Wave Packets in Simple Equilibrated Baroclinic Systems

J. GAVIN ESLER

Centre for Atmospheric Science, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom

(Manuscript received 26 March 1996, in final form 12 May 1997)

ABSTRACT

The zonal modulation of baroclinic disturbances is studied in a quasigeostrophic two-layer periodic channel. The system is relaxed toward an unstable state with a uniform flow in each layer. For small criticality, two weakly nonlinear systems are then developed, which differ in the choice of boundary condition used for the correction to the basic flow. Each system is described by an amplitude equation that determines the evolution of the wave envelope over ''long'' time- and space scales. For the first system the amplitude equation allows wave packet formation. Depending upon the ratio of the length scale of the packets to the channel length, either a steady wave train, stable solitonlike wave packets, or chaotically evolving wave packets are observed. The mechanism that leads to wave packet formation is then discussed with reference to the instability criterion of the amplitude equation. For the second system the amplitude equation is found to allow convergence to a steady, uniform wave train only.

A numerical model is then used to investigate the finite criticality extension of the second weakly nonlinear system. At low criticality, the assumptions that underpin the weakly nonlinear theory are tested by analyzing the convergence to a uniform wave train. As the criticality is increased, the effects of full nonlinearity cause the weakly nonlinear theory to become invalid. Initially, resonant triads of waves that have fixed amplitudes become excited owing to the dissipative nature of the system. As the criticality is increased further, other waves are excited and the system approaches full baroclinic chaos. Wave packet–like structures are then observed that evolve rapidly, growing, decaying, merging, and dividing.

1. Introduction

Baroclinic instability is studied chiefly because it is regarded as the fundamental mechanism behind the formation of midlatitude synoptic-scale weather systems. Modern study of the subject has followed two distinct paths, each of which are outlined in a recent review by Pierrehumbert and Swanson (1995).

The first approach involves the study of nonlinear initial value problems, in which an unstable flow is perturbed by a small disturbance. This results in a cycle of exponential growth, nonlinear saturation, and, depending on the system in question, subsequent nonlinear decay. Simmons and Hoskins (1978) studied the evolution of a wavenumber 6 disturbance on a sphere and observed the baroclinic-growth–barotropic-decay cycle, which has characterized many similar experiments. In a subsequent investigation (Simmons and Hoskins 1979), they introduced full zonal resolution. This study detailed the development of baroclinic waves downstream of a zonally localized initial disturbance in the midlatitudes. More recently, Swanson and Pierrehumbert (1994) considered the evolution of a zonally localized disturbance on an unstable jet in a two-layer channel model. They showed that as the disturbance spreads downstream, modes at the leading edge of the resultant wave packet attain the largest amplitude and eventually dominate.

The second direction of study has considered forced–dissipative baroclinic systems, which are closer to those observed in nature. In these systems there is a statistical equilibrium between the forcing of the flow by an active eddy field, and its relaxation toward an unstable state by external forcing. Therefore the disturbances seldom, if ever, are observed to grow from small perturbations to an unstable jet. The statistical equilibrium of such ''turbulent'' baroclinic systems has been the subject of recent study, for example by Whitaker and Barcilon (1995), who investigated wavenumber selection in the two-layer model.

It has recently been shown, in both observations, by Chang (1993), and in a range of numerical models, by Lee and Held (1993) (LH hereafter), that eddies in forced–dissipative baroclinic systems organize spontaneously into groups or wave packets. Just as the presence of baroclinic waves in such a system may be understood through a normal-mode instability of an unstable flow, one explanation for the formation of these wave packets may be as a secondary instability of the zonally homogeneous steady wave train.

In this paper a simple forced–dissipative baroclinic

Corresponding author address: Dr. J. Gavin Esler, Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Room 54-1721, 77 Massachusetts Ave., Cambridge, MA 02139.

© 1997 American Meteorological Society

model is investigated, using both analytical and numerical techniques. This is primarily to determine the presence of a secondary instability that could lead to wave packets, and to assess its importance by comparing the effects of finite criticality. The model in question, which is described in section 2, is the two-layer quasigeostrophic channel model. The flow is relaxed towards a basic state, which is uniform in each layer. To prevent the formation of barotropic jets the channel width is taken to be of the order of the radius of deformation (e.g., Panetta 1993).

In section 3, a weakly nonlinear theory is developed for small criticality ϵ. Unlike other such analyses (e.g., Pedlosky 1970, 1972; Tan and Liu 1995), an idealized local balance between the wave forcing of the background zonal flow and its restoration by mechanical and radiative forcing is assumed. A parameter ξ, which is the ratio of the criticality ϵ to the Rossby number Ro, determines the boundary conditions for the basic flow correction. Physically, ξ determines the importance of ageostrophic effects in modulating the background flow. Two limiting cases are considered, which lead to two contrasting weakly nonlinear systems. These systems differ in the magnitude of the (meridionally averaged) background flow response, which results along the length of a wave packet. For system A, in which this response is relatively large, the amplitude equation encourages wave packet formation for certain parameter settings. These wave packets exhibit both steady and chaotic wave packet behavior. Analysis of the form of the equation as well as the results of numerical integrations allow some insight into the mechanism by which wave packets in system A are maintained. In system B, however, the response of the background flow to the wave packets was found to have structure so as not to allow wave packet formation. The amplitude equation for this system allows convergence to a zonally symmetric steady wave train for every parameter setting investigated.

In section 4, the model is investigated numerically for finite criticality. This represents the finite ϵ extension of weakly nonlinear system B, in which the secondary instability that leads to wave packet formation is absent. By analyzing the flow as it converges toward a steady wave train, the assumptions that underpin the weakly nonlinear theory are shown to be valid at low criticality. It is then illustrated that fully nonlinear effects at finite criticality can also lead to zonal modulations of the wave train. As the criticality is increased, wave–wave interactions cause dissipative resonant triads to become excited. These are characterized by a wave train that is modulated by a symmetric wave that is close in structure to a free mode of the system. At higher criticality still, several baroclinic waves attain finite amplitude in the time mean. This results in packetlike structures that evolve rapidly, growing, decaying, merging, and dividing.

Any realistic model of the observed atmospheric flow would also require the presence of a jet with strong barotropic shear. This would allow the baroclinic waves to break at the jet edges in the upper layer. The numerical experiments described in section 4 isolate the effects of full nonlinearity in the absence of this process. Briefly, in the conclusion, separate experiments with a barotropic jet are discussed, which will be described in a future paper. Ideas from both sections 3 and 4 appear to be necessary to understand the behavior of this more complicated system.

2. The two-layer model: Linear theory

The quasigeostrophic two-layer model, as derived by Phillips (1951), has been used extensively in the study of baroclinic instability and related phenomena. Its highly truncated vertical structure makes for considerable computational and analytic economy. To appreciate the weakly nonlinear analysis that follows in section 3, it is necessary to understand the linear behavior of the two-layer model, and in particular the way in which this behavior is perturbed by the effects of a small amount of dissipation.

The model is described by two coupled potential vorticity equations of the form,

$$\frac{D_i Q_i}{Dt} = d_i \qquad i = 1, 2, \tag{1}$$

where the subscript i correspond to the ith layer, $i = 1$ is the upper layer, and $i = 2$ the lower. In this equation, Q_i is the potential vorticity, d_i the forcing/dissipation, and D_i/Dt is the quasigeostrophic convective derivative given by

$$\frac{D_i}{Dt} = \frac{\partial}{\partial t} - \Phi_{iy}\frac{\partial}{\partial x} + \Phi_{ix}\frac{\partial}{\partial y}. \tag{2}$$

The quasigeostrophic streamfunction Φ_i is related to the potential vorticity through the relation,

$$Q_i = \beta y + \nabla^2\Phi_i + (-1)^{i+1}F(\Phi_2 - \Phi_1), \tag{3}$$

where F is the nondimensional internal Froude number given by

$$F = \frac{f_0^2 L^2}{g'H}, \tag{4}$$

and β is the nondimensionalized gradient in the Coriolis parameter f. Here, H and L are the vertical and horizontal length scales, respectively, and g' is the reduced gravity due to the density difference between the layers. The reader is referred to Pedlosky (1987) for details.

The terms Φ_i and Q_i are expanded about a basic state corresponding to a uniform flow in each layer, by writing $\Phi_i = \Phi_i^0 + \phi_i$ and $Q_i = Q_i^0 + q_i$, with $\Phi_i^0 = -U_i y$ and $Q_i^0 = \beta_y + (-1)^{i+1}F(U_1 - U_2)y$. If the dissipation, denoted by d_i above, takes the form of Ekman damping in each layer, together with radiative relaxation of the system back toward the basic-state shear, then the resultant equations are given by

$$\left(\frac{\partial}{\partial t}+U_1\frac{\partial}{\partial x}\right)\left[\nabla^2\phi_1+F(\phi_2-\phi_1)\right]+(\beta+FU_s)\phi_{1x}+J_{x,y}(\phi_1,q_1)=-E_1\nabla^2\phi_1+rF(\phi_1-\phi_2)$$

$$\left(\frac{\partial}{\partial t}+U_2\frac{\partial}{\partial x}\right)\left[\nabla^2\phi_2+F(\phi_1-\phi_2)\right]+(\beta-FU_s)\phi_{2x}+J_{x,y}(\phi_2,q_2)=-E_2\nabla^2\phi_2+rF(\phi_2-\phi_1). \quad (5)$$

Here, $J_{x,y}(a, b) = a_x b_y - a_y b_x$ is the Jacobian operator and $U_s = U_1 - U_2$ is the shear in the basic state. These equations are subject to the boundary conditions

$$\phi_{ix} = 0 \quad \text{on} \quad y = 0, L_y, \qquad i = 1, 2, \quad (6)$$

as well as a condition on the zonal mean streamfunction, the derivation of which is reviewed in appendix A. (The two limiting cases that result from this derivation become the boundary conditions that define the two different weakly nonlinear systems that will be presented.) Here, L_y is the nondimensional width of the channel, E_i is the rate of Ekman damping, and r is the rate of radiative relaxation, both scaled with $\beta F^{-1/2}$ where values are quoted.

If we neglect the nonlinear terms, Eqs. (5) have a linear plane-wave solution of the form

$$\phi_1 = \mathrm{Re}[Ae^{ik(x-ct)}\sin ly]$$
$$\phi_2 = \mathrm{Re}[\gamma Ae^{ik(x-ct)}\sin ly], \quad (7)$$

where the phase speed c is given in terms of k and l ($=n\pi/L_y$ for integer n) by the dispersion relation,

$$c = \frac{1}{2a^2(a^2+2F)}\left\langle (U_1+U_2)a^2(a^2+2F) - 2\beta(a^2+F) - \frac{2a^2Fri}{k} - \frac{a^2(a^2+F)(E_1+E_2)i}{k} \pm \left\{ 4\beta^2F^2 - U_s^2a^4(4F^2-a^4) + \frac{a^4}{k^2}[4E_1E_2a^2(a^2+2F)] + \frac{a^4}{k^2}[4F^2(E_1r+E_2r-r^2) - (a^2+F)^2(E_1+E_2)^2] + \frac{2U_sa^2(a^2+2F)(a^2-F)(E_2-E_1)i}{k} + \frac{4\beta a^2F^2(E_2+E_1-2r)i}{k} \right\}^{1/2} \right\rangle, \quad (8)$$

where $a^2 = k^2 + l^2$. An expression for the constant γ controlling the relative phase and amplitude of the disturbance in the second layer relative to the first is given by Eq. (38) in appendix B. Note that with dissipation present γ is generally a complex quantity. Here, U_2 has been taken to be zero without any loss of generality. For consistency with the previous literature, throughout this paper it is taken that $U_s = 1.0$, $F = 0.5$, and β is used as a measure of the criticality of the system. Also, where values are quoted, it is taken that $L_y = 2\pi$, and $n = 1$. Here, E_i and r are both scaled with $\beta F^{-1/2}$. One time unit is defined by $U^{-1}L$ where U is the velocity scale for the shear.

Before proceeding to the nonlinear problem, it is important to appreciate the destabilizing effect of the dissipation, which will be denoted by the vector $\{E_1, E_2, r\}$, on the dispersion relation (8). The real part of the linear frequency of the waves ω_r (not shown) was found to be altered only slightly by this dissipation, at least at values typical of that quoted in this paper. It is the effect of the dissipation on the imaginary part of the frequency ω_i that has important consequences for any weakly nonlinear expansion. Figure 1 is a con-

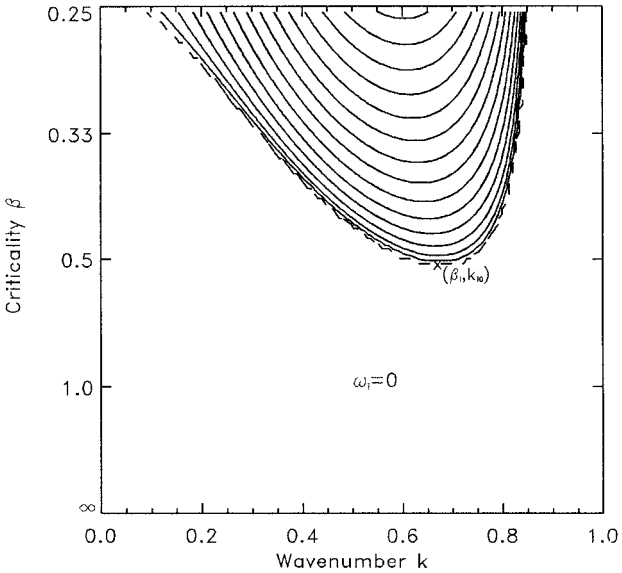

FIG. 1. A contour plot showing the scaled growth rate $\sqrt{2\beta\omega_i}$ against wavenumber and β for the inviscid case. The contour interval is 0.0125 and the dashed line marks the curve of critical β, below which the growth rate is exactly zero.

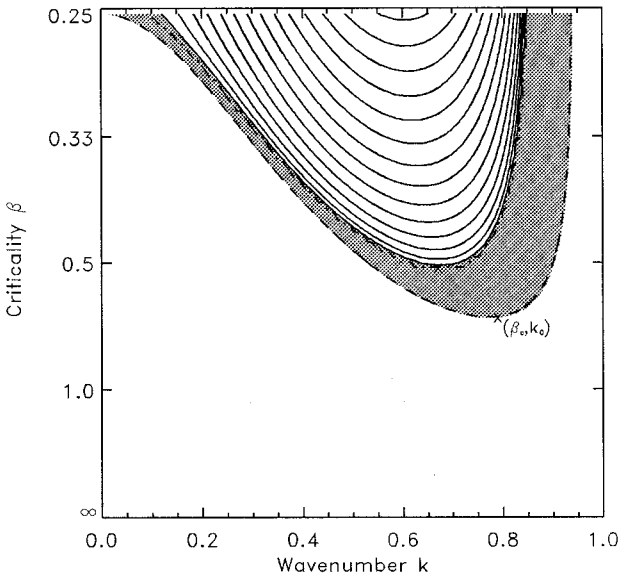

FIG. 2. As in Fig. 1 but for the case of very weak damping $\{E_1 = 0.0005,\ E_2 = 0.0025,\ r = 0.0005\}$. The dashed line marks the new curve of critical β, and the dot–dash line the inviscid curve. In the shaded region between these curves the waves have a very low growth rate, of the order of the damping. Below the lower curve the waves are very weakly decaying.

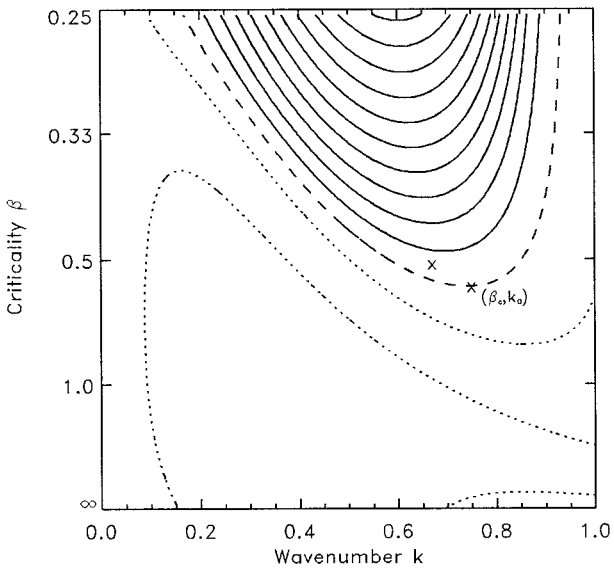

FIG. 3. As in Fig. 1 but for the damping coefficients at setting 1: $\{E_1 = 0.05,\ E_2 = 0.25,\ r = 0.05\}$. The dotted curves denote negative values of $\sqrt{2\beta\omega_i}$. The inviscid point of minimum critical shear is marked by the unlabeled "×."

tour plot of ω_i in the absence of dissipation, against wavenumber k and criticality β. It is characterized by a critical curve, denoted by the dashed line, below which the waves are marginally stable ($\omega_i = 0$) and above which the waves are unstable. The critical value of β below which some waves become unstable is given by $\beta_I = 0.5$, and the wavenumber to first become excited is $k_{I0} = 2^{-1/2}$.

Figure 2 shows how this picture is altered by a small amount of dissipation $\{E_1 = 0.0005,\ E_2 = 0.0025,\ r = 0.0005\}$. In this diagram the inviscid critical curve is now denoted by a dot–dash line, while the actual critical curve is the lower dashed curve. In the shaded region between these two curves the growth rate of the waves is of the order of the dissipation, which is small compared to the growth rate of the waves in the region above the inviscid curve. It is important to point out that the position of the new critical curve is a function of the *form* rather than the magnitude of the damping, see Romea (1977). Any weakly nonlinear expansion about the new critical point (β_c, k_0) will be relevant only to the waves in the shaded region, as the gradient of the growth rate with respect to the criticality in this region is very small.[1]

Figure 3 shows that for more typical values of the damping $\{E_1 = 0.05,\ E_2 = 0.25,\ \tau = 0.05\}$ (hereafter referred to as setting 1) the growth rate ω_i is no longer influenced by the inviscid critical curve, and instead varies smoothly in the region surrounding (β_c, k_0). This means that in this case a weakly nonlinear expansion about the critical point (β_c, k_0) can be expected to have a much larger range of validity than when the damping is very small, and this justifies the weakly nonlinear analysis that follows.

3. Weakly nonlinear behavior

With a uniform flow present in each layer, the two-layer model shows a variety of nonlinear behavior, particularly when no forcing or dissipation is present (e.g., Pedlosky 1970, 1972; Warn and Gauthier 1989). However, to describe a forced baroclinic system by weakly nonlinear theory, it is necessary to adopt a dynamical balance that reflects the competition between the wave forcing of the background zonal flow and its restoration by forcing/damping. The wave forcing of the flow is nonzero even if the waves are not growing, due to the presence of dissipation. If we take the zonal average of (5), we get the equations

$$\frac{\partial \overline{q}_i}{\partial t} + (\overline{v_i' q_i'})_y = -E_i \overline{\phi}_{iyy} + (-1)^{i+1} rF(\overline{\phi}_1 - \overline{\phi}_2), \qquad i = 1, 2, \tag{9}$$

where overbars denote zonal mean quantities, and primes the deviations from these. The divergence of the eddy potential vorticity flux $(\overline{v_i' q_i'})_y$ effectively provides

[1] Boville (1981) showed numerically that in the limit of vanishing dissipation the waves formed for $\beta \le \beta_I$ had identical properties to waves in the inviscid model, and partly for this reason the weakly growing waves formed for values $\beta_c > \beta > \beta_I$, which were first studied by Romea (1977), have been thought of as a peculiarity of the Phillips' Model.

a wave-driven forcing of the zonal mean potential vorticity. If an assumption is made that the background zonal flow varies over time and space scales that are large compared with those of the eddies, the wave-forcing terms $(\overline{v_i' q_i'})_y$ will primarily be balanced locally by the forcing/dissipation terms on the right of the equation. This is the idea that underpins the study of the limiting case that follows. If we take the meridional derivative of (9), the steady-state equation for $\overline{u}$ is

$$(\overline{v_i' q_i'})_{yy} = -E_i \overline{u}_{iyy} + (-1)^{i+1} \tau F(\overline{u}_1 - \overline{u}_2). \quad (10)$$

An important point is that $(\overline{v_i' q_i'})_{yy}$ is not, in general, equal to zero at the boundaries where $\overline{u}_i = 0$. For a steady state to be possible therefore, it is necessary that E_i is nonzero in each layer. This also proves to be necessary in the numerical experiments of section 4 to prevent small scales developing at the boundaries.

The nonlinear balances relevant here were first considered by Romea (1977). The analysis that follows was developed independently from that of Romea and includes the possibility of packet behavior by the inclusion of a ''long'' space variable X and an ''intermediate'' time variable T. Furthermore, a more general form of damping is used, in the form of asymmetric Ekman friction and radiative relaxation, which allows the amplitude equations derived to exhibit a much broader range of behavior.

It is assumed that there exists some critical value of $\beta = \beta_c$ for which there are no growing modes, and for which there exists at least one marginally stable wavenumber k_0, which is unstable for all values of $\beta < \beta_c$. This ensures that $c_i(\beta_c, k_0) = 0$ and that $\partial\omega_i/\partial k(\beta_c, k_0) = 0$.

The multiple long time and space variables[2] are defined as

$$X = \epsilon x, \qquad T = \epsilon t, \qquad T' = \epsilon^2 t, \quad (11)$$

where ϵ is a measure of the criticality defined by the equation, $\beta = \beta_c(1 - \Delta\epsilon^2)$, $\Delta = \pm 1$. Here, Δ is included in order that the amplitude equation covers the case of decaying waves ($\Delta = -1$), as well as growing waves ($\Delta = +1$). We can use the method of separation of scales to derive the full nonlinear asymptotic equations for small ϵ. These are

[2] A long space variable $X' = \epsilon^2 x$ is not included as the first-order behavior of the amplitude equation is captured by the variables X, T, and T'. For example, the channel may be thought of as being periodic in X. However, the higher order behavior in X' may be obtained by making the transformation $T' \mapsto T' - X'/c_g$ in the equations that follow. This is a consequence of the linear expansion.

$$\left(\frac{\partial}{\partial t} + U_s\frac{\partial}{\partial x}\right)[\nabla_0^2\phi_1 + F(\phi_2 - \phi_1)] + (\beta + FU_s)\phi_{1x} + E_1\nabla_0^2\phi_1 - rF(\phi_1 - \phi_2)$$
$$+ \epsilon\left\{\left(\frac{\partial}{\partial T} + U_s\frac{\partial}{\partial X}\right)[\nabla_0^2\phi_1 + F(\phi_2 - \phi_1)] + (\beta + FU_s)\phi_{1X} + \left(E_1 + \frac{\partial}{\partial t} + U_s\frac{\partial}{\partial x}\right)\frac{2\partial^2\phi_1}{\partial x\partial X}\right\}$$
$$+ \epsilon^2\left\{\frac{\partial}{\partial T'}[\nabla_0^2\phi_1 + F(\phi_2 - \phi_1)] + \left(\frac{\partial}{\partial T} + U_s\frac{\partial}{\partial X}\right)\frac{2\partial^2\phi_1}{\partial x\partial X} + \left(E_1 + \frac{\partial}{\partial t} + U_s\frac{\partial}{\partial x}\right)\frac{\partial^2\phi_1}{\partial X^2} - \Delta\beta\phi_{1x}\right\}$$
$$+ J_{x,y}(\phi_1, \nabla_0^2\phi_1 + F(\phi_2 - \phi_1)) + \epsilon J_{X,y}(\phi_1, \nabla_0^2\phi_1 + F(\phi_2 - \phi_1)) + \epsilon J_{x,y}\left(\phi_1, \frac{2\partial^2\phi_1}{\partial x\partial X}\right) = 0, \quad (12)$$

and

$$\frac{\partial}{\partial t}[\nabla_0^2\phi_2 + F(\phi_1 - \phi_2)] + (\beta - FU_s)\phi_{2x} + E_2\nabla_0^2\phi_2 - rF(\phi_2 - \phi_1)$$
$$+ \epsilon\left\{\frac{\partial}{\partial T}[\nabla_0^2\phi_2 + F(\phi_1 - \phi_2)] + (\beta - FU_s)\phi_{2X} + \left(E_2 + \frac{\partial}{\partial t}\right)\frac{2\partial^2\phi_2}{\partial x\partial X}\right\}$$
$$+ \epsilon^2\left\{\frac{\partial}{\partial T'}[\nabla_0^2\phi_2 + F(\phi_1 - \phi_2)] + \left(\frac{\partial}{\partial T}\right)\frac{2\partial^2\phi_2}{\partial x\partial X} + \left(E_2 + \frac{\partial}{\partial t}\right)\frac{\partial^2\phi_2}{\partial X^2} - \Delta\beta\phi_{2x}\right\}$$
$$+ J_{x,y}(\phi_2, \nabla_0^2\phi_2 + F(\phi_1 - \phi_2)) + \epsilon J_{X,y}(\phi_2, \nabla_0^2\phi_2 + F(\phi_1 - \phi_2)) + \epsilon J_{x,y}\left(\phi_2, \frac{2\partial^2\phi_2}{\partial x\partial X}\right) = 0, \quad (13)$$

where

$$\nabla_0^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

is the $O(1)$ part of the Laplace operator.

To solve these we attempt a series expansion in ϵ

$$\phi_i = (\phi_i^{(0)} + \epsilon\phi_i^{(1)} + \epsilon^2\phi_i^{(2)} + \dots).$$

At leading order the solution is simply the linear solution described earlier,

$$\phi_1^{(0)} = \mathrm{Re}(A(X, T, T')e^{ik(x-ct)} \sin ly)$$

$$\phi_2^{(0)} = \mathrm{Re}(\gamma A(X, T, T')e^{ik(x-ct)} \sin ly),$$

where c is in this instance real as we are at the critical value β_c.

At the next order, nonlinear terms due to the $O(1)$ phase difference of the waves induced by the dissipation must be balanced. This is where the nature of the analysis diverges from that of the inviscid problem of Pedlosky (1970), where nonlinear terms enter the problem at a higher order. We must seek a solution of the form,

$$\phi_i^{(1)} = \mathrm{Re}(B_i e^{ik(x-ct)} \sin ly) + \Phi_i^{(1)}(y, X, T, T'), \quad (14)$$

and must equate wavelike and x independent terms separately. The condition for the removal of the secular wavelike terms at this order is found to be the group velocity condition:

$$\left(\frac{\partial}{\partial T} + c_g\frac{\partial}{\partial X}\right)A(X, T, T') = 0. \quad (15)$$

Here $c_g = \partial\omega/\partial k$ is a real quantity, since we are dealing with the marginally stable mode with $k = k_0$, for which $\partial\omega_i/\partial k = 0$. Details are given in appendix B. We infer from this that $A = A(\zeta, T')$, where $\zeta = X - c_g T$. Solving for B_1, B_2, we find after renormalization (see Pedlosky 1970), that,

$$B_1 = 0,$$

$$B_2 = i\kappa A_\zeta, \quad (16)$$

where κ is a complex constant given in appendix B.

Equating the mean flow terms with those from the nonlinear wave forcing, which enters the equation at this order through the Jacobian, we arrive at the equations,

$$E_1\frac{\partial^2\Phi_1^{(1)}}{\partial y^2} + rF(\phi_2^{(1)} - \phi_1^{(1)}) = -\gamma_i F|A|^2 l \sin ly \cos ly$$

$$E_2\frac{\partial^2\Phi_2^{(1)}}{\partial y^2} + rF(\phi_1^{(1)} - \phi_2^{(1)}) = \gamma_i F|A|^2 l \sin ly \cos ly, \quad (17)$$

where $\gamma_i = \mathrm{Im}\gamma$. These equations represent a balance between the wave forcing of the basic-state vorticity field and its restoration due to friction and thermal relaxation, which is analogous to that in equation (10).[3]

At this point the analysis diverges, as it is possible to choose different boundary conditions for these equations. The condition that applies depends upon the value of ξ chosen, the ratio of the criticality ϵ to the Rossby number Ro. This is discussed in detail in appendix A. The two sets of boundary conditions that characterize the most extreme behavior, for $\xi \to 0$ and $\xi \to \infty$, are

(a) $\Phi_{iy}^{(1)} = 0$ on $y = 0, L_y$ (local zonal flow condition), (18)

(b) $\overline{\Phi_{iy}^{(1)}}^X = 0$ and $\Phi_i^{(1)\prime} = 0$ on $y = 0, L_y$ (quasigeostrophic long wave condition) (19)

The prime in condition (b) denotes deviation from the X average, which is denoted by the overbar. The physical difference between the two conditions is essentially that condition (b) assumes that all X variations in the background zonal flow must behave as quasigeostrophic waves. The assumption behind condition (a) is that ageostrophic corrections can act over the long length scale, with the result that the background zonal flow is determined locally in X. It is condition (a) that has been used in the literature (e.g., Tan and Liu 1995), but condition (b) that is equivalent to that used in the standard numerical model featured in section 4. The two conditions effectively lead to two different weakly nonlinear systems that differ in how

[3] The upper-layer Ekman friction should not be regarded as parameterizing a particular physical process, but has instead been included to allow a solution to Eq. (17). An alternative might have been to use interfacial friction, $= (-1)^i E_I(\nabla^2\phi_1 - \nabla^2\phi_2)$. From the solution to Eq. (17) with this friction included, the results would be similar to those obtained when the ratio E_2:E_1 is large.

the structure of the background zonal flow changes along a wave packet. System A exhibits a weakly nonlinear instability that allows wave packet formation, whereas in system B the steady wave train appears to be the stable solution to which the system always converges.

a. System A: The local zonal flow system

Equation (17) may be solved with the boundary conditions (18) to give

$$\Phi_1^{(1)} = \frac{\gamma_i F|A|^2 klL_y^2}{2E_1(4n^2\pi^2 + \alpha^2)}\left\{\sin 2ly - \frac{2n\pi}{\alpha}\left[\sinh\frac{\alpha y}{L_y} - \cosh\frac{\alpha y}{L_y}\left(\frac{\cosh\alpha - 1}{\sinh\alpha}\right)\right]\right\}$$

$$\Phi_2^{(1)} = -\frac{\gamma_i F|A|^2 klL_y^2}{2E_2(4n^2\pi^2 + \alpha^2)}\left\{\sin 2ly - \frac{2n\pi}{\alpha}\left[\sinh\frac{\alpha y}{L_y} - \cosh\frac{\alpha y}{L_y}\left(\frac{\cosh\alpha - 1}{\sinh\alpha}\right)\right]\right\}, \tag{20}$$

where $\alpha^2 = r(E_1 + E_2)FL_y^2/E_1E_2$. The structure of this mean flow correction is of principal importance in determining the zonal stability of the steady wave train solution to the problem.

To derive the amplitude equation we must go to the next order. We may ignore all nonlinear terms that enter the equation at this stage, because they are balanced with the second-order background flow correction terms $\Phi_i^{(2)}$. It only remains to seek a solution,

$$\phi_i^{(2)} = \mathrm{Re}(C_i e^{ik(x-ct)}\sin ly),$$

for the wavelike terms and discover the condition for the removal of those secular terms that would otherwise invalidate the expansion. Details are again given in appendix B. Once we have integrated across the channel to remove y dependence, we arrive at the amplitude equation:

$$A_{T'} + \mu A_{\zeta\zeta} = \Delta\rho A + \nu A|A|^2. \tag{21}$$

This is the Ginsburg-Landau equation, also known as the nonlinear Schröedinger equation with complex coefficients.

The explicit representation of the coefficients $\mu = \mu_r + i\mu_i$, $\rho = \rho_r + i\rho_i$, and $\nu = \nu_r + i\nu_i$ is given in appendix B; however, it is a consequence of the linear expansion that

$$\rho = i\frac{\partial\omega}{\partial\beta} \quad \text{and} \quad \mu = -\frac{i}{2}\frac{\partial^2\omega}{\partial k^2}, \tag{22}$$

both evaluated at the point of marginal stability (β_c, k_0) (Moon et al. 1983). This means that at the point of marginal stability, $\mu_r \le 0$ (as $\partial^2\omega_i/\partial k^2 \le 0$) and $\rho_r \ge 0$ (as $\partial\omega_i/\partial\beta \le 0$) for the marginally stable mode. In all that follows ρ_i will be taken to be zero, as it is easily removed from (21) by the transformation $A \to Ae^{i\rho_i T'}$, which is equivalent to a global $O(\epsilon^2)$ correction to the frequency of the baroclinic waves.

1) PROPERTIES OF THE AMPLITUDE EQUATION FOR SYSTEM A AND THE MECHANISM FOR WAVE PACKET FORMATION

Equation (21) has solutions that remain finite in time only if the condition $\nu_r < 0$ is met. Otherwise the solutions will ''burst'' to infinite magnitude (Hocking and Stewartson 1972). It was found that $\nu_r < 0$ at the point of marginal stability, for every case investigated in this paper. However, Fig. 4 shows a plot of ν_r, as defined by Eq. (51), against wavenumber for different values of β (for setting 1). The picture shows that for values of $\beta > \beta_c$, for which all the waves are linearly stable, ν_r can be greater than zero. This is suggestive of *subcritical instability* (see Lee and Held 1991). Given that $\nu_r < 0$, the qualitative behavior of (21) is determined by the values of (μ_i/μ_r) and (ν_i/ν_r) alone. The equation can be rescaled so that ρ_r^{-1}, $(-\mu_r/\rho_r)^{1/2}$, and $(-\rho_r/\nu_r)^{1/2}$ become the scales for T', ζ, and $|A|$, respectively.

The simplest solution of Eq. (21) is independent of ζ, and the modulus of A evolves according to

$$|A|^2 = \frac{-\rho_r|A_0|^2 e^{2\rho_r T'}}{\nu_r[\mathrm{vb}1 - |A_0|^2(1 - e^{2\rho_r T'})]}. \tag{23}$$

This simply represents convergence to a uniform wave train with $|A| \to (-\rho_r/\nu_r)^{1/2}$ as $T' \to \infty$ and was shown by Romea (1977) to be the solution if no modulation of the wave packets is allowed in the zonal direction. Clearly, the ζ-stability criterion of the above solution is of paramount importance to our investigation of the possibility of wave packets in this model.[4] This was investigated by Newell (1974), who examined the linear behavior of a sideband, and derived the following condition for instability,

[4] The equation is known to support other steady amplitude traveling waves of the form $A = e^{i(q\zeta+\sigma T')}$, the stability of which were investigated by Stuart and Diprima (1978). All were shown to have regions of stability smaller than that given above.

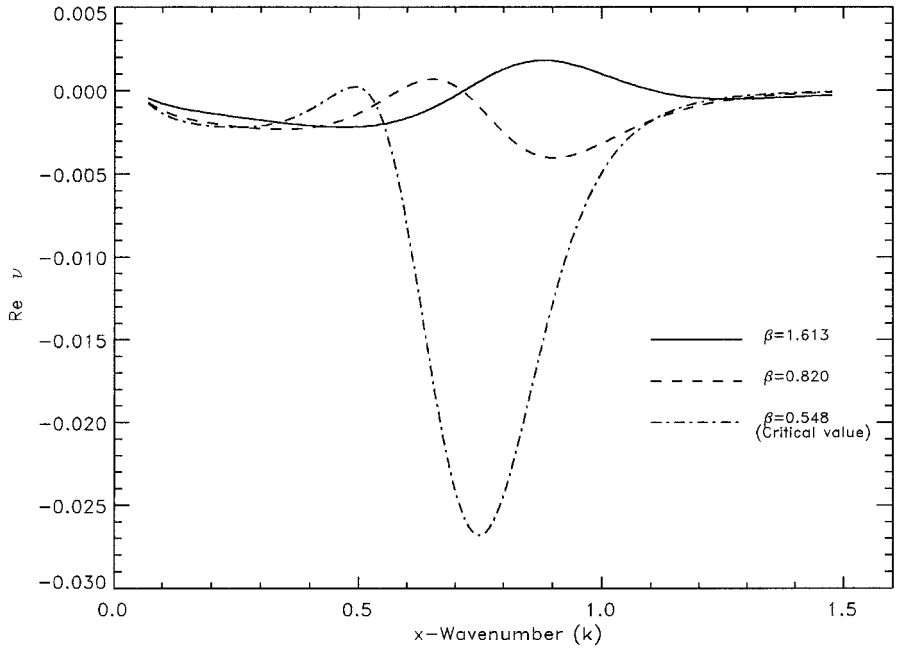

FIG. 4. Here, $\nu_r[(\nu)(r)]$ as calculated from (51), against wavenumber, for different values of β. Damping coefficients are at setting 1, as in Fig. 3.

$$\mu_i\nu_i + \mu_r\nu_r < 0. \tag{24}$$

Moon et al. (1983) among others have studied the transition to chaos that governs the behavior of the unstable sidebands, with the constraint $(\mu_i/\mu_r) = -(\nu_i/\nu_r)$ on the values of their coefficients. Even when such a constraint is placed on the coefficient values, Moon et al. showed that the amplitude equation (21) still exhibits a wide range of behavior. It was therefore decided to integrate (21) numerically in a channel with imposed periodicity, using values of the coefficients actually obtained from the expressions for ρ, μ, and ν in appendix B.

Table 1 shows the values of β_c, k_0, c, c_g, μ, ρ, and ν obtained for a range of different values of the damping coefficients $\{E_1, E_2, r\}$. The system is very sensitive to changes in these parameters; in particular, the linear group velocity $c_g = d\omega/dk$ varies a great deal at the critical point. The variation of those quantities that determine the wave train stability, (μ_i/μ_r) and (ν_i/ν_r), is shown in Fig. 5. The three curves show the changing properties of the wave train as each damping parameter is varied from its setting 1 value. If the ratio E_2:E_1 is high, and the value of r is relatively small, then the coefficients of (21) will fall within the unstable region, according to (24), and wave packet formation will be possible. From (20) it can be shown that if the ratio E_2:E_1 is high, then the correction to the background flow is confined mostly to the upper layer. In this case not only is the shear between the layers reduced but importantly the strength of the mean westerlies (averaged between the layers) is also reduced. Figure 6a shows the typical meridional and vertical structure of the background zonal flow correction when the parameters are at setting 2 $\{E_1 = 0.05, E_2 = 0.05, r = 0.05\}$, a setting for which wave packet formation is possible. By contrast, Fig. 6b shows the structure for $\{E_1 = 0.05, E_2 = 0.05, r = 0.05\}$, a setting for which wave packets do not form.

This result can be understood if one considers the nonlinear *self-focusing* behavior exhibited by the Ginsburg-Landau equation [see the recent review by Balmforth (1995) and references]. The instability criterion for the steady wave train (24) may be intepreted as a competition between this nonlinear self-focusing to form wave packets, which occurs if $\nu_i\mu_i < 0$ (nonlinear defocusing occurs if $\nu_i\mu_i > 0$), against a diffusive tendency to converge to the steady wave train (which is always present as $\nu_r\mu_r > 0$). If the nonlinear Schröedinger limit (μ_i/μ_r), $(-\nu_i/\nu_r) \to \infty$ is considered, the wave packets formed by the self-focusing mechanism are thought to become the well-known steady soliton solutions of that equation (see again Balmforth 1995). Closer to the critical curve defined by (24), wave packets are often observed to be formed by self-focusing and subsequently ''break'' by dissipation after some finite time.

If the nonlinear term in (21) is considered, then the nonlinear coefficient ν can be related to the effective change in the complex frequency $\omega = \omega_r + i\omega_i$ of the waves induced by a mean flow correction with structure given by (20). In a steady wave train this relationship is given by $\delta\omega_i = \epsilon^2|A|^2\nu_r$ and $\delta\omega_r = -\epsilon^2|A|^2\nu_i$. The first of these relations describes the stabilizing effect of the background flow correction on the waves, and the second describes the correction to the frequency of the waves induced by the background flow.

The nonlinear self-focusing mechanism can be under-

TABLE 1. Coefficient values at the critical value of $\beta = \beta_c$, and the marginally stable wave $k = k_0$ for a range of different values of the damping parameters E_1, E_2, and r.

Upper-layer damping E_1	Lower-layer damping E_2	Thermal damping r	Critical value β_c	Fastest growing mode k_0	Phase velocity c	Group velocity c_g	$\mu = -\frac{i}{2}\frac{\partial^2\omega}{\partial k^2}$	Re $\rho = -\frac{\partial\omega_i}{\partial\beta}$	Nonlinear constant ν
0.05	0.05	0.05	0.496	0.67	0.036	0.490	−2.862	0.386	−0.1378
							−2.535*i*		−0.1049*i*
0.05	0.08	0.05	0.506	0.68	0.059	0.550	−1.262	0.217	−0.0690
							−2.186*i*		−0.0687*i*
0.05	0.1	0.05	0.513	0.69	0.071	0.582	−1.307	0.173	−0.0529
							−1.350*i*		−0.0504*i*
0.05 (Setting 1)	0.25	0.05	0.548	0.74	0.135	0.749	−0.647	0.082	−0.0271
							−1.174*i*		+0.0074*i*
0.05 (Setting 2)	0.5	0.05	0.549	0.78	0.221	0.847	−0.396	0.047	−0.0215
							−0.778*i*		+0.0427*i*
0.05	1.0	0.05	0.495	0.80	0.265	0.900	−0.228	0.024	−0.0182
							−0.565*i*		+0.0676*i*
0.01	0.25	0.05	0.728	0.93	0.201	1.025	−0.208	0.034	−0.0077
							−0.704*i*		+0.0424*i*
0.02	0.25	0.05	0.647	0.84	0.168	0.918	−0.351	0.050	−0.0128
							−0.935*i*		+0.0303*i*
0.1	0.25	0.05	0.483	0.68	0.121	0.599	−0.888	0.106	−0.0474
							−1.138*i*		−0.0134*i*
0.25	0.25	0.05	0.403	0.61	0.128	0.396	−0.849	0.080	−0.0850
							−0.704*i*		−0.0388*i*
0.5	0.25	0.05	0.340	0.54	0.159	0.270	−0.505	0.071	−0.1091
							−0.365*i*		−0.0573*i*
0.05 (Setting 3)	0.25	0.0	0.616	0.76	0.122	0.743	−0.516	0.067	−0.0356
							−1.167*i*		+0.0360*i*
0.05	0.25	0.025	0.578	0.75	0.127	0.740	−0.587	0.075	−0.0291
							−1.179*i*		+0.0161*i*
0.05	0.25	0.1	0.504	0.72	0.136	0.742	−0.797	0.100	−0.0266
							−1.156*i*		+0.0018*i*
0.05	0.25	0.5	0.387	0.64	0.090	0.701	−1.112	0.129	−0.0230
							−0.419*i*		+0.0003*i*

stood with the help of these relations as follows. If one considers a wave packet such as that illustrated in Fig. 7a, the effect of positive ν_i is to reduce the frequency of the waves at the center of the packet. (See also Fig. 11, which shows the corrections to the phase speed of the fundamental along an idealized wave packet.) This causes the wave packet to adjust to a state where the wavelength is decreased toward the rear of the packet and increased toward the front, as illustrated schematically in Fig. 7b (see also Fig. 9.) As $\mu_i = -(d^2\omega_r/dk^2)/2)$ is generally negative for the two-layer model at the point of criticality, to first order the group speed $c_g = d\omega_r/dk$ is greater for waves of larger wavenumber. This means that the wave envelope at the rear of the wave packet will radiate faster than at the front, leading to convergence of wave activity at the packet center. This causes steepening of the wave envelope and packet growth.[5] Nonlinear self-focusing can take place (for $\mu_i < 0$) only if the frequency of large amplitude waves is *decreased* by the background flow correction they induce. This explains why wave packets can form when the structure of the mean flow response is of the form shown in Fig. 6a, but not when it has structure such as that shown in Fig. 6b. From the dispersion relation (8), the advection of the waves by the vertically averaged velocity is one of the most important terms in determining their phase speed. In Fig. 6a this is reduced by the background flow correction and so the frequency of the waves will be reduced approximately in proportion to the square of the amplitude of the wave train, allowing nonlinear self-focusing to take place. In Fig. 6b the vertically averaged velocity is unchanged by the background flow correction, and ν_i was found to be negative (see Table 1), indicating nonlinear defocusing of the wave packets.

The packet-forming mechanism has to compete with the diffusive effects. The larger the magnitude of $\mu_r = (d^2\omega_i/dk^2)/2$ (which is always negative), the more the spreading of wavenumbers within a wave packet is inhibited, suppressing packet growth. Similarly, the larger the magnitude of ν_r (also always negative), the more strongly large amplitude waves are stabilized by the correction to the background flow that they induce.

[5] Similarly if $\mu_i > 0$ and $\nu_i < 0$, the wavelength would decrease at the *front* of the packet, and the group speed of the packet would then be greater at the rear of the packet—also leading to growth.

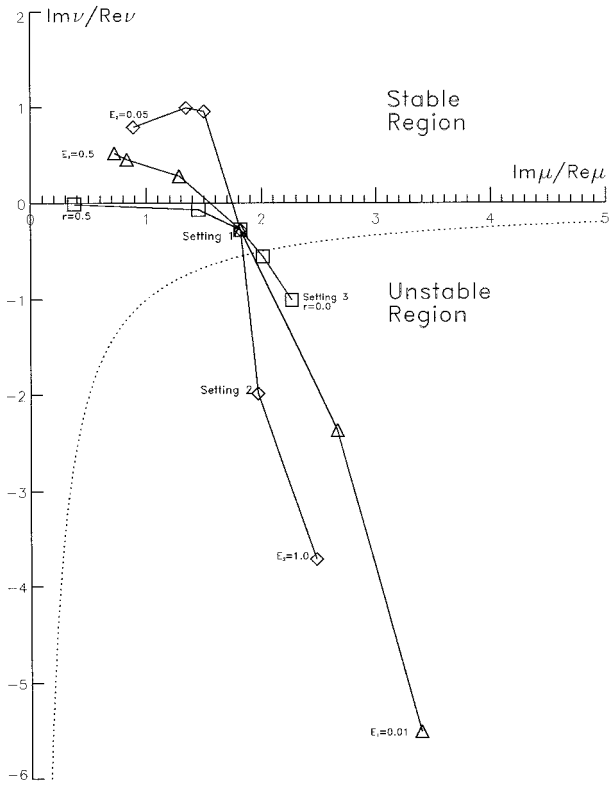

FIG. 5. A diagram showing how the stability properties of Eq. (21) change as each damping coefficient is varied from its setting 1 value $\{E_1 = 0.05, E_2 = 0.25, r = 0.05\}$. Values of E_1 (marked by triangles) are { 0.01, 0.02, 0.05, 0.1, 0.25, 0.5}. Values of E_2 (diamonds) are {0.05, 0.08, 0.1, 0.25, 0.5, 1.0}. Values of r (squares) are {0.0, 0.02, 0.05, 0.1, 0.5}.

2) INTEGRATIONS OF THE AMPLITUDE EQUATION FOR SYSTEM A

Some simple integrations of the amplitude equation have been carried out using an implicit Crank-Nicholson gridpoint scheme (see Taha and Ablowitz 1984). All the integrations are carried out on a grid of 200 points, and a time step of $\delta T' = 2\delta\zeta$ is used. The numerical algorithm was tested for known soliton solutions of the nonlinear Schröedinger equation and was found to be accurate over the number of time steps required. Two different coefficient settings are investigated, settings 2 and 3, which are associated with values of the damping coefficients (see Table 1 and Fig. 5). Periodic boundary conditions are imposed in ζ, with the channel length ζ_0 being a free parameter in the problem. The effect of varying ζ_0 may be thought of as being equivalent to varying the criticality ϵ in the problem. If the physical length of the channel is thought of as being a fixed size L_x, on which the long space variable X, or equivalently ζ, is periodic, then the domain size ζ_0 is given by $\zeta_0 = \epsilon L_x$, so the larger the domain size, effectively the larger ϵ becomes.

Figure 8 shows the evolution of $|A(\zeta, T')|$ in six experiments. Figure 8a corresponds to an experiment with the coefficients at setting 3, and with $\zeta_0 = 50$. The diagram shows a perturbed wave train with maxima ($|A| \approx 1.5$) and minima ($|A| \approx 1.1$) traveling eastward in ζ—that is, propagating more rapidly than the linear group velocity. Another experiment with the dissipation coefficients at setting 3, and with $\zeta_0 = 20$ (not shown) was carried out, and the system was observed to evolve rapidly into a steady wave train with $|A|$ constant. The sideband instability that is excited at this parameter setting therefore has a zonal scale intermediate between these two channel lengths.

Figures 8b–f all correspond to experiments with the coefficients at setting 2. The channel lengths are given by $\zeta_0 = 20$, 24, 26, 28, and 40, respectively. An experiment with $\zeta_0 = 16$ (not shown) was observed to converge rapidly to a steady uniform wave train. Figure 8b shows that when $\zeta_0 = 20$, however, the wave envelope evolves into a single packet that propagates at the linear group velocity. The packet is steady in the sense that it has a fixed amplitude and half-width. As this was the only steady solution obtained, it was investigated further to verify the theory of the nonlinear self-focusing mechanism discussed above. The wave envelope was found to converge to a solution of the form

$$A(\zeta, T') \to \mathcal{A}(\zeta) e^{(i f(\zeta) - i\omega_c T')}. \tag{25}$$

The functions $\mathcal{A}(\zeta)$ and $f(\zeta)$ are shown in Figs. 9a and 9b, respectively. Here, ω_c denotes an $O(\epsilon^2)$ correction to the frequency of the waves, which was found to take the value 0.011. This solution also implies an $O(\epsilon)$ correction to the local wavelength of the waves, of magnitude $f'(\zeta)$. From Fig. 9b this indicates an increase in wavenumber to the rear of the wave packet and a decrease to the front, which is consistent with the nonlinear self-focusing mechanism described in the last section. Figure 9c shows a wave packet constructed with this envelope (using the value $\epsilon = 0.45$) of the form $\mathcal{A}(\zeta) e^{i f(\zeta) - i(kx - \omega t)}$. The corrections to the wavelength can be clearly observed.

When the channel length is increased to $\zeta_0 = 24$ (Fig. 8c), we enter a regime in which packets grow and decay at fixed positions in ζ, (i.e., grow and decay while propagating at the linear group velocity). Over their life cycle these packets map out a path in amplitude half-width space. Figure 10 (top) shows that the packets start with large amplitude and half-width, become narrower, and then decay as they become broader again. The beginning and end of the cycles mark the points at which the packets become distinct from, or merge into, a steady wave train. The group of $\times$'s at small amplitude mark a small packet that appears just as the main packet packet begins to decay. Figure 8d appears to show that the system behaves in a similar fashion for $\zeta_0 = 26$. However Fig. 10 (bottom) shows that the evolution of the packets in amplitude half-width space has become more complex at this channel width. Figure 8e shows that the system supports two packets when $\zeta_0 = 28$, but that these are unstable, and merge into a wave train after a long period, before reforming into two packets again.

Figure 8f shows that if the channel length is increased further any regularity is lost, and the system enters a chaotic packet regime. This was verified in other integrations (not shown) with longer channel length.

b. System B: The quasigeostrophic long wave system

If Eq. (17) is solved with the boundary conditions (19), the leading order basic flow correction is given by

$$\overline{\Phi_1^{(1)}}^X = \frac{\gamma_i F \overline{|A|^2}^X k l L_y^2}{2E_1(4n^2\pi^2 + \alpha^2)}\left\{\sin 2ly - \frac{2n\pi}{\alpha}\left[\sinh\frac{\alpha y}{L_y} - \cosh\frac{\alpha y}{L_y}\left(\frac{\cosh\alpha - 1}{\sinh\alpha}\right)\right]\right\},$$

$$\overline{\Phi_2^{(1)}}^X = -\frac{\gamma_i F \overline{|A|^2}^X k l L_y^2}{2E_2(4n^2\pi^2 + \alpha^2)}\left\{\sin 2ly - \frac{2n\pi}{\alpha}\left[\sinh\frac{\alpha y}{L_y} - \cosh\frac{\alpha y}{L_y}\left(\frac{\cosh\alpha - 1}{\sinh\alpha}\right)\right]\right\}, \tag{26}$$

$$\Phi_1^{(1)'} = \frac{\gamma_i F (|A|^2)' k l L_y^2}{2E_1(4n^2\pi^2 + \alpha^2)} \sin 2ly,$$

$$\Phi_2^{(1)'} = -\frac{\gamma_i F (|A|^2)' k l L_y^2}{2E_2(4n^2\pi^2 + \alpha^2)} \sin 2ly. \tag{27}$$

The vertical and meridional structure of the zonal flows derived from (26) and (27) are illustrated in Fig. 6c (for damping values at setting 2). Following the derivation of Eq. (21), results in the equation

$$A_{T'} + \mu A_{\zeta\zeta} = \Delta\rho A + \nu A\overline{|A|^2}^X + \nu_2 A(|A|^2)', \tag{28}$$

where the complex coefficients μ, ρ, and ν are as for system A, and ν_2, which is closely related to ν, is given in appendix B.

INTEGRATIONS OF THE AMPLITUDE EQUATION FOR SYSTEM B

Integrations of Eq. (28) were carried out for coefficient values that corresponded to the damping parameters at settings 1, 2, and 3. A relatively long channel length of $\zeta_0 = 40$ was used in each case. All the experiments resulted in rapid convergence to a steady wave train solution, with $|A| = (-\rho_r/\nu_r)^{1/2}$. Equation (28) is shown in appendix C to have the same stability properties as Eq. (21) has with $\nu = \nu_2$. This is because it is the zonal gradient in the mean flow correction that is important for packet formation. All the values calculated for ν_2 result in $\mu_r\nu_{2r} + \mu_i\nu_{2i} > 0$ so system B appears to be largely stable to wave packet formation.

In order to help to explain why system A allows wave packet formation while system B does not, it was interesting to compare the gradients in both linear growth rate and linear phase speed of the fundamental that would exist across a wave packet in each system. The amplitude $|A|$ of the idealized wave packet in question is pictured in Fig. 11a. The complex phase speeds were calculated for the zonal flows constructed by adding the mean flow corrections for system A, calculated from (20), and system B, calculated from (26) and (27), to the basic-state flow. To calculate the linear growth rates with respect to a meridionally nonuniform zonal flow, it was necessary to solve the complex eigenvalue problem detailed in Swanson and Pierrehumbert (1994). For each system, setting 2 was chosen for the damping parameters, and $\beta = 0.500$ so that $\epsilon = 0.297$. The maximum amplitude of the wave packet that determines the mean flow correction was chosen so that the X-averaged corrected flow given by $\Phi_i = \Phi_i^0 + \epsilon^2\overline{\Phi_i^{(1)}}^X$ is marginally stable to wave growth (this X-averaged flow is the same for systems A and B). Figure 11b shows the linear growth rates of the fastest growing mode along the wave packet, for each system. The gradient in the linear growth rate along the packet for system A is roughly twice that for system B.[6] This shows that the effects of dissipation on the wave packet are twice as strong for system A compared with system B. Figure 11c shows that the linear phase speeds of the waves are reduced at the packet center for system A, but *increased* for system B. This means that while the wave packet in system A will undergo nonlinear self-focusing, the system B wave packet will undergo nonlinear defocusing. This explains why the symmetric wave train is a stable solution for system B but not system A at this parameter setting.[7]

[6] This is perhaps not surprising as the background shear is reduced much less across the wave packet for system B than for system A, as pressure gradients are not allowed along the sidewalls for system B [cf. the structure of the correction to the background flow in Fig. 6c (dashed line) to Fig. 6a.]

[7] It worth noting that it is possible to recover reasonable approximations for (ν_i/ν_r) and (ν_{2i}/ν_{2r}) from Fig. 11, as they equal $(-\sigma\omega_r/\sigma\omega_i)$ for systems A and B, respectively.

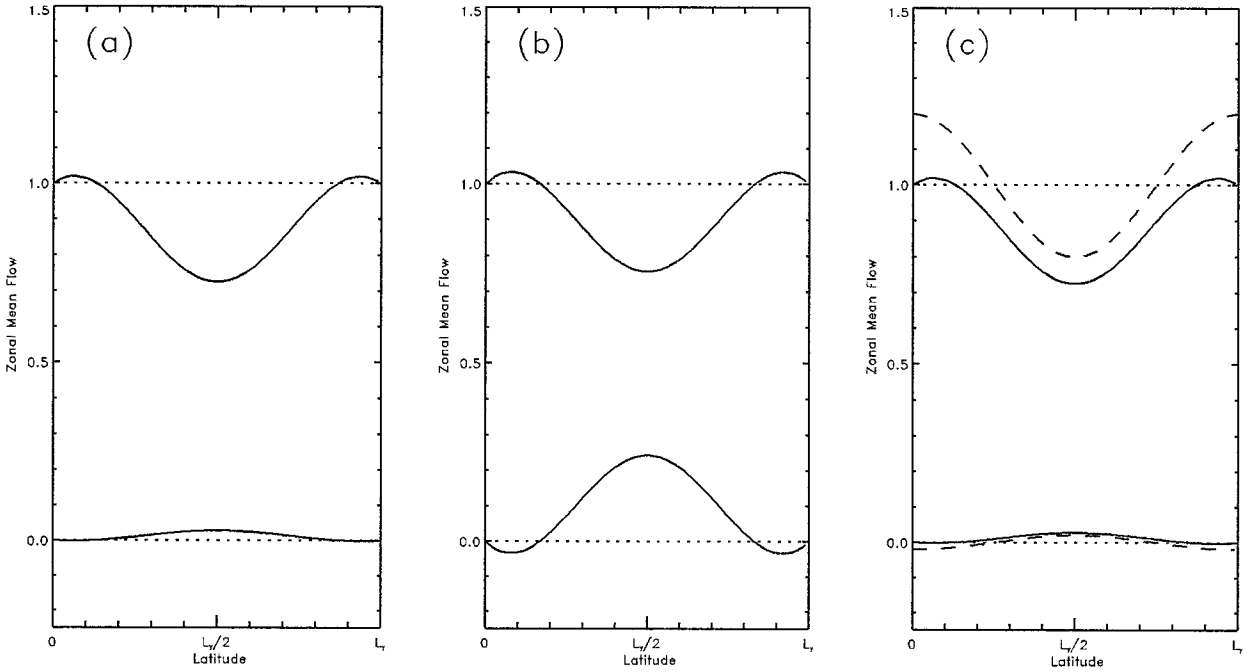

FIG. 6. Showing the vertical and meridional structure of the mean flow response to finite-amplitude waves as calculated from (20) (system A) and (26) and (27) (system B). The upper-layer response is the deviation from the upper dotted line, and the lower-layer response is the deviation from the lower dotted line. (a) The response for system A when the damping is at setting 2 $\{E_1 = 0.05, E_2 = 0.5, r = 0.05\}$ and wave packet formation is possible. (b) The response for system A when the damping has values $\{E_1 = 0.05, E_2 = 0.05, r = 0.05\}$ and wave packet formation is not possible. (c) The response of the X-averaged $\overline{\Phi_{iy}^{(1)}}^{x}$ flow for system B (solid lines), along with the structure of the quasigeostrophic wave response $(\Phi_{iy}^{(1)})'$ (dashed lines), which would be observed along a wave packet. The damping is as in (a).

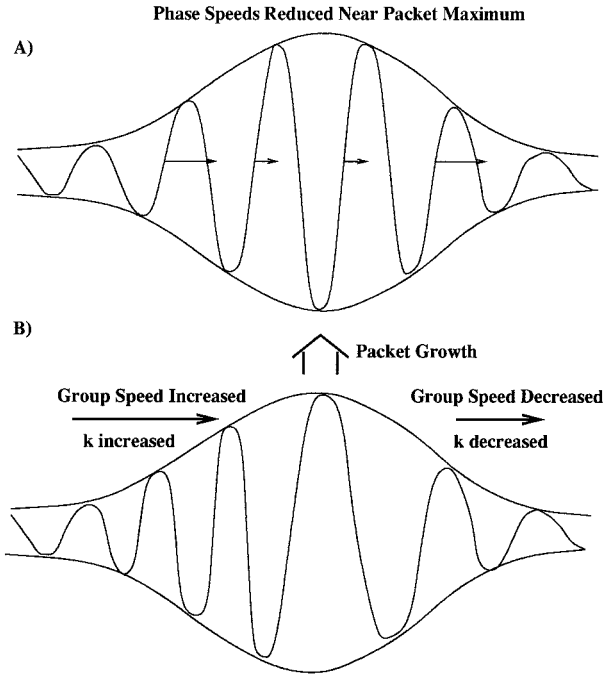

FIG. 7. A schematic picture illustrating the mechanism behind nonlinear self-focusing. See text for description.

4. Transition to full nonlinear baroclinic chaos in system B: Numerical experiments

To integrate Eq. (1) numerically, a higher-order diffusion term is introduced to inhibit any cascade to sub–grid scales,

$$\frac{D_i Q_i}{Dt} = d_i + \nu_0 \nabla^4 \phi_i \qquad i = 1, 2. \tag{29}$$

The boundary conditions applied at the sidewalls are (6) along with

$$\overline{\phi_{i_y}}^{x} = 0 \quad \text{on} \quad y = 0, L_y. \tag{30}$$

This is equivalent to imposing the quasigeostrophic long-wave condition (19) in the weakly nonlinear system (see appendix A.) In the limit of low criticality, the numerical model therefore behaves as weakly nonlinear system B, and must be expected to allow convergence toward a steady wave train for all parameter settings investigated.

The model used is spectral in the zonal direction, with 64 waves, and grid point in the meridional direction, with 100 grid points. For consistency with the analytic work, it is again taken that $U_s = 1.0$, $F = 0.5$, and $L_y = 2\pi$. Here, $L_x = 20\sqrt{2}\pi$ is the channel length. Several integrations with the damping parameters at setting 2, $\{E_1 = 0.05, E_2 = 0.5, r = 0.05\}$, were carried out for different values of the criticality β. Other experiments

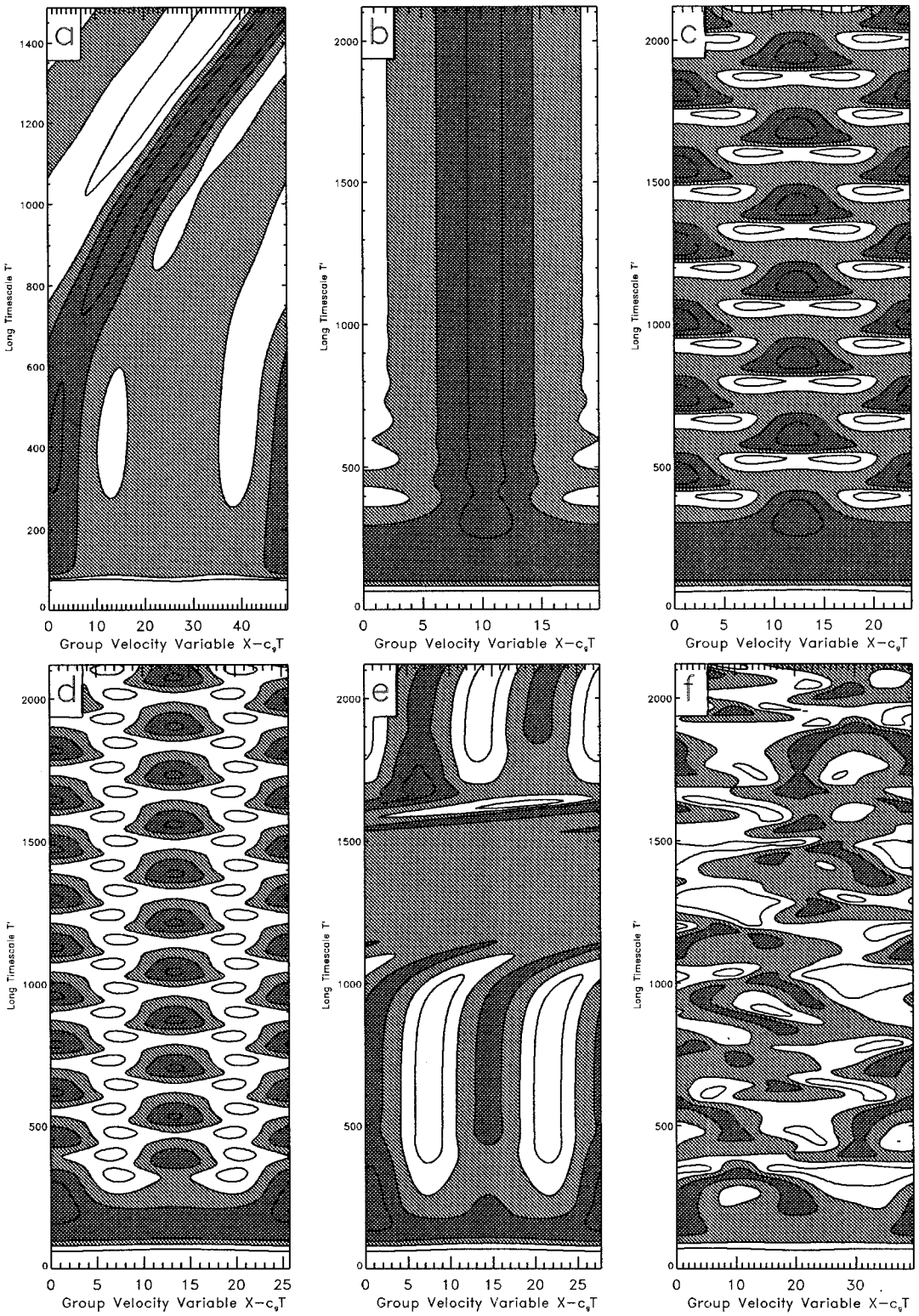

FIG. 8. Contour plots showing the evolution of $|A(\zeta, T')|$. In (a) the damping is set to setting 3, $\{E_1 = 0.05, E_2 = 0.25, r = 0.00\}$ and with the channel length set to $\zeta_0 = 50$. In (b)–(f) the damping is set to setting 2, $\{E_1 = 0.05, E_2 = 0.5, r = 0.05\}$, with channel lengths $\zeta_0 = 20$, 24, 26, 28, and 40, respectively. Contour intervals are (a) 0.1 (values above 1.4 in light shading, and above 1.5 in heavy shading), (b)–(e) 0.4, (0.8, 1.2), and (f) 0.5, (1.0, 1.5).

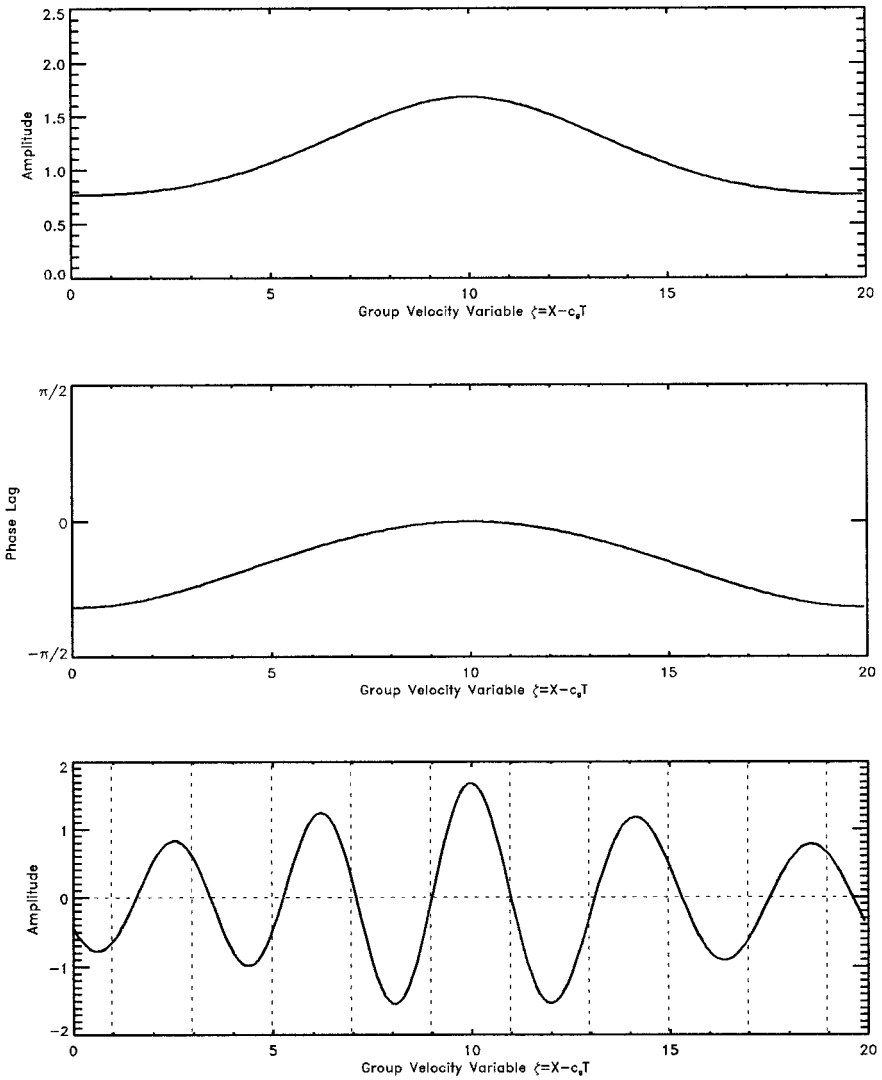

FIG. 9. Plots illustrating the steady solution for the integration with damping at setting 2 and $\zeta_0 = 20$ (see also Fig. 8b). (Above) Packet envelope $\mathcal{A}(\zeta)$, (middle) phase lag $f(\zeta)$, (below) $\mathcal{A}(\zeta)e^{if(\zeta)+i(kx-\omega t)}$ with $\epsilon = 0.45$ so that wave 5 is the dominant wavenumber in the channel.

with a range of different damping parameters were also carried out, and the qualitative behavior for the set presented was found to be representative. For each experiment, sufficient spinup time was allowed until a state of statistical equilibrium was reached before results were analyzed. The different experiments are summarized in Table 2, along with the values of ϵ that correspond to the weakly nonlinear system.

For each experiment, the deviation of the zonal flow field from its time-mean state, denoted by u_i^e, was analyzed using a cross-sectional EOF method. This method is described in appendix D. The intended purpose of this method was to isolate in an objective way the mean structure of the baroclinic waves, which are meridionally antisymmetric in u_i^e, and the primary long waves, which are forced by any zonal asymmetries in the baroclinic waves, which are symmetric in u_i^e. This enables the zonal propagation of these modes to be viewed independently in the form of a longitude–time plot. For comparative purposes, the linear complex phase speeds of the waves in the time-mean zonal flow were also determined by solving the complex eigenvalue problem, which is analogous to that in Swanson and Pierrehumbert (1994).

These numerical results are important for two reasons. First, they allow verification of the assumptions used in the weakly nonlinear theory of section 3 at finite criticality. Second, as the criticality of the system is increased, other fully nonlinear effects such as wave–wave interaction become important. These also lead to modulations of the zonally symmetric wave train. To assess the importance of a weakly nonlinear mechanism for packet formation, the importance of these finite criticality effects must also be evaluated. As system B ex-

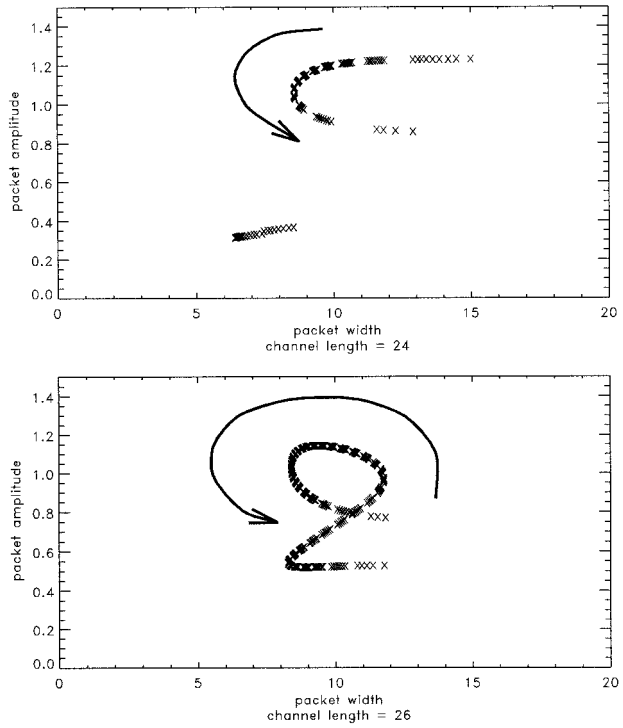

FIG. 10. Shows wave packets moving repeated along the same curves in amplitude/half-width space. Wave packet amplitude and half-width at each time step are marked by an "×." Arrows show the sense of packet evolution in time. Above: experiment c (damping-setting 2, channel length = 24). Below: experiment d (damping setting 2, channel length = 26).

hibits no weakly nonlinear instability, any wave packet–like behavior in these experiments must be due to fully nonlinear effects.

Figure 12 shows a plot of the square of the absolute value of the streamfunction for each zonal wavenumber integrated over both layers for several different experiments. These illustrate the different regimes of behavior discovered. At low criticality, in experiment b, only the baroclinic wave, which is wave 10 ($k = 2^{-1/2}$), is significantly excited. However, before it converges to a steady wave train it can be shown to be modulated in a weakly nonlinear fashion by wave 1. In experiment c, however, waves 3 and 7 are also excited and these form a dissipative resonant triad with wave 10. Experiment f is also characterized by such a triad between wave 5 and 10, although now there is some nonlinear excitation of other waves. Experiment h is much closer to full baroclinic chaos, as waves 7–10 are all baroclinically unstable, and interact to excite many other waves.

a. In a weakly nonlinear regime: Experiments a and b

Figure 13 shows results from experiment a, in which the baroclinic waves converge toward a steady wave train. This is the behavior predicted by Eq. (28). Various fields are shown at a time just before the wave train has reached its final symmetric state, in order that both the equilibrium state and the background flow response to zonal variations in the wave envelope can be tested against the predictions of weakly nonlinear theory.

1) THE AMPLITUDE OF THE WAVE TRAIN

Figure 13a is a snapshot of the upper-layer perturbation streamfunction ϕ_1 at the center of the channel, $y = L_y/2$. Weakly nonlinear theory predicts a value for the amplitude of the wave envelope[8] of $\epsilon|A| = \epsilon(-\rho_r'/\nu_r)^{1/2} = 0.571$, which gives good agreement with the diagram. The slight shortfall is possibly due to the discretization of the wavenumbers in the numerical experiment.

2) THE TIME-MEAN FLOW CORRECTION

Figure 13b shows the time-mean flow correction for the upper and lower layers, denoted by $\overline{u_i}^t$. The meridional structure is close to that predicted by (26) and shown in Fig. 6c (solid lines). According to (26), $\overline{u_1}^t = -(E_2/E_1)\overline{u_2}^t$. For this reason, the dotted line also shows $-E_2/E_1$ times $\overline{u_2}^t$. The vertical structure was found also to be in good agreement with the weakly nonlinear prediction. (It was found that $E_2\overline{u_2}^t \approx -E_1\overline{u_1}^t$ for each of experiments a–i, although there was some variation in the meridional structure of $\overline{u_i}^t$ in those experiments in which more waves were excited.)

3) THE STRUCTURE OF THE FORCED SYMMETRIC WAVE

Figure 13c shows the third cross-sectional EOF for the experiment, which has the structure of the symmetric wave, which is forced by the wavenumber one signal in the wave envelope. Weakly nonlinear theory predicts that it has structure given by $(\Phi_i^{(1)\prime})_y$, calculated from (27), and shown in Fig. 6c (dashed lines). Similarly, one expects $(\Phi_1^{(1)\prime})_y = -(E_2/E_1)(\Phi_2^{(1)\prime})_y$. Agreement with weakly nonlinear theory is not quite so accurate, it predicts slightly too large a response in the lower layer.

4) THE GROUP VELOCITY

Figure 13d is a contour plot of the phase of this EOF as it evolves in longitude and time. This shows that this wave number one signal, which is expected to travel at the linear group velocity, propagates at around 0.69 of the upper-layer basic-state velocity. This compares with $c_g = 0.847$ predicted by linear theory (see Table 1). As mentioned before, however, the group velocity is very

[8] Note that a corrected value of ρ has been used (see appendix B).

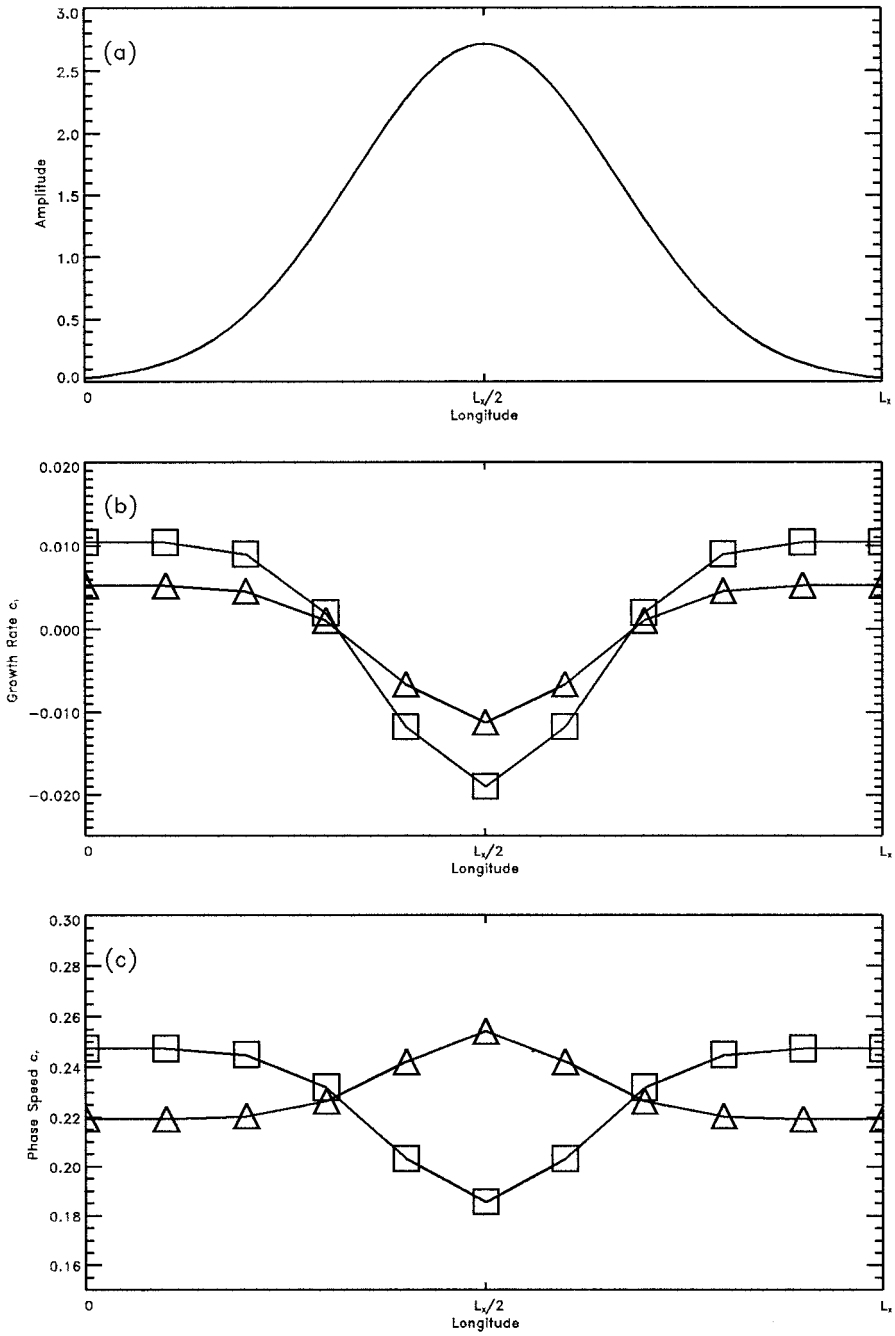

FIG. 11. A diagram showing the variation in the linear growth rate and phase speed of the fastest growing mode along a wave packet in systems A and B. The background flow is calculated at different positions in X, from (20) and from (26) and (27), respectively, with the damping coefficients at setting 2, and the criticality set to $\epsilon = 0.297$ ($\beta = 0.500$). (a) The amplitude of the wave packet against longitude. Its maximum value is set so that the zonal mean flow is marginally stable to wave growth. (b) The imaginary phase speed c_i for system A (squares) and system B (triangles), respectively. (c) The real phase speed c_r for system A (squares) and system B (triangles), respectively.

TABLE 2. A summary of the numerical experiments, with the corresponding value of ϵ in the weakly nonlinear problem.

	β	ϵ	ν_0
(Weakly nonlinear case)	0.548	—	—
Experiment a	0.500	0.296	—
Experiment b	0.455	0.412	—
Experiment c	0.441	0.442	—
Experiment d	0.429	0.466	—
Experiment e	0.417	0.489	—
Experiment f	0.357	0.590	0.001
Experiment g	0.313	0.655	0.002
Experiment h	0.278	0.702	0.002
Experiment i	0.250	0.737	0.005

sensitive to small changes in criticality. The long-wave response is slowly decaying as the system converges to a steady wave train.

5) THE ZONAL LOCATION OF BACKGROUND FLOW RESPONSE WITH RESPECT TO THE WAVE ENVELOPE MAXIMUM

The snapshot in Fig. 13a corresponds to $t = 0$ in panel d. It is possible to see that the wave number one perturbation to the wave envelope is in phase with the long-wave response, and in the sense expected. It was confirmed that the maximum of the background flow response coincided with the wave envelope minimum throughout the convergence to a steady wave train.

A similar agreement with weakly nonlinear theory was found for experiment b. This shows that the assumptions made in section 3 are relevant for a finite range of the critical parameter β. If the criticality is increased further, a new regime of wave–wave interaction is entered. In other experiments with a longer channel length, wave–wave interaction was found to occur for much smaller criticality.

b. Excitation of dissipative resonant triads: Experiments c–f

A relatively small increase in the criticality of the system from that for experiment b is all that is required for the system to enter a different regime of behavior. Weakly nonlinear resonant triads (e.g., Pedlosky 1975) consist of sets of three waves that have the properties:

$$k_1 + k_2 + k_3 = 0,$$
$$l_1 + l_2 + l_3 = 0,$$
$$\omega_1 + \omega_2 + \omega_3 = 0, \tag{31}$$

where k_i, l_i, ω_i, $i = 1, 2, 3$ are, respectively, the zonal and meridional wavenumbers and the frequencies of the three waves. The consequence of these conditions being met is that any pair of waves will then force only the other, through the nonlinear $J(\phi_i, q_i)$ terms in the equation, ensuring that the same three waves can persist independently of any others, provided that two are initially excited.

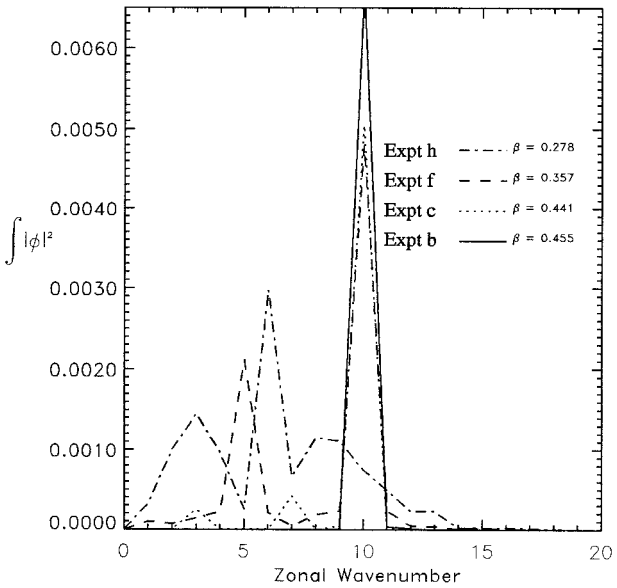

FIG. 12. A plot of the time-mean integral of the absolute value squared of the streamfunction against zonal wavenumber over the channel. Four different numerical experiments are shown: experiments b, c, f, and h.

The reason the resonant triads that have been identified in these experiments are described as dissipative is that the linear frequencies of the waves in the time-mean flow do not add to zero. This is because two of the waves concerned are decaying by an $O(1)$ amount and the other is slightly unstable. This means that it is necessary for $O(1)$ forcing from the waves to alter the phase speeds of each other, by wave–wave interaction, before resonance can occur. The equilibrium amplitudes of each waves in the triad are therefore fixed in an a priori sense by the system, as any pair must provide sufficient forcing to render the third wave marginally stable. Tables 3 and 4 are a comparison of linear wave speeds calculated from the time-mean flow and the actual frequencies measured in the model experiments c and f, respectively.

Although the frequencies and structure of the waves have been altered from their linear state by wave–wave interaction, they remain relatively close to that state when compared with the forced long waves of section 4a.

Figures 14 and 15 show the most of the significant variability of the zonal flow from its time-mean state for experiments c and f. The baroclinic waves are modulated by the symmetric waves, and the peaks in the baroclinic wave train are seen to move at the same velocity as the phase speed of the symmetric waves. The linear group speed of the baroclinic waves is apparently unimportant in this regime. The results for experiments d and e were very similar to those for experiment c.

The question of the prediction of wavenumber selec-

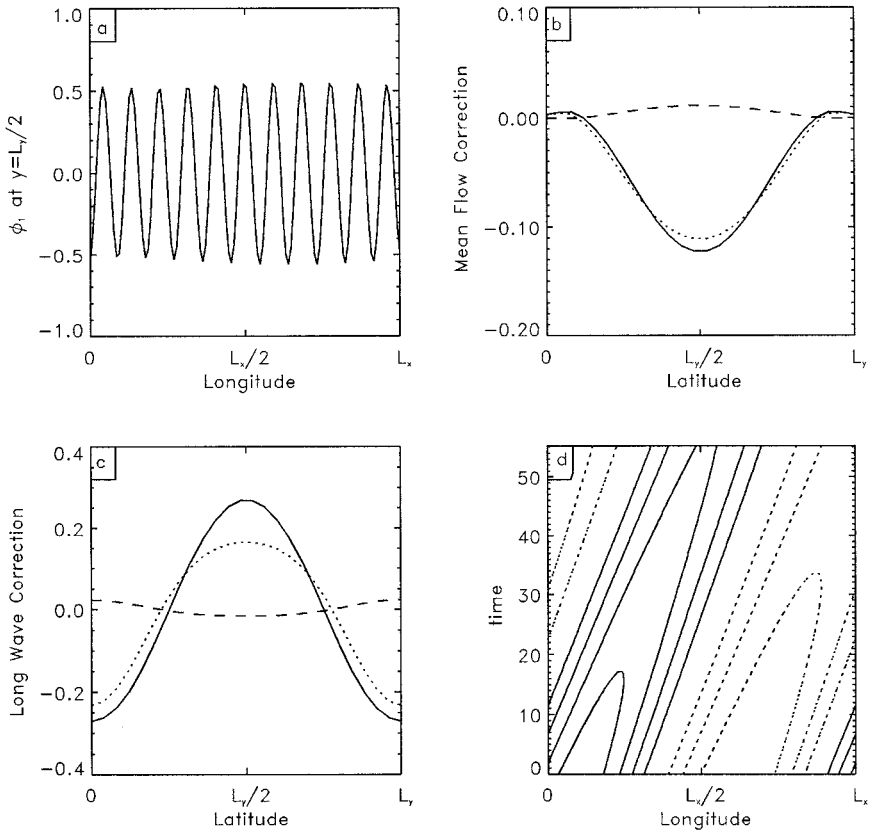

FIG. 13. (a) A snapshot of the upper-layer streamfunction ϕ_1 at the center of the channel ($y = L_y/2$) in experiment a. (b) The time-mean flow correction $\overline{u}_i^t$ in each layer. The solid line corresponds to the upper layer ($i = 1$), and the dashed line the lower ($i = 2$). The dotted line shows $-E_2\overline{u}_2^t/E_1$. (c) The structure of the principal symmetric EOF over this period. Again the solid line corresponds to the upper layer, the dashed line the lower, and the dotted line $-E_2/E_1$ times the lower-layer structure. The y-scale is arbitrary, and has been chosen so that when multiplied by the phase given in panel (d), the contribution of the EOF to the zonal velocity u^e is obtained at that longitude and time. (d) A longitude–time plot showing the evolution of the phase of the EOF shown in panel (c). The snapshot in panel (a) corresponds to $t = 0$ in this diagram.

tion in these triads appears to be a difficult one to answer. Figure 16 shows the real and imaginary phase speeds of the waves with meridional modes 1 and 2, in both the time-mean and radiatively balanced flows, for both experiments c and f. There is nothing remarkable about the phase speeds of those waves that form the resonant triad. In fact $\omega_1 + \omega_2 + \omega_3$ is closer to zero for alternative sets of three waves in both cases. It can only be concluded that in order to correctly predict which waves will be selected, it is necessary to take account of how the growth of any pair of waves affects the frequency of the other by nonlinear means as they grow to $O(1)$ size.

c. *Transition to full baroclinic chaos: Experiments g–i*

Figures 17 and 18 show the evolution of the u_i^e variability for experiments g and i. Wave packet–like structures

TABLE 3. A comparison of the linear frequencies calculated with respect to the time-mean flow, with the actual frequencies of the three waves with finite amplitude in experiment c.

Experiment c	Meridional mode (n)	Linear frequency	Actual frequency
Zonal wave 10	1	0.122 + 0.004i	0.125
Zonal wave 7	2	0.189 − 0.016i	0.203
Zonal wave 3	1	−0.016 − 0.040i	−0.078

TABLE 4. A comparison of the linear frequencies calculated with respect to the time-mean flow, with the actual frequencies of the three waves with finite amplitude in experiment f.

Experiment f	Meridional mode (n)	Linear frequency	Actual frequency
Zonal wave 10	1	0.148 + 0.014i	0.163
Zonal wave 5	2	0.138 − 0.013i	0.151
Zonal wave 5	1	0.009 − 0.019i	0.011

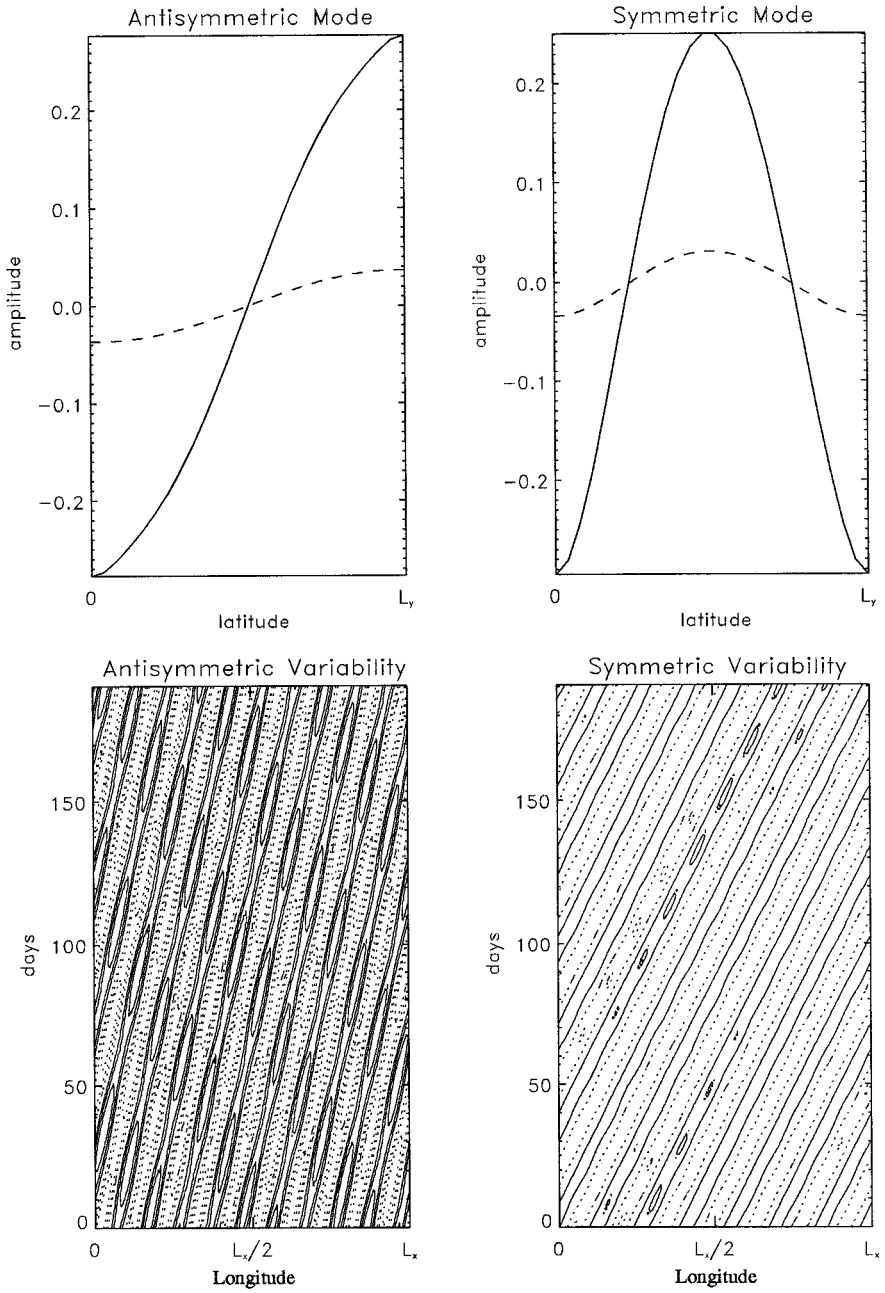

FIG. 14. The top panels show the structure of the first two EOFs for experiment c. The solid line corresponds to the upper-layer structure, and the dashed line the lower. The longitude–time plots show the evolution of the phase of each EOF. The contour interval is 0.4 in both plots.

are present in each experiment, but persist over a short timescale, particularly in experiment i. In both experiments, but more noticeably in experiment i, the wave packets undergo seemingly random phenomenological behavior, merging, growing, and decaying in a spontaneous fashion. This is an equilibrium state that appears qualitatively similar to the integrations of the weakly nonlinear Eq. (21). The structure of the forced waves generated, as in experiments c–f, however, shows that the dynamics are somewhat different. This is because any wave packet structures forced in these experiments are a result of fully nonlinear effects, as no weakly nonlinear packet forming instability is present. In experiment i, any wave packet structure would have a typical lifetime of approximately 50–100 model days (1 model day is defined by $U^{-1}L$). A crude comparison with the equivalent weakly nonlinear experiment for system A (see Fig. 8f) reveals that packets persist for 100–300 long time units, which equates to 200–600

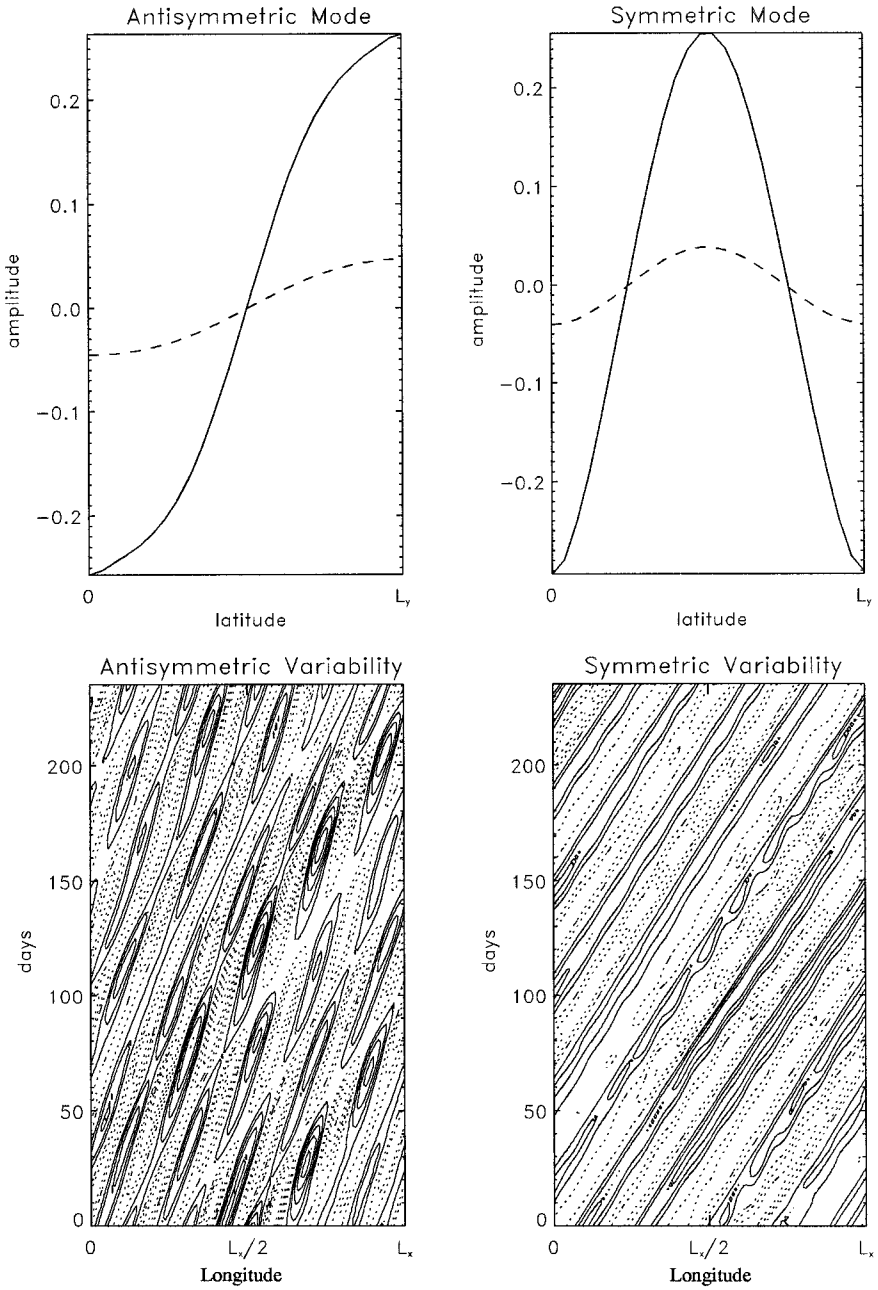

FIG. 15. As in Fig. 14 but for experiment f. The contour interval is 0.5 in both plots.

model days. The rapid evolution in experiment i can be explained because the zonal variations in the wave envelope are simply due to several baroclinically unstable waves attaining finite amplitude in the time-mean state. The wave envelope will then evolve rapidly, as there is little feedback from the background flow to help maintain wave packet structure for any length of time. The relatively long lifetime of packets in the weakly nonlinear system A is because the background flow exerts a feedback on the baroclinic waves to help maintain packet structures.

Figure 19 shows snapshots of some dynamical fields from experiment h. The upper-layer meridional velocity and potential vorticity plots show wave packets in the channel. The lower-layer potential vorticity diagram in particular shows some evidence of strong nonlinearity, with much deformation of the contours and the presence of cutoff vortices. There is also some evidence of wave breaking in the upper layer in those regions where the waves have large amplitude.

Plots similar to Fig. 10, showing the evolution of wave packets in amplitude half-width space, were also prepared for experiments g and i. However, despite some

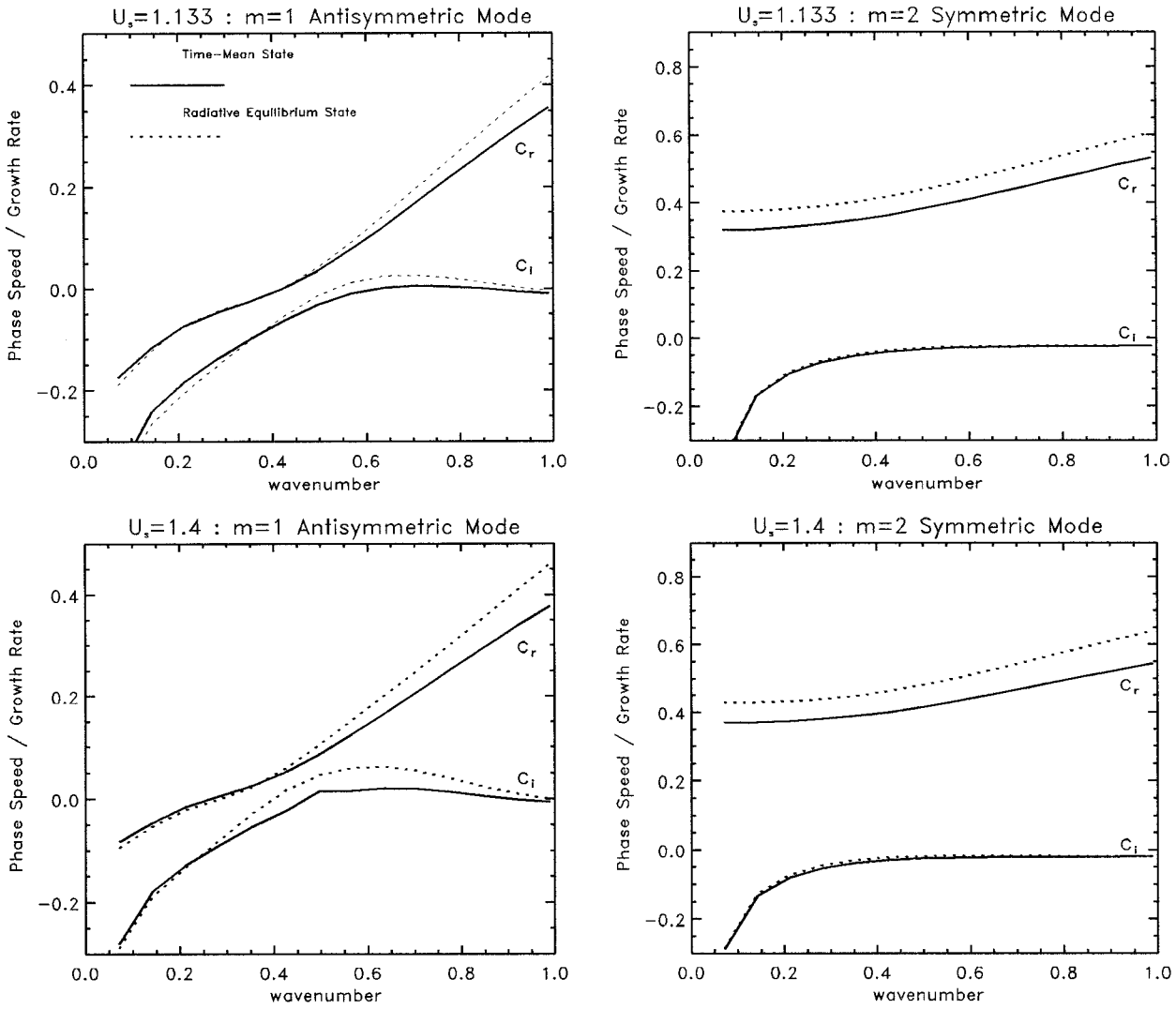

FIG. 16. Linear phase speeds and growth rates against wavenumber, for waves in the radiative equilibrium state and the time mean state for experiments c and f. The diagrams on the left show the fundamental ($n = 1$), and on the right show the second meridional harmonic ($n = 2$).

clustering, especially for experiment g, there was no evidence for the wave packets moving along fixed paths, they rather meander around in a random fashion. This is in contrast with the low resolution numerical results from the upper-layer jet experiments of LH (see their Fig. 9).

5. Summary and conclusions

In order to gain a better understanding of wave packet formation in an idealized forced, dissipative baroclinic system, the Phillips' two-layer model with appropriate forcing was investigated. In section 3, a weakly nonlinear balance was sought that represented a balance between the stabilizing effects of the wave forcing, and the restorative effects of the forcing/dissipation on the mean state. The boundary conditions for the basic flow correction were found to depend upon a parameter ξ, the ratio of the criticality ϵ to the Rossby number Ro. Two particular limits, $\xi \to 0$ and $\xi \to \infty$ were investigated, and these led to two separate weakly nonlinear systems, namely system A and system B.

Each system was shown to be governed by an amplitude equation that determined the evolution of the wave envelope over long time and space scales. The tendency of the equations to exhibit wave packet behavior was investigated. For system A, it was found that when the forcing/dissipation had a certain form, the steady wave train solution to the equation was unstable to small perturbations. In these cases, as the channel length increased, the system equilibrated to either a soliton type steady wave packet solution or unstable packets that followed well-defined curves in amplitude–half-width space, or chaotically behaving unstable wave packets.

These results identify a fundamental weakly nonlinear

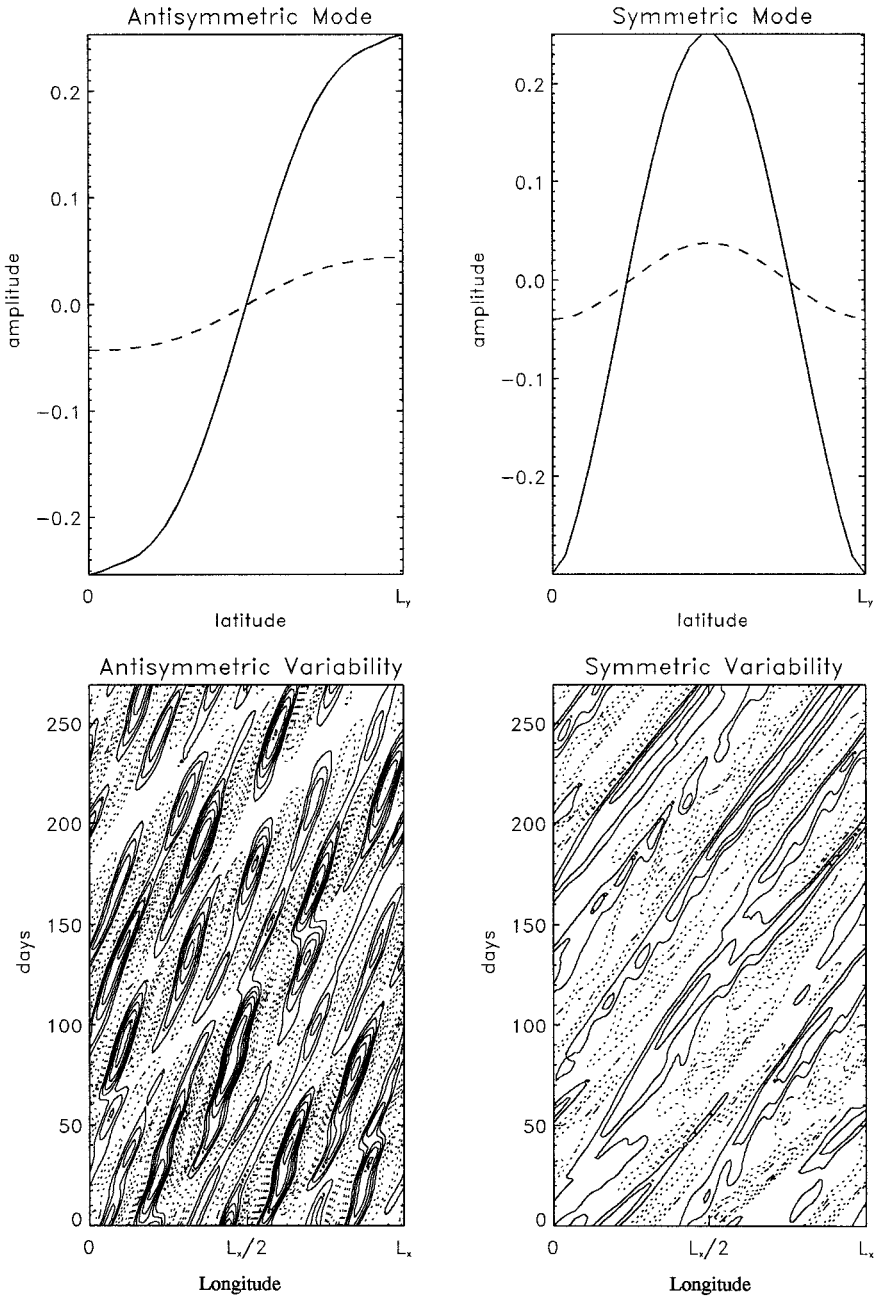

FIG. 17. As in Fig. 14 but for experiment g. The contour interval is 0.5 in both plots.

instability that could be claimed to lie at the root of atmospheric baroclinic wave packet formation. The mechanism, here termed *nonlinear self-focusing,* occurs when wave packets induce a background flow correction that has the feedback effect of reducing the frequency of the waves at the packet maximum. This causes the packet to adjust to a state where the waves at the rear or upstream side have a shorter wavelength than those at the front or downstream side. The enhanced group velocity of the shorter waves at the rear of the packet causes a convergence of wave activity at the packet maximum, and thus is a potential source of packet growth, which can oppose dissipation.

It is possible to draw parallels between this unstable weakly nonlinear system, and experiments (to be reported in a future paper) with an upper-layer jet present, such as those in LH. If the upper-level waveguide is well developed, in the sense that a region of strong PV gradient has formed, surrounded by well-mixed critical-layer regions, then nonlinear self-focusing of wave packets can be detected. This is perhaps not very surprising as there are some dynamical similarities between

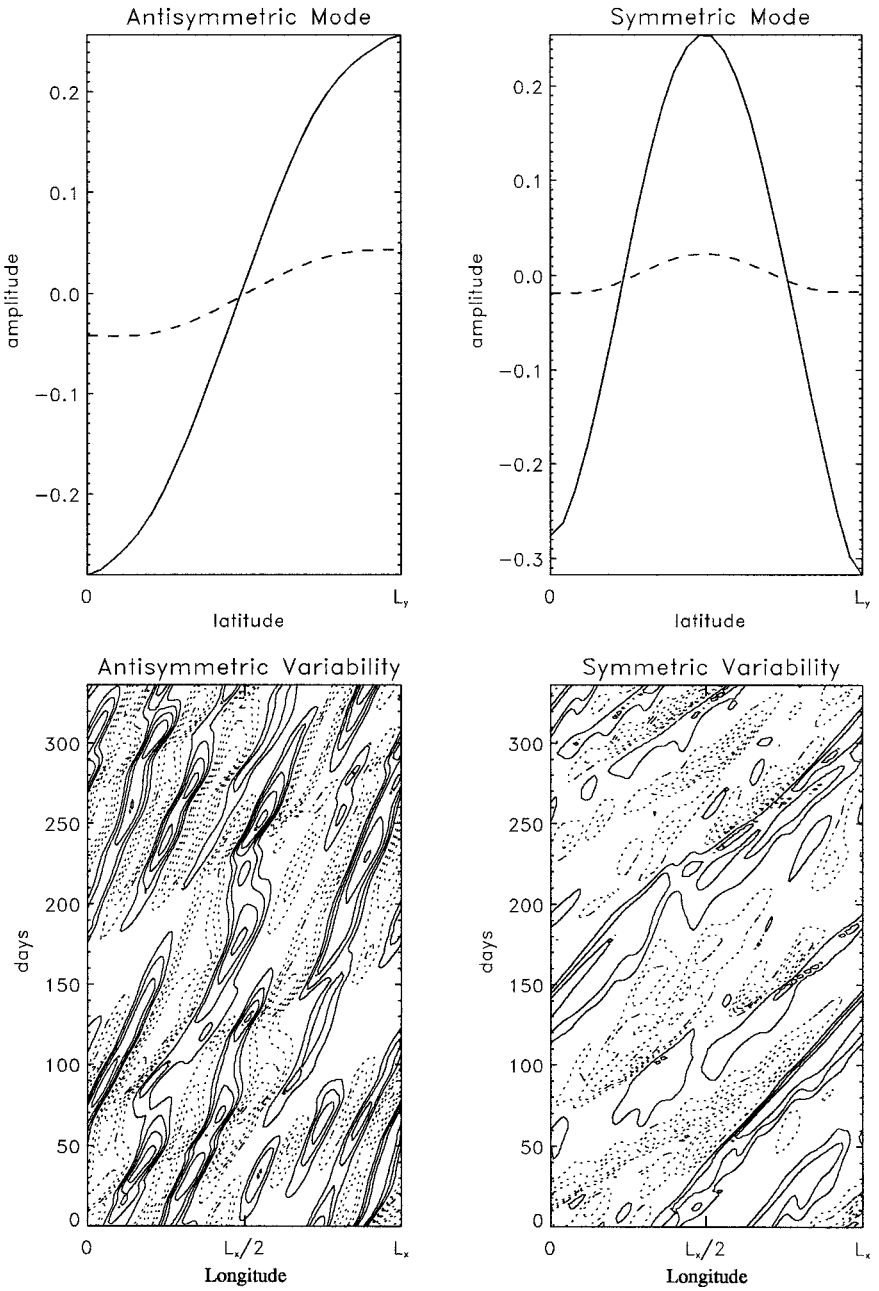

FIG. 18. As in Fig. 14 but for experiment i. Contour intervals are left: 0.8, right: 0.6.

such a jet experiment and weakly nonlinear system A. In each case waves are confined largely to a region of strong PV gradient in the upper layer, in the jet experiment by the well-mixed critical-layer regions and in the weakly nonlinear case by the sidewalls. However, the background flow is unconstrained in both cases, as in system A pressure gradients on the sidewall are supported by the ageostrophic effects of friction.

The numerical experiments of section 4 then explored the finite criticality extension of system B. At low criticality, convergence toward a uniform wave train was observed, as predicted by weakly nonlinear theory. However, as the criticality of the system is increased in experiments c–f, resonant triads of waves are excited, which owing to the dissipative nature of the system, equilibrate at a fixed amplitude. The nonlinear forcing of any pair of waves contribute to force the third wave. The amplitudes of the first two waves are fixed at a value so that the frequency of the third wave becomes real and equal to minus the sum of the first two waves. Wave–wave interaction therefore causes the breakdown of weakly nonlinear theory.

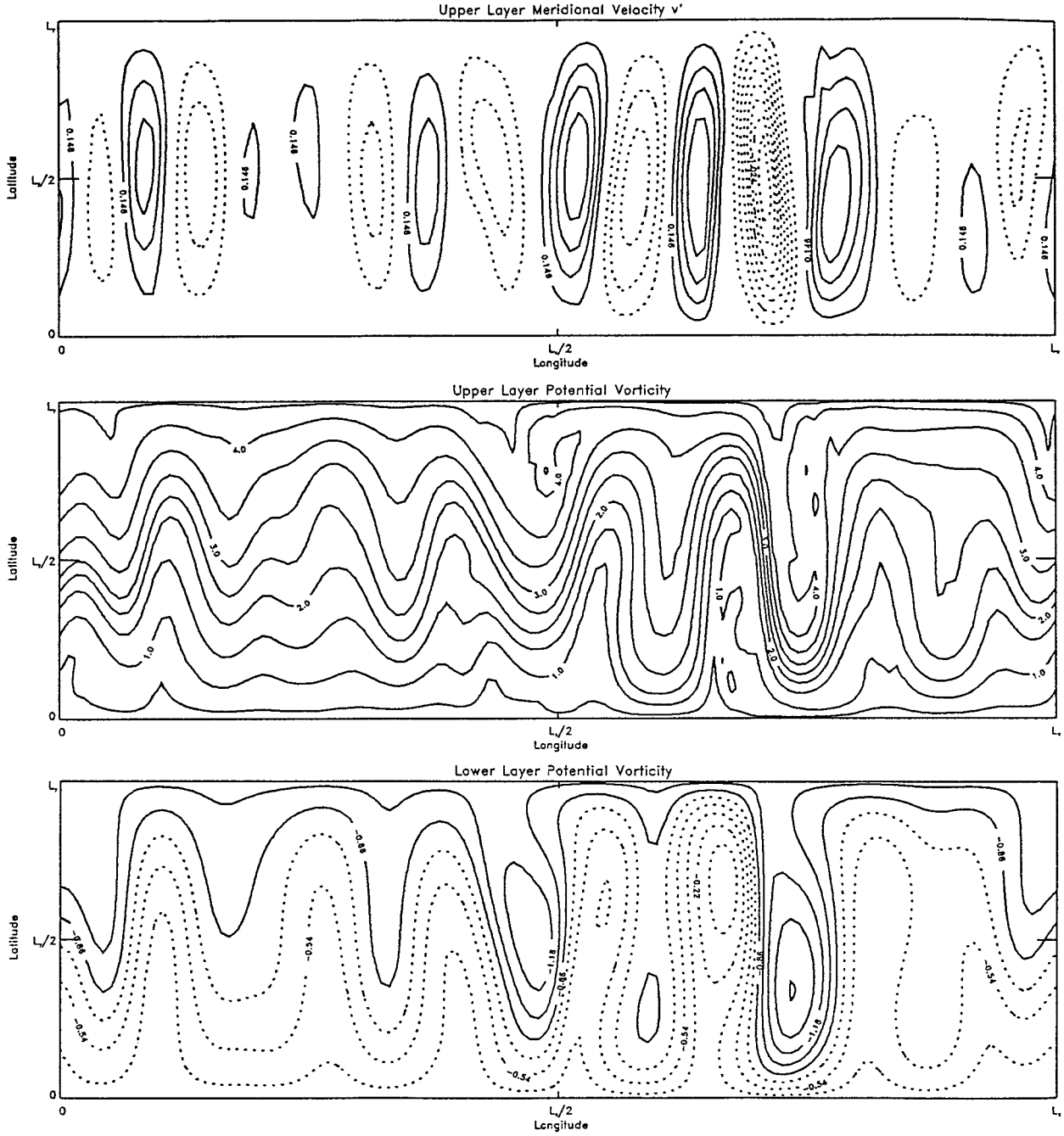

FIG. 19. A snapshot of (a) meridional velocity, (b) upper-layer potential vorticity, and (c) lower-layer potential vorticity in experiment h. Contour intervals are (a) 0.146, (b) 0.5, and (c) 0.16, respectively. It should be borne in mind that the axes are not in the correct ratio of $20\sqrt{2}$:2.

When β is further reduced, the interactions of the baroclinic waves that are excited force nearly all the waves in the system, and a state of baroclinic chaos is approached. In these experiments (g–i), packetlike structures are seen to decay and grow, as well as merge and divide. These wave packet structures are likely to be a consequence of several unstable baroclinic waves reaching finite amplitude in the time-mean state. This results in a zonally varying wave envelope that evolves rapidly. This rapid evolution can be explained by the absence of a feedback from the background flow upon the waves that could maintain the wave packets for a longer time.

In summary, this work has shown that two mechanisms can lead to zonal modulations in the envelope of baroclinic waves. The first of these, which is illustrated by weakly nonlinear system A, requires the background zonal flow forced by a wave packet to exert a feedback

on the baroclinic waves, in order to maintain the packet structure. The second mechanism is the effect of full nonlinearity, which was illustrated in section 4. It was shown that at high criticality, a rapidly varying wave envelope would result, regardless of interactions with the background flow. However, it seems baroclinic waves must feel some sort of feedback from the background flow, such as that described by weakly nonlinear system A, for wave packets to remain well defined and persist for long periods.

Chang (1993) and LH have recently published observations of baroclinic wave packets in the atmosphere in both the Northern and Southern Hemispheres. These packets are observed to undergo some phenomenology such as that discussed in section 4, but are usually distinct and well defined. To understand them fully, it is necessary not only to take account of barotropic shear, but also the effects of spherical geometry, storm track structure, tropospheric variability, and their interaction with low latitudes. This requires a hierarchy of models, of which those presented in this paper may form a useful part.

Acknowledgments. This work was funded by the Natural Environment Research Council (UK), with additional financial support from the UK Meteorological Office through a CASE studentship. The author would also like to acknowledge his Ph.D. supervisor Peter Haynes, as well as the three anonymous referees, for many ideas and comments relating to the work. Thanks are also due to John Scinocca and David Sankey who proofread the paper, Mark Taylor for the complex eigenvalue solver, and the rest of the atmospheric dynamics group at DAMTP for much advice over the last three years.

APPENDIX A

The Application of the Sidewall Boundary Conditions in the Weakly Nonlinear Analysis

The sidewall boundary conditions for Eq. (17) stem from the boundary conditions for the quasigeostrophic streamfunction ϕ_i. The first of these is $\phi_{ix} = 0$ on $y = 0, L_y$, [condition (6) in section 2]. A further boundary condition for the zonal-mean streamfunction is needed, which was first discussed by Phillips (1954). To understand the application of this boundary condition at higher order in ϵ for a weakly nonlinear problem, it is necessary to review the derivation of Eq. (1). It is derived by expanding the full two-layer shallow-water equations in powers of the Rossby number $\mathrm{Ro} = U/f_0L$ (Pedlosky 1970, 1987). The full perturbation pressure p_i^F and horizontal velocity (u_i^F, v_i^F) are expanded in Ro (e.g., $u_i^F = u_i^{[0]} + \mathrm{Ro}\, u_i^{[1]} + \ldots$). When p_i^F is suitably scaled, the momentum equations become

$$-v_i^{[0]} + \frac{\partial p_i^{[0]}}{\partial x} + \mathrm{Ro}\left\{\frac{\partial u_i^{[0]}}{\partial t} + U_i\frac{\partial u_i^{[0]}}{\partial x} + \frac{\partial}{\partial x}\left[\frac{1}{2}(u_i^{[0]})^2\right] + v_i^{[0]}\frac{\partial u_i^{[0]}}{\partial y} - v_i^{[1]} - \beta y v_i^{[0]} + \frac{\partial p_i^{[1]}}{\partial x} + E_i u_i^{[0]},\right\}$$
$$+ O(\mathrm{Ro}^2) = 0, \qquad i = 1, 2, \tag{32}$$

and

$$u_i^{[0]} + \frac{\partial p_i^{[0]}}{\partial y} + O(\mathrm{Ro}) = 0 \qquad i = 1, 2. \tag{33}$$

From the leading order form of (32) and (33), it is clear that the quasigeostrophic streamfunction $\phi_i \equiv p_i^{[0]}$. Condition (6) follows from the fact that $v_i^{[0]} = 0$ on the rigid boundaries.

To determine how the boundary conditions apply to the streamfunction of the basic flow correction, $\Phi_i^{(1)}$, it is necessary to make some assumption about the relative magnitudes of the two small parameters in the weakly nonlinear problem, ϵ and Ro. This can be done without loss of generality by writing

$$\epsilon = \xi\mathrm{Ro}. \tag{34}$$

It turns out that ξ is a measure of the importance of ageostrophic effects in modulating the response of the zonal flow to wave forcing, which varies on the long length scale X. After the transformations

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + \epsilon\frac{\partial}{\partial T} + \epsilon^2\frac{\partial}{\partial T'} \quad \text{and}$$
$$\frac{\partial}{\partial x} \to \frac{\partial}{\partial x} + \epsilon\frac{\partial}{\partial X}, \tag{35}$$

the x-average of the $O(\mathrm{Ro})$ part of (32) becomes

$$\frac{\partial \overline{u_i^{[0]}}^x}{\partial t} - \overline{v_i^{[1]}}^x + \xi\frac{\partial \overline{p_i^{[0]}}^x}{\partial X} + E_i\overline{u_i^{[0]}}^x = 0, \qquad i = 1, 2. \tag{36}$$

When we insert the mean flow correction $\epsilon^2\Phi_i^{(1)}(y, X - c_gT, T')$ for $\overline{p_i^{[0]}}^x$, we get the general sidewall boundary condition for $\Phi_i^{(1)}$,

$$\xi\Phi_{iX}^{(1)} + E_i\Phi_{iy}^{(1)} = 0, \quad \text{on} \quad y = 0, L_y, \; i = 1, 2. \tag{37}$$

Although Eq. (17) can be solved with (37), we have

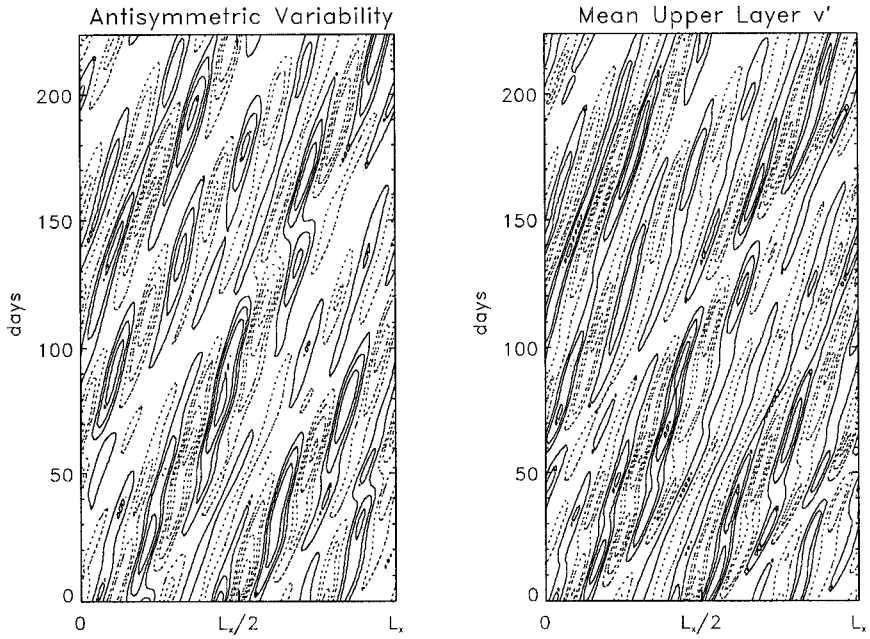

FIG. A1. A longitude–time plot showing the evolution of the principal antisymmetric cross-sectional EOF of the zonal wind u^e, beside a similar plot of the upper-layer meridional wind v_1^e, averaged across the channel. The two plots have similar structure, particularly if the $\pi/2$ phase lag is taken into account, as they both measure the signal from the baroclinic waves. Contour intervals are left: 0.8, right: 0.16.

restricted ourselves in this paper to the two special cases of $\xi \to 0$ and $\xi \to \infty$. The limit $\xi \to 0$ describes the case where ageostrophic effects allow the background flow to be determined locally in X, and (37) becomes condition (18),

$$\Phi_{iy}^{(1)} = 0 \quad \text{on} \quad y = 0, L_y, \qquad \text{(local zonal flow condition).}$$

This is the traditional boundary condition used in this problem (e.g., Tan and Liu 1995). Physically it can also be understood as a high viscosity limit, if (37) is interpreted as a balance between frictional effects and the pressure perturbations on the long length scale. As ageostrophic effects are important in the allowing the zonal flow to be determined locally, this boundary condition is *not* correct for the standard numerical model described in section 4. As there is no formal separation of scales in the numerical model, all the waves must be quasigeostrophic. This means that for the weakly nonlinear analysis to correspond to the numerical model, the appropriate limit is $\xi \to \infty$ (which can similarly be understood as a low viscosity limit), which leads to the boundary condition (19),

$$\overline{\Phi_{iy}^{(1)}}^{X} = 0 \quad \text{and} \quad \Phi_i^{(1)\prime} = 0 \quad \text{on} \quad y = 0, L_y, \qquad \text{(quasigeostrophic long wave condition).}$$

The prime in this equation denotes deviation from the X-average, and the condition forces all X variations in the background flow to fit the boundary conditions of quasigeostrophic waves.

APPENDIX B

Details of the Derivation of the Amplitude Equations and the Explicit Representation of their Coefficients

In this appendix we include the details of the derivation of Eqs. (21) and (28) in section 3, as well as the explicit representations of the quantities referred to in sections 2 and 3. Where appropriate, they have been shown numerically to be equal to their linear formulations.

First, in section 2, the complex constant giving the relative phase and amplitude of the linear baroclinic waves in the lower layer to those in the upper, γ, is given by

$$\gamma = \frac{(c - U_s)(a^2 + F) + \beta + FU_s + \dfrac{rFi}{k} + \dfrac{E_1 a^2 i}{k}}{(c - U_s)F + \dfrac{rFi}{k}} = \frac{cF + \dfrac{rFi}{k}}{c(a^2 + F) + \beta - FU_s + \dfrac{rFi}{k} + \dfrac{E_2 a^2 i}{k}}. \tag{38}$$

In section 3, Eq. (15) is derived from the second-order [$O(\epsilon^2)$] balance for the wavelike terms in Eqs. (12) and (13), which is given by

$$\begin{aligned} &-ik\mathcal{E}(B_2 - \gamma B_1) + (\gamma F - a^2 - F)A_T \\ &\quad + U_s(\gamma F - a^2 - F)A_X + (\beta + FU_s)A_X \\ &\quad + 2(ikE_1 + k^2c - k^2U_s)A_X = 0, \end{aligned} \tag{39}$$

$$\begin{aligned} &-ik\mathcal{F}(B_2 - \gamma B_1) + [F - \gamma(a^2 + F)]A_T \\ &\quad + (\beta - FU_s)\gamma A_X + 2(ikE_2 + k^2c)A_X = 0, \end{aligned} \tag{40}$$

where

$$\mathcal{E} = (c - U_s)F + \frac{rFi}{k}, \tag{41}$$

and

$$\mathcal{F} = c(a^2 + F) + \beta - FU_s + \frac{rFi}{k} + \frac{E_2a^2i}{k}. \tag{42}$$

If we add $\mathcal{F} \times$ (39) and $\mathcal{E} \times$ (40), we arrive at Eq. (15), where the group velocity c_g is given by

$$c_g = \frac{\mathcal{F}\left[U_s(\gamma F - a^2) + 2k^2\left(c - U_s + \frac{E_1i}{k}\right) + \beta\right] + \mathcal{E}\left[2\left(c + \frac{E_2i}{k}\right)k^2 + \gamma(\beta - FU_s)\right]}{\mathcal{F}(\gamma F - a^2 - F) + \mathcal{E}[F - \gamma(a^2 + F)]}. \tag{43}$$

It is a consequence of the linear expansion that $c_g = \partial\omega/\partial k$, which is real at the point of minimum critical shear. This can be shown by implicit differentiation, with repect to k, of the dispersion relation when it is in the form

$$\begin{aligned} &[(\omega - U_sk)(a^2 + F) + (\beta + FU_s)k + rFi + E_1a^2i] \\ &\quad \times [\omega(a^2 + F) + (\beta - FU_s)k + rFi + E_2a^2i] \\ &\quad - [\omega + rFi][\omega - U_sk + rFi]F^2 = 0. \end{aligned} \tag{44}$$

Without loss of generality, we may take $B_1 = 0$ [as $A(X, T, T')$ may be renormalized], so this leaves the equation for B_2 [Eq. (16)], in which the complex coefficient κ is given by

$$\kappa = \frac{2k^2\gamma\left(c + \frac{E_2i}{k}\right) + \gamma(\beta + FU_s) - c_g(F - \gamma(a^2 + F))}{\mathcal{F}k}. \tag{45}$$

Equation (21) is derived from the third-order balance [$O(\epsilon^3)$] for the wavelike terms in Eqs. (12) and (13), which (including y dependence) is given by

$$\begin{aligned} &-ik\mathcal{E}(C_2 - \gamma C_1)\sin ly + (\gamma F - a^2 - F)A_{T'}\sin ly - \Delta i\beta kA\sin ly \\ &\quad + \left[i\kappa F(U_s - c_g) + 2ik(U_s - c_g) - k^2\left(U_s - c - \frac{E_1i}{k}\right)\right]A_{\zeta\zeta}\sin ly - i\frac{\gamma_iFn^2\pi^2k^2|A|^2A}{(4n^2\pi^2 + \alpha^2)E_1E_2} \\ &\quad \times \left\{[E_2(4l^2 - a^2 + \gamma F) + E_1F]\sin ly\cos 2ly - \left[E_2\left(-\frac{\alpha^2}{L_y^2} - a^2 + \gamma F\right) + E_1F\right]\right. \\ &\qquad \left.\times \left[\cosh\frac{\alpha y}{L_y} - \sinh\frac{\alpha y}{Ly}\left(\frac{\cosh\alpha - 1}{\sinh\alpha}\right)\right]\right\}\sin ly = 0, \end{aligned} \tag{46}$$

and

$$ik\mathcal{F}(C_2 - \gamma C_1)\sin ly + [F - \gamma(a^2 + F)]A_{T'}\sin ly - \Delta i\beta k\gamma A \sin ly$$
$$+ \left[i\kappa(a^2 + F)c_g + ik(\beta - FU_s) - ik\left(2c_g + c + \frac{E_2 i}{k}\right)\gamma + k^2\left(c + \frac{E_2 i}{k}\right)\right]A_{\zeta\zeta}\sin ly$$
$$+ i\frac{\gamma_i F n^2\pi^2 k^2|A|^2 A}{(4n^2\pi^2 + \alpha^2)E_1E_2}\cdot\left\{[E_1(4l^2\gamma - a^2\gamma + F) + E_2F)\sin ly\cos 2ly\right.$$
$$\left. - \left[E_1\left(-\frac{\alpha^2\gamma}{L_y^2} + F - a^2\gamma\right) + E_2F\right]\left[\cosh\frac{\alpha y}{L_y} - \sinh\frac{\alpha y}{L_y}\left(\frac{\cosh\alpha - 1}{\sinh\alpha}\right)\right]\right\}\sin ly = 0. \quad (47)$$

Similarly, if we add $\mathcal{F}\times$ (46) and $\mathcal{E}\times$ (47), once the y dependence has been removed by multiplying each equation by $\sin ly$ and integrating across the channel, we arrive at Eq. (21). The coefficients in Eq. (21) are given by

$$\mu = i\frac{\mathcal{F}[(U_s - c_g)\kappa F + k(3U_s - \mathcal{D}_1)] + \mathcal{E}\kappa\left[c_g(a^2 + F) + \beta - FU_s + 2k^2\left(c + \frac{E_2 i}{k}\right) - \frac{k\gamma\mathcal{D}_2}{\kappa}\right]}{\mathcal{F}(\gamma F - a^2 - F) + \mathcal{E}[F - \gamma(a^2 + F)]}, \quad (48)$$

where

$$\mathcal{D}_i = c + 2c_g + \frac{E_i i}{k},$$

and

$$\rho = i\frac{\beta k(\mathcal{F} - \mathcal{E}\gamma)}{\mathcal{F}(\gamma F - a^2 - F) + \mathcal{E}[F - \gamma(a^2 + F)]}. \quad (49)$$

In the calculation of the amplitude of the wave envelope corresponding to that in Fig. 13, a corrected expression for ρ was used in order to allow for the scaling of the damping coefficients, $\{E_1, E_2, r\}$ with β, this value is

$$\rho' = \frac{(i\beta k - a^2E_1 - (1 - \gamma)rF)\mathcal{F} - [i\beta k\gamma - \gamma a^2E_2 + (1 - \gamma)rF]\mathcal{E}}{\mathcal{F}(\gamma F - a^2 - F) + \mathcal{E}[F - \gamma(a^2 + F)]}. \quad (50)$$

Finally, the nonlinear coefficient ν is given by

$$\nu = -i\frac{\gamma_i F n^2\pi^2 k^2}{2(4n^2\pi^2 + \alpha^2)E_1E_2} \times \frac{\mathcal{F}\mathcal{P} + \mathcal{E}\mathcal{Q}}{\mathcal{F}(\gamma F - a^2 - F) + \mathcal{E}[F - \gamma(a^2 + F)]}. \quad (51)$$

In this expression the quanities $\mathcal{P}$ and $\mathcal{Q}$ are given by

$$\mathcal{P} = E_2(4l^2 + \gamma F - a^2) + E_1F + \chi\left[E_2\left(-\frac{\alpha^2}{L_y^2} + \gamma F - a^2\right) + E_1F\right]$$
$$\mathcal{Q} = -E_1(4l^2\gamma + F - a^2\gamma) - E_2\gamma F + \chi\left[E_1\left(\frac{\alpha^2\gamma}{L_y^2} - F + a^2\gamma\right) + E_2\gamma F\right]$$

and

$$\chi = \frac{16n^2\pi^2(\cosh\alpha - 1)}{\alpha\sinh\alpha(4n^2\pi^2 + \alpha^2)}.$$

The quantities μ and ρ have been shown numerically to be equal to their linear formulations, which are given in Eq. (22).

In the derivation of the amplitude equation for system B, which follows that for system A exactly, the quantity ν_2 is given by (51) with the coefficient $\mathcal{X}$ set to zero in the expressions for $\mathcal{P}$ and $\mathcal{Q}$.

APPENDIX C

The Stability Properties of the System B Equation

In this appendix the relationship between the stability properties of Eqs. (21) and (28) is elucidated. Also, expressions for the linear dispersion relations of sideband modulations of the symmetric wave train are presented. Equation (28) may be written in the form

$$A_{T'} + (\mu_r + i\mu_i)A_{\zeta\zeta} = \rho_r A + (\nu_{2r} + i\nu_{2i})A|A|^2 + [(\nu_r - \nu_{2r}) + i(\nu_r - \nu_{2r})]A\overline{|A|^2}^X, \quad (52)$$

where ν_r, ν_{2r}, $\mu_r < 0$, and $\rho_r > 0$. Equation (21) can be recovered as a special case by setting ν_2 equal to ν. The ζ-independent Stokes solution for (52) is given by

$$A_{ST} = \left(-\frac{\rho_r}{\nu_r}\right)^{1/2} e^{-i(\rho_r\nu_i/\nu_r)T'}. \quad (53)$$

To determine the ζ-stability of this solution, a sideband solution of the form

$$A = \left(-\frac{\rho_r}{\nu_r}\right)^{1/2} e^{-i(\rho_r\nu_i/\nu_r)T'} \times (1 + A_1 e^{i(q\zeta - \sigma T')} + A_2 e^{-i(q\zeta - \sigma^* T')}) \quad (54)$$

may be inserted into (52) (where α^* denotes the complex conjugate of α), and then linearized with respect to A_1, A_2 in order to obtain a dispersion relation of the form $\sigma = \sigma(q)$. Note that after linearization,

$$|A|^2 = -\frac{\rho_r}{\nu_r}[1 + (A_1 + A_2^*)e^{i(q\zeta - \sigma T')} + (A_1^* + A_2)e^{-i(q\zeta - \sigma^* T')}] \quad (55)$$

and

$$\overline{|A|^2}^X = -\frac{\rho_r}{\nu_r}. \quad (56)$$

After inserting (54) into (52), and linearizing further, the following equations for A_1 and A_2 are obtained:

$$i\sigma A_1 + q^2(\mu_r + i\mu_i)A_1 - (A_1 + A_2^*)\rho_r\left(\frac{\nu_{2r} + i\nu_{2i}}{\nu_r}\right) = 0, \quad (57)$$

and

$$-i\sigma A_2 + q^2(\mu_r + i\mu_i)A_2 - (A_2 + A_1^*)\rho_r\left(\frac{\nu_{2r} + i\nu_{2i}}{\nu_r}\right) = 0. \quad (58)$$

The complex conjugate of (58) is

$$i\sigma A_2^* + q^2(\mu_r - i\mu_i)A_2^* - (A_2^* + A_1)\rho_r\left(\frac{\nu_{2r} - i\nu_{2i}}{\nu_r}\right) = 0. \quad (59)$$

As the two equations in A_1, A_2^* must be self-consistent, the following condition must hold

$$\sigma^2 + 2i\left(\frac{\nu_{2r}\rho_r}{\nu_r} - q^2\mu_r\right)\sigma - q^4(\mu_r^2 + \mu_i^2) + 2\frac{\rho_r}{\nu_r}(\nu_{2r}\mu_r + \nu_{2i}\mu_i)q^2 = 0. \quad (60)$$

Solving for σ,

$$\sigma = -i\left(\frac{\nu_{2r}\rho_r}{\nu_r} - q^2\mu_r\right) \pm \left[-\left(\frac{\nu_{2r}\rho_r}{\nu_r} - q^2\mu_r\right)^2 + q^4(\mu_r^2 + \mu_i^2) - 2\frac{\rho_r}{\nu_r}(\nu_{2r}\mu_r + \nu_{2i}\mu_i)q^2\right]^{1/2}. \quad (61)$$

Instability is only possible for $\sigma_i > 0$. As $\mu_r < 0$ we know that

$$\left(\frac{\nu_{2r}\rho_r}{\nu_r} - q^2\mu_r\right) > 0, \quad (62)$$

so for $\sigma_i > 0$ to hold, we require

$$q^4(\mu_r^2 + \mu_i^2) - 2\frac{\rho_r}{\nu_r}(\nu_{2r}\mu_r + \nu_{2i}\mu_i)q^2 < 0 \quad (63)$$

for some q. It is clear that as $\nu_r < 0$ and $\rho_\tau > 0$ this condition will be met (for some small q) if and only if

$$\nu_{2r}\mu_r + \nu_{2i}\mu_i < 0. \quad (64)$$

By setting $\nu_2 = \nu$ in (52) Newell's criterion (24) for the system A Eq. (21) is recovered. The dispersion relation for sideband instabilities for system A is also obtained:

$$\sigma = -i(\rho_r - q^2\mu_r) \pm \left[-(\rho_r - q^2\mu_r)^2 + q^4(\mu_r^2 + \mu_i^2) + 2\frac{\rho_r}{\nu_r}(\nu_r\mu_r + \nu_i\mu_i)q^2\right]^{1/2}. \quad (65)$$

APPENDIX D

A Method of Cross-Sectional EOFs

The method of empirical orthogonal functions (EOFs) (North 1982), was adapted and exploited in order to investigate the principal variance of the zonal flow in

the experiments of section 4. In the experiments, the quasigeostrophic two-layer model was forced in such a way that in the long time mean, the model is zonally invariant. This means that some of the statistical methods often used to investigate propagating signals, such as complex EOFs (CEOFs) or principal oscillation patterns (POPs), have certain limitations. Specifically, the problem is that any accurately evaluated signal using these methods will retain some zonal symmetry in common with the system. We are interested in wave packets, which are a zonally asymmetric phenomenon on an intermediate timescale, and would be averaged out by the above methods. An alternative statistical method was devised and is described as follows.

For brevity, some familiarity with the method of evaluating EOFs is assumed (e.g., North 1982). The basic premise of the method is to relax the constraint that when evaluating the covariance matrix S, the matrix must give the covariance between every spatial point of each dimension. Instead, only the covariance between spatial points in the meridional-height plane are taken, with the zonal coordinate x being treated as an alternative time coordinate. Therefore the components of S are given by

$$S_{ij} = \sum_{x,t} u^e_{(j)}(x, t)u^e_{(k)}(x, t), \tag{66}$$

where $u^e_{(j)}$ is the variation from the time-mean zonal flow at a position j in the meridional-height plane. Analogously to standard EOFs, the eigenvectors of this matrix are orthogonal structures in the meridional-height plane, which describe optimal amounts of the variance of u^e in time and the x direction, a measure of this variance being given by the corresponding eigenvalue. Maps of what are termed the principal components (denoted by $\mathcal{P}$), showing the propagation of these ''optimal'' structures in longitude and time, are then easily obtained by exploiting their orthogonality,

$$\mathcal{P}^{(\alpha)}(x, t) = \sum_j e^{(\alpha)}_{(j)}u^e_{(j)}(x, t), \tag{67}$$

where $e^{(\alpha)}$ is the normalized eigenvector of interest.

These are the longitude–time plots given in section 4. In each of Figs. 14, 15, 17, and 18, over 90% of the variance of u^e in the (x, t) plane is described by the two signals shown. To support the claim that the principal components of the first eigenvector are a reliable measure of the baroclinic waves, Fig. A1 is a comparison with a similar plot, for experiment f, of the meridional velocity averaged across the channel. The latter is a more typical measure of baroclinic waves in geophysical flows (e.g., in LH), and the two methods are observed to identify similar wave packet structures with waves that are out of phase by $\pi/2$. The signals are out of phase because v^e is a ''differentiated'' quantity when compared with u^e.

REFERENCES

Balmforth, N. J., 1995: Solitary waves and homoclinic orbits. *Annu. Rev. Fluid. Mech.,* **27,** 335–373.

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

Chang, E. M., 1993: Downstream development of baroclinic waves as inferred from regression analysis. *J. Atmos. Sci.,* **50,** 2038–2053.

Hocking, L. M., and K. Stewartson, 1972: On the non-linear response of a marginally unstable plane parallel flow to a two dimensional disturbance. *Proc. Roy. Soc. London Ser. A,* **326,** 289–313.

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

Lee, S., and I. M. Held, 1991: On sub-critical instability and hysterises in the two-layer model. *J. Atmos. Sci.,* **48,** 1071–1077.

——, and ——, 1993: Baroclinic wave packets in models and observations. *J. Atmos. Sci.,* **50,** 1413–1428.

Moon, H. T., P. Huerre, and L. G. Redekop, 1983: Transition to chaos in the Ginsburg–Landau equation. *Physica D,* **7,** 135–150.

Newell, A. C., 1974: Nonlinear wave motion. *Lect. Appl. Math.,* **15,** 157–163.

North, G. R., 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

Panetta, R. L., 1993: Zonal jets in wide baroclinically unstable regions. *J. Atmos. Sci.,* **50,** 2073–2106.

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

——, 1972: Finite amplitude baroclinic wave packets. *J. Atmos. Sci.,* **29,** 680–686.

——, 1975: The amplitude of baroclinic wave triads and mesoscale motion in the ocean. *J. Phys. Oceanogr.,* **5,** 608–614.

——, 1987: *Geophysical Fluid Dynamics.* 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. 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.

Simmons, A. J., and B. J. Hoskins, 1978: The life cycles of nonlinear baroclinic waves. *J. Atmos. Sci.,* **35,** 1364–1373.

——, and ——, 1979: The downstream and upstream development of unstable baroclinic waves. *J. Atmos. Sci.,* **36,** 1239–1254.

Stuart, J. T., and R. C. Diprima, 1978: The Ekhaus and Benjamin-Fier resonance mechanisms. *Proc. Roy. Soc. London Ser. A,* **362,** 27–41.

Swanson, K., and R. T. Pierrehumbert, 1994: Nonlinear wave packet evolution on a baroclinically unstable jet. *J. Atmos. Sci.,* **51,** 384–396.

Taha, T. R., and M. J. Ablowitz, 1984: Analytical and numerical aspects of certain non-linear evolution equations. Part II: Numerical nonlinear Schroedinger equation. *J. Comput. Phys.,* **55,** 192–202.

Tan, B., and S. Liu, 1995: Collision interactions of solitons in a baroclinic atmosphere. *J. Atmos. Sci.,* **52,** 1501–1512.

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

Whitaker, J. S., and A. Barcilon, 1995: Low-frequency variability and wavenumber selection in models with zonally symmetric forcing. *J. Atmos. Sci.,* **52,** 491–503.