J. Fluid Mech. (2001), *vol.* 428, *pp.* 387–408. Printed in the United Kingdom

Waves and resonance in free-boundary baroclinic instability

By P. RIPA

CICESE, Km. 107 carretera Tijuana-Ensenada (22860) Ensenada, BC, México

(Received 11 November 1999 and in revised form 2 August 2000)

Eady's model of baroclinic instability has been generalized by including β (the meridional gradient of planetary potential vorticity) while assuming that total potential vorticity is uniform. Moreover, the problems of Eady and of Phillips have been enriched by including a fixed topography or a free boundary (which implies a flow-dependent geostrophic topography). The most general cases (with β, fixed topography and a free boundary) of both problems are shown to have nearly identical stability properties, mainly determined by two Charney numbers: the planetary one and a topographic one. The question of whether this generalized baroclinic instability problem can be described by wave resonance or component 'resonance' is addressed. By *waves* are meant physical modes, which could freely propagate by themselves but are effectively coupled by an independent basic shear, producing the instability. *Components*, on the other hand, are mathematical modes for which the shear is also crucial for their existence, not just for their coupling, hence the quotation marks around 'resonance'. In this paper it is shown that both scenarios, components 'resonance' and waves resonance, cast light on the free-boundary baroclinic instability problem by providing explanations of the instability onset (at minimum shear) and maximum growth rate cases, respectively. The importance of the mode pseudomomentum for the fulfillment of both mechanisms is also stressed.

1. Introduction

Eady's (1949) three-dimensional baroclinic instability problem on the f-plane was enriched by Blumsack & Gierasch (1972) by adding the possibility of a fixed topography, whilst Fukamachi, McCreary & Proehl (1995), in a particular case, and more generally Beron-Vera & Ripa (1997) and Ripa (2000, hereafter denoted by R2K) allowed one of the boundaries to be free – which implies a flow-dependent geostrophic topography.† On the other hand, Lindzen (1994) included the planetary β-effect under the simplifying assumption of uniform potential vorticity. Similarly, Phillips' (1951) two-layer baroclinic instability model was modified by Bretherton (1966), who made $\beta = 0$ and assumed both boundaries to have the same slope, and by Olascoaga & Ripa (1999, hereafter denoted by O99) who made the lower boundary free. The most general case (with β, fixed topography and/or a free boundary) is shown in this paper to have nearly identical instability properties in both problems (three-dimensional or two layers). The nature of this general baroclinic instability is also investigated

† By 'topography' it is meant here a non-horizontal boundary felt by the perturbation. It might be fixed or due to the basic shear. It could also be located at the top, since a freely evolving QG system is invariant under up/down reflection.

here. Given all parameters of the problem, it is not difficult to obtain the dispersion relation. The goal here is to isolate the physically relevant parameters, to seek scenarios for the instability onset and maximum growth rate cases, and to unify the work of different authors. Typical questions addressed here are: Does a sloping top boundary add anything qualitatively different? Can the bounds on the growth of a finite-amplitude perturbation derived by other authors be used with the generalized problem studied here? Is the instability produced by some kind of resonance?

Many types of shear flow instabilities are found near the frequency/wavenumber coincidence of two elementary modes, whose coupling is parameterized by a basic velocity variation U_s (see for instance Hayashi & Young 1987; Sakai 1989; Ripa 1992; Iga 1993, 1997). Thus, in Eady's classical baroclinic instability problem on the f-plane, those modes are zero potential vorticity perturbations, driven by the density gradient $\nabla\rho$ in each horizontal boundary (Bretherton 1966). The thermal wind shear associated with $\nabla\rho$ Doppler shifts each component differently, allowing for a 'resonance' at the non-dimensional horizontal wavenumber $\kappa = 2.0653$ (instability actually occurs for $0 < \kappa < 2.3994$, whereas maximum growth rate is attained at $\kappa = 1.6061$). Baines & Mitsudera (1994) have discussed a similar mechanism in non-rotating systems with piecewise uniform basic density and (horizontal) vorticity profiles: the instability is associated with a couple of modes trapped at two vorticity interfaces (or one vorticity and one density interface). If the modes are phase-locked, with the correct phase, the advection of each perturbation by the velocity field due to the other one produces the instability. In all these cases, the components that 'resonate' do not travel freely without the shear; instead, this is an essential part of the restoring mechanism that allows the modes to propagate. In some cases, even though these components coincide with the waves of some physical system, this is quite different from that of the (instability) problem under consideration (see Ripa 1992).

One may wonder whether baroclinic instability can be associated with the resonance of true waves, i.e. physical modes which can freely propagate in the absence of the basic shear. Ripa & Marinone (1983), for instance, found an equatorial jet instability mode whose dispersion curve looked like the merging of the Yanai and first meridional mode Rossby waves; however, on including only those two waves in the expansion of the perturbation, the instability disappeared. However, Rossby waves resonance is indeed associated with baroclinic instability when two contributions to the gradient of ambient potential vorticity are present: the planetary β and a topographic β_T (the latter might be due to a fixed boundary slope and/or to the geostrophic slope of a free boundary). The enhancement of baroclinic instability near $\beta_T = -\beta$ and its relationship to Rossby waves has been analysed under the simplifying assumption that both the basic flow and the perturbation have a linear structure with depth (Ripa 1999*b*). This approximated model as well as the generalized Phillips problem have only two vertical modes. The generalized Eady system used here, on the other hand, has an infinity number of vertical modes: the instability is nevertheless shown to be mainly associated with the first two Rossby waves; higher modes do not matter much.

Neither component 'resonance' nor wave resonance is sufficient for instability: the phase-locked modes must also have opposite signs of pseudomomentum. Although instability has been associated with negative energy (Cairns 1979; Morrison & Kotschenreuther 1990) or, rather, pseudoenergy (Ripa 1990), the pseudomomentum appears to be more useful because it is frame-independent and it can be linked to the over-reflection instability scenario (Takehiro & Hayashi 1992). In closing this Introduction, it is also worth pointing out that not all instabilities are associated with

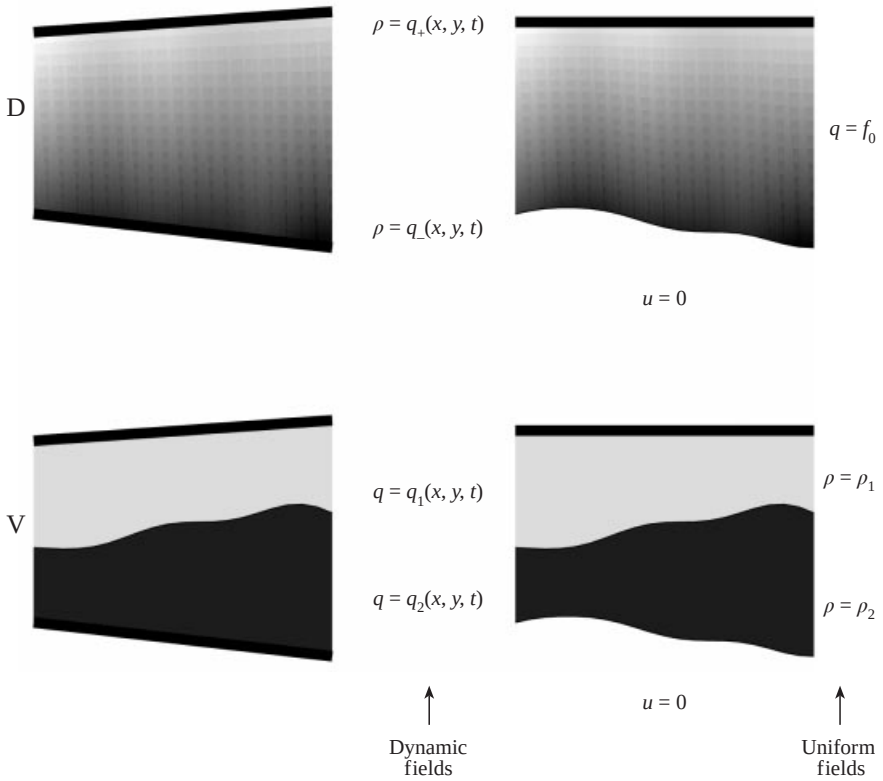

FIGURE 1. Two-field models of baroclinic instability. Vertical structure of the two types of model (D and V) (see table 1). Left-hand graphs: classical models (rigid boundaries). Right-hand graphs: the top line is the ocean surface and the bottom boundary, which is free, represents the connection with the deep ocean.

nearby neutral modes (see for instance Balmforth, Del-Castillo-Negrete & Young 1997; Samelson 1999).

The rest of this paper is organized as follows. Baroclinic instability for both models used here, three-dimensional and two-layer with β, topography and/or a free boundary, is presented in § 2, where the component 'resonance' interpretation is in particular discussed. In § 3 the unstable shear is not included and the restriction of uniform potential vorticity is relaxed, in order to study the Rossby waves of the three-dimensional model. The properties of these waves are interesting *per se* and are used, in particular, to verify the wave resonance mechanism for the enhancement of baroclinic instability. The Appendix is devoted to the derivation of the maximum growth rate case. The final discussion is presented in § 4.

2. Models

Several quasi-geostrophic models of baroclinic instability are describable by two fields. If the vertical domain is written as

$$z_- \leqslant z \leqslant z_+,$$

	Top and bottom boundaries		
	Rigid horizontal	Rigid sloping	Free
Model D	Eady (1949)	Blumsack & Gierasch (1972)	Beron-Vera & Ripa (1997)
Model V	Phillips (1951)	Bretherton (1966)	Olascoaga & Ripa (1999)
Boundary slope	$v = 0$	$v \neq 0$	$v \neq 0$
Interior strat.	$s = 0$	$s = 0$	$s \neq 0$

TABLE 1. Two-field models of baroclinic instability (see figure 1). Model D: there is one uniform potential vorticity layer and the dynamical fields are the *density* in each boundary q_+ and q_-. Model V: there are two uniform density layers and the dynamical fields are the potential *vorticity* in each layer q_1 and q_2. The non-dimensional boundary slope v is proportional to the 'topographic β' whereas s is the ratio of the interior stratification to the buoyancy jump at the free boundary.

different assumptions are made about the top $z = z_+$ and bottom $z = z_-$ boundaries and the vertical structure (see table 1 and figure 1). The simplest horizontal setting is that of a zonal channel and a basic flow such that $Q_n' (= \mathrm{d}Q_n/\mathrm{d}y) = \text{const.}$, where y is the across-channel coordinate. The interest here lies in the possibility of making one boundary free ($s > 0$), i.e. allowing it to be modified by the motion itself; such a boundary is interpreted to be the interface to a deep and motionless ocean (the reduced gravity setting) although it could also be the top one.

For the structure of Model D, buoyancy is written as $\vartheta = \Theta(z - \zeta)$ in terms of a reference profile $\Theta(z)$ and the isopycnal vertical displacement ζ. The buoyancy jump at the interface in the reference state, $g_b = \Theta(-H + 0) - \Theta(-H - 0)$, and the stratification, $N^2 = \Theta'(z)$, are used to define the 'external' and 'internal' length scales by

$$R = \frac{\sqrt{g_b H}}{|f_0|}, \quad L = \frac{NH}{|f_0|}.$$

For simplicity N^2 will be assumed to be constant, unless otherwise noticed. As in R2K, the interior stratification is parametrized by

$$s = \frac{N^2 H}{g_b} \equiv \frac{L^2}{R^2}.$$

Dynamical fields are the quasi-geostrophic (QG) potential vorticity and the densities at the top and bottom boundaries,

$$\begin{aligned} q(\boldsymbol{x}, z, t) &= f + \hat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{u} - f_0 \partial_z \zeta \quad && (z_- < z < z_+), \\ q_\pm(\boldsymbol{x}, t) &= \mp f_0 H^{-1} (z - \zeta) && (z = z_\pm), \end{aligned}$$

where the factors $\mp f_0/H$ are introduced for convenience. In a β-plane geometry, the horizontal velocity and dynamical fields are calculated from the streamfunction as $\boldsymbol{u} = \hat{\boldsymbol{z}} \times \nabla\psi$ and

$$\left.\begin{aligned} q &= f_0 + \beta y + \nabla^2 \psi + \partial_z (f_0^2 N^{-2} \partial_z \psi) && (-H < z < 0), \\ q_+ &= -f_0 H^{-1} (z_+ + f_0 N^{-2} \partial_z \psi) && (z = 0), \\ q_- &= f_0 H^{-1} (z_- + f_0 N^{-2} \partial_z \psi) && (z = -H). \end{aligned}\right\} \tag{2.1}$$

Note that even though the water column extends from a soft boundary to a rigid one,

$$z_- = -H - f_0 g_b^{-1} \psi_-, \quad z_+ = Z'_+ y,$$

the calculations are made in $-H \leqslant z \leqslant 0$, in a manner that is typical of the quasi-geostrophic approximation; hereafter the subscripts $+$ and $-$ mean evaluated (or defined) at $z = 0$ and $z = -H$, respectively. Following Lindzen (1994), a uniform potential vorticity field is assumed here for the interior, namely

$$q \equiv f_0 \tag{2.2}$$

even for $\beta \neq 0$. This constraint (which will be relaxed in § 3) leaves $q_\pm$ as the only dynamical fields, whose evolution is controlled by density conservation at both boundaries

$$(\partial_t + \boldsymbol{u}_\pm \cdot \nabla) q_\pm = 0,$$

equations which conserve energy, momentum, and a family of Casimirs:

$$\mathscr{E} = -\tfrac{1}{2} \langle \psi_+ q_+ + \psi_- q_- \rangle, \quad \mathscr{M} = \langle y q_+ + y q_- \rangle, \quad \mathscr{C}_\pm^{(n)} = \langle q_\pm^n \rangle, \tag{2.3}$$

where $\langle \cdots \rangle$ denotes an horizontal average. (These definitions are made modulo Kelvin circulations at any depth at each coast, which are also Casimirs and therefore integrals of motion.)

A zonal basic current is defined as having a vanishing horizontal shear, with top and bottom velocities equal to

$$\{U_+, U_-\} = \{U_b + U_s, U_b\},$$

and with a vertical curvature $U''(z) = \beta N^2 f_0^{-2}$ in order to satisfy (2.2):

$$U(z) = U_b + \left(1 + \frac{z}{H}\right) U_s + \tfrac{1}{2} \beta L^2 \left(1 + \frac{z}{H}\right) \frac{z}{H}. \tag{2.4}$$

The dynamical fields in the basic state have the form $Q_\pm = Q'_\pm y$. Let top and bottom boundaries have slopes $Z'_\pm$ in the basic state. Each slope may be prescribed or determined by the basic current, depending upon whether the boundary is rigid or soft. The driving gradients $Q'_\pm$ are functions of those slopes, β, and the mean shear U_s, namely

$$Q'_\pm = \pm U_s L^{-2} \mp f_0 H^{-1} Z'_\pm + \tfrac{1}{2} \beta. \tag{2.5}$$

The equations for the perturbation fields $\delta q_\pm := q_\pm - Q_\pm$ are

$$(\partial_t + U_\pm \partial_x) \delta q_\pm + \delta v_\pm Q'_\pm = -\delta \boldsymbol{u}_\pm \cdot \nabla \delta q_\pm. \tag{2.6}$$

From these equations or (2.3) it follows that the pseudomomentum

$$\mathscr{M}_p[\delta q] = -\frac{1}{2} \left\langle \frac{\delta q_+^2}{Q'_+} + \frac{\delta q_-^2}{Q'_-} \right\rangle \tag{2.7}$$

$= \mathscr{M} - \tfrac{1}{2} \mathscr{C}_+^{(2)} / Q'_+ - \tfrac{1}{2} \mathscr{C}_-^{(2)} / Q'_-$, is an exact nonlinear integral of motion of the evolution equations (2.6). Consequently if $Q'_+ Q'_- > 0$ (density increases in opposite directions at the top and bottom boundary) there is normed stability (and hence normal mode stability) in the boundary density variance norm $\|\delta q\|^2 = \langle \delta q_+^2 + \delta q_-^2 \rangle$. Defining $a_\pm$ by writing (2.5) as

$$Q'_\pm = \pm (U_s - a_\pm) L^{-2}, \tag{2.8}$$

the normed stability condition $Q'_+ Q'_- > 0$ translates into $(U_s - a_+)(a_- - U_s) > 0$,

i.e. U_s between the critical shears $a_\pm$. The width of the probable stable shears is $a_- - a_+ = (\beta_T + \beta)L^2$, where

$$\beta_T = f_0 H^{-1}(Z'_- - Z'_+)$$

is the topographic beta parameter. Consequently, if there is a cancellation of both beta effects, $\beta_T + \beta = 0$, no shear is probably stable by pseudomomentum conservation.

A very interesting result is that the two-layer Model V has the same type of evolution equations and conservation laws. The dynamical fields are

$$q_1 = f_0 + \beta y + \nabla^2\psi_1 + 2L^{-2}(\psi_1 - \psi_2),$$
$$q_2 = f_0 + \beta y + \nabla^2\psi_2 + 2L^{-2}(\psi_2 - \psi_1) + 2R^{-2}\psi_2,$$

where now $L^2 = g'H/f_0^2$, with g' the buoyancy jump between both active layers (see O99). Thus, for a basic state with uniform currents $\{U_1, U_2\} = \{U_b + U_s, U_b\}$, the values $\{Q'_1, Q'_2\}$ are found to be exactly twice those of $\{Q'_+, Q'_-\}$ from (2.5). Unless otherwise noted, all the results in the rest of this section are valid for either Model D or V (see table 1 and figure 1). The critical shear $U_s = a_-$ [$U_s = a_+$] corresponds to an isopycnal lower [upper] boundary, for Model D, or an uniform potential vorticity lower [upper] layer, for Model V.

Even though there are three wave restoring parameters, β and $Z'_\pm$, it will be shown that the dispersion relation depends only on two combinations of them: the critical shears a_+ and a_-. Non-dimensional parameters are obtained dividing these by the destabilizing parameter U_s, say

$$\begin{pmatrix} a_+ \\ a_- \end{pmatrix} = \begin{pmatrix} -b/2 \\ v + b/2 \end{pmatrix} U_s. \tag{2.9}$$

The results derived by different authors for Models D or V can then be made equivalent by choosing equal values of b and v. Thus, the particular case of Model V with rigid boundaries ($s = 0$) studied by Bretherton (1966) has $\beta = 0$ and uniform total depth ($Z'_+ = Z'_-$) which imply $a_+ = a_-$: this is then equivalent to the problem studied in O99 ($\beta > 0$, $Z'_+ = 0$, $Z'_- \neq 0$) along the line $b = -v$, which, incidentally, is where the maximum growth occurs for small s. Also, O99 proved that there is formal stability outside the wedge $-2 \leqslant b \leqslant 2 - 2v$ for Model V, a result which is also valid for Model D defined here. Results derived from pseudomomentum conservation for Models D or V can also be made equivalent by choosing the same values of

$$\frac{Q'_+}{Q'_-} \equiv \frac{Q'_1}{Q'_2} = \frac{b+2}{b+2v-2}.$$

Thus, finite amplitude bounds on the wavy part of the perturbation derived for the classical Phillips problem (Shepherd 1988), which corresponds to $v = 0$, can be used in the present one making $b \mapsto (b + v)/(1 - v/2)$. Similarly, bounds on the total perturbation derived for the present problem assuming $b = 0$ (Ripa 1999*a*) can be used for the more general case studied here making the transformation $v \mapsto (b + v)/(1 + b/2)$.

Since only two parameters are important, from here on an horizontal 'rigid lid' surface and a soft interface to the deep ocean will be chosen. This yields

$$\left.\begin{aligned} Z'_+ &= 0 \\ Z'_- &= f_0 g_b^{-1} U_b \end{aligned}\right\} : b = \frac{\beta L^2}{U_s}, \quad v = \frac{\beta_T L^2}{U_s} \equiv \frac{sU_b}{U_s},$$

namely v and b are the two Charney numbers v and b, defined in O99 and R2K.

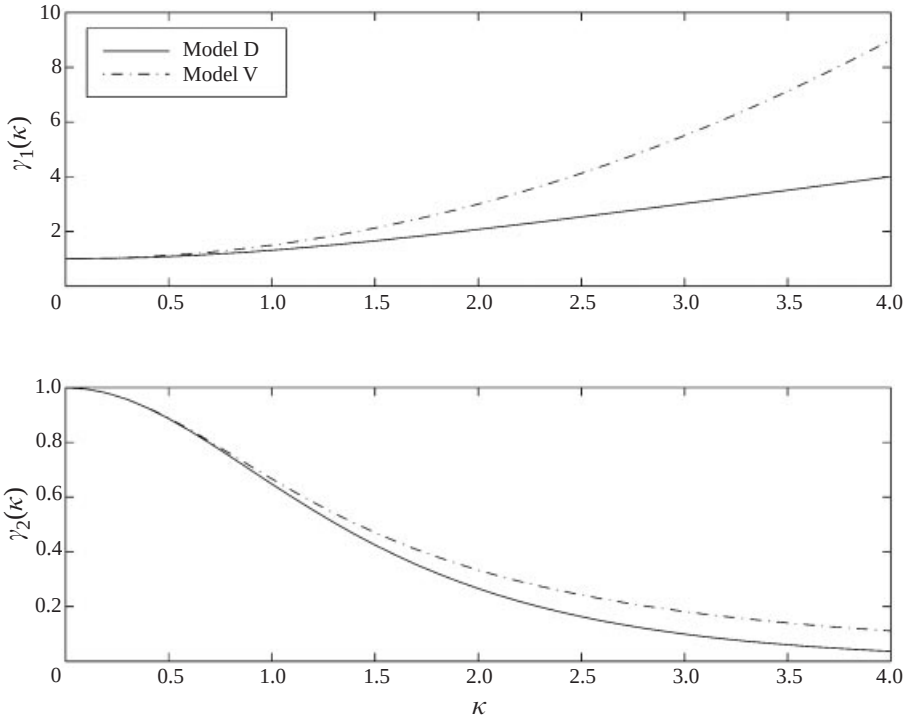

FIGURE 2. Parameters γ_1 and γ_2 that define the energy matrix (2.10) in wavenumber space; γ_1 gives a scale for the importance of the stratification parameter s, whereas γ_2 determines the coupling beween the two q modes (see table 1).

In order to integrate the evolution equations (2.6), it is necessary to calculate the velocity fields, $\boldsymbol{u}_j = \hat{\boldsymbol{z}} \times \nabla\psi_j$, from the dynamical fields q_j, where $j = \pm$ for Model D and $j = 1, 2$ for Model V. In the first case, $\psi_\pm$ are calculated by inverting the last two equations in (2.1), subject to uniform potential vorticity constraint (2.2) and the conditions of zero normal flow and constant Kelvin circulation at each coast (see R2K). Thus, defining

$$\delta\boldsymbol{q}(\boldsymbol{x}, t) = \begin{pmatrix} \delta q_+ \\ \delta q_- \end{pmatrix} = \sum_{k,l} \boldsymbol{q}_{k,l}(t) \mathrm{e}^{\mathrm{i}kx} \sin ly,$$

in R2K it is shown that

$$\begin{pmatrix} \delta\psi_+ \\ \delta\psi_- \end{pmatrix} = -L^2 \sum_{k,l} \mathbb{E}(s, \kappa) \boldsymbol{q}_{k,l}(t) \mathrm{e}^{\mathrm{i}kx} \sin ly$$

where the matrix †

$$\mathbb{E}(s, \kappa) = \frac{1}{\mu} \begin{pmatrix} 1 + s/\gamma_1 & \gamma_2 \\ \gamma_2 & 1 \end{pmatrix} \tag{2.10}$$

gives the energy in $\delta\boldsymbol{q}$ space for each Fourier component,

$$\mathscr{E}[\delta\boldsymbol{q}] = \tfrac{1}{4} L^2 \sum_{k,l} \mathrm{Re}(\boldsymbol{q}_{k,l}^\dagger \mathbb{E} \boldsymbol{q}_{k,l}).$$

† If $k \neq 0$; for $k = 0$ the matrix $\mathbb{E}$ is different because it has also the contribution of zero-frequency Kelvin modes, included to ensure constancy of Kelvin circulations (see R2K).

Similar equations hold for Model V (see O99); the only difference between the two types of model is in the definitions of γ_1 and γ_2, namely (see figure 2)

$$\text{Model} \begin{cases} \text{D}: & \gamma_1 = \kappa/\tanh\kappa, \quad \gamma_2 = 1/\cosh\kappa \\ \text{V}: & \gamma_1 = 1 + \tfrac{1}{2}\kappa^2, \quad \gamma_2 = \gamma_1^{-1}, \end{cases}$$

$$\text{Both models}: \quad \mu = \gamma_1(1-\gamma_2^2) + s,$$

where $\kappa = \sqrt{k^2+l^2}L$ is the non-dimensional horizontal wavenumber. Both models coincide for $\kappa \to 0$ but differ for $\kappa \gg 1$, particularly because γ_2 decreases as an exponential [power] with κ for Model D [V]. A rigid lower boundary corresponds to making $s = 0$ in $\mathbb{E}$; a finite value of v then implies that such a rigid boundary has a finite slope (see table 1). Eady's problem corresponds to Model D's $\mathbb{E}(\kappa)$ with $s = 0$, $v = 0$, and $b = 0$, whereas Phillips' problem corresponds to the Model V's $\mathbb{E}(\kappa)$ with $s = 0$ and $v = 0$.

Nonlinear dynamics has two quadratic integrals of motion: the pseudomomentum (2.7) and the pseudoenergy $\mathscr{E}_p = \mathscr{E} - \frac{1}{2}\mathscr{C}_+^{(2)}U_+/Q'_+ - \frac{1}{2}\mathscr{C}_-^{(2)}U_-/Q'_-$. It is convenient to combine them into the Hamiltonian

$$\mathscr{H}_\alpha[\delta\boldsymbol{q}] = \mathscr{E}_p - \alpha\mathscr{M}_p = \tfrac{1}{4}L^2\sum_{k,l}\mathrm{Re}[\boldsymbol{q}_{k,l}^\dagger\,\mathbb{H}\,\boldsymbol{q}_{k,l}],$$

where α is arbitrary,

$$\mathbb{H}(\alpha) = \mathbb{E}(s,\kappa) + \mathbb{C}(\hat{\alpha},v,b),$$

and the Casimir matrix is

$$\mathbb{C} = \begin{pmatrix} \dfrac{\alpha-U_+}{U_s-a_+} & 0 \\ 0 & \dfrac{\alpha-U_-}{a_--U_s} \end{pmatrix} \equiv \begin{pmatrix} \dfrac{\hat{\alpha}-1}{1+b/2} & 0 \\ 0 & \dfrac{\hat{\alpha}}{v+b/2-1} \end{pmatrix},$$

with $\hat{\alpha} := (\alpha - U_b)/U_s$. (The integral of motion $\mathscr{H}_\alpha$ may be called the 'free energy' measured in a frame moving with velocity α along x.)

If $\mathbb{H}(\alpha)$ is sign definite for some α, then the basic flow (characterized by s, v, and b) is stable to normal modes with wavenumber κ. Conversely, the instability region – where $\mathbb{H}(\alpha)$ is indefinite for any α – is bounded by the solutions of $P(\alpha) = P'(\alpha) = 0$, where $P(\alpha) = \det\mathbb{H}(\alpha)$ is quadratic in α. This is given by

$$\Delta := (\gamma_1 b_1 - b_2(\mu - \tfrac{1}{2}s))^2 + \mu\gamma_1\gamma_2^2(b_2^2 - \gamma_1^2) < 0, \tag{2.11}$$

where $b_1 = v + b + \frac{1}{2}s$, $b_2 = b + 2 - \gamma_1$. For fixed (s,κ), condition (2.11) represent an ellipse in the (v,b)-plane (see figure 3*a*). Equation (2.11) can also be seen as an implicit definition of short- and a long-wave cutoffs, namely (see figure 4)

$$\Delta < 0 \Longleftrightarrow \kappa_L(s,v,b) < \kappa < \kappa_S(s,v,b).$$

The short-wave cutoff can be made arbitrarily large, but $\kappa_S \to \infty$ results in rapidly shrinking ellipses in the (v,b)-plane (see figure 3*b*). The long-wave cutoff vanishes in the limit $\kappa \to 0$ of (2.11), namely

$$\kappa_L = 0 \Longleftrightarrow \Delta_0 := (v + b - \tfrac{1}{2}bs)^2 + s(b+1)^2 - s \leqslant 0,$$

which is the ellipse shown in figure 3(*b*). The ellipses (2.11) for all κ fill the wedge $-2 \leqslant b \leqslant 2 - 2v$, except for its tip, which is chopped at the segment of $\Delta_0 = 0$ that goes from $v = 2 - s$ in $b = -2$ to $v = 2 - s/(1+s)$ in $b = 2 - 2v$ (see figure 3*b*).

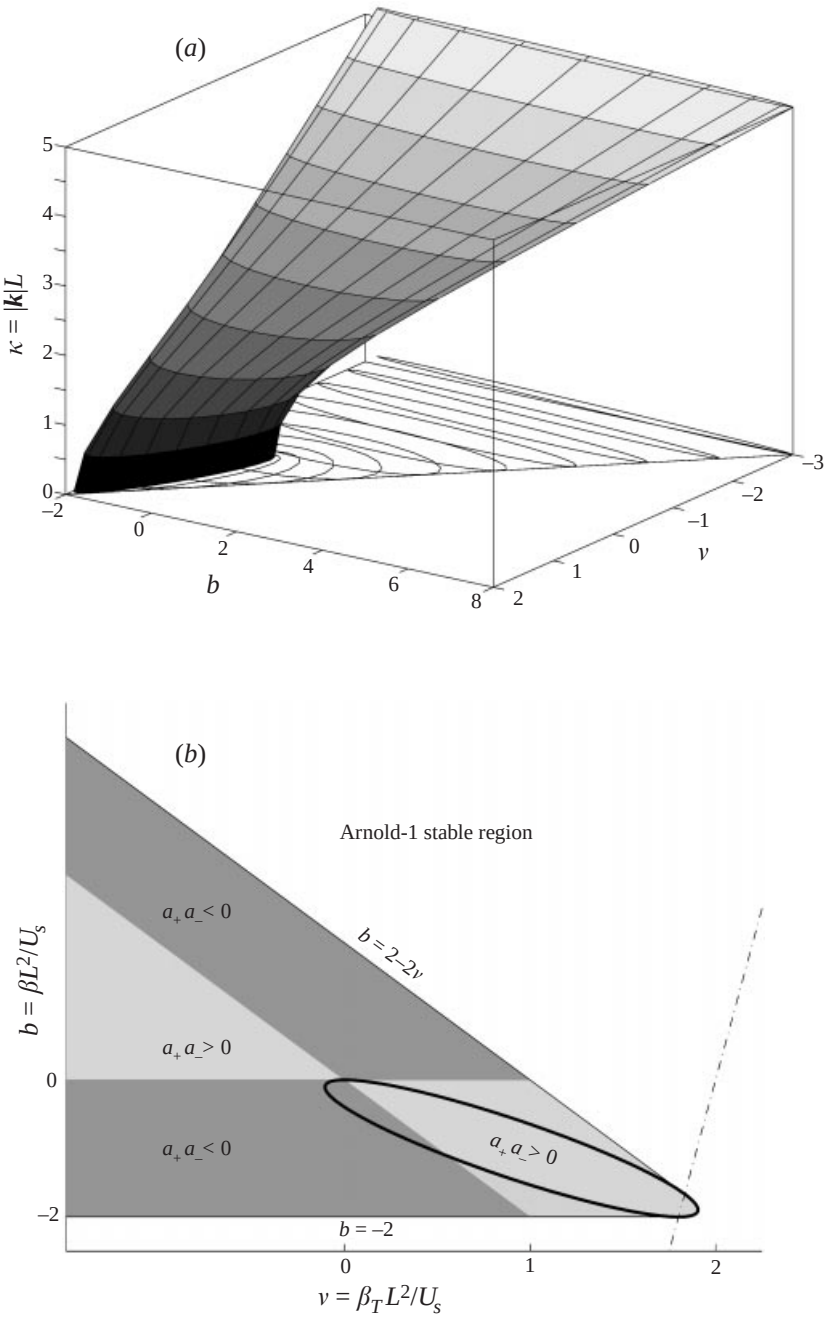

FIGURE 3. (*a*) Locus of indefinite Hamiltonians, in the space of planetary b and topographic ν Charney numbers, and non-dimensional wavenumber κ. This is also the region of unstable normal modes. (*b*) Normal mode instability regions ($s = 0.2$). In the (b, ν)-subspace, the instability is restricted to the wedge $-2 \leqslant b \leqslant 2-2\nu$. The shading indicates the relative sign of the critical shears $a_{\pm}$, defined in (2.5) and (2.8), which determines the type of dispersion relation, as shown in figure 4.

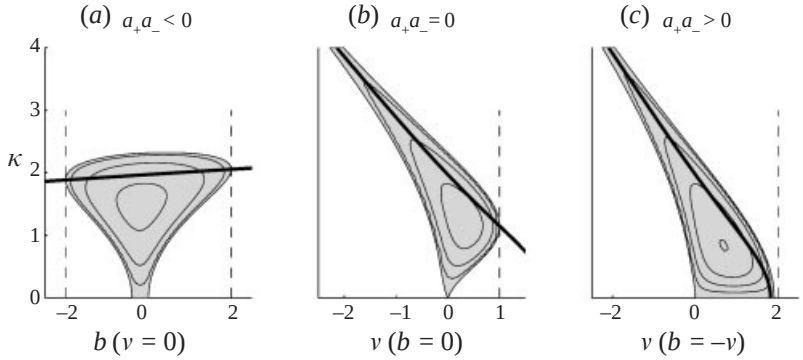

FIGURE 4. Instability region (shaded), growth rate (contours), critical shears (dashed), and 'resonance' of top and bottom boundary components (thick curve), for different directions in the (v, b)-plane (see figure 3). The stratification parameter equals $s = 0.2$. For $s \ll 1$ or for $s = 0$ (rigid lower boundary) the instability region and 'resonance' curve in (c) go all the way to the double critical point $v = -b = 2$.

The eigensolutions of the linearized equations, (2.6) with vanishing right-hand side, are obtained from

$$\text{Normal modes}: \begin{cases} \delta\boldsymbol{q} = \hat{\boldsymbol{q}}\mathrm{e}^{\mathrm{i}k(x-ct)}\sin ly \\ \mathbb{H}(c)\hat{\boldsymbol{q}} = \boldsymbol{0}, \end{cases}$$

where c is the (complex) phase speed of the perturbation. From $\det\mathbb{H}(c) = 0$ it is found that

$$\frac{c - U_b}{U_s} = \frac{1}{2} - \frac{\gamma_1(v+b) + s\left(1 + \frac{1}{2}b\right) \pm \sqrt{\Delta}}{2\mu\gamma_1}.$$

This dispersion relation can, of course, be obtained directly by replacing the normal mode structure in the linearized evolution equations, i.e. with no mention of the conservation laws. The way it is done here shows clearly that growing normal modes, $\Delta < 0$, correspond to indefinite pseudoenergy and pseudomomentum integrals of motion. Furthermore, figures 3(b) and 4 show that the onset of the instability, $\Delta = 0$, corresponds to those critical shears ($U_s = a_+$ or $U_s = a_-$) where the pseudomomentum changes from definite to indefinite (except at the 'chopped tip', where the onset is at $\kappa = 0$). (Recall that U_s growing away from zero, with either sign and at fixed values of a_+ and a_-, corresponds to moving towards the origin in the (v, b)-plane, along some direction $b/v = \text{const.}$) Basic states in the other wedge, $2 - 2v \leqslant b \leqslant -2$, are stable even though the pseudomomentum is sign indefinite, because $\mathscr{H}_\alpha$ is positive definite for $(\alpha - U_b - U_s)/U_s > 0$ (see O99).

Figure 4 illustrates how the shape of the instability region depends on the relative sign of a_+ and a_-, indicated by different shading in figure 3(b). If $a_+a_- < 0$ (figure 4a) the instability region is compact in $[U_s^{-1}, \kappa]$ space; Phillips' problem (with $a_\pm/U_s = \mp b/2$) is an example of this case. Figure 4(b) illustrates the cases with a horizontal top boundary and $\beta = 0$ (and therefore $a_+ = 0$) studied by Blumsack & Gierasch (1972) and R2K, with a rigid sloping bottom ($s = 0$ but finite a_-) or with a free lower boundary, respectively: the instability region has a peculiar tail towards $\kappa \to \infty$ and $v\ (= a_-/U_s) \to -\infty$. Finally, figure 4$(c)$ corresponds to $a_+ = a_-$, i.e. $\beta_T + \beta = 0$, which is here associated with maximum growth rate. If $s \to 0$, then in figure 3(b) the 'chopped tip' is $O(s)$ and the Δ_0 ellipse shrinks to the line that goes from the origin to

the double critical point $v = -b = 2$. Therefore, in figure 4(a) the width of the long wave ($\kappa \to 0$) instability region vanishes ($a_+ a_- < 0$), whereas in figure 4(c) ($a_+ = a_-$) it extends all the way to $v = 2$.

The thick curves in figure 4 show that component 'resonance' (defined next) explains the instability onset, at the critical shears $b = -2$ or $b = 2 - 2v$, but misses the maximum growth rate point; the latter will be explained as Rossby wave resonance in the following section. Non-diagonal terms in $\mathbb{H}(c)$ come from the advection of $Q_\pm$ by the velocity field produced by $\delta q_\mp$ (for Model D, or changing $_+$ and $_-$ for $_1$ and $_2$, for Model V) in equation (2.6). Neglecting these 'cross'-advections, the dispersion relation $\det \mathbb{H}(c) = 0$ yields

$$\gamma_2/\mu \mapsto 0 : c = \begin{cases} c_+ = U_b + U_s - (U_s - a_+)(1 + s/\gamma_1)/\mu \\ c_- = U_b - (a_- - U_s)/\mu. \end{cases} \tag{2.12}$$

These modes are *not* physical solutions of the problem (except in the limit $\kappa \gg 1$): the effect of U_s is not just to produce a difference in the advection of the modes, it is also an important part of both driving terms. Coincidence of the two eigenvalues

$$c_+ = c_- \Leftrightarrow \gamma_1 v - \tfrac{1}{2} s b = \gamma_1 (2 - \mu) + s$$

implicitly defines the 'resonance' wavenumber $\kappa = \kappa_r(s, v, b)$, indicated by thick lines in the examples of figure 4 (κ_r exists for all values of (v, b) left of the dash-dotted line in figure 3(b), which corresponds to $\kappa_r = 0$).

Including the interaction between these modes (i.e. switching on the non-diagonal elements of $\mathbb{H}$, equal to γ_2/μ), the real dispersion relation is found as the roots of

$$(c - c_+)(c - c_-) = \frac{\gamma_2^2}{\mu^2}(U_s - a_+)(a_- - U_s). \tag{2.13}$$

Neither 'resonance' at a certain κ_r nor a negative right-hand side for some U_s (the components have opposite sign pseudomomenta; see (2.7) and (2.8)) is sufficient for instability, but *both* conditions together guarantee instability in a neighborhood of (κ_r, U_s). For model types D and V studied here, the q-component 'resonance' curve goes through both points of instability onset (see figure 4), except for $s > 0$ at the 'chopped tip' in figure 3(b). Furthermore, equation (2.13) with

$$c_\pm \approx c_r + c'_\pm(\kappa^2 - \kappa_r^2)$$

predicts instability in the classical 'square root' shaped region

$$(\kappa^2 - \kappa_r^2)^2 < \frac{4\gamma_2^2(U_s - a_+)(a_- - U_s)}{\mu^2(c'_+ - c'_-)^2}.$$

This a very good estimate of the instability region near onset, for all cases in figure 4, even the one in 4(c), near the point $\kappa_r = 0$.

Pichevin (1998) derived another type of resonance curve, by writing down the normal mode equations using $\delta\psi_j$ as elementary components, instead of the δq_j, and zeroing the non-diagonal terms of the corresponding matrix: $\mathrm{diag}(Q')\mathbb{H}\mathbb{E}^{-1}$. In order to compare both possibilities, consider the classical Phillips problem ($v = s = 0$, $\gamma_1 = 1 + \frac{1}{2}\kappa^2$, and $\gamma_2 = \gamma_1^{-1}$). 'Resonance' of the q components gives $|b| < 2$ and $\kappa = \sqrt[4]{8}$, which is the wavenumber for instability onset, at $|b| = 2$. On the other hand, 'resonance' of the ψ components – as in Pichevin – gives $|b| < 2$ and $\kappa = \sqrt{2}$; this passes closer to the maximum growth rate point ($\kappa = 2(\sqrt{2} - 1)^{1/2}$, $b = 0$), but

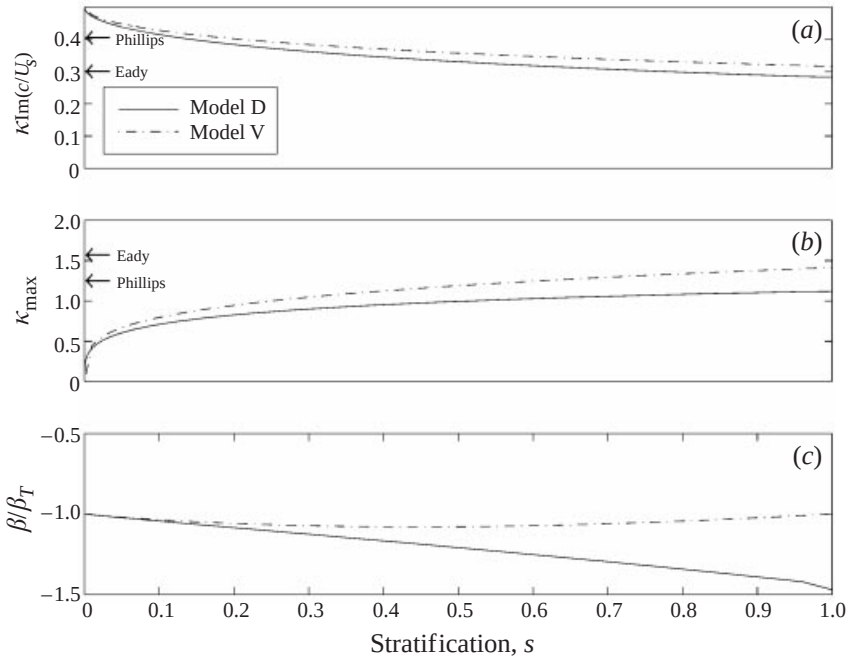

FIGURE 5. (*a*) Maximum non-dimensional growth rate and corresponding parameters – (*b*) wavenumber for maximum growth rate and (*c*) planetary/topographic ambient potential vorticity gradients – for Models D and V (see figure 1) and as a function of *s*. The limiting cases of Eady's (D: $s = v = b = 0$) and Phillips' (V: $s = v = 0$) problems are also indicated.

goes outside the instability region for $\sqrt{3} < |b| < 2$. The failure of ψ-component resonance to predict correctly the instability onset may be related to the fact that the pseudomomentum is *not* diagonal in $\delta\psi_j$ components, unlike the case of δq_j components: see (2.7).

The maximum growth rate (figure 5 and the Appendix) is reached at the centre of one the ellipses (2.11), i.e. for $b_1 = b_2 = 0$ which, in dimensional variables, read

$$\beta + \beta_T = -\tfrac{1}{2} U_s / R^2 = \tfrac{1}{2} s\beta(2 - \gamma_1(\kappa_{\max})),$$

where $\kappa_{\max}(s)$ is given in (A 1) and shown in figure 5(*b*). At a moderate stratification *s*, maximum growth rate is reached at small or moderate wavenumbers κ, which figure 2 relates to $\gamma_1 \approx 1$, and therefore $\beta + \beta_T \approx \frac{1}{2} s\beta$. Figure 5 and (A 2) show that the fastest growth rate actually corresponds to weak stratifications, intermediate scales, and a near cancellation of both beta effects:

$$\max_{s,v,b,\kappa} \left(\frac{\kappa \, \mathrm{Im} c}{U_s} \right) : \begin{cases} s \to 0 \\ \kappa = O\left(\sqrt[4]{s}\right) \\ \beta + \beta_T \ll \beta. \end{cases} \tag{2.14}$$

In Ripa (1999*b*) the condition $\beta + \beta_T \ll \beta$ for the basic state with the maximum growth rate was related to a resonance of the first two Rossby waves. This explanation is further tested in the following section, where the uniform potential vorticity constraint (2.2) is relaxed and, consequently, there is an infinite set of vertical modes for the Rossby waves, not just two.

3. Rossby waves

The study of the Rossby waves superimposed on the state with uniform flow is of interest *per se*. Their dispersion relation and orthogonality conditions (suitable to expand any perturbation into waves) are developed here and used, in particular, to shed further light on the nature of baroclinic instability. Eliminating all shear in (2.4), i.e. for $U = U_b$, potential vorticity conservation,

$$(\partial_t + U_b\partial_x)\delta q + \beta\delta v = -\delta\boldsymbol{u}\cdot\nabla\delta q \quad (-H < z < 0),$$

must be used instead of (2.2), since β is not balanced by the basic flow. This equation and (2.6) are linearized using $\delta\psi = \mathrm{Re}(\varepsilon F(z)\mathrm{e}^{\mathrm{i}k(x-ct)})\sin ly + O(\varepsilon^2)$. For the sake of generality, let us for a moment assume that N^2 is not necessarily constant, that the top boundary slope Z'_+ does not necessarily vanish, and that the bottom boundary is free (characterized by $s = L_-^2/R^2$). Recall that $R^2 = g_bH/f_0^2$, whereas keeping the definition $L^2 = N^2H^2/f_0^2$ makes L potentially a function of z. The prescribed top slope as well as the geostrophic bottom slope (due to the current U_b) are used to define two topographic beta parameters as

$$\begin{aligned}\beta_+ &= -f_0H^{-1}Z'_+,\\ \beta_- &= U_bR^{-2}.\end{aligned}$$

The linearized model equations yield

$$\left.\begin{aligned}(U_b - c)[H^2(L^{-2}F')' - \boldsymbol{k}^2F] + \beta F = 0 \quad (-H < z < 0),\\ (U_b - c)L^{-2}HF' - \beta_+F = 0 \quad (z = 0),\\ (U_b - c)L^{-2}(HF' - sF) + \beta_-F = 0 \quad (z = -H),\end{aligned}\right\} \tag{3.1}$$

where $\boldsymbol{k}^2 = k^2 + l^2$. The solutions are pairs of eigenvalues and eigenfunctions $\{c_m, F_m(z)\}$. Multiplying the first equation by another eigenfunction $F_n(z)$ and integrating, the following orthogonality conditions are obtained:

$$\left.\begin{aligned}\left\lfloor \boldsymbol{k}^2F_mF_n + f_0^2N^{-2}F'_mF'_n\right\rceil^z + R^{-2}(F_mF_n)_- = \delta_{mn},\\ \beta\left\lfloor F_mF_n\right\rceil^z + \beta_+(F_mF_n)_+ + \beta_-(F_mF_n)_- = (U_b - c_m)\delta_{mn},\end{aligned}\right\} \tag{3.2}$$

where $\lfloor\cdots\rceil^z$ denotes a vertical average and the F_m have been normalized appropriately. These equations imply that on a Rossby wave basis, both the pseudoenergy and the pseudomomentum of the wave field (without any basic shear) have a diagonal representation, namely, if

$$\delta\psi = \sum_a A_a\,\mathrm{e}^{\mathrm{i}kx}\,\sin ly\,F_m(z),$$

with $a = \{k \neq 0, l, m\}$, then

$$\mathscr{E}_p - U_b\mathscr{M}_p = -\tfrac{1}{2}\langle\lfloor\delta\psi\delta q\rceil^z + \delta\psi_+\delta q_+ + \delta\psi_-\delta q_-\rangle = \sum_a |A_a|^2, \tag{3.3}$$

$$\mathscr{M}_p = -\frac{1}{2}\left\langle \frac{\lfloor\delta q^2\rceil^z}{\beta} + \frac{\delta q_+^2}{\beta_+} + \frac{\delta q_-^2}{\beta_-}\right\rangle = \sum_a \frac{|A_a|^2}{c_a - U_b}, \tag{3.4}$$

are conserved. The functional

$$\Gamma[F] = U_b - \frac{\beta\lfloor F^2\rceil^z + \beta_+F_+^2 + \beta_-F_-^2}{\left\lfloor \boldsymbol{k}^2F^2 + f_0^2N^{-2}F'^2\right\rceil^z + R^{-2}F_-^2}, \tag{3.5}$$

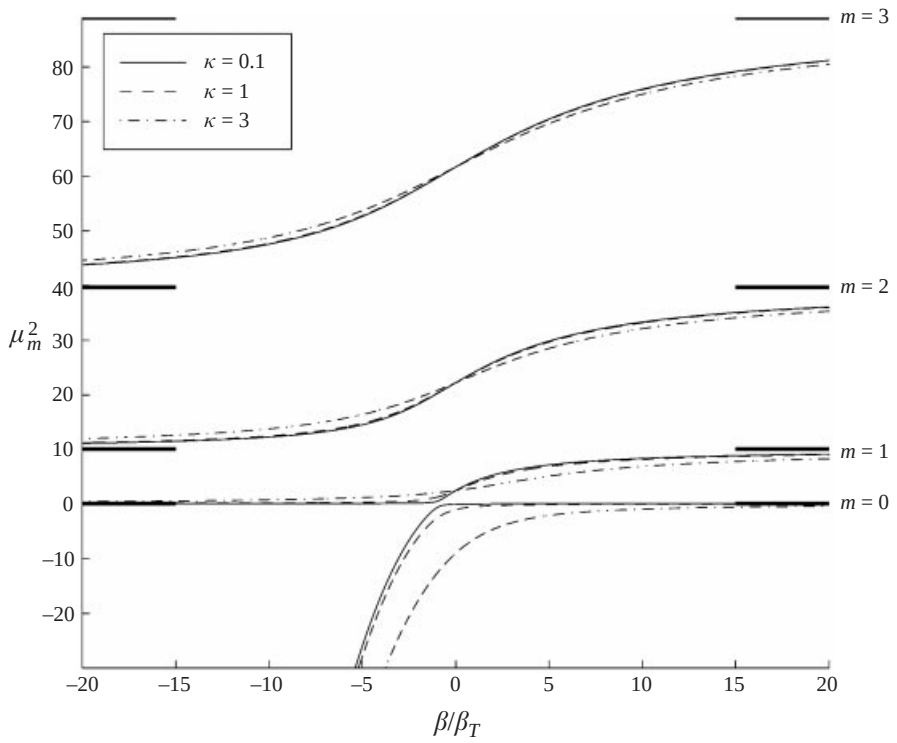

FIGURE 6. Square of the vertical wavenumber of Rossby waves, subject to the effects of a variable Coriolis parameter (β) and of an uniform flow U_b, responsible for a topographic $\beta_T = U_b/R^2 = sU_b/L^2$. The top boundary is horizontal. $s = 0.1$.

can be used to obtain the Rossby waves $\{c_m, F_m\}$ for arbitrary $N^2(z)$, from the variational principle $\Gamma[F_m + \varepsilon\delta F] = c_m + O(\varepsilon^2)$ for arbitrary δF. Numerical approximations derived with trial functions $F(z)$ in a reduced space (e.g. polynomials) also satisfy the orthogonality condition (3.2) (see for instance Ripa 1986). The numerator in (3.5) shows that the restoring mechanisms are the planetary (β) or topographic (β_+ and β_-) potential vorticity gradients; one of them might be more important than the others for a particular mode, as shown below. The softness of the bottom boundary enters in two places: the term $R^{-2}F_-^2 = s(L^{-2}F^2)_-$ in the denominator (related to the 'external' potential energy) and β_- in the numerator. The Rossby waves for a rigid sloping bottom (instead of a soft one), studied by Rhines (1970), correspond to making $s = 0$ while keeping β_- finite.

For simplicity, let us consider the case studied in R2K, with uniform N^2 and a horizontal lid ($\beta_+ = 0$ and therefore $\beta_- = \beta_T$). The eigenfunctions have the form

$$F_m(z) = N_m^{-1/2}\cos(\mu_m z/H),$$

which already satisfies the top boundary condition. From the potential vorticity equation and the bottom boundary condition it follows that

$$\frac{c - U_b}{L^2} = \frac{-\beta}{\kappa^2 + \mu^2} = \frac{\beta_T}{\mu\tan\mu - s},$$

and therefore the values of μ are the roots of the transcendental equation $(\kappa^2+\mu^2)/(s-\mu\tan\mu) = \beta/\beta_T$. Examples of the values of μ_n as a function of β/β_T are shown in figure 6; the lowest branch ($m = 0$) corresponds to the 'equivalent barotropic' or

external mode, and the other branches ($m > 0$) to the baroclinic or internal modes. The general properties of planetary and topographic Rossby modes are the following:

(i) If $\beta_T = 0$ and $\beta \neq 0$, then $\mu_m = \tilde{\mu}_m(st)$ are the roots of $\boxed{\tilde{\mu}\tan\tilde{\mu} = s}$ ($m\pi < \tilde{\mu}_m < \left(m + \frac{1}{2}\right)\pi$, $m = 0, 1, \ldots$) indicated by short horizontal lines in figure 6. These correspond to the regular vertical modes with no currents in the primitive equation model (Ripa 1995). Notice that if $\beta/\beta_T \to +\infty$ then $\mu_m \to \tilde{\mu}_m$, whereas for $\beta/\beta_T \to -\infty$ then $\mu_0 \to -\mathrm{i}\beta/\beta_T$ (which represents a mode trapped at the bottom boundary) and $\mu_m \to \tilde{\mu}_{m-1}$ for $m > 0$.

(ii) If $\beta_T \neq 0$ and $\beta = 0$, then $\mu_0 = \mathrm{i}\kappa$ and $\mu_m = \left(m - \frac{1}{2}\right)\pi$ for $m > 0$. Notice that in this case the eigenvalue of all baroclinic modes is $c_m = U_b$ ($m > 0$), whereas the barotropic mode has $c_0 = U_b\kappa/(\kappa + s/\tau)$ with $\tau = \tanh\kappa$; with a free boundary ($s \neq 0$), both types of modes have different phase speeds. The eigenfunction normalization (used below) is $N_0 = L^{-2}(\kappa\tau + s)\cosh^2\kappa$ and $N_m = \frac{1}{2}L^{-2}((m - \frac{1}{2})^2\pi^2 + \kappa^2)$ for $m > 0$.

The uniform flow U_b affects the barotropic more than the baroclinic modes. Consider, for instance, eigensolutions near $\beta/\beta_T = 0$, with $s, \kappa^2 \ll \frac{1}{4}\pi^2$. For the equivalent barotropic mode, using the trial function $F(z) = 1$ in (3.5) yields the dispersion relation

$$c_0 \approx U_b - \frac{\beta + \beta_T}{\boldsymbol{k}^2 + R^{-2}}. \tag{3.6}$$

Thus the mean current U_b affects barotropic Rossby waves both through a Doppler shift and a topographic β-term. Baroclinic Rossby waves, on the other hand, mainly suffer a Doppler shift: using the trial function $F(z) = \cos(m - 1/2)\pi z/H$ in (3.5) yields the dispersion relation

$$c_m \approx U_b - \frac{\beta}{\boldsymbol{k}^2 + \left(m - \frac{1}{2}\right)^2\pi^2 L^{-2}} \quad (m \geqslant 1). \tag{3.7}$$

These approximate expressions imply a crossing of the barotropic and baroclinic modes:

$$c_0 \approx c_1 \Rightarrow \frac{\beta}{\beta_T} = -1 + \frac{\kappa^2 + s}{\kappa^2 + \frac{1}{4}\pi^2}, \tag{3.8}$$

whereas figure 7 shows that the $m = 0$ and $m = 1$ phase speeds get close at about $\beta/\beta_T = -1$ and 'exchange characteristics'. Away from the crossing, equations (3.6) and (3.7) give a reasonable approximation of the phase speeds in figure 7.

3.1. *Rossby waves and baroclinic instability*

The normal modes for $U = U_b$ can be used to analyse the free-boundary baroclinic instability problem, i.e. the normal modes for (2.4). In Ripa (1999*b*) it was argued that the enhancement of instability for $\beta > 0$, which happens near $\beta + \beta_T = 0$, is due to a true resonance of the first two Rossby waves. The approximate model of that paper and the $2\frac{1}{2}$-layer model of O99 have only two vertical modes (i.e. two Rossby waves for each horizontal wavenumber). The present model, on the other hand, has an infinite number of vertical modes, and therefore the hypothesis of the resonance of the $m = 0$ and $m = 1$ waves can be subjected to a stricter test.

Equation (A 2) shows that for weak stratifications ($s \ll 1$), all three models reach their absolute maximum growth at similar values of $\kappa = O\left(\sqrt[4]{s}\right)$ and for b/v ($\equiv \beta/\beta_T$) $= -1 + O(s)$. Figure 8 illustrates how closely all three models behave near the point of maximum instability. Since $s, \kappa^2 \ll \frac{1}{4}\pi^2$, then (3.8) shows that the resonance $c_0 \approx c_1$ does in fact require $\beta + \beta_T \approx 0$, as discussed in Ripa (1999*b*). However, the near

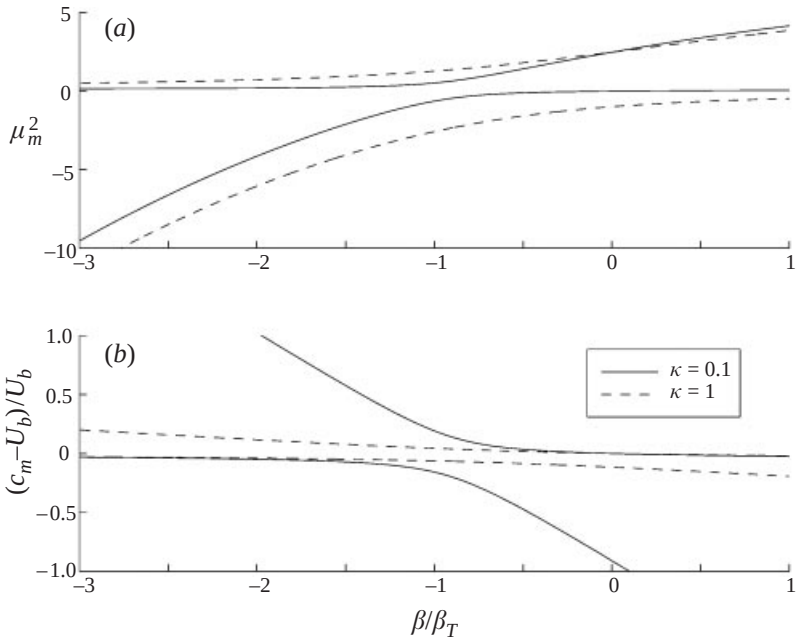

FIGURE 7. (*a*) Detail of figure 6, for the first two Rossby waves ($m = 0, 1$); (*b*) corresponding phase speed. $s = 0.1$.

cancellation of both beta-terms makes $c_0 \approx c_m$ for *all* $m \geqslant 1$, and therefore this explanation does not single out just the first baroclinic wave. However, figures 6 and 7 show that the vertical structure of the $m = 0$ and $m = 1$ waves is also similar at $\beta/\beta_T \approx -1$, and quite different from that of the higher baroclinic waves ($m > 1$). This is why the maximum instability is due, even in the present model, to a resonance of just the first two Rossby waves.

The perturbations of the $\beta = 0$ case, studied by Fukamachi *et al*. (1995) Beron-Vera & Ripa (1997), and R2K (which includes Eady's problem as a particular case), can also be satisfactorily described by the corresponding first two Rossby waves, as shown next. Both non-singular perturbations of the free-boundary baroclinic instability problem ($U_s \neq 0$) with $\beta = 0$, have the form

$$\delta\psi = \mathrm{e}^{\mathrm{i}kx} \sin ly\, G(z),$$
$$G(z) = (c - U_b - U_s)\kappa \cosh \frac{\kappa z}{H} - U_s \sinh \frac{\kappa z}{H}$$

(Beron-Vera & Ripa 1997). In the limit of a vanishing shear, $U_s \to 0$, the eigenvalue and eigenfunction of one of these two normal modes tend to those of the 'equivalent barotropic' Rossby wave, whereas the other one makes $c \to U_b$ (i.e. the eigenvalue of all baroclinic Rossby waves) and $G(z) \to \cosh \kappa(z/H + 1)$, which is a linear combination of all the topographic baroclinic Rossby modes: $F_m(z)$ for $m > 0$. In Eady's problem ($s \to 0$ with κ fixed) both limiting modes are equivalent, each one trapped at one of the horizontal boundaries, whereas for $s > 0$ there is a clear vertical asymmetry (see R2K). The vertical structure function of growing and decaying perturbations, with $U_s \neq 0$, can be expanded on the Rossby wave basis (i.e. the waves for $\beta_T \neq 0$ and $\beta = 0$), namely

$$G(z) = \sum_{m=0}^{\infty} A_m F_m(z).$$

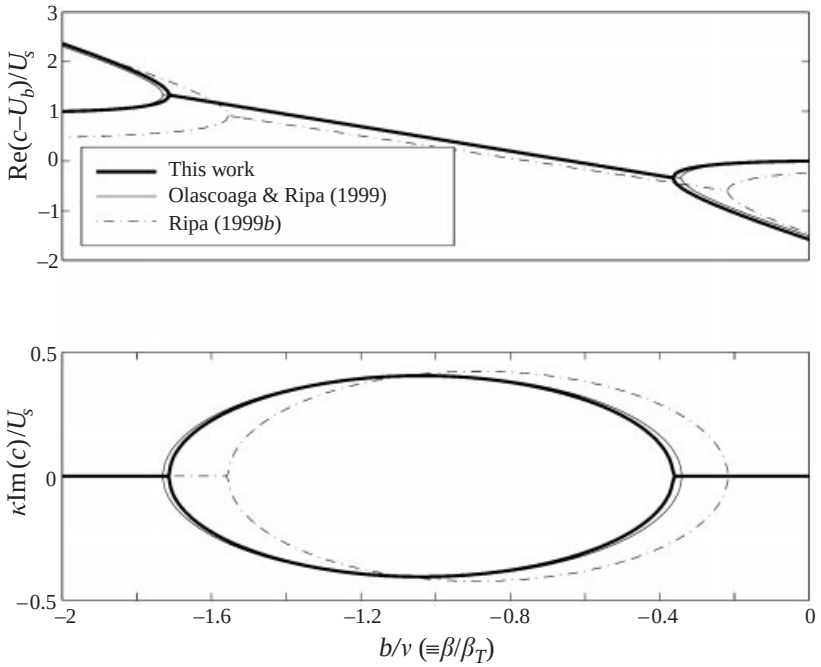

FIGURE 8. Resonance of the first two Rossby waves from figure 7, due to a finite shear U_s. $\kappa = 0.6, s = 0.1, v = 1$.

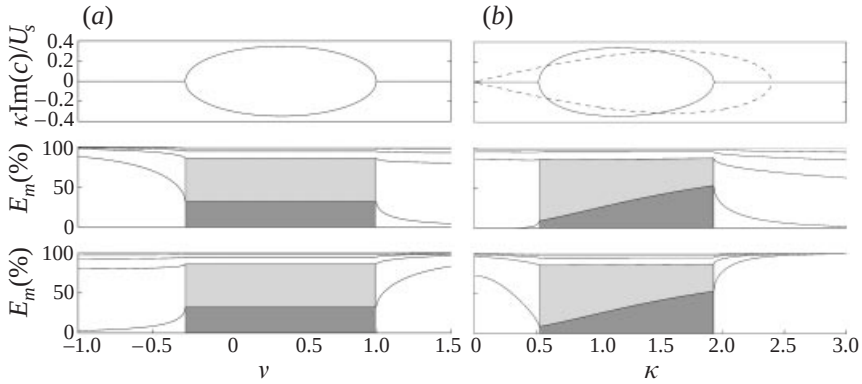

FIGURE 9. Non-dimensional growth rate (top graphs) and cumulative energy partition $\sum_{m=0}^{N} |A_m|^2$ from the expansion (3.3) of the two baroclinic instability normal modes into the Rossby waves of the problem without shear (middle and lower graphs). Dark (light) shading highlights the contribution of the $m = 0$ ($m = 1$) wave in the region where $\mathrm{Im}\, c \neq 0$ (where middle and lower graphs correspond to growing an decaying normal modes). $b = 0, s = 0.1$ and (*a*) $\kappa = 1.2$, (*b*) $v = 0.5$.

The coefficients, obtained using the orthogonality conditions (3.2), are

$$A_0 = N_0^{1/2}[(c - U_b - U_s)\kappa + U_s\tau],$$
$$A_m = -N_m^{-1/2} U_s \kappa \quad (m > 0).$$

As a way to analyse the relative importance of the different waves, the free energy of the perturbation (kinetic + internal potential + external potential) is expanded

as in (3.3):

$$\left\lfloor \kappa^2 |G|^2 + H^2 |G'| \right\rfloor^z + s |G_-|^2 = \sum_{m=0}^{\infty} |A_m|^2 .$$

Figure 9 shows examples of the cumulative energies of both perturbations, as a function of ν and κ. The equivalent barotropic and first baroclinic waves account for about 90% of the energy of a growing perturbation.

The first two Rossby waves with $\beta_- = U_b/R^2$, $\beta_+ = 0$ and $\beta \geqslant 0$ provide an explanation for the maximum growth rate occurring at $\beta + \beta_T \approx 0$ and $s, \kappa^2 \ll \frac{1}{4}\pi^2$ as due to their resonance, figures 7 and 8, and account for most of the energy of growing perturbations, figure 9. However, their pseudomomentum (3.4) is not defined, because $\beta_+ = 0$. This deficiency may corrected using the Rossby waves in a uniform potential vorticity ambient, i.e. the normal modes superimposed on the basic flow (2.4) with $U_s = 0$ but keeping $U''(z) = \beta L^2/H^2$. The pseudomomentum conserved for this Rossby wave field, given by

$$\mathscr{M}_p = -\frac{1}{2}\left\langle \frac{\delta q_+^2}{\beta_+ + \frac{1}{2}\beta} + \frac{\delta q_-^2}{\beta_- + \frac{1}{2}\beta} \right\rangle,$$

is well defined in a horizontal 'rigid lid' approximation, $\beta_+ = 0$ (as long as $\beta > 0$), and is indefinite in the range

$$-2 < \frac{\beta}{\beta_T} < 0$$

(light shaded region in figure 3*b*); the critical case $\beta/\beta_T = -1$ falls right at the middle of this range. The Hamiltonian $\mathscr{H}_\alpha = \mathscr{E}_p - \alpha \mathscr{M}_p$ is positive definite for any β/β_T and some α; consequently, these Rossby waves are stable as expected. At each wavenumber there are two Rossby waves, which have asymptotically the same frequency for $s, \kappa \ll 1$. In sum, the maximum growth rate is associated with the near resonance of two waves which have opposite-sign pseudomomenta even though their pseudoenergies have the same sign.

4. Discussion

Rossby waves (RW) in the presence of a uniform current U_b are derived for the three-dimensional quasi-geostrophic model with a free boundary (R2K). The RW are driven by the gradient of planetary potential vorticity β and also by the topographic one due to the geostrophic slope of the free boundary $\beta_T = U_b/R^2$, where R is the external deformation radius. The latter is important for the $m = 0$ or 'equivalent barotropic' wave but not so for the $m > 0$ or baroclinic waves. The barotropic wave has a deformation radius generally larger than those of the baroclinic waves, and therefore there is usually a large frequency gap between the two types of modes, except near $\beta/\beta_T = -1$ (figures 6 and 7). The RW form a complete basis, used to expand the perturbations of the free-boundary baroclinic instability problem, with shear $U_s \neq 0$ and $\beta = 0$ (Beron-Vera & Ripa 1997, this has Eady's problem as a particular case). A large fraction of the energy of growing perturbations is found to be explained by the just the first two RW (figure 9).

Three different types of models are compared in the description of the free-boundary baroclinic instability problem with $\beta > 0$. The first one (Ripa 1999*b*) is an approximation, in the sense that both the basic flow and the perturbation are assumed to have at most a linear variation with depth. The second one (O99) is

a generalization of Phillips' problem (allowing for a free boundary), which is then described by two uniform-density layers with space/time-varying potential vorticity fields. Finally, the third one is that of R2K in which the basic velocity is assumed to have a curvature $U''(z) = \beta N^2/f_0^2$ (as in Lindzen 1994), so that potential vorticity is uniform and the system is fully describable by varying density fields at the top and bottom boundaries. (The properties of last two types of models are given in table 1 and figure 1.)

The three models are, in principle, quite different: in the first two β is not balanced by the basic flow curvature, as it is in the third one, and the second and third systems have opposite choices of dynamical and uniform variables. Nevertheless, the three models have nearly identical instability properties and, in particular, attain maximum instability at similar values of the basic state parameters (figure 5); this happens close to the cancellation of the two beta effects, $\beta_T \approx -\beta$, a condition related to a near *resonance* (figure 8) of the barotropic and first baroclinic waves, namely they have not only similar frequencies but also similar vertical structure (figures 6 and 7).

The word resonance is traditionally associated with a system forced close to one of its characteristic frequencies. The forcing could be external or an internal feedback, as in the case of a resonant triad or in the following linear prototype. Consider the canonical Hamiltonian

$$\mathscr{H} = \mathscr{H}_1 + \mathscr{H}_2 + \mathscr{H}_I, \tag{4.1}$$

with

$$\mathscr{H}_1 = \tfrac{1}{2}\omega_1(q_1^2 + p_1^2), \quad \mathscr{H}_2 = -\tfrac{1}{2}\omega_2(q_2^2 + p_2^2), \quad \mathscr{H}_I = \tfrac{1}{2}\alpha(p_1 q_2 + q_1 p_2). \tag{4.2}$$

Solving the canonical equations,

$$\dot{q}_j = \frac{\partial \mathscr{H}}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial \mathscr{H}}{\partial q_j},$$

it is easily seen that: (i) the first two parts, $\mathscr{H}_1$ and $\mathscr{H}_2$, are responsible for independent oscillations of the pairs of variables (q_1, p_1) and (q_2, p_2), with frequencies $|\omega_1|$ and $|\omega_2|$ – the same result is obtained by changing the sign of $\mathscr{H}_1$ and/or $\mathscr{H}_2$; (ii) the third part, $\mathscr{H}_I$, produces an interaction between both oscillators, with the parameter α switching on an instability for

$$\alpha^2 > (\omega_1 - \omega_2)^2. \tag{4.3}$$

If $\omega_1 = \omega_2$, both oscillators have matching frequencies (i.e. there is resonance) and $\mathscr{H}_1\mathscr{H}_2 < 0$ (the modes have opposite sign 'energies'), leading to instability $\forall\alpha \neq 0$. Many hydrodynamic instability problems can be cast in the form (4.1), where the equivalent of the destabilizing parameter α (the velocity variation U_s in the basic flow) may be either present in just the third term or in all three terms. The first possibility is like the case of wave resonance discussed above. The second possibility may be envisioned by the system (4.1)/(4.2) where ω_1 and ω_2 are functions of α, and corresponds to the components 'resonance' scenario, discussed next.

These components are normal modes of the top and bottom boundary densities $\{q_+, q_-\}$, for Model D, or of the top and bottom layer potential vorticities $\{q_1, q_2\}$, for Model V. Their 'resonance' (figure 4) refers to the coincidence of phase speeds, when (arbitrarily) neglecting the advection of q_+ by the velocity field due to q_- and vice versa. This is not a true resonance in the sense that the modes are both driven and coupled by U_s (see also Ripa 1992). In a more formal description (Ripa 1999*a*), the problem linearized in the wave amplitudes $B_\pm(t) \propto \langle \mathrm{e}^{-\mathrm{i}kx} \sin ly\, \delta q_\pm \rangle$ can be described

by a Hamiltonian like (4.1), where $\mathscr{H}_1$ and $\mathscr{H}_2$ are responsible for the uncoupled evolution, with phase speeds c_+ and c_- from (2.12), and $\mathscr{H}_I$ (which is also linear) triggers the instability when $\mathscr{H}_1\mathscr{H}_2 < 0$. As pointed out by Baines & Mitsudera (1994), the right phase between $B_+(t)$ and $B_-(t)$ is crucial for the perturbation to grow. Allowing for nonlinearities in a minimal way compatible with the integrals of motion, the relative phase eventually changes sign and the perturbation decays back to its original value, rounding out a baroclinic cycle (Ripa 1999*a*).

It is worth mentioning that Morrison & Kotschenreuther (1990) have shown that linearly stable states with indefinite free energy (similar to $\mathscr{H}_1\mathscr{H}_2 < 0$ used here) can also be made unstable through nonlinearity (say, resonant interaction with a third wave) or the inclusion of dissipation (i.e. structural instability).

This work has been supported by CICESE core funding and by CONACyT (México) under grant 26670-T. Conversations with Josefina Olascoaga were very useful in clarifying many concepts of this paper.

Appendix. Maximum growth rate

Minimizing Δ from (2.11) with respect to (b, v), at fixed (s, κ), it is found that the maximum non-dimensional growth rate is reached at $b_1 = b_2 = 0$, i.e. at the centre of the ellipse, which gives

$$\Gamma(s,\kappa) := \max_{v,b}\left(\frac{\kappa \mathrm{Im} c}{U_s}\right) = \frac{\kappa\gamma_2}{2}\sqrt{\frac{\gamma_1}{\mu}} < \tfrac{1}{2}.$$

Further, maximizing Γ yields $\kappa = \kappa_{\max}(s)$ which is model dependent, namely the root of

$$\left.\begin{aligned} &\text{Model V}: s = \tfrac{1}{4}\kappa^4, \\ &\text{Model D}: s = \frac{2\kappa\tau(\kappa-\tau)}{3\tau-\kappa-\kappa\tau^2}, \end{aligned}\right\} \tag{A 1}$$

with $\tau = \tanh\kappa$. Finally, figure 5 shows that $\max_s \Gamma(s, \kappa_{\max}(s)) = \frac{1}{2}$, and correspond to $s \to 0$.

In Ripa (1999*b*) the restriction $\beta = 0$ of R2K was relaxed, but assuming that both the basic flow and the perturbation had a linear structure with depth (instead of the curvature $U''(z) = \beta N^2/f_0^2$ used here), which is asymptotically correct as $\kappa \to 0$. The corresponding dispersion relation is given by the root of $a_0C^2 + a_1C + a_2 = 0$, where $C = (c - U_b)/U_s$ and

$$\begin{aligned} a_0 &= 12(\kappa^2 + s) + \kappa^2(\kappa^2 + 4s), \\ a_1 &= 12(b + v) + 2s(2b + v) + \kappa^2(6v + 2b - 12 - \kappa^2 - 2s), \\ a_2 &= 2v^2 + b(b + 6v - 6 + s) + \tfrac{1}{6}\kappa^2(24 - 24v - 6b - 4s + \kappa^2). \end{aligned}$$

For this approximate model, the maximum non-dimensional growth rate

$$\max_{v,b,\kappa}(\kappa \mathrm{Im} C) = \frac{1 - s/6}{2\sqrt{1 + \sqrt{s/3} + s/3}}$$

is found at $\kappa_{\max} = \sqrt[4]{12s}$, $(1 - s/6)v_{\max} = 1 - \sqrt{s/3}(1 + s/6) - s(1 - s/9)/2$, and $(1 - s/6)b_{\max} = -1 + \sqrt{3s}(1 - s/6) - s(1 + s/12)/3$.

Consequently, at weak stratification all three models give

$$s \to 0 : \max_{v,b,\kappa} \left(\frac{\kappa \mathrm{Im} c}{U_s} \right) = \tfrac{1}{2} - O(\sqrt{s}) \quad \text{at} : \begin{cases} \kappa_{\max} = O\left(\sqrt[4]{s}\right) \\ v_{\max} = 1 + O\left(\sqrt{s}\right) \\ b_{\max} = -1 + O\left(\sqrt{s}\right) \end{cases} \tag{A 2}$$

and $v_{\max} + b_{\max} = O(s)$.

REFERENCES

Baines, P. & Mitsudera, H. 1994 On the mechanism of shear flow instability. *J. Fluid Mech.* **276**, 327–342.

Balmforth, N. J., Del-Castillo-Negrete, D. & Young, W. R. 1997 Dynamics of vorticity defects in shear. *J. Fluid Mech.* **333**, 197–230.

Beron-Vera, F. J. & Ripa, P. 1997 Free boundary effects on baroclinic instability. *J. Fluid Mech.* **352**, 245–264.

Blumsack, S. L. & Gierasch, P. J. 1972 Mars: The effects of topography on baroclinic instability. *J. Atmos. Sci.* **29**, 1081–1089.

Bretherton, F. 1966 Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. *Q. J. R. Met. Soc.* **92**, 335–345.

Cairns, R. 1979 The role of negative energy waves in some instabilities of parallel flows. *J. Fluid Mech.* **92**, 1–14.

Eady, E. T. 1949 Long waves and cyclone waves. *Tellus* **1**, 33–52.

Fukamachi, Y., McCreary, J. & Proehl, J. 1995 Instability of density fronts in layer and continuously staratified models. *J. Geophys. Res* **100**, 2559–2577.

Hayashi, Y. & Young, W. 1987 Stable and unstable shear modes of rotating parallel flows in shallow water. *J. Fluid Mech.* **184**, 477–504.

Iga, K. 1993 Reconsideration of Orlanski's instability theory of frontal waves. *J. Fluid Mech* **255**, 213–236.

Iga, K. 1997 Instability of a front with a layer of uniform potencial vorticity. *J. Met. Soc. Japan* **75**, 1–11.

Lindzen, R. S. 1994 The Eady problem for a basic state with zero PV gradient but $\beta \neq 0$. *J. Atmos. Sci.* **51**, 3221–3226.

Morrison, P. J. & Kotschenreuther, M. 1990 The free energy principle, negative energy modes, and stability. In *Nonlinear World: IV International Workshop on Nonlinear and Turbulent Processes in Physics, Kiev, USSR, 1989* (ed. V. G. Bar'yakhtar, V. M. Chernousenko, N. S. Erokhin, A. G. Sitenko & V. E. Zakharov), pp. 910–932.

Olascoaga, M. J. & Ripa, P. 1999 Baroclinic instability in a two-layer model with a free boundary and β effect. *J. Geophys. Res.* **104**, 23,357–23,366 (referred to herein as O99).

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

Pichevin, T. 1998 Baroclinic instbility in a three layer flow: a wave approach. *Dyn. Atmos. Oceans* **28**, 179–204.

Rhines, P. 1970 Edge-, bottom-, and Rossby waves in a rotating stratified fluid. *Geophys. Astrophys. Fluid Dyn.* **1**, 273–302.

Ripa, P. 1986 Evaluation of vertical structure functions for the analysis of oceanic data. *J. Phys. Oceanogr.* **16**, 223–232.

Ripa, P. 1990 Positive, negative and zero wave energy and the flow stability problem, in the Eulerian and Lagrangian-Eulerian descriptions. *Pure Appl. Geophys.* **133**, 713–732.

Ripa, P. 1992 Instability of a solid-body-rotating vortex in a two layer model. *J. Fluid Mech.* **242**, 395–417.

Ripa, P. 1995 On improving a one-layer ocean model with thermodynamics. *J. Fluid Mech.* **303**, 169–201.

Ripa, P. 1999*a* A minimal nonlinear model of free boundary baroclinic instability. *Proc. 12th Conf. on Atmospheric and Oceanic Fluid Dynamics*, pp. 249–252. American Meteorological Society.

Ripa, P. 1999*b* On the validity of layered models of ocean dynamics and thermodynamics with reduced vertical resolution. *Dyn. Atmos. Oceans* **29**, 1–40.

Ripa, P. 2000 Baroclinic instability in a reduced gravity, three-dimensional, quasi-geostrophic model. *J. Fluid Mech.* **403**, 1–22 (referred to herein as R2K).

Ripa, P. & Marinone, S. 1983 The effect of zonal currents on equatorial waves. In *Hydrodynamics of the Equatorial Ocean.* (ed. J. C. J. Nihoul), pp. 291–317. Elsevier.

Sakai, S. 1989 Rossby-Kelvin instability: a new type of ageostrophic instability caused by a resonance between Rossby waves and gravity waves. *J. Fluid Mech.* **202**, 149–176.

Samelson, R. 1999 Note on a baroclinic analogue of vorticity defects in shear. *J. Fluid Mech.* **382**, 367–373.

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

Takehiro, S. & Hayashi, Y. 1992 Overreflection and shear instability in a shallow water. *J. Fluid Mech.* **236**, 259–279.