Nonlinear Baroclinic Equilibration in the Presence of Ekman Friction

B. T. WILLCOCKS AND J. G. ESLER

Department of Mathematics, University College London, London, United Kingdom

(Manuscript received 18 June 2011, in final form 25 August 2011)

ABSTRACT

Two theories for the nonlinear equilibration of baroclinic waves in a two-layer fluid in a β channel are tested by comparison with high-resolution numerical simulations. Predictions are tested for a range of parameters (β, κ), where the inverse criticality β measures the degree of instability and the quasigeostrophic Ekman number κ measures the strength of Ekman friction. The first theory, from Warn, Gauthier, and Pedlosky (WGP), is formally valid for marginally unstable waves at $\kappa = 0$. The second, from Romea, is formally valid for nonzero κ and for waves that are marginally stable with respect to a different criterion, which enters because of the dissipative destabilization of otherwise stable waves by Ekman friction. The predictions of the two theories are in conflict in the limit $\kappa \to 0$. When κ is slightly greater than zero, it is found that the WGP accurately predicts the maximum wave amplitude attained during a baroclinic life cycle across a significant range of parameter space. By contrast, accurate predictions of the long-time asymptotic wave amplitude are obtained only from Romea's theory, even for those cases where WGP describes the initial behavior during the life cycle accurately. The results first indicate the importance of understanding the nonlinear equilibration mechanism of dissipatively destabilized waves. Second, it follows that baroclinic adjustment theories formulated from inviscid and frictionless stability criterion make demonstrably incorrect predictions for the equilibrated state, even in the limit of vanishing Ekman friction.

1. Introduction

In their otherwise comprehensive review article on baroclinic instability, Pierrehumbert and Swanson (1995) briefly credit the weakly nonlinear theory of baroclinic instability as ''providing an island of certainty in a bewildering sea of nonlinear behaviors.'' A survey of the literature reveals the reasons why such analytical theories do not feature heavily in their review. Even for arguably the simplest fluid dynamical model of baroclinic instability, the Phillips two-layer model (Phillips 1951, 1954), both linear and nonlinear theories are subject to a number of confounding features. Chief among these are the following:

- Critical layer formation in the lower layer: The issues surrounding finite-amplitude equilibration in the absence of Ekman friction are covered in a companion paper (Esler and Willcocks 2011, hereafter EW11). Briefly summarizing, following disagreement with the numerical results of Boville (1981), the influential weakly nonlinear analysis of the Phillips model near minimum critical shear of Pedlosky (1970) was found to be incorrect. Pedlosky's original solution breaks down as all harmonics of the fundamental are resonantly excited in a lower-layer critical layer. The first attempts (Pedlosky 1982a,b) at a corrected solution including the harmonics resulted in an infinite dimensional system that could only be analyzed numerically. Subsequently, Warn and Gauthier (1989) derived a new solution in which an evolution equation for the lower-layer critical layer is coupled to an amplitude equation for the (upper layer) fundamental wave. An implicit expression for the time evolution of the wave amplitude was found in terms of the quadrature of a two-dimensional definite integral. This solution to the frictionless problem will be referred to here as the Warn–Gauthier–Pedlosky (WGP) solution.
- Dissipative destabilization of stable linear waves by Ekman friction: It is a natural step to investigate the effects of Ekman friction, due to Ekman layers at horizontal boundaries, within the framework of the Phillips model. Holopainen (1961) discovered that the presence of the Ekman friction introduces a new instability mechanism, the ''Holopainen mechanism'' (see also Swaters 2010). Waves that are linearly stable

Corresponding author address: J. G. Esler, Department of Mathematics, University College London, 25 Gower Street, London WC1E 6BT, United Kingdom.
E-mail: gavin@math.ucl.ac.uk

DOI: 10.1175/JPO-D-11-0112.1

when no Ekman friction is present can become ''dissipatively destabilized'' by the Holopainen mechanism [Krechetnikov and Marsden (2009) discuss dissipative destabilization by the Holopainen mechanism in the framework of dynamical systems theory], with the result that the location of the curve of marginal stability is displaced by an order one distance in parameter space, even where the Ekman number is infinitesimal. The growth rates of the destabilized waves near the new marginal stability curve are of the order of the Ekman number.

- An alternative theory for nonlinear equilibration in the presence of Ekman friction: A weakly nonlinear theory for the dissipatively destabilized waves has been given by Romea (1977), based on a weakly nonlinear expansion about the marginal stability curve found in Holopainen's analysis. A lower-layer critical layer does not form in Romea's problem, and a (typical) Landau-type amplitude equation is the result. Boville (1981) has suggested that Romea's theory may only be applicable below the frictionless stability criterion (i.e., for the dissipatively destabilized waves only), but to our knowledge this has not been tested. Nor is it clear why Romea's theory should not be accurate for larger values of the Ekman number, when the baroclinic instability mechanism is strongly modified by Ekman friction.

At best then, rather than Pierrehumbert and Swanson's ''island of certainty,'' there are two such islands associated with the WGP and Romea theories; each of which has an uncertain domain of influence over the wider sea of parameter space. To understand the wider relevance of either theory, it seems necessary to turn to numerical experiments. Here, the predictions of each will be tested numerically against two-layer baroclinic life cycles in a β channel, initialized with a uniform unstable zonal flow plus (near) infinitesimal noise. Ekman friction acts in each fluid layer. Following Romea (1977), for simplicity the magnitude of the Ekman friction in each layer will be taken to be of equal magnitude, although the qualitative form of the theory changes little if this assumption is relaxed or indeed if interfacial friction is included (e.g., Esler 1997). The degree of instability of the flow is controlled by the inverse criticality parameter β to be defined below. The baroclinic life cycles that result, involving linear growth and nonlinear equilibration, are relevant as simple models of flows in experiments (i.e., in the rotating annulus) and in oceanic channels subject to surface forcing by wind stress. The primary aim of the present work is to explore (β, κ) parameter space in order to answer the following questions:

(i) Are the WGP predictions relevant when Ekman friction is present ($\kappa \neq 0$)? Are the dynamics of nonlinear equilibration fundamentally altered by dissipative destabilization?
(ii) Are the Romea predictions relevant to flows satisfying the frictionless criterion for instability ($\beta < \frac{1}{2}$) (i.e., those where wave growth is primarily due to baroclinic instability rather than the Holopainen mechanism)? What happens in 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 baroclinic equilibration? For example, some authors define ''baroclinic adjustment'' to be a process during which regions of an unstable flow where the latitudinal potential vorticity (PV) gradient is negative are homogenized by PV mixing, rendering the flow stable according to the Charney–Stern–Pedlosky criterion (see, e.g., section 12.6.2 of Vallis 2006). Is this frictionless baroclinic adjustment idea ever relevant when $\kappa \neq 0$?

The plan of the paper is as follows: In section 2, the physical problem under investigation will be described in detail together with the numerical implementation of the equations of motion. Linear theory and dissipative destabilization by the Holopainen mechanism are discussed in section 3a, the predictions of the WGP theory and the results of EW11 are briefly summarized in section 3b, and the Romea theory is covered in detail in section 3c. The results of the numerical parameter sweep are given in section 4. In section 5, conclusions are presented.

2. The physical problem and its numerical implementation

a. The physical problem

The model of baroclinic instability adopted here is the quasigeostrophic (QG), two-layer Phillips's model (Phillips 1951, 1954), which describes fluid motion in a recirculating channel between rigid sidewalls at the fixed latitudes $y = \pm L_y/2$. The channel is taken to be periodic in the x direction $x \in [0, L_x]$, as in the physical situation of a recirculating atmosphere, a rotating annulus, or the Southern Ocean. Periodicity is not strictly necessary for the results to be relevant, but the situation of an open channel requires some extra analysis to ensure that the time scale for the nonlinear development is shorter than the time scale for propagation out of the domain of interest, so for definiteness the periodic situation will be assumed here. Under the β-plane approximation, the channel is subject to differential rotation at rate $f/2$, where $f = f_0 + \beta^* y$. The upper and lower layers, which are denoted with subscripts 1 and 2, respectively, are each of undisturbed depth H and have densities ρ_1 and ρ_2 ($\rho_1 < \rho_2$). The

Boussinesq approximation $\Delta = 2(\rho_2 - \rho_1)/(\rho_1 + \rho_2) \ll 1$ is adopted, meaning that an effective "reduced" gravity $g' = g\Delta$ acts on the interface between the two layers. The equations of motion are then identical whether the upper layer is bounded by a rigid lid, as in an annulus experiment, or has a free surface, as in the oceanic situation.

The basic flow to be considered is a uniform vertical shear with zonal velocity U in the upper layer and no flow in the lower layer. Ekman layers are present at the upper and lower boundaries, because of no-slip boundary conditions, and penetrate into the fluid interior because of vertical (eddy[1]) viscosity $\mathcal{A}$ within the fluid. Nondimensionalizing with horizontal length scale equal to the internal Rossby radius of deformation [see Eq. (5.191) of Vallis 2006] $L_D = \sqrt{g'H/2f_0^2}$; vertical length scale H; and vertical and horizontal velocity scales U and UH/L_D, respectively, results in equations of motion (e.g., Pedlosky 1987),

$$(\partial_t - \psi_{iy}\partial_x + \psi_{ix}\partial_y)q_i = -\kappa\nabla^2\psi_i$$

$$q_i = \beta y + \nabla^2\psi_i + \frac{1}{2}(-1)^i(\psi_1 - \psi_2). \tag{1}$$

Here, q_i is the PV in each layer ($i = 1, 2$), ψ_i is the geostrophic streamfunction, and the geostrophic velocities are given by $\mathbf{u}_i = -\nabla \times \psi_i\mathbf{k}$. For simplicity, the possibility of Stewartson layers at the sidewalls is neglected, in favor of no-normal flow and Phillips's boundary conditions,

$$\psi_{ix} = 0 \quad \text{on} \quad y = \pm\frac{1}{2}W, \tag{2}$$

$$\overline{\psi}_{iyt} + \kappa(\overline{\psi}_{iy} + \delta_{1i}) = 0 \quad \text{on} \quad y = \pm\frac{1}{2}W, \quad \text{and} \tag{3}$$

$$\psi(x, y) = \psi(x + L, y), \tag{4}$$

where δ_{ij} is the Kronecker delta.

The four parameters appearing in (1)–(4) are

$$\beta = \frac{\beta^* L_D^2}{U}, \quad \kappa = \frac{\mathcal{A}^{1/2} f_0^{1/2} L_D}{2UH},$$

$$W = \frac{L_y}{L_D}, \quad \text{and} \quad L = \frac{L_x}{L_D}. \tag{5}$$

The inverse criticality β is a measure of the degree of instability of the flow, with the Charney–Stern–Pedlosky criterion for instability in the absence of Ekman friction being $\beta \leq \frac{1}{2}$. The quasigeostrophic Ekman number κ is a measure of the influence of the Ekman layers on the fluid interior. In terms of the usual Ekman number $\mathrm{Ek} = \mathcal{A}/f_0H^2$, the quasigeostrophic Ekman number is given by $\kappa = \mathrm{Ek}^{1/2}/2\mathrm{Ro}$. The remaining parameters (L, W) are the nondimensional channel dimensions.

The experiment setting is envisaged to be a long channel; hence, $L \gg W$, where W is order unity. Following EW11, the initial conditions will be taken to consist of the basic flow plus an infinitesimal perturbation, in order that the subsequent evolution is dominated by the fastest growing mode with zonal wavenumber k_m. Unlike in EW11, where the dependence of baroclinic life cycles on the width parameter W was investigated, the channel width parameter will be set here to a fixed value $W = 2^{3/4}\pi$. The value $W = 2^{3/4}\pi$ is significant because, in the absence of Ekman friction and at minimum critical shear $\beta = \frac{1}{2}$, the fastest growing mode is exactly isotropic ($k_m = \pi/W$). All of the simulations to be shown below will therefore be for nearly isotropic waves, a simplification that to a certain extent can be justified on physical grounds, because waves that are approximately isotropic emerge spontaneously in most annulus experiment atmospheric and oceanic flows.

It is important to note that the nondimensionalization adopted above differs from that of previous authors (e.g., Pedlosky 1987; Shepherd 1988; Warn and Gauthier 1989) in that the internal Rossby radius, rather than the channel width, has been used as the horizontal length scale. The advantage of the new approach, as discussed in EW11 (see their appendix A), is that the channel width parameter W scales trivially in much of the resulting analysis. The fundamental dynamics of the current problem therefore depends primarily on the parameters β and κ, and a parameter sweep over (β, κ) space will be the focus below.

b. *Numerical implementation and experimental setup*

The quasigeostrophic equations in (1) are integrated numerically using the numerical model described in detail in EW11, adapted from that used in Esler and Haynes (1999). The discretization is pseudospectral in the x direction and grid point in the y direction. Numerical stability is maintained throughout the integrations by means of a PV diffusion $\nu\nabla^2 q_i$ on the right-hand side of (1). At a given numerical resolution, the value of ν is chosen to be sufficiently large to ensure that there is no spurious buildup of enstrophy at the grid scale. The enstrophy spectrum in each layer is monitored throughout each numerical experiment, and experiments are discarded if at late times the enstrophy spectrum does not decay superalgebraically within a finite band of wavenumbers close to the largest resolved wavenumber. The numerical time step δt is then chosen to satisfy the Courant–Friedrichs–Lewy criterion. Details of the spatial discretizations used, the corresponding diffusivities ν, and time steps δt are given in Table 1.

[1] In the experimental setting where the flow is near laminar, $\mathcal{A}$ can be taken to be the kinematic viscosity, whereas in the atmosphere and ocean $\mathcal{A}$ will be an eddy viscosity that depends on the surface roughness.

TABLE 1. Details of the three different resolutions used for the numerical simulations.

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

Simulations are repeated at more than one numerical resolution to ensure that results do not depend significantly on the value of the PV diffusivity ν. In EW11, it is shown that high-resolution (HR) experiments with $\kappa = 0$ conserve total energy to a high degree of accuracy over the time scale of the nonlinear life cycle of the baroclinic waves and that the small amount of energy dissipation that does take place in the medium-resolution (MR) and low-resolution (LR) experiments has little quantitative bearing on the results.

Simulations are initialized, as described above, with an initial state

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = -\begin{pmatrix} y \\ 0 \end{pmatrix} + \text{“noise”}, \tag{6}$$

where noise refers to random perturbations that in practice are $O(10^{-6})$ in amplitude. Because the fastest growing mode, with zonal wavelength $2\pi/k_m$, emerges from such a setup in a long zonal channel, periodicity can be enforced in the numerical model by setting $L = 2\pi/k_m$.

A single parameter sweep of (β, κ) parameter space is performed, with the channel width parameter fixed at $W = 2^{3/4}\pi$ throughout and k_m calculated separately for each simulation using the results of the linear theory of section 3a. A total of 330 experiments is performed for each parameter sweep, within the range of parameters $\beta \in [0, 0.55]$, $\kappa \in [0, 0.2]$ for which there is instability.

3. Linear and nonlinear theories

a. Linear theory and dissipative destabilization

To set the scene for the nonlinear theories of WGP and Romea, it is helpful to first revisit the linear theory for the Phillips model in the presence of Ekman friction (Holopainen 1961). The key point is that, because of the dissipative destabilization mechanism, the marginal stability curve when Ekman friction is absent ($\kappa = 0$) and the marginal stability curve when Ekman friction is vanishingly small ($\kappa \to 0$) are different.

Linear stability is investigated using the ansatz

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = -\begin{pmatrix} y \\ 0 \end{pmatrix} + \text{Re}\,\epsilon\begin{pmatrix} 1 \\ \gamma \end{pmatrix} e^{ik(x-ct)} \sin l\left(y - \frac{W}{2}\right), \tag{7}$$

where the first term is the uniform flow ϵ streamfunction, k and l are zonal and meridional wavenumbers, $\tilde{\epsilon}$ is the wave amplitude, and γ is the complex-valued streamfunction ratio between the two layers. Hence, $|\gamma|$ is the amplitude ratio and $\text{Arg}(\gamma)$ is the phase difference between the streamfunctions in each layer. Note that the values of $l = n\pi/W$ (n integer) are discretized because of the finite channel width and, because of the restriction to the fastest growing mode adopted above, only the gravest mode ($n = 1$) need be considered here. Inserting (7) into (1) and neglecting terms of $O(\tilde{\epsilon}^2)$ leads to the dispersion relation for the complex phase speed,

$$c^{\pm} = \frac{1}{2} - \frac{\left(a^2 + \frac{1}{2}\right)}{a^2(a^2+1)}\beta - i\kappa\frac{a^2 + \frac{1}{2}}{k(a^2+1)} \pm \frac{1}{2a^2(a^2+1)}\left[\beta^2 + a^4(a^4-1) - \frac{a^4\kappa^2}{k^2} + 2i\frac{a^2\kappa}{k}\beta\right]^{1/2}, \tag{8}$$

where $a = (k^2 + l^2)^{1/2}$ is the total wavenumber. The solution also determines the complex streamfunction ratio

$$\gamma^{\pm} = 2\left(a^2 + \frac{1}{2}\right) + 2\left(\frac{\beta + \frac{1}{2} + i\frac{a^2\kappa}{k}}{c^{\pm} - 1}\right). \tag{9}$$

Instability corresponds to the imaginary part of the phase speed being positive, which occurs for the positive branch of the dispersion curve (i.e., Im $c^+ > 0$).

In the absence of Ekman friction ($\kappa = 0$), the condition for instability is

$$\beta < \beta_c^I(a) \equiv a^2(1 - a^4)^{1/2}. \tag{10}$$

For $\kappa \neq 0$, the condition for instability is found to be

$$\beta < \beta_c^R(a, \kappa) \equiv \left(a^2 + \frac{1}{2}\right)\left[a^2(1-a^2) - \frac{4a^4}{k^2}\kappa^2\right]^{1/2}, \tag{11}$$

where $\beta_c^R(a, \kappa)$ is the frictional marginal stability curve. Taking the limit $\kappa \to 0$ in (11), the marginal stability curve is given by

$$\beta_c^R(a, 0) = \left(a^2 + \frac{1}{2}\right)[a^2(1-a^2)]^{1/2}, \tag{12}$$

which differs from the frictionless value $\beta_c^I(a)$ by a factor $(a^2 + 1/2)/a(a^2+1)^{1/2} > 1$ (e.g., Fig. 2 of Romea 1977). It is clear that $\beta_c^R(a, 0) < \beta_c^I(a)$, meaning that an infinitesimal amount of Ekman friction causes the marginal stability curve to shift by an order one amount (as first shown by Holopainen 1961). Waves with total wavenumber a satisfying $\beta_c^R(a, 0) < \beta < \beta_c^I(a)$ are said to be dissipatively destabilized by the presence of Ekman friction. The marginal

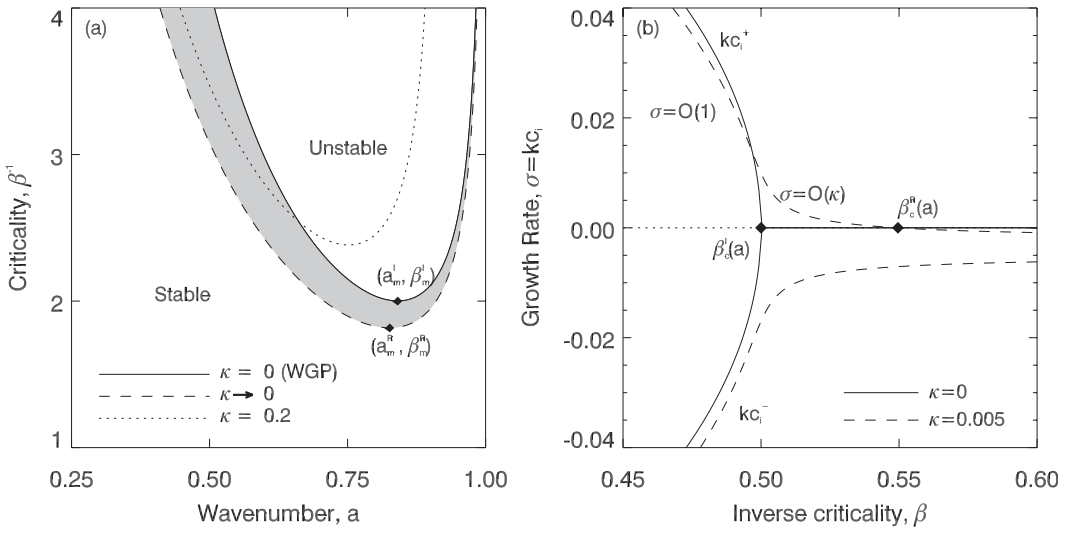

FIG. 1. (a) Criticality β^{-1} at which the system is marginally stable to linear instability (Im $c^+ = 0^+$), plotted as a function of total wavenumber a. Solid curve is $\kappa = 0$ [$\beta_c^I(a)$; (10)]. Dashed curve is $\kappa \to 0$ [$\beta_c^R(a, 0)$; (12)]. Dotted curve is $\kappa = 0.2$ [$\beta_c^R(a, 1/5)$; (11)]. The shaded region shows waves that are dissipatively destabilized by the addition of a small amount of Ekman friction, and the filled points show the minimum critical values of $\beta = \beta_m$ and $a = a_m$ in each case. (b) Growth rate (σ = Im $kc^{\pm}$) as a function of inverse criticality β for fixed total wavenumber $a = 2^{-1/4}$, for $\kappa = 0$ (solid curve) and $\kappa = 0.005$ (dashed curves). Positive and negative branches of the dispersion curve are shown. Solid points show the location of the point of marginal stability in each case.

stability curves (10) and (11) are plotted in Fig. 1a for $\kappa = 0$, $\kappa \to 0$ (dashed line), and $\kappa = 0.2$ (dotted line). The shaded region in Fig. 1a, bounded above and below by the marginal stability curves $\beta_c^I(a)$ and $\beta_c^I(a, 0)$, respectively, shows the dissipatively destabilized waves. As discussed above, the mechanism of instability for the dissipatively destabilized waves will be referred to here as the Holopainen instability mechanism (for discussion, see, e.g., Swaters 2010). Dissipative destabilization corresponds to a situation where the frictionally induced phase shift between the waves in each layer allows more energy to be released from the sloping interface than is dissipated by Ekman friction. It is notable from Fig. 1a that the Holopainen instability mechanism has a greater effect at low wavenumbers, with the explanation being that Ekman friction has a stronger effect on long waves, as can be argued on scaling grounds from (1).

The $O(1)$ difference in the $\kappa = 0$ and $\kappa \to 0$ marginal stability curves can perhaps best be understood by examination of how the growth rate of a wave of fixed wavenumber a, considered as a function of inverse criticality β, changes when Ekman friction is introduced. Hence, growth rates (Im $kc^{\pm}$) calculated from (8) are plotted in Fig. 1b as a function of β, for fixed total wavenumber $a = 2^{-1/4}$, for the no Ekman friction case (solid curve, $\kappa = 0$), and for weak Ekman friction ($\kappa = 0.005$). The $\kappa = 0$ curve shows a bifurcation at the frictionless minimum critical shear ($\beta = \frac{1}{2}$) corresponding to the coalescence of the two normal modes of the system, which have distinct real phase speeds for stable flow ($\beta > \frac{1}{2}$) and complex conjugate phase speeds for unstable flow ($\beta > \frac{1}{2}$), as can be seen from (8). The mode coalescence corresponds to the vanishing of the PV gradient in the lower layer at $\beta = \frac{1}{2}$, meaning that at marginal criticality the system can support only a single Rossby wave in the upper layer, as opposed to two coupled Rossby waves in the general case.

The addition of Ekman friction results in complex-valued perturbations to the $\kappa = 0$ phase speeds and therefore also the growth rates by an amount linear in κ. Significantly, the perturbation is different for each normal mode, so the bifurcation structure associated with mode coalescence at $\beta = \frac{1}{2}$ is destroyed for all values of κ, however small. The new curve of marginal stability β_c^R is associated with only one normal mode, and for $\beta_c^I < \beta < \beta_c^R$ the dissipatively destabilized waves have $O(\kappa)$ (slow) positive growth rates.

Below, nonlinear theories will be developed for marginally critical flows with inverse criticality $\beta = \beta_m$, where β_m is defined as the maximum value of β at which instability occurs (see points marked in Fig. 1a). The total wavenumber of the marginally stable wave at $\beta = \beta_m$ will be labeled a_m, and the corresponding zonal wavenumber $k_m = (a_m^2 - \pi^2/W^2)^{1/2}$. For frictionless flow ($\kappa = 0$), it is straightforward to show from (10) that $\beta = \frac{1}{2}$ and $a_m = 2^{-1/4}$. When $\kappa \neq 0$, β_m and a_m are determined by first finding

$$\beta_m(\kappa) = \sup_{a > \pi/W} \{\beta_c^R(a, \kappa)\}, \tag{13}$$

with $\beta_c^R(a,\kappa)$ given by (11). The supremum occurs at $a = a_m$.

b. The Warn–Gauthier–Pedlosky theory

Next, the main results of the WGP nonlinear theory (Warn and Gauthier 1989; Pedlosky 1982a,b) are briefly reviewed. The WGP theory is formally valid in the limit of vanishing supercriticality $\beta \to \beta_m = ½$ in the absence of Ekman friction, for marginally stable waves with $a_m = 2^{-1/4}$. In EW11, an appraisal was made of the accuracy of the WGP theory at finite criticality, including a detailed exploration of (β, W) parameter space. However, the question of how the WGP solution is perturbed by Ekman friction was not addressed.

The WGP solution is based on a perturbation theory, in which the frictionless supercriticality ϵ, defined to be

$$\epsilon = \left(\frac{1}{2} - \beta\right)^{1/2}, \tag{14}$$

is taken to be formally small ($\epsilon \ll 1$). A slow time scale $\tau = \epsilon t$ is introduced to eliminate secular terms in the solution. The multiple scales substitution $\partial_t \to \partial_t + \epsilon\partial_\tau$ is then made in the frictionless ($\kappa = 0$) form of (1), and the expansion

$$\begin{aligned} \psi_i &= -y\delta_{i1} + \epsilon(\psi_i^{(0)} + \epsilon\psi_i^{(1)} + \epsilon^2\psi_i^{(2)} + \cdots) \\ q_i &= (\delta_{1i} - \epsilon^2)y + \epsilon(q_i^{(0)} + \epsilon q_i^{(1)} + \epsilon^2 q_i^{(2)} + \cdots), \end{aligned} \tag{15}$$

where

$$q_i^{(j)} = \nabla^2\psi_i^{(j)} + (-1)^i\frac{1}{2}(\psi_1^{(j)} - \psi_2^{(j)}) \quad \text{and}$$

δ_{ij} is the Kronecker delta, is inserted. The leading order solution in ϵ, emerging from infinitesimal noise, is given by

$$\begin{pmatrix} \psi_1^{(0)} \\ \psi_2^{(0)} \end{pmatrix} = \mathrm{Re}\,\epsilon A(\tau)\begin{pmatrix} 1 \\ \gamma_m \end{pmatrix} e^{ik_m x}\cos\frac{\pi y}{W}. \tag{16}$$

This is the linear solution for the marginally stable wave, with arbitrary complex amplitude $A(\tau)$ and zonal wavenumber $k_m = (2^{-1/2} - \pi^2/W^2)^{1/2}$. It has the gravest meridional structure $\sim\cos(\pi y/W)$ [$n = 1$ in Eq. (7)] and vertical structure associated with a linear growing disturbance at minimum critical shear ($\gamma = \gamma_m = \sqrt{2} - 1$).

The nonlinear evolution takes place on the time scale τ. On this time scale the complex amplitude $A(\tau)$ of the leading order solution can be shown to satisfy the WGP governing equations [see EW11 for a fast route to (17) and (18)],

$$|A(\tau)|^2 = \frac{8}{LW}\int_{\mathcal{D}} y(Q + y)\,d\mathbf{x} \quad \text{and} \tag{17}$$

$$\begin{aligned} &(\partial_\eta - \Psi_y\partial_x + \Psi_x\partial_y)Q = 0, \quad \text{where} \\ &\Psi(x,y) = \cos(k_m x)\cos(\pi y/W), \end{aligned} \tag{18}$$

where the domain of integration $\mathcal{D}$ in (17) denotes the horizontal domain of the channel. Here, $Q(x,y,t) = q_2^{(1)} - y$ is the total PV at the leading order $O(\epsilon^2)$ in the lower layer. The initial condition for (18) is $Q(x, y, 0) = -y$ and

$$\eta(\tau) = \gamma_m\int_0^\tau |A(\tilde{\tau})|\,d\tilde{\tau} \tag{19}$$

serves as a rescaled time variable. Warn and Gauthier (1989) (see also appendix B of EW11) realized that the evolution Eq. (18) can be solved as a passive tracer advection equation. Inserting the solution into (17) leads to the following implicit form solution for the upper-layer wave amplitude:

$$|A(\tau)|^2 = W^2\left[\frac{2}{3} - \frac{16}{\pi^4}I\left(\frac{k_m\pi}{W}\eta\right)\right]. \tag{20}$$

Here, $I(z)$ denotes the single parameter family of definite integrals

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

where $\mathrm{sn}(\cdot|\cdot)$ is a Jacobi elliptic function in standard notation and $K(m)$ is the complete elliptic integral of the first kind.

Certain properties of the function $I(z)$ allow important characteristics of the solution (20) to be deduced. First, $I(z)$ attains its global minimum at $z_m \approx 5.55$, where $I(z_m) \approx -2.257$, giving $|A|^2_{\max} \approx 1.04W^2$. Second, $I(z) \to 0$ as $z \to \infty$, and hence $|A|^2_{\mathrm{eq}} = 2W^2/3$. The WGP predictions for maximum and equilibrated amplitudes are therefore

$$\begin{aligned} \epsilon|A|_{\max} &\approx 1.02W\left(\frac{1}{2} - \beta\right)^{1/2} \quad \text{and} \\ \epsilon|A|_{\mathrm{eq}} &= \left(\frac{2}{3}\right)^{1/2} W\left(\frac{1}{2} - \beta\right)^{1/2}. \end{aligned} \tag{21}$$

The quantitative successes of the predictions of the WGP theory at low to moderate supercriticality ($0.3 \lesssim \beta < 0.5$), reported in EW11, prompt the question of whether WGP remains relevant when the system is (weakly) perturbed by Ekman friction. It is tempting to introduce Ekman friction directly into the WGP analysis at $O(\epsilon)$. This approach fails in the sense that a system of equations predicting unbounded wave growth results. The reason is that even infinitesimal Ekman friction shifts the fundamental wave from marginal stability [$O(\epsilon)$ supercriticality] to outright instability [$O(1)$ supercriticality], because of dissipative destabilization. To understand nonlinear equilibration in the presence of Ekman friction, a different approach is needed.

c. The Romea theory

Next, an alternative nonlinear theory that is valid when Ekman friction is present, from Romea (1977), is reviewed. Romea's analysis is updated using the current nondimensionalization and adopting a more modern mathematical approach (following, e.g., Warn and Gauthier 1989), and a new simplified expression for the equilibrated wave amplitude is obtained. Romea's theory will be formulated here for flows close to marginal criticality $\beta = \beta_m$ and for waves with the marginally stable total wavenumber a_m, as determined from (13).[2]

To facilitate the perturbation expansion a new small parameter, the "Ekman friction supercriticality" ε is introduced as,

$$\varepsilon = \sqrt{\beta_m - \beta}. \tag{22}$$

Note that ε is distinct from the supercriticality ϵ in the WGP problem above, because in general $\beta_m \neq \frac{1}{2}$, the frictionless value. The natural slow time scale when Ekman friction is present will turn out to be $T = \varepsilon^2 t$. Note that T has a slower (quadratic) dependence on supercriticality compared to the linear dependence in the WGP problem above.

Romea's solution is found by first making the multiple scales substitution $\partial_t \to \partial_t + \varepsilon^2 \partial_T$ in (1), and then inserting the ansatz

$$\begin{aligned} \psi_i &= -y\delta_{i1} + \varepsilon(\psi_i^{(0)} + \varepsilon\psi_i^{(1)} + \varepsilon^2\psi_i^{(2)} + \cdots), \\ q_i &= \left[\beta_m - \varepsilon^2 + (-1)^{i+1}\frac{1}{2}\right]y \\ &\quad + \varepsilon(q_i^{(0)} + \varepsilon q_i^{(1)} + \varepsilon^2 q_i^{(2)} + \cdots), \end{aligned} \tag{23}$$

where

$$q_i^{(j)} = \nabla^2\psi_i^{(j)} + (-1)^i\frac{1}{2}(\psi_1^{(j)} - \psi_2^{(j)}),$$

which, like the expansion (15) for WGP, consists of the basic uniform flow plus perturbation terms.

At leading order in ε, Eq. (1) becomes

$$\mathcal{L}\boldsymbol{\Psi}^{(\mathbf{0})} = 0, \tag{24}$$

where the matrix differential operator $\mathcal{L}$, which is defined to be

$$\mathcal{L} \equiv \begin{pmatrix} (\partial_t + \partial_x)\left(\nabla^2 - \frac{1}{2}\right) + \left(\beta_m + \frac{1}{2}\right)\partial_x + \kappa\nabla^2 & \frac{1}{2}(\partial_t + \partial_x) \\ \frac{1}{2}\partial_t & \partial_t\left(\nabla^2 - \frac{1}{2}\right) + \left(\beta_m - \frac{1}{2}\right)\partial_x + \kappa\nabla^2 \end{pmatrix}, \tag{25}$$

acts on the streamfunction vector $\boldsymbol{\Psi}^{(\mathbf{0})} = (\psi_1^{(0)}\ \psi_2^{(0)})^\dagger$. The solution of (24) is the linear plane wave solution at minimum critical shear, with arbitrary complex amplitude $A(T)$,

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

where $k_m = (a_m^2 - \pi^2/W^2)^{1/2}$ is the zonal wavenumber of the marginally stable wave and c_m and γ_m denote the real phase speed and complex amplitude ratios, respectively. From (8) and (9), these are

$$\begin{aligned} c_m &= \frac{1}{2} - \frac{\beta_m}{a_m^2 + \frac{1}{2}}, \\ \gamma_m &= 2\left(a_m^2 + \frac{1}{2}\right) - 2\left[\frac{\beta_m + \frac{1}{2} + \frac{ia_m^2\kappa}{k_m}}{\frac{1}{2} + \beta_m \Big/ \left(a_m^2 + \frac{1}{2}\right)}\right]. \end{aligned} \tag{26}$$

The complex amplitude $A(T)$ has an undetermined dependence on the slow time scale T to be uncovered at higher order in the expansion. Note that, unlike in the WGP analysis above, γ_m has a nonzero imaginary part; that is, there is a leading order phase shift between the upper and lower layer because of Ekman fiction. In the

[2] In fact, in contrast with WGP, Romea's theory can be formulated for any wavenumber on the marginal stability curve [(11)].

WGP problem, γ_m is strictly real at leading order, and the phase shift between the upper- and lower-layer waves enters at higher order and is proportional to $dA/d\tau$, as opposed to A here. This fundamentally changes the asymptotics and is the reason why the long time scale T in the Romea problem is $O(\varepsilon^{-2})$ in contrast to WGP, where it is $O(\epsilon^{-1})$. The phase shift between the streamfunction in the two layers leads to a nonzero meridional PV flux in each layer, which is necessary for momentum exchange between the layers. In the WGP solution momentum is exchanged between the layers only in association with wave growth and decay, whereas in the Romea solution momentum exchange is proportional to $|A|^2$, but its effects are opposed by momentum restoration because of the Ekman friction in each layer.

At $O(\varepsilon^2)$ in the expansion, the equation is found to be

$$\mathcal{L}\boldsymbol{\Psi}^{(1)} = \mathbf{f}^{(1)}, \tag{27}$$

where the forcing term

$$\mathbf{f}^{(1)} = \begin{pmatrix} -J(\psi_1^{(0)}, q_1^{(0)}) \\ -J(\psi_2^{(0)}, q_2^{(0)}) \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{k_m \pi \gamma_i}{4W} |A|^2 \sin\left(\frac{2\pi y}{W}\right), \tag{28}$$

where $J(\cdot,\cdot)$ is the usual Jacobian operator. Here, $\gamma_i = \mathrm{Im}\,\gamma_m$ is a measure of the phase difference between each layer and is proportional to the Ekman friction κ; hence, the Jacobian terms in (28) are nonzero only when $\kappa \neq 0$, and the Romea solution breaks down when $\kappa = 0$ (although it will be shown below that the limit $\kappa \to 0$ is well defined). The forcing $\mathbf{f}^{(1)}$ is easily shown to satisfy the solvability condition to be discussed below, so a particular integral for $\boldsymbol{\Psi}^{(1)}$, satisfying the Phillips boundary condition (3) on $y = \pm W/2$, is straightforward to find as

$$\boldsymbol{\Psi}^{(1)} = \begin{pmatrix} \psi_1^{(1)} \\ \psi_2^{(1)} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{a_m^2 \left(a_m^2 + \frac{1}{2}\right) W}{4\pi \left(2\beta_m + a_m^2 + \frac{1}{2}\right)} |A|^2 \times \left[\sin\left(\frac{2\pi y}{W}\right) + \frac{2\pi}{W} y\right]. \tag{29}$$

The solution (29) represents the mean flow correction resulting from a competition between eddy fluxes and Ekman friction. Both the eddy fluxes and Ekman friction act to force the mean flow correction and both are linear in κ. Consequently, $\boldsymbol{\Psi}^{(1)}$, which represents the balance between the two effects, has only a relatively weak dependence on κ, because of β_m, a_m, and k_m, depending on κ through (13).

At the next order $O(\varepsilon^3)$, Eq. (1) takes the form

$$\mathcal{L}\boldsymbol{\Psi}^{(2)} = \mathbf{f}^{(2)}, \tag{30}$$

where

$$\mathbf{f}^{(2)} = \begin{pmatrix} -q_{1T}^{(0)} + \psi_{1x}^{(0)} - J(\psi_1^{(0)}, q_1^{(1)}) - J(\psi_1^{(1)}, q_1^{(0)}) \\ -q_{2T}^{(0)} + \psi_{2x}^{(0)} - J(\psi_2^{(0)}, q_2^{(1)}) - J(\psi_2^{(1)}, q_2^{(0)}) \end{pmatrix}. \tag{31}$$

It is not necessary to solve (30) explicitly in order to derive the amplitude evolution equation for $A(T)$. Instead, the amplitude equation follows from a solvability condition, which arises because the operator $\mathcal{L}$ has a nontrivial kernel. If the forcing $\mathbf{f}^{(2)}$ were to project onto the kernel of $\mathcal{L}$, then the response in $\boldsymbol{\Psi}^{(2)}$ would include so-called secular (i.e., resonantly excited) terms that would grow linearly in time. The presence of secular terms would cause the expansion of (1) to become disordered on times $t \sim \varepsilon^{-2}$ [$T \sim O(1)$]. For the expansion to remain uniformly valid on the long time scale of interest, the solvability condition must therefore be imposed to remove the secular terms from the solution.

An elegant method to derive the solvability condition is to consider the adjoint of the operator $\mathcal{L}$. The adjoint is defined with reference to an inner product, which acts on two pairs of "streamfunction like" functions $\boldsymbol{\Phi} = (\phi_1\ \phi_2)^\dagger$ and $\boldsymbol{\Psi} = (\psi_1\ \psi_2)^\dagger$ as follows:

$$\langle \boldsymbol{\Phi},\ \boldsymbol{\Psi} \rangle = \frac{1}{\overline{T}} \int_0^{\overline{T}} \int_{\mathcal{D}} \phi_1 \psi_1 + \phi_2 \psi_2 \, d\mathbf{x}\, dt,$$

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

$$\langle \boldsymbol{\Phi}, \mathcal{L}\boldsymbol{\Psi} \rangle = \langle \mathcal{L}^\dagger \boldsymbol{\Phi}, \boldsymbol{\Psi} \rangle,$$

and it is straightforward to show using integration by parts that

$$\mathcal{L}^\dagger \equiv \begin{pmatrix} -(\partial_t + \partial_x)\left(\nabla^2 - \frac{1}{2}\right) - \left(\beta_m + \frac{1}{2}\right)\partial_x + \kappa\nabla^2 & -\frac{1}{2}\partial_t \\ -\frac{1}{2}(\partial_t + \partial_x) & -\partial_t\left(\nabla^2 - \frac{1}{2}\right) - \left(\beta_m - \frac{1}{2}\right)\partial_x + \kappa\nabla^2 \end{pmatrix}. \tag{32}$$

Consider now any pair of functions $\mathbf{\Phi_0} = (\phi_{0,1}\ \phi_{0,2})^\dagger$ belonging to the kernel of $\mathcal{L}^\dagger$; that is, $\mathcal{L}^\dagger\mathbf{\Phi_0} = 0$. Taking the inner product of Eq. (30) with $\mathbf{\Phi_0}$ gives

$$\langle \mathbf{\Phi_0}, \mathcal{L}\mathbf{\Psi^{(1)}} \rangle = \langle \mathcal{L}^\dagger\mathbf{\Phi_0}, \mathbf{\Psi^{(1)}} \rangle = 0 = \langle \mathbf{\Phi_0}, \mathbf{f^{(2)}} \rangle.$$

The solvability condition is therefore simply $\langle \mathbf{\Phi_0}, \mathbf{f^{(2)}} \rangle = 0$. It turns out that the relevant solution of $\mathcal{L}^\dagger\mathbf{\Phi_0} = 0$; that is, the solution that results in a nonvanishing inner product with $\mathbf{f^{(2)}}$ is

$$\mathbf{\Phi_0} = \mathrm{Re}\begin{pmatrix} -c_m \\ (1 - c_m)\gamma_m^* \end{pmatrix} e^{ik_m(x-c_m t)} \cos\frac{\pi y}{W}, \quad (33)$$

where the superscript asterisk denotes the complex conjugate. The resulting inner product $\langle \mathbf{\Phi_0}, \mathbf{f^{(2)}} \rangle$ is straightforward to evaluate and yields the amplitude equation of Romea (1977),

$$A_T = (\sigma_r + i\sigma_i)k_m A + (\mu_r + i\mu_i)k_m |A|^2 A, \quad (34)$$

where after some algebra the coefficients simplify to

$$\sigma = \sigma_r + i\sigma_i = \frac{\kappa}{k_m}\frac{4a_m^2\left[a_m^2 + \frac{1}{2}\right]\left[\beta_m - 4\left(a_m^2 + \frac{1}{2}\right)^2(\kappa/k_m)i\right]}{\beta_m^2 + 16a_m^4\left(a_m^2 + \frac{1}{2}\right)^4(\kappa/k_m)^2}, \quad \text{and}$$

$$\mu = \mu_r + i\mu_i = \frac{a_m^2\left(a_m^2 + \frac{1}{2}\right)^2\left[3a_m^2(1 - a_m^2) + 4\left(a_m^2 - \frac{1}{2}\right)\pi^2/W^2\right]}{2\left(2\beta_m + a_m^2 + \frac{1}{2}\right)\left[\beta_m^2 + 16a_m^4\left(a_m^2 + \frac{1}{2}\right)^4(\kappa/k_m)^2\right]}\left[4a_m^2\left(a_m^2 + \frac{1}{2}\right)\frac{\kappa}{k_m} + i\beta_m\right]. \quad (35)$$

The expressions (35) are much more explicit than the corresponding expressions in Romea [1977, his (A1) and (A2)] and will be used below to make clear the parameter dependencies of the solution.

Forming the product $[A^* \times (34)] + [A \times (34)^*]$ results in the real amplitude equation

$$\frac{d|A|^2}{dT} = 2k_m|A|^2(\sigma_r - \mu_r|A|^2). \quad (36)$$

Equation (36), which was first suggested by Landau (1944), is generic for dynamical systems near marginal stability and has been derived explicitly for numerous fluid dynamical systems beginning with Stuart (1960) for plane Poiseuille flow at relatively low Reynolds number. The usual situation of linear instability and nonlinear equilibration requires σ_r and $\mu_r > 0$, which are both satisfied here for supercritical flow. As is typical, the linear growth rate in time (i.e., $2k_m\sigma_r\varepsilon^2$) can be obtained directly from the dispersion relation (8) by a Taylor series expansion in β about $\beta = \beta_m$. The nonlinear coefficient quantifies the stabilizing effect of the mean flow correction (29) on the fundamental wave and requires explicit evaluation.

Equation (36) can be integrated to give the solution

$$|A(T)|^2 = \frac{|A(0)|^2 e^{2k_m\sigma_r T}}{1 + (\mu_r/\sigma_r)|A(0)|^2(e^{2k_m\sigma_r T} - 1)}. \quad (37)$$

For a linearly unstable wave ($\sigma_r > 0$), the wave amplitude is a monotonic increasing function of the slow time scale T and approaches an asymptotic limit for the wave amplitude given by

$$\varepsilon|A|_{\mathrm{eq}} = \lim_{T\to\infty}\varepsilon|A(T)| = (\beta_m - \beta)^{1/2}\left(\frac{\sigma_r}{\mu_r}\right)^{1/2} = (\beta_m - \beta)^{1/2}\left\{\frac{2\beta_m\left(2\beta_m + a_m^2 + \frac{1}{2}\right)W^2}{a_m^2\left(a_m^2 + \frac{1}{2}\right)^3\left[3a_m^2(1 - a_m^2)W^2 + 4\left(a_m^2 - \frac{1}{2}\right)\pi^2\right]}\right\}^{1/2}. \quad (38)$$

Equation (38) reveals the parameter dependency of the amplitude of the equilibrated solution. One surprising aspect is that there is no explicit dependence on κ whatsoever; that is, the only effect of Ekman friction on the equilibrium state is through changes in β_m and a_m, the inverse criticality and total wavenumber at marginal

TABLE 2. Summary of selected simulations.

Simulation	Inverse criticality β	QG Ekman number κ	Channel width W	Description
Expt I	0.53	0.04	$2^{3/4}\pi$	Dissipatively destabilized; Romea is accurate.
Expt II	0.31	0.16	$2^{3/4}\pi$	Strong Ekman friction; Romea is accurate.
Expt III	0.50	0.04	$2^{3/4}\pi$	Dissipatively destabilized; illustrates Holopainen mechanism.
Expt IV	0.495	0	$2^{3/4}\pi$	Weak baroclinic instability; WGP is accurate.
Expt V	0.48	0.0025	$2^{3/4}\pi$	Weak baroclinic instability and weak Ekman friction; WGP predicts early times, and Romea predicts equilibrium.
Expt VIa	0.47	0.005	$2^{3/4}\pi$	As in Expt V, but at greater supercriticality and stronger Ekman friction.
Expt VIb	0.47	PV damping $r = 0.005$	$2^{3/4}\pi$	Weak baroclinic instability with PV damping; forced WGP is accurate.

stability defined by (13). Because β_m and a_m are given implicitly by (13), it is instructive to examine the limit $\kappa \to 0$ for which explicit values can be found. In the limit $\kappa \to 0$, it is straightforward to show from (12) and (13) that the inverse criticality and total wavenumber at marginal stability are

$$\beta_m = \frac{\sqrt[4]{3}(3+\sqrt{3})}{8\sqrt{2}}, \quad \text{and} \quad a_m = \frac{1}{2}(1+\sqrt{3})^{1/2}.$$

Inserting these values into (38), the following expression for the equilibrated amplitude is obtained,

$$\lim_{\kappa\to 0} \varepsilon|A|_{\text{eq}} = (\beta_m - \beta)^{1/2} \times \left[\frac{32\sqrt[4]{3}(\sqrt{2}+\sqrt[4]{3})W^2}{8(3+\sqrt{3})\pi^2 + 9(2+\sqrt{3})W^2}\right]^{1/2}. \quad (39)$$

The result (39) can be compared with the equilibrated WGP amplitude (21), noting the different supercriticalities ($\beta_m \neq \frac{1}{2}$). The most significant difference between the two formulas is that Romea predicts an equilibrated amplitude that is independent of W in the wide channel limit ($W \to \infty$) whereas WGP predicts an amplitude proportional to W. Given that it is necessary that $W > 2^{1/4}\pi$ for (frictionless) instability, the coefficient multiplying the supercriticality in Romea is always smaller than that in WGP. The predictions (38) and (39) for the equilibrated amplitude will be compared with the results of numerical simulations next.

4. Numerical results: Behavior at finite supercriticality

a. Typical behavior near marginal stability

A natural question to ask is, under what circumstances, if any, is the analytic solution (37) accurate? The question can be addressed numerically, just as the corresponding question for the WGP solution was addressed (for $\kappa = 0$) in EW11. The (β, κ) parameter sweep of simulations that forms the basis for our results was described in section 2b. From that set of simulations, six in particular, which are labeled experiments (Expts) I–VIa and detailed in Table 2, will be examined in detail as representative of typical model behavior at low supercriticality.

Expts I and II are simulations for which the Romea theory might be expected to be accurate. They differ in that Expt I with $(\beta, \kappa) = (0.53, 0.04)$ is dissipatively destabilized; that is, the only instability mechanism in operation is the Holopainen mechanism and the flow would be stable to baroclinic instability in the absence of Ekman friction ($\kappa = 0$). Expt II with $(\beta, \kappa) = (0.31, 0.16)$, by contrast, is weakly supercritical because of the action of strong Ekman friction, without which it would be strongly unstable to baroclinic instability.

Figure 2a shows the time evolution of the upper wave amplitudes for both Expts I and II, with the simulations results plotted as solid curves and the corresponding results from Romea's equation (37) plotted as dashed curves. The amplitudes $|A_1^s|$ are obtained from the upper-layer streamfunction[3] in the simulations, from the complex-valued wave amplitude A_i^s defined to be

$$A_i^s(t) = \frac{4}{LW}\int_{\mathcal{D}} \psi_i(x,y,t)e^{-ik_m x}\cos\left(\frac{\pi y}{W}\right)d\mathbf{x}, \quad i = 1,2. \quad (40)$$

The amplitude $A_1^s(t)$ has been defined in order that it can be directly identified with $\varepsilon A(T)$ in Romea's theory. It is clear that the Romea theory is accurate for both Expts I and II, with the maximum discrepancy in wave amplitude being around 12% (during Expt I) and with the equilibrated amplitude accurate to within 3% in each case. Figure 2b shows the upper-layer zonal-mean winds

[3] Note that, in EW11, wave amplitudes were defined using the PV field rather than the streamfunction, in order to exploit results derived from pseudomomentum bounds.

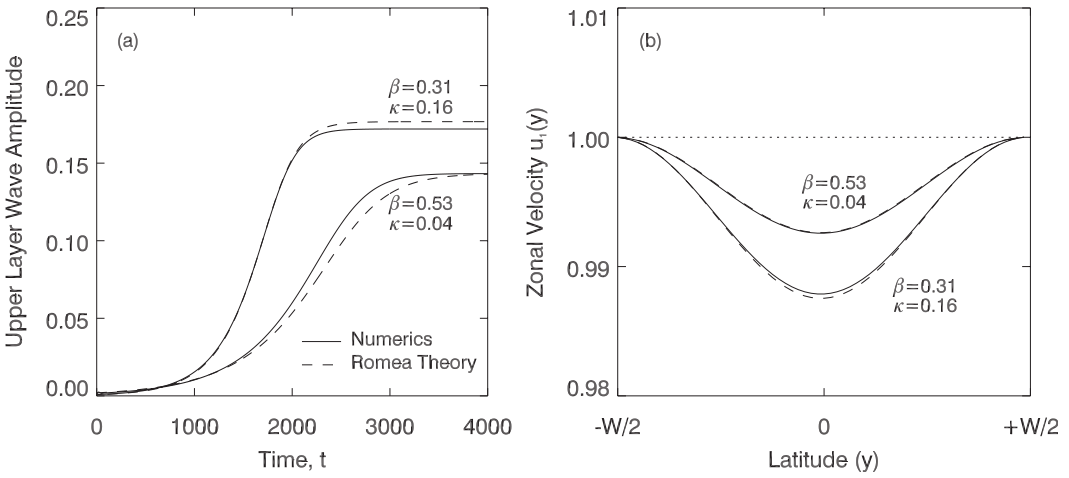

FIG. 2. (a) Evolution of the fundamental wave amplitudes $|A_1^s|$ for Expt I $[(\beta, \kappa) = (0.53, 0.04)]$ and Expt II $[(\beta, \kappa) = (0.31, 0.16)]$. Simulation results are solid curves and Romea's theoretical prediction [(37)] are the dashed curves. (b) Solid curves are equilibrated upper-layer zonal-mean flow corrections $-\overline{\psi}_{1y}$ as a function of latitude y for Expts I and II. Dashed curves are predictions from Romea's theory $1 - \varepsilon^2\psi_{1y}^{(1)}$ [Eq. (29)].

$\overline{u}_1 = -\overline{\psi}_{1y}$ (solid curves) for Expts I and II at late times in the simulations when the flow is steady. Also plotted is the zonal-mean flow correction predicted from Romea's theory $-\varepsilon^2\psi_{1y}^{(1)}$ (dashed curves) given by (29). It is clear that the nonlinear mean flow correction is well captured by the Romea theory for both Expts I and II, with the upper-layer zonal velocity decreasing across the channel and with the largest adjustment observed at the channel center.

Figure 3 shows the development of the lower-layer PV field in two further simulations Expts III and IV detailed in Table 2. The two simulations have been chosen as ''clean'' illustrations of a Romea-type equilibration (Expt III) and a frictionless WGP-type equilibration (Expt IV). The development of the lower-layer PV is key to understanding the equilibration mechanism in each case.

The qualitative dynamical behavior in Expt III is similar to that of Expt I, featuring dissipatively destabilized waves that grow because of the Holopainen instability, except that the flow is somewhat more unstable. However, from a PV perspective, Expt III is particularly instructive, because there is no background gradient of PV in the lower-layer initial flow, because inverse criticality $\beta = \frac{1}{2}$ corresponds to marginal stability of the frictionless system. Because Ekman friction acts to relax the system toward the initial zonal flow throughout the evolution, 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.

In Fig. 3a, the linear growing mode is present at small amplitude. The perturbation is amplified during the linear growth stage (approximately up to $t = 400$; Fig. 3b), during which time the PV distribution remains close to the plane wave solution (7) and consequently has the form in the lower layer of a pair of vortices spanning the channel. During the linear growth stage, the lower-layer vortices grow exponentially because of the Holopainen mechanism. Physically, Ekman friction acts on the lower-layer flow that is induced by the Rossby wave present in the upper layer, causing the growth of the vortices within each lower-layer circulation cell. The upper-layer Rossby wave and lower-layer vortex train are phase locked, so that the velocity fields they induce in the opposite layer are mutually reinforcing, in a manner reminiscent of the counterpropagating Rossby wave (CRW) mechanism of baroclinic instability (Bretherton 1966). As discussed elsewhere (Willcocks 2011), there remain significant differences between the CRW mechanism and the Holopainen mechanism, because of the differing nature of the PV structure in the lower layer (Rossby wave versus train of forced vortices). Eventually, the vortex train in the lower layer reaches finite amplitude and a significant mean flow feedback develops. The changes to the mean flow act to advect both the upper-layer Rossby wave and lower-layer vortex train, disrupting the phase locking and inhibiting growth. The Holopainen instability is equilibrated and the flow approaches a final steady state. In the lower layer, the final state consists of finite-amplitude vortices, together with a weak positive PV gradient associated with the mean flow feedback, as shown in Fig. 3c.

The evolution of the lower-layer PV field during Expt III can be contrasted with that in Expt IV, $(\beta, \kappa) = (0.495, 0)$, in Figs. 3d–f. Expt IV is weakly unstable to baroclinic instability; hence, at early times (Fig. 3d) there is a uniform negative PV gradient in the lower layer. No Ekman friction is acting; hence, the lower-layer PV is

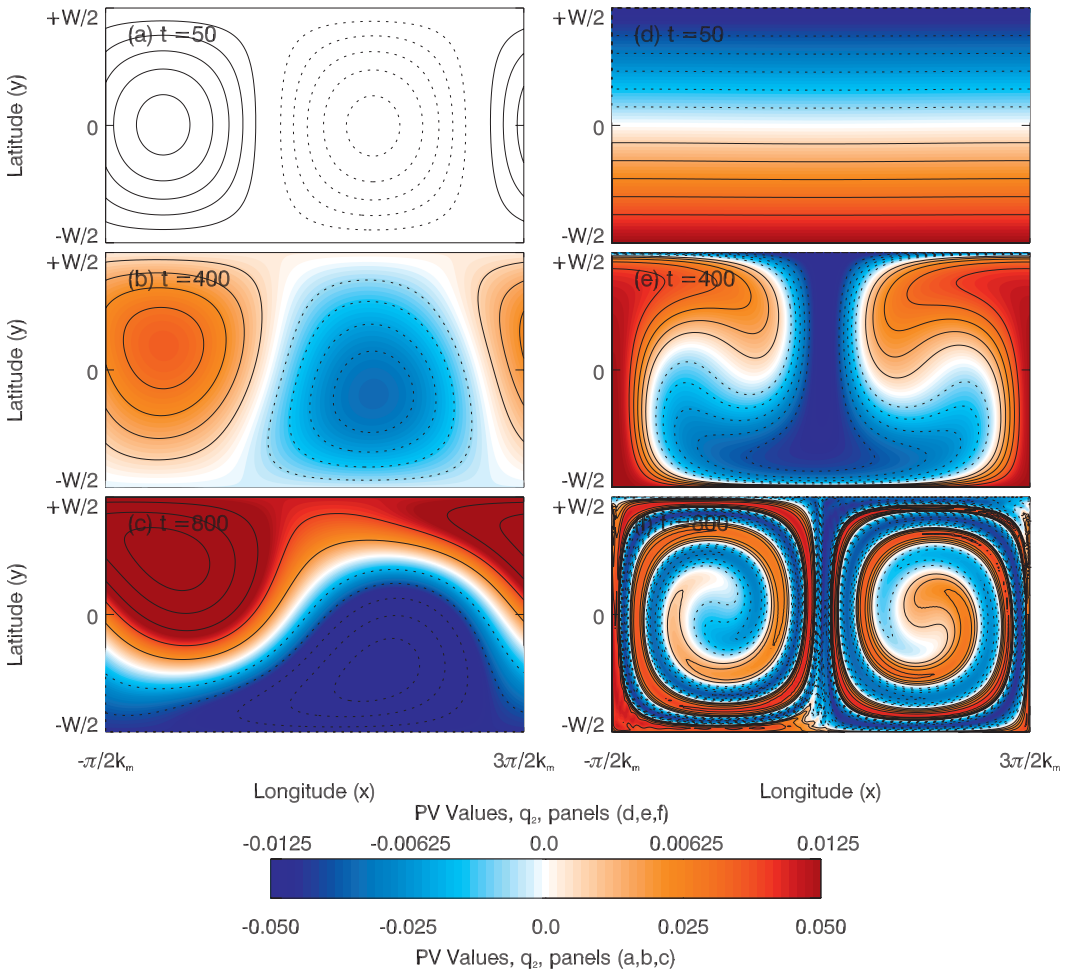

FIG. 3. (a)–(c) Snapshots of the lower-layer PV field at three times (t = 0, 400, and 800) for Expt III (β, κ) = (0.5, 0.04). Blue/dashed lines correspond to negative PV values and red/solid lines correspond to positive PV values. Contour intervals are (a) 1.5×10^{-3}, (b) 7.5×10^{-3}, and (c) 2×10^{-2}. (d)–(f) As in (a)–(c), but for Expt IV, (β, κ) = (0.495, 0), showing the corresponding PV evolution in the presence of baroclinic instability but the absence of Ekman friction. Contour intervals for (d)–(f) are 8×10^{-3}.

conserved following fluid parcels. Figures 3e–f show that equilibration nevertheless occurs by the removal of the negative PV gradient, via PV stirring throughout the entire lower critical layer and eventual (coarse grain) homogenization, exactly as predicted by the WGP theory (see Warn and Gauthier 1989; Fig. 2 of EW). The elimination of the negative lower-layer PV gradient was found to occur in all of the frictionless baroclinic life cycles examined in EW11, although the mechanism by which the gradient is removed changes somewhat as supercriticality increases.

Expts I–III belong to a subset of (β, κ) parameter space that can be described as "Romea like," which can be contrasted with the "WGP like" evolution of Expt IV. Both Romea (1977) and Boville (1981) suggest that the Romea-like region of (β, κ) parameter space might be restricted to the dissipatively destabilized region with $\frac{1}{2} < \beta < \beta_m$. Expt II, however, demonstrates that the Romea-like regime extends far into the region of frictionless baroclinic instability $\beta < \frac{1}{2}$, provided that Ekman friction κ is sufficiently strong.

Another means to distinguish between WGP-like and Romea-like evolution is to examine the qualitative evolution of the upper-layer wave amplitude. In the WGP life cycle, the wave amplitude reaches a maximum that is around 25% greater than the final equilibrated value attained at long times. By contrast, Romea-like flows feature wave growth that is monotonic up to the maximum equilibrated value as seen in Romea's solution (37). The full extent of the Romea-like and WGP-like regions of parameter space will be determined next.

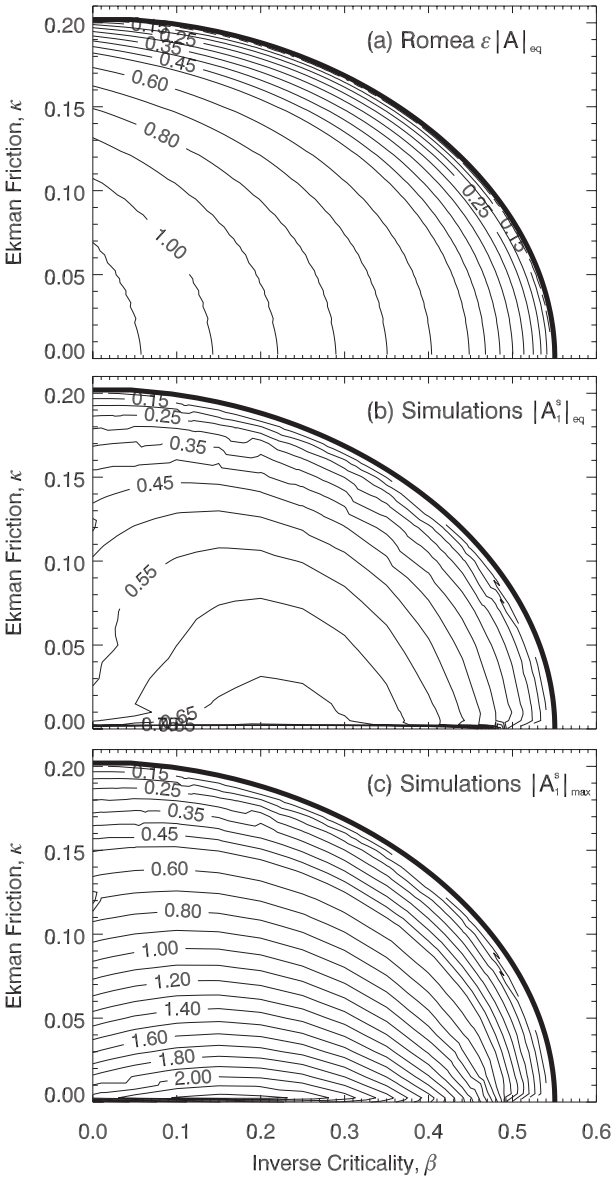

FIG. 4. (a) Equilibrated wave amplitude predicted by Romea's theory [$\varepsilon|A|_{eq}$; Eq. (38)] as a function of inverse criticality β and QG Ekman number κ. (b) Actual equilibrated amplitude attained in the simulations [calculated using (40)]. (c) Maximum amplitude attained during the simulations.

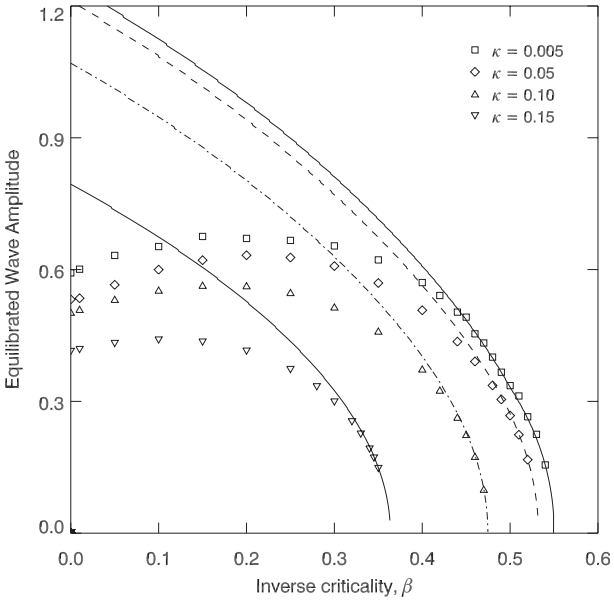

FIG. 5. Equilibrated amplitude $|A_1^s|_{eq}$ as a function of inverse criticality β for κ = 0.005, 0.05, 0.1, and 0.15. Simulations are open points (as labeled), whereas the curves show the Romea's predictions [from Eq. (38)].

b. *Exploration of (β, κ) parameter space*

Here, the main results from the full set of simulations across (β, κ) parameter space will be described. Figures 4a,b contrast Romea's prediction for the equilibrated wave amplitude $\varepsilon|A|_{eq}$ [Eq. (38)] with the actual wave amplitudes attained in the simulations [calculated using (40)]. The thick solid curve in each panel of Fig. 4 shows the inverse criticality $\beta_m = \beta_m(\kappa)$ associated with marginal stability. The first region of parameter space to notice is the Romea-like regime lying close to this curve: that is, where the value of $\varepsilon = (\beta_m - \beta)^{1/2}$ is small. Agreement between Romea and the simulations within this region, where Expts I–III are located, is good. Maximum wave amplitudes, shown in Fig. 4c, are close to the equilibrated values within this region as expected.

A second region of parameter space to be seen in Fig. 4b is a "boundary layer" at $\kappa = 0$, for $\beta \in [0, 0.5)$. The discontinuity in $|A_1^s|_{eq}$ at the boundary layer is strong evidence that the equilibrium wave amplitude in the limit $\kappa \to 0$ differs from that for the frictionless problem ($\kappa = 0$). This point will be discussed further below.

A third region of (β, κ) parameter space is located far from both the marginal stability curve and away from $\kappa = 0$, which will be termed the strongly nonlinear region. Romea overestimates $|A_1^s|_{eq}$ here by a factor of 2 or more, and $|A_1^s|_{max} > |A_1^s|_{eq}$ throughout. In fact, the dependence of $|A_1^s|_{eq}$ on (β, κ) within much of this region is relatively weak.

Figure 5 shows the extent to which Romea's theory remains accurate as the supercriticality increases and the system moves from the Romea-like region of parameter space to the strongly nonlinear region. The equilibrated amplitudes $|A_1^s|_{eq}$ from the simulations are plotted as a function of β along with the corresponding predictions from Romea, for κ = 0.005, 0.05, 0.1, and 0.15. The equilibrated amplitudes remain close to the predicted values for $\beta_m - \beta \lesssim 0.06$ ($\varepsilon \lesssim 0.25$) for each value of κ. As the supercriticality increases, the equilibrated amplitudes

then maximize at $\beta_m - \beta \approx 0.25$ ($\varepsilon \approx 0.5$), before decreasing. Interestingly, at very low inverse criticalities $\{\beta \in [0, 0.1]\}$, equilibrated amplitudes are almost independent of κ, suggesting the possibility of a different theoretical treatment for this region.

A fourth region of (β, κ) parameter space, roughly occupying $0.4 \lesssim \beta < 0.5$ and $0 < \kappa \lesssim 0.05$ will be termed the mixed behavior region. Here, Romea accurately predicts the equilibrated amplitude (see Figs. 4a,b), but the maximum amplitude (see Fig. 4c) is much greater. The evolution cannot be described as Romea like, because Romea's solution (37) predicts that the equilibrated and maximum amplitudes are identical. Understanding the behavior of the simulations in the mixed behavior region will be the focus next.

c. *Evolution in the mixed behavior region of parameter space*

A region of (β, κ) parameter space roughly occupying $0.4 \lesssim \beta < 0.5$ and $0 < \kappa \lesssim 0.05$ was identified above and termed the mixed behavior region. The flow in this region is unstable to baroclinic instability ($\beta < \frac{1}{2}$) and Ekman friction is low ($\kappa \to 0$). Romea theory's makes an accurate prediction of the equilibrated amplitude $|A_1^s|_{\text{eq}}$, but the baroclinic life cycle is not Romea like, because the maximum wave amplitude attained $|A_1^s|_{\max}$ is significantly greater than $|A_1^s|_{\text{eq}}$. Interestingly, the mixed behavior region of parameter space is a region where the WGP theory (valid for $\kappa = 0$) might be expected to be relevant. Most flows in this region have the property that the time scale $T_R \sim \kappa^{-1}(\varepsilon)^{-2}$ for the development of the Romea solution is much longer than $T_I \sim \varepsilon^{-1}$, the time scale for the frictionless baroclinic life cycle. Recall that ε is the frictionless supercriticality given by (14) and ε given by (22) is the Ekman friction supercriticality. Should flows in the mixed behavior region with $T_I \ll T_R$ be expected to show features of the frictionless solution at short times?

Expt V with $(\beta, \kappa) = (0.48, 0.0025)$ (see also Table 2) is precisely such a flow. Figure 6a (solid curve) shows the evolution of the amplitude of the fundamental wave during Expt V. The dashed line shows the corresponding evolution of the WGP solution (from 20), and the dotted lines show the Romea solution (37). Both the initial growth rate and the maximum amplitude of the wave are captured reasonably well by the WGP theory up to $t \approx 400$ when the wave amplitude begins to oscillate about the WGP equilibrated amplitude as described in EW11. The wave amplitude then decays on the much longer time scale T_R toward an equilibrated value, which is seen to be well predicted by (38). Similar behavior can be seen in Fig. 6b for Expt Via, for which the supercriticality is slightly greater. The contrast with Expt VIb, in which Ekman friction is replaced by PV damping, is discussed further below.

Figure 7 shows snapshots of the lower-layer PV evolution during Expt V, which can be contrasted with the Romea-like behavior reported for Expt III (Figs. 3a–c) and the frictionless WGP-like behavior of Expt IV (Figs. 3d–f; see also Figs. 4 and 5 of EW11). The initial state, as can been seen in Fig. 7a, is a negative lower-layer PV gradient, consistent with baroclinic instability. As the disturbance grows (Figs. 7b,c), the evolution of the lower-layer PV remains very close to the critical layer behavior exhibited by the corresponding frictionless flow and predicted by the WGP theory, demonstrating that at this time the flow is essentially undergoing a frictionless baroclinic life cycle. Ekman friction is not strong enough to act on the baroclinic life cycle time scale as $T_I \ll T_R$. However, slight asymmetries are apparent between the two circulation cells, with negative PV becoming concentrated in the cyclonic (anticlockwise) cell and positive PV becoming concentrated in the anticyclonic cell. Interestingly, the finite criticality effect identified in EW for frictionless flows acts in exactly the opposite sense in that positive PV becomes concentrated within the cyclonic cell. By $t = 225$–275 (Figs. 7d,e), coherent Romea-like vortices have organized within the lower layer because of the cumulative effect of the Ekman friction, and these continue to grow by the Holopainen mechanism (cf. Figs. 3b,c and surrounding discussion). By $t = 2000$, a significant positive PV gradient has emerged because of the induced mean flow, which acts to stabilize the system.

In summary, Expt V undergoes a (near) frictionless baroclinic life cycle on a time scale T_I, during which time the initial negative PV gradient in the lower layer is (partially) homogenized by stirring. The life cycle is followed by a much longer adjustment on time scale T_R, during which forced vortices due to Ekman friction emerge within the lower layer and subsequently generate a positive background PV gradient associated with the mean flow correction. An equilibrium determined by the Ekman friction is eventually reached.

The behavior seen in Expt V is typical of that throughout the mixed behavior region of parameter space identified above. Figure 8 shows maximum amplitudes $|A_1^s|_{\text{eq}}$ (solid points) and equilibrated amplitudes $|A_1^s|_{\text{eq}}$ (unfilled points) for simulations with $\kappa = 0.0025, 0.005$, and 0.02 and $\beta \in [0.2, 0.55]$. The solid curves show the predictions of WGP for the maximum amplitude [(21)] and Romea [(39)] for the equilibrated amplitude in the limit $\kappa \to 0$. For $\beta > 0.5$, as discussed in the context of Expt I above, the maximum and equilibrated amplitudes are approximately equal, and Romea's theory makes reasonable predictions of both, particularly as $\kappa \to 0$. At $\beta = 0.5$, there is a pivot point. The equilibrated amplitude continues to be accurately predicted by Romea's theory, but the maximum amplitude is now much better predicted by the WGP

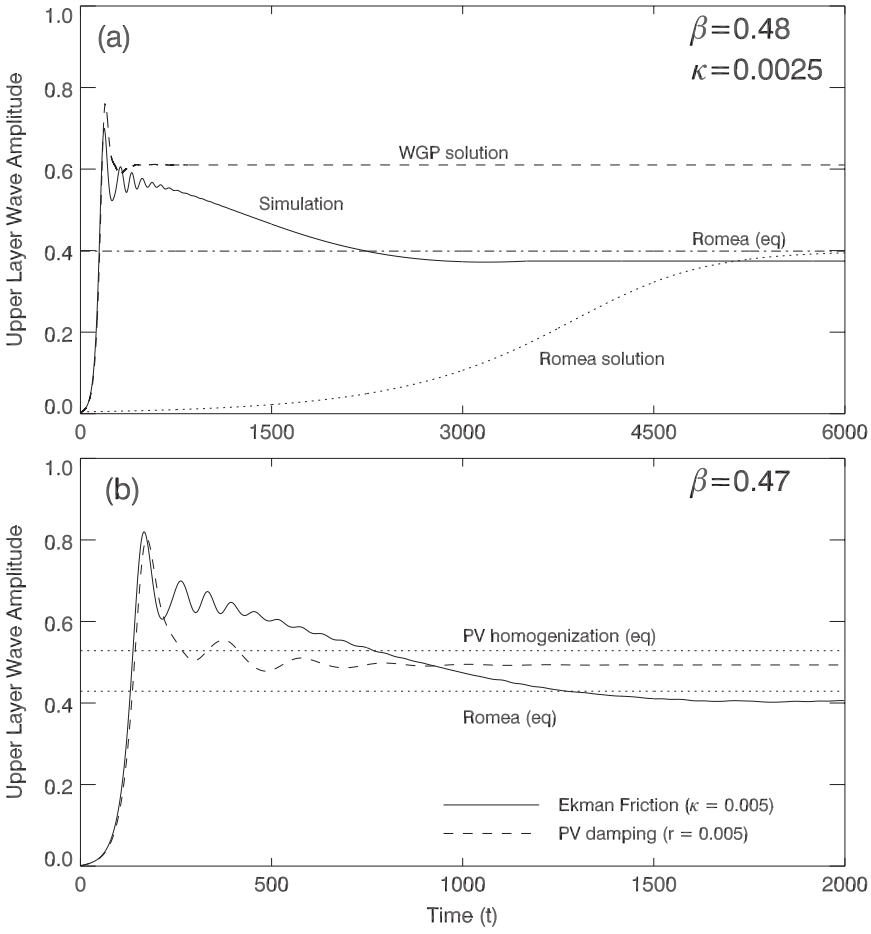

FIG. 6. (a) Evolution in time of wave amplitude $|A_1^s|$ during Expt V [$(\beta, \kappa) = (0.48, 0.0025)$; solid curve]. The dashed curve shows the corresponding WGP solution [for $\kappa = 0$; see (20)], the dotted curve shows the Romea solution, [see (37)] and the dotted line shows the asymptotic value for Romea's solution. (b) As in (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).

theory. The relevance of Romea's theory for $\beta < ½$—and in the mixed behavior region in particular—is a new and surprising result and is one of the main findings of this work.

5. Conclusions

The main objective of this work has been to determine the relevance of two different asymptotic theories, with each formally valid under different assumptions, to baroclinic life cycles in a β channel at finite supercriticality and in the presence of Ekman friction. The results presented allow the three questions posed in the introduction to be answered as follows:

(i) The WGP theory accurately describes the early time behavior of the baroclinic life cycle for flows that are unstable according to the frictionless criterion ($\beta < ½$), provided that the supercriticality is not too large [$\epsilon = (1/2 - \beta)^{1/2} \ll 1$] and Ekman friction is relatively weak ($\kappa \lesssim 0.05$). The time scale for the wave to equilibrate, via the formation and rollup of a nonlinear critical layer in the lower layer, is $T_I \sim \varepsilon^{-1}$. The maximum wave amplitude during the life cycle is attained on this time scale and is well predicted by WGP. Typically thereafter, as seen in Expt V described in section 4c, the upper-layer wave amplitude then briefly oscillates around the value predicted by WGP as reported in EW11. On a longer time scale $T_R = \kappa^{-1}\varepsilon^{-2}$, where ε is the Ekman friction supercriticality [(22)], the wave amplitude decays from the WGP predicted value. Isolated vortices intensify in the lower layer and obliterate all trace of the early time critical layer evolution. The late time behavior is reminiscent

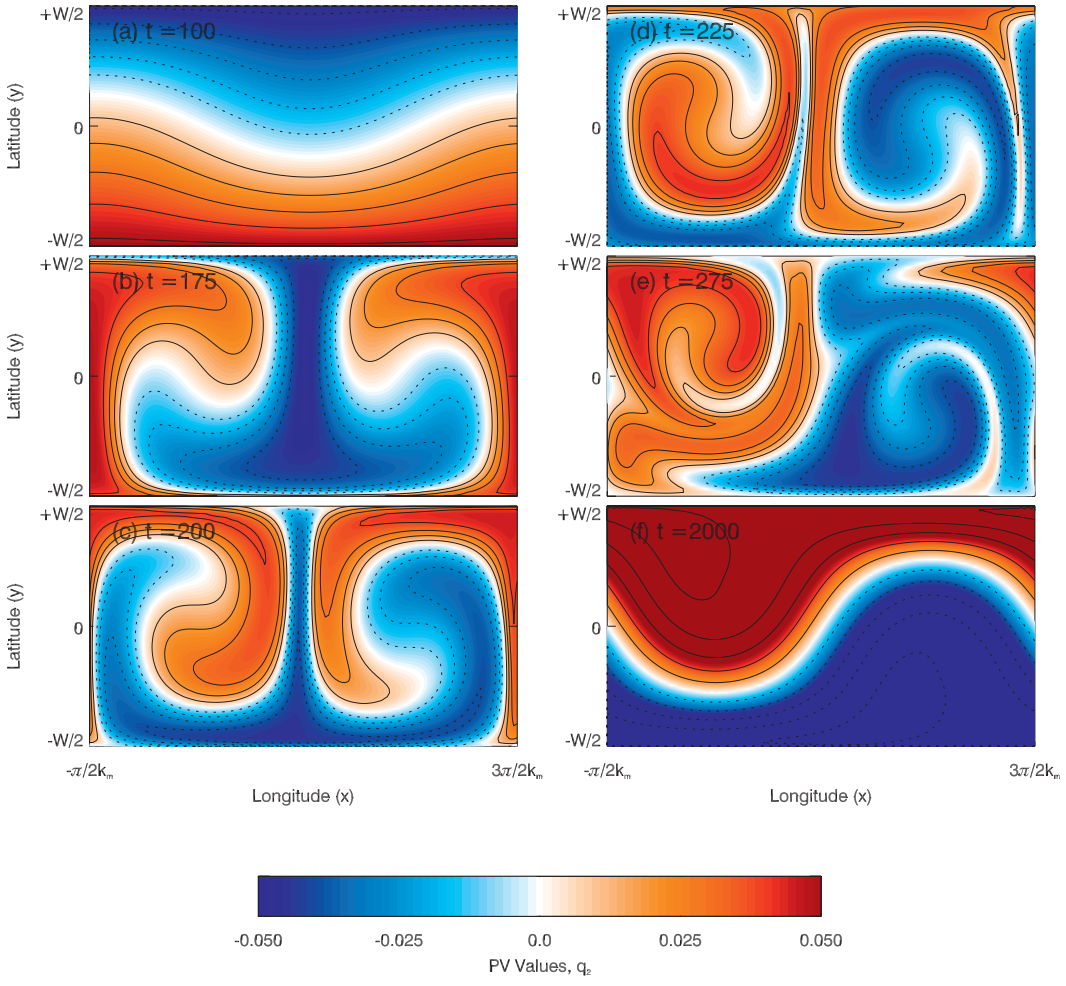

FIG. 7. Snapshots of the lower-layer PV field for Expt V [$(\beta, \kappa) = (0.48, 0.0025)$] at $t = 100, 175, 200, 225, 275$, and 2000. Blue/dashed contours correspond to negative PV values, and red/solid contours correspond to positive PV. Contour intervals are (a)–(e) 0.01 and (f) 0.02.

instead of the equilibration of dissipatively destabilized waves.

(ii) The Romea theory accurately describes the nonlinear evolution of dissipatively destabilized waves that grow because of the Holopainen instability but are stable to frictionless baroclinic instability ($1/2 < \beta < \beta_m$) for low Ekman friction ($\kappa \lesssim 0.05$): for example, Expt I in section 4a. At higher Ekman numbers ($0.05 \lesssim \kappa \lesssim 0.2$) for which $\beta_m < \frac{1}{2}$ (e.g., Expt II), Romea is also found to be accurate near marginal stability [i.e., for $\varepsilon = (\beta_c(\kappa) - \beta)^{1/2} \ll 1$]. For flows subject to baroclinic instability at low Ekman number ($\beta < \frac{1}{2}$ and $\kappa \ll 1$; i.e., for which Romea's theory has not been thought relevant; e.g., Boville 1981), the equilibrium wave amplitude and zonal flow structure are nevertheless well predicted by Romea. The equilibrium state is established on the time scale T_R, which throughout much of parameter space is much longer than the WGP time scale T_I. Hence, there is no contradiction between the short-time behavior of such flows being well described by WGP, whereas the equilibrium behavior is captured by Romea.

(iii) The frictionless baroclinic adjustment hypotheses for nonlinear equilibration, Vallis (see, e.g., section 12.6.2 of Vallis 2006) hypothesizes that in general flows adjust through PV homogenization to a state that is stable under the Charney–Stern–Pedlosky criterion (i.e., a state with single-signed PV gradients throughout the domain). The equilibrium state in the WGP solution is an example of an "adjusted" state of this type (as pointed out by Shepherd 1993), because the lower-layer PV is completely homogenized at late times in the WGP solution. Clearly, however, in the presence of Ekman friction the equilibrated states differs significantly from the

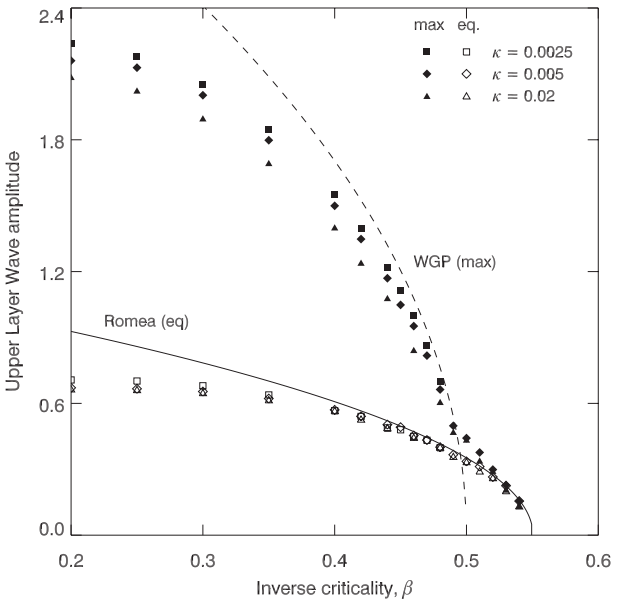

FIG. 8. Equilibrated amplitude $|A_1^s|_{eq}$ (unfilled points) and maximum amplitude $|A_1^s|_{max}$ (solid points) as a function of β in simulations with $\kappa = 0.0025$, 0.005, and 0.02 (as labeled). The dashed curve shows WGP prediction for the maximum amplitude [Eq. (21)], and the solid curve shows Romea's prediction for the equilibrated amplitude for $\kappa \to 0$ [Eq. (39)].

adjusted states, because, for example, Fig. 7f reveals the equilibrated state of Expt V has a significant positive PV gradient in the lower layer, indicating that ''overadjustment'' has taken place. Hence, nonlinear baroclinic equilibration in the presence of Ekman friction presents a clear example of a flow for which the baroclinic adjustment hypotheses fails.

The conclusions above raise wider questions about the validity of adjustment arguments for predicting equilibrated states. Dynamical systems for which the adjustment hypothesis will be valid are those for which (weakly) forced-dissipative experiments equilibrate at a state close to a stable equilibrium of the corresponding unforced system. Examples include weakly forced convection (through radiation or boundary conditions), and indeed the two-layer β-channel flows studied here in the special case where the forcing/dissipation is applied directly to the PV. The case of PV damping is a singular situation (for further discussion, see, e.g., Pedlosky 1971; Warn and Gauthier 1989) in which the Holopainen instability vanishes, in contrast to other forms of damping, which might quantitatively change both the Holopainen instability and its nonlinear equilibration but leave the qualitative behavior essentially unaltered (e.g., Esler 1997). Expt VIb is an example of a flow with PV damping, and in Fig. 6b it is shown to convergence to an equilibrium state close to that predicted by the forced-dissipative formulation of the WGP theory (e.g., section 4 of Warn and Gauthier 1989), with the prediction being $\epsilon W/\sqrt{3}$ (valid in the limit of weak PV damping $r \to 0$). The equilibrated state of Expt VIb can be described as an adjusted state as the lower-layer PV (not shown) is homogenized, just as occurs for the unforced life cycles discussed in EW11.

The feature of baroclinic instability in the presence of Ekman friction that prevents adjustment arguments from being correct is that the flow must also be rendered stable to the dissipatively destabilized waves. In this respect, it is important to note that dissipative destabilization is not a feature peculiar to the Phillips model. In the two-layer model, a recent energetics budget study by Lee (2010) indicates that dissipative destabilization provides an eddy energy source in a more realistic jet flow. In the Eady model, Weng (1990) and Weng and Barcilon (1991) have shown that dissipative destabilization analogous to that in the Phillips model occurs when sloping boundaries (the so-called topographic β effect) are present. The situation in other models appears to be more complicated (e.g., Lin and Pierrehumbert 1988; Panetta et al. 1988), but it can be concluded that a better understanding the mechanism of the dissipative destabilization could provide an important step toward understanding baroclinic equilibration in a wider context. Swaters (2010) has made a step forward in this regard with a modal interpretation of the instability mechanism.

The robust feature of the present study is that it gives a clear example of how equilibration can occur by a mechanism distinct from that hypothesized by adjustment arguments. It is this feature that is likely to be universal. The fact that Romea's theory has been shown to have considerable predictive power for the equilibrated state across much of parameter space suggests a similar approach may bring insight into a wider range of flow scenarios, with the most relevant physical situations being wind-driven flow in an oceanic channel and rotating annulus experiments.

Acknowledgments. JGE acknowledges support from the UK Natural Environment Research Council Grant NE/G003122/1, and BTW acknowledges the support of an EPSRC studentship.

REFERENCES

Boville, B. A., 1981: Amplitude vacillation on a β-plane. *J. Atmos. Sci.,* **38,** 609–618.

Bretherton, F. P., 1966: Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. *Quart. J. Roy. Meteor. Soc.,* **92,** 335–345.

Esler, J. G., 1997: Simple models of wave packets in an equilibrated baroclinic system. *J. Atmos. Sci.,* **54,** 2820–2849.

——, and P. H. Haynes, 1999: Mechanisms for wave packet formation and maintenance in a quasigeostrophic two-layer model. *J. Atmos. Sci.,* **56,** 2457–2489.

——, and B. T. Willcocks, 2011: Nonlinear baroclinic equilibration at finite supercriticality. *Geophys. Astrophys. Fluid Dyn.,* in press.

Holopainen, E., 1961: On the effect of friction in baroclinic waves. *Tellus,* **13,** 363–367.

Krechetnikov, R., and J. E. Marsden, 2009: Dissipation-induced instability phenomena in infinite-dimensional systems. *Arch. Ration. Mech. Anal.,* **194,** 611–668.

Landau, L. D., 1944: On the problem of turbulence. *Dokl. Akad. Nauk SSSR,* **44,** 311.

Lee, S., 2010: Dissipative energization of baroclinic waves by surface Ekman pumping. *J. Atmos. Sci.,* **67,** 2251–2259.

Lin, S.-J., and R. T. Pierrehumbert, 1988: Does Ekman friction suppress baroclinic instability? *J. Atmos. Sci.,* **45,** 2920–2933.

Panetta, R. L., I. M. Held, and R. T. Pierrehumbert, 1988: External Rossby waves in the 2-layer model. *J. Atmos. Sci.,* **44,** 2924–2933.

Pedlosky, J., 1970: Finite amplitude baroclinic waves. *J. Atmos. Sci.,* **27,** 15–30.

——, 1971: Finite-amplitude baroclinic waves with small dissipation. *J. Atmos. Sci.,* **28,** 587–597.

——, 1982a: Finite-amplitude baroclinic waves at minimum critical shear. *J. Atmos. Sci.,* **39,** 555–562.

——, 1982b: A simple model for nonlinear critical layers in an unstable baroclinic wave. *J. Atmos. Sci.,* **39,** 2119–2127.

——, 1987: *Geophysical Fluid Dynamics*. 2nd ed. Springer-Verlag, 710 pp.

Phillips, N. A., 1951: A simple three-dimensional model for the study of large-scale extratropical flow patterns. *J. Meteor.,* **8,** 381–394.

——, 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level quasi-geostrophic model. *Tellus,* **6,** 273–286.

Pierrehumbert, R. T., and K. L. Swanson, 1995: Baroclinic instability. *Annu. Rev. Fluid Mech.,* **27,** 419–467.

Romea, R. D., 1977: The effects of friction and β on finite amplitude baroclinic waves. *J. Atmos. Sci.,* **34,** 1689–1695.

Shepherd, T. G., 1988: Nonlinear saturation of baroclinic instability. Part I: The two-layer model. *J. Atmos. Sci.,* **45,** 2014–2025.

——, 1993: Nonlinear saturation of baroclinic instability. Part III: Bounds on the energy. *J. Atmos. Sci.,* **50,** 2697–2709.

Stuart, J. T., 1960: On the nonlinear mechanics of wave disturbances in stable and unstable parallel flows. Part 1: The basic behaviour in plane Poiseuille flow. *J. Fluid Mech.,* **9,** 353–370.

Swaters, G. E., 2010: Modal interpretation for the Ekman destabilization of inviscidly stable baroclinic flow in the Phillips model. *J. Phys. Oceanogr.,* **40,** 830–839.

Vallis, G. K., 2006: *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. Cambridge University Press, 745 pp.

Warn, T., and P. Gauthier, 1989: Potential vorticity mixing by marginally unstable baroclinic disturbances. *Tellus,* **41A,** 115–131.

Weng, H.-Y., 1990: The effects of oppositely sloping boundaries with Ekman dissipation in a nonlinear baroclinic system. *Quart. J. Roy. Meteor. Soc.,* **116,** 1–29.

——, and A. Barcilon, 1991: Asymmetric Ekman dissipation, sloping boundaries and linear baroclinic instability. *Geophys. Astrophys. Fluid Dyn.,* **59,** 1–24.

Willcocks, B. T., 2011: Instability and nonlinear equilibration of baroclinic flows. Ph.D. thesis, University College London, 261 pp.