the EPVH-A hypothesis, is that which minimizes the available potential energy. Panel (a) of Fig. 5.6.2 shows that the single mixing region solution is the relevant one over the whole parameter space.

(B) There are also two solutions for EPVH-B, however in this case, according to the hypothesis of section 5.4.4, the relevant solution is the one that maximizes the total width of the mixing regions. This is the only EPVH-B solution plotted in Fig. 5.6.1 (solid lines) and it has two upper layer mixing regions.

(C) The EPVH-C equations also allow two types of solution, either two upper layer mixing regions or one upper layer mixing region. In Fig. 5.6.1 the solution plotted (dashed line) has one mixing region in the upper layer

> as this is the global minimum of the total enstrophy. The EPVH-A and EPVH-C solutions, global minimums of potential energy and enstrophy respectively, give precisely the same predictions in Fig. 5.6.1.

The dashed-dotted lines in Fig. 5.6.1 are the critical lines calculated from linear theory in section 5.4.6. The critical lines are located at the latitudes where linear theory breaks down, $U_1(y) - c_r = 0$. For weak shears ($\Gamma \leq 0.0153$) there are three critical lines, two to the northern flank and one on the southern flank of the jet. The northern critical lines (dashed-dotted curve) in Fig. 5.6.1 coalesce at $\Gamma = 0.153$, preceding the disappearance of turbulent mixing on that flank as indicated by the solid points. The southern critical line coincides with the mixing region on the southern flank of the jet over the whole parameter range $\Gamma \in [0, 0.15]$.

All three hypotheses predict the lower layer PV mixing region relatively well (bottom panel of Fig. 5.6.1).EPVH-B predicts that the lower layer mixing region is slightly wider than that predicted by EPVH-A/C (the same difference is seen for symmetric jets). In the upper layer, the latitudes predicted by all three of the EPVH hypotheses are inaccurate. None of the three hypotheses correctly predicts the disappearance of the northern mixing region observed in numerical simulations. For the 'weak shear' solutions, there is also little agreement with the northward migration of both mixing regions. EPVH-A/C predict mixing on the correct flank of the jet at large shears and also the movement of the mixing region towards the jet core as the shear gets stronger.

EPVH predictions of the total available potential energy, enstrophy and maximum jet strength in each layer are compared to numerical simulations

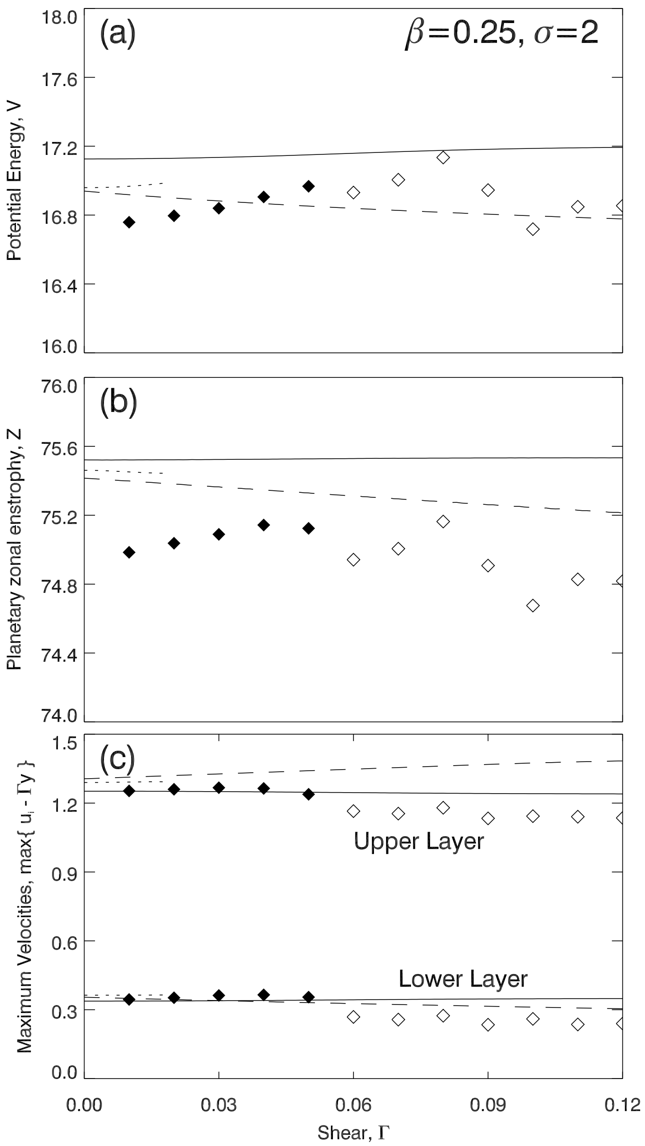

Figure 5.6.2: Total available potential energy V, total enstrophy Z and maximum jet strength in the upper and lower layers $\max\{u_i\}$ for $i = 1, 2$. Predictions made by the EPVH-A and EPVH-C hypothesis are identical, plotted with a dashed line. An alternative (not minimum) potential energy solution for EPVH-A is also plotted (dotted lines). Predictions using the EPVH-B hypothesis are solid lines. All curves correspond to the latitude solutions of Fig. 5.6.1. The physical quantities calculated from the equilibrated state of numerical simulations are plotted as filled (‘weak’ shear) and unfilled (‘strong’ shear) points.

in Fig. 5.6.2. Each curve corresponds to a solution in Fig. 5.6.1. Two curves are plotted for EPVH-A, dotted (two upper layer mixing zones) and dashed (one upper layer mixing zone). As mentioned above, panel (a) of Fig. 5.6.2 shows that the latter solution for EPVH-A is the global minimum of available potential energy and therefore the relevant prediction under the EPVH-A hypothesis.

Accepting that none of the EPVH hypotheses successfully predicts the qualitative change in behaviour of the flow when the northern mixing region in the upper layer disappears, the EPVH predictions can be tested to see if they are successful in any subdomain of the parameter space. The EPVH-A/C hypotheses predict a single mixing region in the upper layer and so are compared to the 'strong shear' numerical simulations (unfilled points). The EPVH-B hypothesis predicts two mixing regions in the upper layer and is compared to the 'weak shear' numerical simulations where two mixing regions are observed. Panels (a) and (b) show the available potential energy and enstrophy respectively, which the EPVH-A/C and EPVH-B hypotheses predict reasonably closely. However, slight changes in the potential energy and enstrophy observed in numerical simulations are not captured. Panel (c) of Fig. 5.6.2 shows the maximum upper and lower layer jet strengths. The 'weak shear' jet strengths in each layer are very close to the predictions of EPVH-B. EPVH-A/C greatly over-predicts the upper layer jet strengths at 'strong shears' (unfilled points).

Jet profiles are shown in Fig. 5.6.3 for $\Gamma = 0.02, 0.04, 0.08, 0.12$ with $(\beta, \sigma) = (0.25, 2.0)$, which includes numerical simulations (solid curves) and

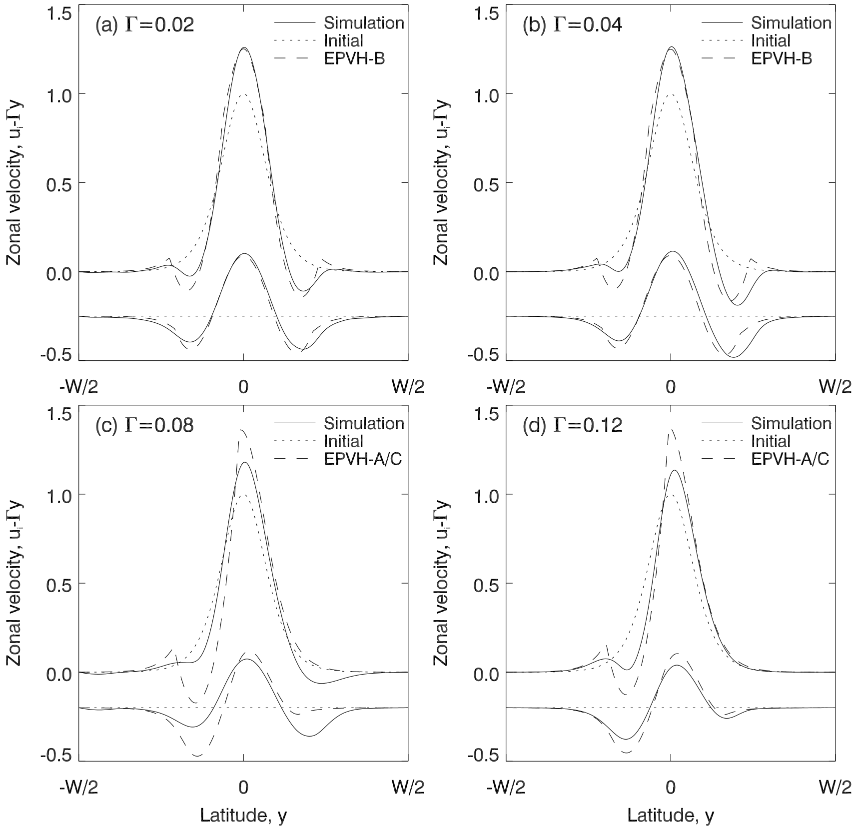

Figure 5.6.3: Zonal mean velocity profiles in each layer as predicted by EPVH hypotheses (broken lines) and from numerical simulations (solid lines). The background shear ($= \Gamma y$) has been taken away from the velocity profiles to allow for an easier comparison between each panel. Panels show four different shear profiles: (a) $\Gamma = 0.02$, (b) $\Gamma = 0.04$, (c) $\Gamma = 0.08$ and (d) $\Gamma = 0.12$. The broken curves correspond to the EPVH hypotheses as marked in the figure and the dotted lines show the initial jet profiles.

EPVH predictions (broken curves, as labelled). The velocity profiles are plotted with the background shear removed $\overline{u}_i - \Gamma y$ to allow an easy comparison between flows with different shear strengths. Panels (a) and (b) of Fig. 5.6.3 show the equilibrated velocity profile for 'weak' shears' $\Gamma = 0.02, 0.04$ where the mixing region on the northern flank of the jet is wider than that on the southern flank and the westward flow is stronger as a result. Also plotted are the profiles predicted by EPVH-B. The EPVH-B hypothesis captures the shape of the jet well and accurately predicts the strength of the main eastward jet and the difference in strength of the westward flow on each flank.

Panels (c) and (d) show velocity profiles for $\Gamma = 0.08$ and $\Gamma = 0.12$ respectively, which from Fig. 5.6.1 are flows where PV homogenization only occurs on the southern flank of the jet. The shape of the jet for $\Gamma = 0.08$ shows that EPVH-A/C greatly over-predicts the strength of the upper layer jet and does not capture the shape of the jet in either layer: in simulations there is a stronger westward flow to the northern flank of the jet than that on the southern flank, which EPVH-A/C fails to predict. EPVH-A/C captures the shape the $\Gamma = 0.12$ jet in panel (d) slightly better, though the strength of the central eastward jet is still greatly over-predicted.

It might be argued that the EPVH-B hypothesis is reasonably successful for the weakest shears (panels a and b of Fig. 5.6.3). However, consideration of the latitude solutions in Fig. 5.6.1 shows that this hypothesis does not capture the trend of the mixing regions, i.e. it fails to predict the disappearance of the northern mixing region or the slight increase with shear of potential energy and enstrophy displayed in Fig. 5.6.2. It is difficult to argue therefore that the accuracy of the jet shapes in panels (a-b) of Fig. 5.6.3 is due to the success of EPVH-B in capturing the features of the equilibrated flow, but rather it is due to the flow being sufficiently close to the symmetric profile $\Gamma = 0$ where all three EPVH hypotheses have been shown to be accurate.

5.7 Conclusions

Most theories of 'adjustment' for unstable baroclinic jets focus on changes to the mean flow forced by eddies, which act to stabilize the jet to further linear normal mode disturbances. A major problem with such theories is that they do not explain the formation of upgradient momentum fluxes and subsequent acceleration of the jet [119, 57]. Nakamura [84, 85] attempted to adapt adjustment theories to allow for the upgradient momentum fluxes, however the resulting theories were not able to make definite predictions. A theory by Esler [29] (E08) has been reviewed, which makes definite predictions for the equilibrated flow of an initially unstable symmetric jet in the two-layer, quasi-geostrophic model. The theory, named 'Equilibration via Potential Vorticity Homogenization' (EPVH-A, here) accounts for the upgradient momentum fluxes, latitudinal heat fluxes and the shape of the final jet over a range of the parameter space.

It was reported in E08 that the EPVH-A theory is successful partly due to a correct treatment on the global dynamical constraints of total momentum and energy, but also due to the two assumptions that (i) PV homogenization occurs within well delineated regions, and (ii), upon completion of the baroclinic lifecycle, the available potential energy is minimized.

The first assumption, that of PV homogenization is widely supported by observations and numerical simulations in the geophysical fluid dynamics literature (e.g. Rhines [101, 102], Dritschel & McIntyre [23]). The idea that homogenization of PV occurs in well-defined regions is observed in numerical simulations of the flow and can be thought of in terms of 'Rossby-wave elasticity' [80]: in regions where latitudinal PV gradients are strong, Rossby wave-like motions predominate, whereas in areas where the PV gradient is weaker, turbulent motions can overcome the PV gradient and mixing can occur. The constraint that PV mixing occurs in this way is then a kinematic constraint on the development of the flow.

The second assumption is that available potential energy (APE) is minimized and is argued to be a requirement that the instability proceeds to completion. The flow is then stable to further disturbances. It is this second assumption that has been investigated in the first part of this chapter as two further assumptions have been suggested as a requirement that the instability has proceeded to completion.

The first, EPVH-B, is that the area of the regions of PV homogenization is maximized over both layers of channel in the final state of the jet. This is a consequence of Rossby-wave elasticity [80] and can also be related to maximum mixing entropy principles in statistical mechanics. It was demonstrated that EPVH-B also accurately captures the magnitude and shape of the final jet, though tends to predict narrower mixing regions in the upper layer and wider mixing regions in the lower layer. This results in more accurate predictions for the weaker westward jets that form on the flanks of the main eastward jet, but EPVH-B under-predicts the main jet strength where it is slightly less accurate than EPVH-A.

The third hypothesis, labelled EPVH-C, is that the zonal enstrophy is minimized in the final jet state. The minimum enstrophy state results from the cascade of enstrophy to small scales in two-dimensional turbulence where it is removed by viscous effects [16, 69]. Consequently, the system evolves to a state of minimum zonal enstrophy subject to the conservation of energy, which decays on time scales much longer than enstrophy. Curiously, EPVH-C makes precisely the same predictions as EPVH-A, though it is not obvious why. The fact that each of the hypotheses EPVH-A, EPVH-B and EPVH-C capture the final jet state relatively well suggests that the correct treatment of the global constraints is important in this problem. As pointed out in E08, maximum entropy theories [82, 105] also treat conservation of total energy and momentum correctly, but predict final states which are far less accurate than EPVH. It is inferred, therefore, that the assumption that PV is completely homogenized within well de-lineated regions is important in constraining the predictions.

In the second part of this chapter, the restriction that the initial jet is symmetric is relaxed by introducing a linear, barotropic shear on top of the basic jet profile. The asymmetric lifecycles provide a further test of EPVH to assess the performance of the three different hypotheses more comprehensively. Attention was focused to jets with parameters $(\beta, \sigma) = (0.25, 2)$ and the barotropic shear was parameterized by $\Gamma \in [0, 0.12]$. The same qualitative behaviour is seen in other parts of the parameter space. As the strength of the linear background shear is increased, numerical simulations show that the region of PV mixing on the northern flank of the upper layer jet vanishes (at $\Gamma \approx 0.05$. The disappearance of the mixing region is consistent with the disappearance of the two northern critical lines of the flow although this occurs at a weaker shear ($\Gamma \approx 0.0153$). None of the three EPVH hypotheses successfully predict the qualitative change in the mixing region behaviour emphasized by Fig. 5.6.1.

EPVH-B (maximization of mixing area) predicts two mixing regions over the entire parameter space. For ‘weak’ shears where this behaviour is observed, the hypotheses are fairly successful at predicting the maximum jet strengths and also the sense of the westward flow at the flanks of the main jet. However, consideration of physical quantities such as available potential energy and enstrophy suggest that the success is due to the flow profile being close to symmetric rather than EPVH accurately capturing the properties of the flow. For ‘strong’ shears, the one upper layer mixing region predicted by EPVH-A and EPVH-C largely over-predict the strength of the jet, which seems to be a consequence of over-estimation of the PV mixing region in the upper layer.

It was asserted that the success of EPVH is due to the correct treatment of the global constraints and the diagnosis of de-lineated regions where PV is completely homogenized. The introduction of a horizontal shear to the background flow fundamentally alters where PV is homogenized, which EPVH is unable to predict. The key aspect of the problem it would seem is to ascertain where wave breaking and subsequent PV mixing occurs within the channel. The linear theory correctly predicts that closed streamfunction contours (associated with turbulent flow) disappear on the correct flank of the jet. It may be possible to construct a ‘quasi-linear’ variational principle by varying the phase speed and amplitude of the fastest growing linear mode subject to momentum and energy conservation. The regions of PV homogenization in the equilibrated flow would then be predicted by the latitudinal extent of the closed streamfunction contours. This is an avenue for future research.

5.A Formulation of variational problems

5.A.1 EPVH-A: Minimization of available potential energy

When a background shear is added to the basic flow, the streamfunction may be written in terms of the symmetric part ψ_i^s and the additional barotropic shear $\psi_i = \psi_i^s - \frac{1}{2}\Gamma y^2$. The potential vorticity gradients in each layer dq_1/dy and dq_2/dy are unchanged by the addition of the background shear. The total energy of the system is then

$$\begin{aligned} E &= \frac{1}{2}\int_D |\nabla\left(\psi_1^s - \frac{1}{2}\Gamma y^2\right)|^2 + |\nabla\left(\psi_2^s - \frac{1}{2}\Gamma y^2\right)|^2 + \frac{1}{2}\left(\psi_1^s - \psi_2^s\right)^2 d^2\mathbf{x}, \\ &= \frac{\Gamma W^3 L}{24} + \frac{\Gamma}{2}\int_D y^2\left(q_1 + q_2\right) d^2\mathbf{x} - \frac{1}{2}\int_D \psi_1^s q_1 + \psi_2^s q_2 d^2\mathbf{x}. \qquad (5.A.38) \end{aligned}$$

where the second step is done via an integration by parts. The momentum M and potential energy $\overline{V}$ are unchanged by the background shear.

Using (5.4.20) and differentiating with respect to y_1, y_2 etc. in turn, gives

$$\begin{aligned}
\frac{\partial F}{\partial y_1} &= \left(\overline{Q}_1^{(1,2)} - Q_1(y_1)\right)\left(-\left(\phi(y_1) - \overline{\phi}^{(1,2)}\right) + 2\lambda\left(\psi_1(y_1) - \overline{\psi}_1^{(1,2)}\right)\right. \\
&\quad \left. + \mu\left(y_2 - y_1\right)\right) + \frac{\Gamma}{6}\left(\overline{Q}_1^{(1,2)} - Q_1(y_1)\right)\left(2y_1 + y_2\right)\left(y_2 - y_1\right), \\
\frac{\partial F}{\partial y_2} &= \left(\overline{Q}_1^{(1,2)} - Q_1(y_2)\right)\left(\left(\phi(y_2) - \overline{\phi}^{(1,2)}\right) - 2\lambda\left(\psi_1(y_2) - \overline{\psi}_1^{(1,2)}\right)\right. \\
&\quad \left. + \mu\left(y_2 - y_1\right)\right) + \frac{\Gamma}{6}\left(\overline{Q}_1^{(1,2)} - Q_1(y_2)\right)\left(2y_2 + y_1\right)\left(y_2 - y_1\right), \\
\frac{\partial F}{\partial y_3} &= \left(\overline{Q}_2^{(-3,3)} - Q_2(y_3)\right)\left(-\left(\phi(y_3) - \overline{\phi}^{(-3,3)}\right) + 2\lambda\left(\psi_2(y_3) - \overline{\psi}_2^{(-3,3)}\right)\right. \\
&\quad \left. + \mu\left(y_3 - y_{-3}\right)\right) + \frac{\Gamma}{6}\left(\overline{Q}_2^{(-3,3)} - Q_2(y_3)\right)\left(2y_3 + y_{-3}\right)\left(y_3 - y_{-3}\right), \\
\frac{\partial F}{\partial y_{-2}} &= \left(\overline{Q}_1^{(-2,-1)} - Q_1(y_{-2})\right)\left(-\left(\phi(y_{-2}) - \overline{\phi}^{(-2,-1)}\right) + 2\lambda\left(\psi_1(y_{-2}) - \overline{\psi}_1^{(-2,-1)}\right)\right. \\
&\quad \left. + \mu\left(y_{-1} - y_{-2}\right)\right) + \frac{\Gamma}{6}\left(\overline{Q}_1^{(-2,-1)} - Q_1(y_{-1})\right)\left(2y_{-1} + y_{-2}\right)\left(y_{-1} - y_{-2}\right), \\
\frac{\partial F}{\partial y_{-1}} &= \left(\overline{Q}_1^{(-2,-1)} - Q(y_{-1})\right)\left(\left(\phi(y_{-1}) - \overline{\phi}^{(-2,-1)}\right) - 2\lambda\left(\psi_1(y_{-1}) - \overline{\psi}_1^{(-2,-1)}\right)\right. \\
&\quad \left. + \mu\left(y_{-1} - y_{-2}\right)\right) + \frac{\Gamma}{6}\left(\overline{Q}_1^{(-2,-1)} - Q_1(y_{-2})\right)\left(y_{-1} + 2y_{-2}\right)\left(y_{-1} - y_{-2}\right), \\
\frac{\partial F}{\partial y_{-3}} &= \left(\overline{Q}_2^{(-3,3)} - Q_2(y_{-3})\right)\left(\left(\phi(y_{-3}) - \overline{\phi}^{(-3,3)}\right) - 2\lambda\left(\psi_2(y_{-3}) - \overline{\psi}_2^{(-3,3)}\right)\right. \\
&\quad \left. + \mu\left(y_3 - y_{-3}\right)\right) + \frac{\Gamma}{6}\left(\overline{Q}_2^{(-3,3)} - Q_2(y_{-3})\right)\left(y_3 + 2y_{-3}\right)\left(y_3 - y_{-3}\right),
\end{aligned} \tag{5.A.39}$$

where ϕ is defined by

$$\phi_{yy} - \phi = \psi_1 - \psi_2, \quad \phi_y = 0, \quad \text{on} \quad y = \pm\frac{W}{2} \tag{5.A.40}$$

and $Q_i^{(j,k)}$, $\overline{\psi}_i^{(j,k)}$ and $\overline{\phi}^{(j,k)}$ are defined by (5.4.18), repeated here for convenience.

$$\overline{f}_i^{(j,k)} = \frac{1}{y_k - y_j}\int_{y_j}^{y_k} f_i(y)dy,$$

for a general function $f_i(y)$.

5.A.2 PV inversion with a background shear

The zonal mean PV inversion problem can be recast through the introduction of Dirac delta functions in the potential vorticity to restore the original boundary conditions of the symmetric ($\Gamma = 0$) problem as follows:

$$\begin{aligned} \frac{\partial \psi_i}{\partial x} &= 0, && (5.A.41) \\ \frac{\partial \overline{\psi}_i}{\partial y} &= 0 \quad \text{on} \quad y = \pm\frac{W}{2} \quad i = 1, 2, && (5.A.42) \end{aligned}$$

with

$$q_i \equiv \beta y + \frac{\Gamma W}{2}\left(\delta(y - z^+) - \delta(y - z^-)\right) + \nabla^2\psi_i + (-1)^i\left(\frac{\psi_1 - \psi_2}{2}\right). \quad (5.A.43)$$

Taking the limits $z^+ \to W/2$ and $z^- \to -W/2$ and replacing (5.2.3) and (5.2.4) with (5.A.41) and (5.A.42) allows the results from E08 to be applied directly.

5.A.3 EPVH-B: Maximization of mixing zone width

Differentiation of (5.4.26) with respect to y_1, y_2 etc. yields

$$\begin{aligned} \frac{\partial G}{\partial y_1} &= -1 + \lambda\frac{\partial \overline{E}}{\partial y_1} + \mu\frac{\partial M}{\partial y_1}, \\ \frac{\partial G}{\partial y_2} &= 1 + \lambda\frac{\partial \overline{E}}{\partial y_2} + \mu\frac{\partial M}{\partial y_2}, \\ \frac{\partial G}{\partial y_3} &= 1 + \lambda\frac{\partial \overline{E}}{\partial y_3} + \mu\frac{\partial M}{\partial y_3}, \\ \frac{\partial G}{\partial y_{-1}} &= -1 + \lambda\frac{\partial \overline{E}}{\partial y_{-1}} + \mu\frac{\partial M}{\partial y_{-1}}, \\ \frac{\partial G}{\partial y_{-2}} &= 1 + \lambda\frac{\partial \overline{E}}{\partial y_{-2}} + \mu\frac{\partial M}{\partial y_{-2}}, \\ \frac{\partial G}{\partial y_{-3}} &= -1 + \lambda\frac{\partial \overline{E}}{\partial y_{-3}} + \mu\frac{\partial M}{\partial y_{-3}}, && (5.A.44) \end{aligned}$$

where $\partial\overline{E}/\partial y_i$ and $\partial M/\partial y_i$ for $i = \pm1, \pm2, \pm3$ are precisely the same as for section 5.A.1.

5.A.4 EPVH-C: Minimization of potential enstrophy

Using (5.4.28) and differentiating with respect to y_1, y_2 etc. in turn, gives

$$\begin{aligned}
\frac{\partial H}{\partial y_1} &= \frac{1}{2}\left(\overline{Q}_1^{(1,2)} - Q_1(y_1)\right)^2 + \lambda\frac{\partial \overline{E}}{\partial y_1} + \mu\frac{\partial M}{\partial y_1},\\
\frac{\partial H}{\partial y_2} &= -\frac{1}{2}\left(\overline{Q}_1^{(1,2)} - Q_1(y_2)\right)^2 + \lambda\frac{\partial \overline{E}}{\partial y_2} + \mu\frac{\partial M}{\partial y_2},\\
\frac{\partial H}{\partial y_3} &= -\frac{1}{2}\left(\overline{Q}_2^{(-3,3)} - Q_2(y_3)\right)^2 + \lambda\frac{\partial \overline{E}}{\partial y_3} + \mu\frac{\partial M}{\partial y_3},\\
\frac{\partial H}{\partial y_{-1}} &= -\frac{1}{2}\left(\overline{Q}_1^{(-2,-1)} - Q_1(y_{-1})\right)^2 + \lambda\frac{\partial \overline{E}}{\partial y_{-1}} + \mu\frac{\partial M}{\partial y_{-1}},\\
\frac{\partial H}{\partial y_{-1}} &= \frac{1}{2}\left(\overline{Q}_1^{(-2,-1)} - Q_1(y_{-2})\right)^2 + \lambda\frac{\partial \overline{E}}{\partial y_{-2}} + \mu\frac{\partial M}{\partial y_{-2}},\\
\frac{\partial H}{\partial y_{-3}} &= \frac{1}{2}\left(\overline{Q}_2^{(-3,3)} - Q_2(y_{-3})\right)^2 + \lambda\frac{\partial \overline{E}}{\partial y_{-3}} + \mu\frac{\partial M}{\partial y_{-3}}.
\end{aligned} \tag{5.A.45}$$

where again $\partial\overline{E}/\partial y_i$ and $\partial M/\partial y_i$ for $i = \pm1, \pm2, \pm3$ are precisely the same as for section 5.A.1.

Chapter 6

Conclusions

The goal of this work has been to bridge the gap in understanding between analytical theories and more realistic flows in the form of high resolution numerical simulations. The two-layer, quasi-geostrophic model has been used to investigate two different baroclinically unstable flows, a uniform upper layer flow and an isolated jet in the upper layer.

Chapters 2 3 and 4 investigate baroclinic instability of a uniform upper layer flow focusing in particular on the contrast between instability in the absence of, and presence of, Ekman friction at the top and bottom of the channel. Since the first investigation by Phillips (1951) [95] and notably the highly influential weakly nonlinear theory put forward by Pedlosky in 1970 [91], the frictionless behaviour of the two-layer model has been widely studied as an example of baroclinic instability due to its tractable mathematics. Later Warn & Gauthier [134] produced an analytic solution (referred to here as the WGP solution), correct at minimum critical shear, and there has not been an investigation into the relevance of this solution to more realistic flows. The high resolution numerical experiments presented here are therefore a useful addition to the literature. It has been demonstrated that the weakly nonlinear theory is relatively accurate in predicting the fundamental wave amplitude and PV fields in these experiments for a range of channel widths. Furthermore, an investigation over the full range of supercriticalities for which the flow is unstable has diagnosed two different mechanisms of nonlinear equilibration. The first, as in the weakly nonlinear regime, is via the homogenization of potential vorticity in the lower layer, which is a critical layer. The second, occurring at higher supercriticalities and especially for a narrower channel width, occurs via the roll up of oppositely signed vortices in the lower layer.

In chapter 4, a suite of numerical simulations was carried out in the presence of Ekman friction in each layer. The curve of marginal stability in the presence of even an infinitesimal amount of Ekman friction is drastically altered from the frictionless case. Romea [109] put forward a weakly nonlinear theory for these flows, which has not previously been assessed (though Lovegrove et al. [73] do discuss Romea's theory in the context of other models). Romea's solution was found to be accurate at small supercriticality over a range of values of the Ekman number. A variety of different behaviours were identified, including fully nonlinear flows away from the marginal stability curve and a 'boundary layer' for weak Ekman friction where the flow behaves more like the fully nonlinear frictionless flow. A notable achievement is the discovery of an intermediate flow regime, which exhibits the WGP weakly nonlinear behaviour at early times, but whose equilibrated amplitude is entirely determined on much longer timescales by Ekman friction and Romea's corresponding weakly nonlinear theory. This is significant because the equilibrated state of the flow is always determined by the presence of Ekman friction, no matter how weak. This result has implications for inviscid theories of baroclinic adjustment in which the equilibrated flow is determined by the state in which the flow is stable to the frictionless criterion for instability. In this example, the presence of Ekman friction means that stability is governed by an entirely different criterion due to the presence of Ekman friction, which demonstrates a failure of the adjustment hypothesis.

The results of chapters 2 and 4 are a step from mathematical results in the form of weakly nonlinear theories towards more realistic numerical simulations. The biggest question that remains from this work is whether the results can be reproduced in laboratory experiments or observed in atmospheric or oceanic flows. The most likely candidates for experimental investigation are the two-layer, mechanically driven annulus where the Phillips model results might be expected to be applicable or the differentially heated annulus where the Eady model may be appropriate. The weakly nonlinear theories provide predictions of the equilibrated flow profile and wave amplitude. Ideally, one would be able to calculate the relevant supercriticality and Ekman number of annulus experiments in order to quantitatively assess weakly nonlinear predictions of the maximum amplitudes reached by unstable waves and the strength and shape of the equilibrated flow profile. It is also of interest to investigate whether critical layers of the form seen in the WGP solution are observed in reality.

There are complications involved in experiments such as additional sources of friction (e.g. from the sidewalls, between the two layers etc.) and the difficulty of achieving the weakly nonlinear flow regime. A brief calculation in chapter 4 showed that weakly nonlinear behaviour may be realizable in practice. However, it occupies a relatively small region of the parameter space, which may be difficult to access. These potential difficulties highlight the necessity for comparison and refinement of the idealized weakly nonlinear results and numerical simulations presented here to real flows.

The mechanisms in chapter 3 provide a new perspective on the instability of the two-layer model when driven solely by the presence of Ekman friction. This original description in terms of the potential vorticity in each layer of the channel is particularly useful in that it emphasises the contrast with the classic counter-propagating Rossby wave description of baroclinic instability first put forward by Bretherton for the two-layer and Eady models [14, 15]. An advantage of this approach is the ease of application to more realistic models. The CRW mechanism has been demonstrated in the Charney model by Heifetz et al [44] and Held &Pierrehumbert [48] showed that dissipative destabilization can also occur in this model, albeit in a slightly different way to the Phillips and Eady models. There is no reason why the method of chapter 3 could not be extended to the Charney model to investigate whether precisely the same destabilizing mechanism is at work. This should include extending the dissipative destabilizing mechanism in the two-layer, Eady and Charney models to include other types of dissipation such as interfacial friction and radiative damping, which themselves may be of greater importance to more realistic flows.

In chapter 5 the equilibration of an isolated baroclinic jet was considered.

It was shown that a predictive theory (EPVH) due to Esler [29] was unable to account for a barotropic shear where the associated eddy momentum fluxes are important in determining the flow behaviour. As the background horizontal shear varies, the regions in which potential vorticity is mixed alter dramatically. This study has highlighted the importance of being able to anticipate mixing behaviour and consequently regions where potential vorticity homogenization occurs. Three alternatives for a predictive theory were put forward, though none were successful and it is not clear whether any further adaptations of the EPVH theory will be able to overcome these shortcomings.

The isolated jet experiments offer a step towards more realistic flows from the uniform flow of chapters 2-4. Dissipative energization, similar to the dissipative destabilization of uniform flows, has been observed in the two-layer jet by Lee [67]. It would be a useful addition to address the mechanism by which this occurs using the PV-thinking approach of chapter 3. The question of adjusting weakly nonlinear predictions to jet-like flows might then be addressed, though this is unlikely to be entirely analytical as even the linear theory of the jet problem requires numerical calculation.

Bibliography

[1] M Abramowitz and I Stegun. *Handbook of mathematical functions.* Dover, New York, 1972.

[2] D J Acheson. *Elementary Fluid Dynamics.* Oxford University Press, Oxford, 1990.

[3] B K Arbic and G R Flierl. Baroclinically unstable geostrophic turbulence in the limit of strong and weak bottom ekman friction: Application to midocean eddies. *J. Phys. Oceanogr.*, 64:2257–2273, 2004.

[4] V I Arnol'd. On an a priori estimate in the theory of hydrodynamic stability. *Izv. Vyssh. Uchebn. Zaved. Matematika*, 54:3–5, 1966. English translation: Amer. Math. Soc. Transl., Series 2, 79, 267-269.

[5] P G Baines. A unified description of two-layer flow over topography. *J. Fluid Mech.*, 146:127–167, 1984.

[6] V Barcilon. Role of the Ekman layers in the stability of the symmetric regime obtained in a rotating annulus. *J. Atmos. Sci.*, 21:291–299, 1964.

[7] G K Batchelor. *Introduction to fluid dynamics.* Cambridge University Press, Cambridge, 1967.

[8] D Benney and R Bergeron. A new class of nonlinear waves in parallel flows. *Stud. Appl. Math.*, 48:181–204, 1969.

[9] C H Bishop. On the behaviour of baroclinic waves undergoing horizontal deformation. I: The 'RT' phase diagram. *Q. J. R. Met. Soc.*, 119:221–240, 1993.

[10] C H Bishop. On the behaviour of baroclinic waves undergoing horizontal deformation. II: Error-bound amplification and Rossby wave diagnostics. *Q. J. R. Met. Soc.*, 119:241–267, 1993.

[11] B Boville. Amplitude vacillation on a β-plane. *J. Atmos. Sci.*, 38:609–618, 1981.

[12] J P Boyd. *Chebyshev and Fourier Spectral Methods.* New York, 2001.

[13] J Bradford, A S Berman, and T S Lundgren. Nongeostrophic baroclinic instability in a two-layer rotating system. *J. Atmos. Sci.*, 38:1376–1389, 1981.

[14] F P Bretherton. Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. *Q. J. R. Met. Soc.*, 92:335–345, 1966.

[15] F P Bretherton. Critical-layer instability in baroclinic flows. *Q. J. R. Met. Soc.*, 92:325–334, 1966.

[16] F P Bretherton and D B Haidvogel. Two-dimensional turbulence above topography. *J. Fluid Mech.*, 78:129–154, 1976.

[17] J G Charney. The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.*, 4:135–162, 1947.

[18] J G Charney and M E Stern. On the stability of internal baroclinic jets ia a rotating atmosphere. *J. Atmos. Sci.*, 19:159–172, 1962.

[19] H C Davies and C H Bishop. Eady edge waves and rapid development. *J. Atmos. Sci.*, 51:1930–1946, 1994.

[20] R Davis. On the high Reynolds number flow over a wavy boundary. *J. Fluid Mech.*, 36:337–346, 1969.

[21] P G Drazin. Non-linear baroclinic instability of a continuous zonal flow. *Q. J. R. Met. Soc.*, 96:667–676, 1970.

[22] P G Drazin and W H Reid. *Hydrodynamic Stability.* Cambridge University Press, Cambridge, 2004.

[23] D G Dritschel and M E McIntyre. Multiple jets as PV staircases: the Phillips effect and the resilience of eddy transport barriers. *J. Atmos. Sci.*, 65:855–874, 2008.

[24] E J Eady. Long waves and cyclone waves. *Tellus*, 1:33–52, 1949.

[25] V W Ekman. On the influence of the earth's rotation on ocean currents. *Ark. Math. Astr. Fys.*, 2:1–52, 1905.

[26] H Ertel. Ein neuer hydrodynamischer wirbelsatz (a new hydrodynamic eddy theorem. *Meteorolol, Z.*, 59:277–281, 1942.

[27] J G Esler. Simple models of wave packets in an equilibrated baroclinic system. *J. Atmos. Sci.*, 54:2820–2849, 1997.

[28] J G Esler. Robust and leaky transport barriers in unstable baroclinic flows. *Phys. Fluids*, 2008.

[29] J G Esler. The turbulent equilibration of an unstable baroclinic jet. *J. Fluid Mech.*, 599:241–268, 2008.

[30] J G Esler and P H Haynes. Mechanisms for wave packet formation and maintenance in a quasi-geostrophic two-layer model. *J. Atmos. Sci.*, 56:2457–2489, 1999.

[31] J G Esler and B T Willcocks. Nonlinear baroclinic equilibration at finite supercriticality. *Geophys. Astrophys. Fluid Dyn.*, accepted, 2011.

[32] B Farrell. Transient growth of damped baroclinic waves. *J. Atmos. Sci.*, 42:2718–2727, 1985.

[33] B Farrell. Optimal excitation of baroclinic waves. *J. Atmos. Sci.*, 46:1193–1206, 1989.

[34] S B Feldstein and I M Held. Barotropic decay of baroclinic waves in a two-layer beta-plane model. *J. Atmos. Sci.*, 46:3416–3430, 1989.

[35] W L Gates and coauthors. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, 80:29–55, 1999.

[36] P Gauthier. Effect of detuning on the development of marginally unstable baroclinic vortices. *J. Atmos. Sci.*, 47:999–1011, 1990.

[37] P Glendinning. *Stability, Instability and Chaos.* Cambridge University Press, Cambridge, 1994.

[38] J E Hart. A laboratory study of baroclinic instability. *Geophys. Fluid Dyn.*, 3:181–209, 1972.

[39] J E Hart. A laboratory study of baroclinic chaos on the f-plane. *Tellus*, 37A:286–296, 1985.

[40] P H Haynes. Nonlinear instability of a Rossby-wave critical layer. *J. Fluid Mech.*, 161:493–511, 1985.

[41] P H Haynes. The effect of barotropic instability on the nonlinear evolution of a Rossby-wave critical layer. *J. Fluid Mech.*, 207:231–266, 1989.

[42] E Heifetz, C H Bishop, and P Alpert. Counter-propagating Rossby waves in the barotropic Rayleigh model of shear instability. *Q. J. R. Met. Soc.*, 130:211–231, 2000.

[43] E Heifetz, C H Bishop, B J Hoskins, and J Methven. The counter-propagating Rossby-wave perspective on baroclinic instability, I: mathematical basis. *Q. J. R. Met. Soc.*, 130:211–231, 2004.

[44] E Heifetz, J Methven, B J Hoskins, and C H Bishop. The counter-propagating Rossby-wave perspective on baroclinic instability, II: Application to the Charney model. *Q. J. R. Met. Soc.*, 130:233–258, 2004.

[45] I M Held. The gap between simulation and understanding in climate modelling. *Bull. Amer. Meteor. Soc.*, 86:1609–1614, 2005.

[46] I M Held. Progress and problems in large-scale atmospheric dynamics. In *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*, chapter 1. Princeton UP, 2007.

[47] I M Held and V D Larichev. A scaling theory for horizontally homogeneous, baroclinically unstable flow on a β-plane. *J. Atmos. Sci.*, 53:946–952, 1996.

[48] I M Held and R T Pierrehumbert. Dissipative destabilization of external Rossby waves. *J. Atmos. Sci.*, 43:388–396, 1986.

[49] R Hide. The viscous boundary layer at the free surface of a rotating baroclinic fluid. *Tellus*, 16:523–529, 1964.

[50] R Hide. Some laboratory experiments on free thermal convection in a rotating fluid subject to a horizontal temperature gradient and their relation to the theory of the global atmospheric circulation. *Roy. Meteor. Soc. London*, pages 169–221, 1969.

[51] R Hide and P J Mason. Sloping convection in a rotating fluid. *Adv. Phys.*, 24:47–100, 1975.

[52] E J Hinch. *Perturbation Methods.* Cambridge University Press, Cambridge, 1991.

[53] E Holopainen. On the effect of friction in baroclinic waves. *Tellus*, 13:363–367, 1961.

[54] B J Hoskins. Dynamical processes and the use of models. *Q. J. R. Met. Soc.*, 109:1–21, 1983.

[55] B J Hoskins, M E McIntyre, and W Robinson. On the use and significance of isentropic potential vorticity maps. *Q. J. R. Met. Soc.*, 111:877–946, 1985.

[56] B J Hoskins and P J Valdes. On the existence of storm tracks. *J. Atmos. Sci.*, 47:1854–1864, 1990.

[57] I N James. Suppression of baroclinic instability in horizonally sheared flows. *J. Fluid Mech.*, 44:3710–3720, 1987.

[58] I N James. *Introduction to Circulating Atmospheres*. Cambridge University Press, Cambridge, 1995.

[59] I N James and L J Gray. Concerning the effect of surface drag on the circulation of a baroclinic planetary atmosphere. *Q. J. R. Met. Soc.*, 112:1231–1250, 1986.

[60] P D Killworth and M E McIntyre. Do Rossby-wave critical layers absorb, reflect or over-reflect? *J. Fluid Mech.*, 161:449–492, 1985.

[61] J C King. An experimental study of baroclinic wave interactions in a two-layer system. *Geophys. Astrophys. Fluid Dyn.*, 13:153–167, 1979.

[62] R Krechetnikov and J E Marsden. Dissipation-induced instabilities in finite dimensions. *Rev. Mod. Physics.*, 79:519–553, 2007.

[63] R Krechetnikov and J E Marsden. Dissipation-induced instability phenomena in infinite-dimensional systems. *Arch. Rat. Mech. Anal.*, 194:611–668, 2009.

[64] L D Landau. On the problem of turbulence. *C. R. Acad. Sci. U.R.S.S.*, 44:311, 1944.

[65] G Lapeyre and I M Held. Diffusivity, kinetic energy dissipation, and closure theories for the poleward eddy heat flux. *J. Atmos. Sci.*, 60:2907–2916, 2003.

[66] S Lee. Finite-amplitude equilibration of baroclinic waves on a jet. *J. Atmos. Sci.*, 67:434–451, 2009.

[67] S Lee. Dissipative energization of baroclinic wave by surface ekman pumping. *J. Atmos. Sci.*, 67:2251–2259, 2010.

[68] S Lee and I M Held. Baroclinic wave packets in models and observations. *J. Atmos. Sci.*, 50:1413–1428, 1993.

[69] C E Leith. Minimum enstrophy vortices. *Phys. Fluids*, 27:1388–1395, 1984.

[70] M J Lighthill and G B Whitham. On kinematic waves: I. Flood movement in long rivers; II.Theory of traffic flow on long crowded roads. *Proc. Roy. Soc. Lond.*, 229:281–345, 1955.

[71] S-J Lin and R T Pierrehumbert. Does ekman friction suppress baroclinic instability? *J. Atmos. Sci.*, 45:2920–2933, 1988.

[72] J Liu and T Schneider. Mechanisms of jet formation on the giant planets. *J. Atmos. Sci.*, 67:3652–3672, 2010.

[73] A F Lovegrove, I M Moroz, and P L Read. Bifurcations and instabilities in rotating, two-layer fluids: Ii. β-plane. *Nonlin. Proc. in Geophys.*, 9:289–309, 2002.

[74] A F Lovegrove, P L Read, and C J Richards. Generation of inertia-gravity waves in a baroclinically unstable fluid. *Q. J. R. Met. Soc.*, 126:3233–3254, 2000.

[75] A J Majda and X Wang. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows.* Cambridge University Press, Cambridge, 2006.

[76] S Maslowe. Critical layers in shear flows. *Ann. Rev. Fluid Mech.*, 18:405–432, 1986.

[77] P J Mason. Baroclinic waves in a container with sloping end walls. *Philos. Trans. R. Soc. Lon. A.*, 278:397–445, 1975.

[78] W H Matthaeus and D Montgomery. Selective decay hypothesis at high mechanical and magnetic reynolds numbers. *Annals. N.Y. Acad. Sci.*, 357:203–222, 1980.

[79] M A Maximenko, B Bang, and H Sasaki. Observational evidence of alternative zonal jets in the world ocean. *Geophys. Res. Lett.*, 32:L12607, 2005.

[80] M E McIntyre. *The quasi-biennial oscillation (QBO): some points about the terrestrial QBO and the possibility of related phenomena in the solar interior. In The solar engine and its influence on the terrestrial atmosphere and climate (Vol. 25 NATO ASI Subseries I, Global environmental change).* Springer-Verlag, 1994.

[81] J C McWilliams. *Fundamentals of Geophysical Fluid Dynamics.* Cambridge University Press, Cambridge, 2006.

[82] J Miller. Statistical mechanics of of Euler's equations in two dimensions. *Phys. Rev. Lett.*, 65:2137–2140, 1990.

[83] N Nakamura. An illustrative model of instabilities in meridionally and vertically sheared flows. *J. Atmos. Sci.*, 50:357–376, 1993.

[84] N Nakamura. Momentum flux, flow symmetry, and the nonlinear barotropic governor. *J. Atmos. Sci.*, 50:2159–2179, 1993.

[85] N Nakamura. Baroclinic-barotropic adjustments in a meridionally wide domain. *J. Atmos. Sci.*, 56:2246–2260, 1999.

[86] E O'Brien. Optimal growth rates in the quasigeostrophic initial value problem. *J. Atmos. Sci.*, 49:1557–1570, 1992.

[87] D R Ohlsen and J E Hart. Transitions to baroclinic chaos on the beta-plane. *J. Fluid Mech.*, 203:23–50, 1989.

[88] R L Panetta, I M Held, and R T Pierrehumbert. External rossby waves in the 2-layer model. *J. Atmos. Sci.*, 44:2924–2933, 1988.

[89] V Pavan and I M Held. The diffusive approximation for eddy fluxes in baroclinically unstable jets. *J. Atmos. Sci.*, 52:1262–1272, 1996.

[90] J Pedlosky. The stability of currents in the atmosphere and the ocean: Part I. *J. Atmos. Sci.*, 21:201–219, 1964.

[91] J Pedlosky. Finite amplitude baroclinic waves. *J. Atmos. Sci.*, 27:15–30, 1970.

[92] J Pedlosky. Finite-amplitude baroclinic waves at minimum critical shear. *J. Atmos. Sci.*, 39:555–562, 1982.

[93] J Pedlosky. A simple model for nonlinear critical layers in an unstable baroclinic wave. *J. Atmos. Sci.*, 39:2119–2127, 1982.

[94] J Pedlosky. *Geophysical Fluid Dynamics.* Springer-Verlag, New York, 1987.

[95] N A Phillips. A simple three-dimensional model for the study of large-scale extratropical flow patterns. *J. Met.*, 8:381–394, 1951.

[96] N A Phillips. Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level quasi-geostrophic model. *Tellus*, 6:273–286, 1954.

[97] N A Phillips. The general circulation of the atmosphere: a numerical experiment. *Q. J. R. Met. Soc.*, 82:123–164, 1956.

[98] R T Pierrehumbert and K L Swanson. Baroclinic instability. *Ann. Rev. Fluid Mech.*, 27:419–467, 1995.

[99] W H Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical Recipes in Fortran 77, Second Edition.* Cambridge University Press, Cambridge, 1996.

[100] Rayleigh. On the stability or instability of certain fluid motions. *London Math. Soc.*, 9:57–70, 1880.

[101] P Rhines. Waves and turbulence on the beta plane. *J. Fluid Mech.*, 69:417–443, 1975.

[102] P Rhines. Jets. *Chaos*, 4:313–339, 1994.

[103] K J Richards, M A Maximenko, F O Brien, and H Sasaki. Zonal jets in the pacific ocean. *Geophys. Res. Lett.*, 33:L03605, 2006.

[104] K F Riley, M P Hobson, and S J Bence. *Mathematical Methods for Physics and Engineering.* Cambridge University Press, Cambridge, 2002.

[105] R Robert. A maximum-entropy principle for two-dimensional perfect fluid dynamics. *J. Statist. Phys.*, 65:531–533, 1991.

[106] R Robert and J Sommeria. Statistical equilibrium states for two dimensional flows. *J. Fluid Mech.*, 229:291–310, 1991.

[107] A R Robinson. *Eddies in Marine Science.* Springer-Verlag, 1984.

[108] W A Robinson. Dissipation dependence of the jet latitude. *J. Climate*, 10:176–182, 1997.

[109] R D Romea. The effects of friction and β on finite amplitude baroclinic waves. *J. Atmos. Sci.*, 34:1689–1695, 1977.

[110] C G Rossby. Dynamics of steady ocean current in the light of experimental fluid dynamics. *Papers Phys. Oceanog. Meteor.*, 5:1–43, 1936.

[111] C G Rossby. On the mutual adjustment on pressure and velocity distributions in certain simple current systems ii. *J. Mar. Res.*, 5:239–263, 1938.

[112] C G Rossby. Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Marine Res.*, 2:38–55, 1939.

[113] C G Rossby. Planetary flow patterns in the atmosphere. *Q. J. R. Met. Soc.*, 66:68–87, 1940.

[114] T Schneider. The thermal stratification of the extratropical troposphere. In *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*, chapter 3. Princeton UP, 2007.

[115] R K Scott and L M Polvani. Forced-dissipative shallow-water turbulence on the sphere and the atmospheric circulation of the giant planets. *J. Atmos. Sci.*, 65:3158–3176, 2007.

[116] T G Shepherd. Nonlinear saturation of baroclinic instability. Part I: The two-layer model. *J. Atmos. Sci.*, 45:2014–2025, 1988.

[117] T G Shepherd. Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Adv. Geophys.*, 32:287–338, 1990.

[118] T G Shepherd. Nonlinear saturation of baroclinic instability. Part III: Bounds on the energy. *J. Atmos. Sci.*, 50:2697–2709, 1993.

[119] A J Simmons and B J Hoskins. The lifecycles of some nonlinear waves. *J. Atmos. Sci.*, 45:414–432, 1978.

[120] A J Simmons and B J Hoskins. The downstream and upstream development of unstable baroclinic waves. *J. Atmos. Sci.*, 36:1239–1254, 1979.

[121] A J Simmons and B J Hoskins. Barotropic influences on the growth and decay of nonlinear baroclinic waves. *J. Atmos. Sci.*, 37:1679–1684, 1980.

[122] T J Simons. On the theory of atmospheric development. *Mon. Wea. Rev.*, 100:145–152, 1972.

[123] P H Stone. A simplified radiative-dynamical model for the static stability of rotating atmospheres. *J. Atmos. Sci.*, 29:405–418, 1972.

[124] P H Stone. Baroclinic adjustment. *J. Atmos. Sci.*, 35:561–571, 1978.

[125] J T Stuart. On the nonlinear mechanics of wave disturbances in stable and unstable parallel flows. i. *J. Fluid Mech.*, 9:353–370, 1960.

[126] S Sukoriansky, N Dikovskaya, and B Galperin. On the arrest of inverse energy cascade and the rhines scale. *J. Atmos. Sci.*, 64:3312–3327, 2007.

[127] K L Swanson. Storm track dynamics. In *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*, chapter 4. Princeton UP, 2007.

[128] K L Swanson and R T Pierrehumbert. Nonlinear wave packet evolution on a baroclinically unstable jet. *J. Atmos. Sci.*, 51:384–396, 1994.

[129] K L Swanson and R T Pierrehumbert. Lower-tropospheric heat transport in the pacific storm track. *J. Atmos. Sci.*, 54:1533–1543, 1997.

[130] G E Swaters. Modal interpretation for the Ekman destabilization of inviscidly stable baroclinic flow in the Phillips model. *J. Phys. Oceanogr.*, 40:830–839, 2009.

[131] A F Thompson and W R Young. Scaling Baroclinic Eddy Fluxes: Vortices and Energy Balance. *J. Phys. Oceanogr.*, 36:720–738, 2006.

[132] C D Thorncroft, B J Hoskins, and M E McIntyre. Two paradigms of baroclinic wave lifecycle behaviour. *Q. J. R. Met. Soc.*, 119:17–55, 1993.

[133] G K Vallis. *Atmospheric and Oceanic Fluid Dynamics.* Cambridge University Press, Cambridge, 2006.

[134] T Warn and P Gauthier. Potential vorticity mixing by marginally unstable baroclinic disturbances. *Tellus*, 41A:115–131, 1989.

[135] H-Y Weng. The effects of oppositely sloping boundaries with ekman dissipation in a nonlinear baroclinic system. *Q. J. R. Met. Soc.*, 116:1–29, 1990.

[136] H-Y Weng and A Barcilon. Asymmetric ekman dissipation, sloping boundaries and linear baroclinic instability. *Geophys. Astrophys. Fluid Dyn.*, 59:1–24, 1991.

[137] G B Whitham. *Linear and nonlinear waves.* John Wiley and Sons, 1974.

[138] B T Willcocks and J G Esler. Nonlinear baroclinic equilibration in the presence of Ekman friction. *J. Phys. Oceanogr.*, accepted, 2011.

[139] G P Williams. Jovian dynamics. part iii: Multiple, migrating, and equatorial jets. *J. Atmos. Sci.*, 60:1270–1296, 2003.

[140] P D Williams, T W N Haine, and P L Read. On the generation mechanisms of short-scale, unbalanced modes in rotating two-layer flows with vertical shear. *J. Fluid Mech.*, 528:1–22, 2005.

[141] P D Williams, P L Read, and T W N Haine. Spontaneous generation and impact of inertia-gravity waves in a stratified, two-layer shear flow. *Geophys. Res. Lett.*, 30:doi:10.1029/2003GL018498, 2003.

[142] R B Wood and M E McIntyre. A general theorem on angular-momentum changes due to potential vorticity mixing and on potential-energy changes due to buoyancy mixing. *J. Atmos. Sci.*, 67:1261–1274, 2010.

[143] W R Young. Selective decay of enstrophy and the excitation of barotropic waves in a channel. *J. Atmos. Sci.*, 44:2804–2812, 1987.

[144] G C Yuan, L J Pratt, and C K R T Jones. Cross-jet lagrangian transport and mixing in a 2 1/2-layer model. *J. Phys. Oceanogr.*, 34:1991–2005, 2004.

[145] P Zurita-Gotor. The relation between baroclinic adjustment and turbulent diffusion in the two layer model. *J. Atmos. Sci.*, 64:1284–1300, 2007.

[146] P Zurita-Gotor and R S Lindzen. Theories of baroclinic adjustment and eddy equilibration. In *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*, chapter 2. Princeton UP, 2007.