J. Fluid Mech. (1987), *vol.* 184, *pp.* 477–504
Printed in Great Britain

Stable and unstable shear modes of rotating parallel flows in shallow water

By Y.-Y. HAYASHI
Geophysical Institute, University of Tokyo, Tokyo 113, Japan

AND W. R. YOUNG
Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02143, USA

(Received 27 September 1985 and in revised form 22 April 1987)

This article considers the instabilities of rotating, shallow-water, shear flows on an equatorial β-plane. Because of the free surface, the motion is horizontally divergent and the energy density is cubic in the field variables (i.e. in standard notation the kinetic energy density is $\frac{1}{2}h(u^2+v^2)$). Marinone & Ripa (1984) observed that as a consequence of this the wave energy is no longer positive definite (there is a cross-term $Uh'u'$). A wave with negative wave energy can grow by transferring energy to the mean flow. Of course total (mean plus wave) energy is conserved in this process. Further, when the basic state has constant potential vorticity, we show that there are no exchanges of energy and momentum between a growing wave and the mean flow. Consequently when the basic state has no potential vorticity gradients an unstable wave has zero wave energy and the mean flow is modified so that its energy is unchanged. This result strikingly shows that energy and momentum exchanges between a growing wave and the mean flow are not generally characteristic of, or essential to, instability.

A useful conceptual tool in understanding these counterintuitive results is that of disturbance energy (or pseudoenergy) of a shear mode. This is the amount of energy in the fluid when the mode is excited minus the amount in the unperturbed medium. Equivalently, the disturbance energy is the sum of the wave energy and that in the modified mean flow. The disturbance momentum (or pseudomomentum) is defined analogously.

For an unstable mode, which grows without external sources, the disturbance energy must be zero. On the other hand the wave energy may increase to plus infinity, remain zero, or decrease to minus infinity. Thus there is a tripartite classification of instabilities. We suggest that one common feature in all three cases is that the unstable shear mode is roughly a linear combination of resonating shear modes each of which would be stable if the other were somehow suppressed. The two resonating constituents must have opposite-signed disturbance energies in order that the unstable alliance has zero disturbance energy. The instability is a transfer of disturbance energy from the member with negative disturbance energy to the one with positive disturbance energy.

1. Introduction

The large-scale circulation of an ocean or atmosphere is continually disturbed and modified by spontaneous transient waves and eddies whose size and shape bear no relation to the pattern of external forcing. Since Charney and Eady's discovery of baroclinic instability this has been understood as the result of scale-selective instability. Their results, essentially based on the quasi-geostrophic approximation, are the basis of much of our intuition about the stability of large-scale geophysical flows. A central ingredient of the usual physical interpretation of these non-dissipative instabilities is an analysis of the energy balance. A basic result is that as an unstable wave grows, the wave energy increases. (We show below that this is not tautological.) The growing wave also modifies the mean flow so that the mean energy decreases. Total (wave plus mean) energy is conserved because the gain of wave energy is balanced by the loss of mean energy. Analogous results are well known for unstratified shear flows whose linearized stability is decided by the Rayleigh equation. One is tempted to conclude that the mean flow is a 'source' of energy for instability.

Here we argue that this conclusion is misleading in the broader context of ageostrophic, horizontal divergent, shallow-water instabilities. Specifically, Marinone & Ripa (1984) have shown that once the quasi-geostrophic approximation is abandoned the wave energy is no longer positive definite and may indeed be negative. A growing wave may have decreasing energy (i.e. its energy becomes more negative). The growing wave modifies the mean flow so that the mean energy increases. Once again total energy is conserved but in this case the loss of wave energy is balanced by the gain of mean energy.

Is one to conclude that in this 'negative-wave-energy instability' the growing wave is a source of energy? Instead we suggest that it is best to abandon the notion that instabilities require a source of energy. Spontaneously growing modal waves in a non-dissipative and unforced system are possible if the growing waves and their associated alterations in the mean flow (collectively referred to as the 'disturbance') do not change the total energy and momentum of the fluid. Thus if one can prove that all disturbances change the energy or momentum of a basic state then the flow is stable. A trivial example is a fluid at rest whose energy must increase if disturbed. But less obvious examples exist. For instance, Rayleigh's inflexion-point criterion can be physically interpreted as momentum conservation (Taylor 1915). If a parallel, non-divergent shear flow has no inflexion points then all disturbances change the total momentum of the fluid. Consequently the shear flow is stable. An analogous interpretation in terms of energy exists for the Fjortoft theorem.

Further, and this is perhaps the most surprising result of the present work, instabilities can occur that do not entail any transfer of energy between the growing wave and the alterations in mean flow. A growing wave may have zero energy and distort the mean flow so that mean energy is unchanged. Again, total energy is conserved. In §2 we show that this is true for instabilities on any basic state with uniform potential vorticity. This is because the transfer of energy, and momentum, between wave and mean flow is proportional to potential vorticity flux. If the basic state has no potential vorticity gradients then this flux is zero. (Analogous results have been found for non-rotating stratified shear flows with piecewise-constant profiles of velocity and density by Cairns 1979.) This result seems so counterintuitive that a thorough study of the instabilities that exist on flows with uniform potential vorticity is suggested and this is the topic of §3.

These 'zero-wave-energy' instabilities stand on the border between the familiar, positive-wave-energy instabilities described by the Rayleigh equation and the negative-wave-energy instabilities discussed by Marinone & Ripa (1984). Admittedly the basic states that support them are very special but it is precisely for this reason that they are so interesting. The instabilities that exist on states with uniform potential vorticity vividly show that energy and momentum transfer, in either direction, between a growing wave and a mean flow is not essential to, or diagnostic of, instability.

The basic state whose stability is investigated is a one-layer, zonal, geostrophically balanced, zero-potential-vorticity, β-plane current which is bounded by two fronts at which the layer depth vanishes. Because the potential vorticity is uniform (in fact zero) its gradient does not change sign and yet the flow is unstable. As a special case we recover the unstable f-plane gravity current treated extensively by Griffiths, Killworth & Stern (1982, GKS hereinafter). Also closely related are the unstable gravity waves which exist on a non-rotating, shallow-water, Couette flow (Satomura 1981 *a*, *b*). The possibility of instability in these simple, prototypical basic states can be anticipated from Ripa's (1983) theorem (a generalization of the Rayleigh–Fjortoft theorem) which states that a sufficient condition for stability is that there is some constant α such that

$$(\alpha - U)^2 \leqslant gH \quad \text{and} \quad (\alpha - U)\, Q_y \geqslant 0 \qquad (1.1\,a, b)$$

for all y. (Here $U(y)$ is the zonal velocity, $H(y)$ the depth and $Q(y) = (f - U_y)/H$ the potential vorticity of the basic state.) Flows with uniform potential vorticity sometimes violate (1.1 *a*) and calculation reveals that they are indeed unstable. Ripa's proof of the stability condition in (1.1) relies on energy and momentum conservation in the fashion previously mentioned. If the basic state satisfies (1.1) then all disturbances alter either the energy or momentum of the fluid and so the flow is stable.

The choice of a basic state with zero potential vorticity is motivated by the integral arguments in §2 which show that in this case there are no transfers of energy and momentum between a growing wave and the mean flow. Our philosophy is to direct attention at this counterintuitive part of parameter space and in so doing illustrate a variety of theoretical ideas such as negative- and zero-wave-energy instability, Ripa's theorem and destabilizing resonances between geophysical waves.

Additional motivation for an analysis of a zero-potential-vorticity currents is provided by Sardeshmukh & Hoskins (1985) study of vorticity balances in the tropical troposphere. In the upper troposphere, at 150 mb, they find a symmetric band with zero potential vorticity (e.g. their figure 4 *a*). However, direct application of the stability calculation here is dubious because the tropospheric region of zero potential vorticity is bounded by a strong positive potential vorticity gradient, rather than a 'front' where the thickness of the troposphere vanishes. Hoskins, McIntyre & Robertson (1985) present isentropic potential-vorticity maps which show this boundary very clearly: the equatorial belt with zero potential vorticity abuts a poleward region whose high potential vorticity reflects its polar origins. The poleward region, which is not included in our basic state, is very active. Nonetheless, as a baseline for future calculations with realistic basic states, it is worthwhile to exhibit the large-scale instabilities of a flow with zero potential vorticity in their purest form. We think it is likely that the unusual energy balances illustrated here will survive an elaboration of the basic state. Our calculations also illustrate stated limitations of the 'invertibility principle' discussed by Hoskins *et al.* (1985). The instability discussed in §3 has no potential-vorticity signature and moreover is ageostrophic. These complications would make it extremely difficult to detect, and analyse dynamically, using isopycnal potential-vorticity maps.

The plan of this paper is as follows. Section 2 discusses some general results concerning energy and momentum transfers between waves and the associated modification of the mean flow. No restrictive assumptions about the form of the potential-vorticity distribution are made in this section. Section 3 formulates the linearized stability problem for perturbations on a parallel flow with zero potential vorticity. In §4 we discuss and illustrate the stable and unstable modes which exist on this profile. The instabilities are characterized as resonant interactions between shear modes. Section 5 analyses non-divergent instabilities characterized by the Rayleigh equation using this concept of resonating shear modes. Section 6 is the conclusion.

2. General results concerning mass energy and momentum transfers

2.1. *The equatorial β-plane*

We consider one active layer of fluid with a moving, free boundary (the front) which lies on an equatorial β-plane. The dimensional equations of motion are then

$$\frac{\mathrm{D}u}{\mathrm{D}t}-\beta yv=-gh_x, \tag{2.1a}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t}+\beta yu=-gh_y, \tag{2.1b}$$

$$h_t+\nabla\cdot(h\boldsymbol{u})=0, \tag{2.1c}$$

where g is the acceleration due to gravity and β the north–south gradient of the Coriolis parameter. Special cases of (2.1) include non-divergent, non-rotating two dimensional flow and f-plane motion.

The non-dimensional version of the above set is

$$\frac{\mathrm{D}u}{\mathrm{D}t}-yv=-h_x, \tag{2.2a}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t}+yu=-h_y, \tag{2.2b}$$

$$h_t+\nabla\cdot(h\boldsymbol{u})=0, \tag{2.2c}$$

and to recover dimensional results from non-dimensional

$$t\rightarrow t\beta L, \tag{2.3a}$$

$$\left.\begin{aligned}(x,y)&\rightarrow\frac{(x,y)}{L},\\ (u,v)&\rightarrow\frac{(u,v)}{\beta L^2},\end{aligned}\right\} \tag{2.3b}$$

$$h\rightarrow\frac{h}{\beta^2L^4/g}, \tag{2.3c}$$

where L is a lengthscale which we specify below.

There are three important, exact consequences of (2.1) or (2.2). First there is conservation of potential vorticity:

$$q\equiv\frac{y+v_x-u_y}{h},\quad \frac{\mathrm{D}q}{\mathrm{D}t}=0. \tag{2.4a, b}$$

Secondly there is energy conservation:

$$E \equiv \frac{1}{2}\int h^2 + h(u^2+v^2)\,\mathrm{d}a, \quad E_t = 0, \tag{2.5a, b}$$

where the integration is over the area occupied by the fluid. Finally there is x-momentum conservation:

$$M \equiv \int h(u-\tfrac{1}{2}y^2)\,\mathrm{d}a, \quad M_t = 0. \tag{2.6a, b}$$

2.2. *Weakly nonlinear changes in the mean flow and potential-vorticity fluxes*

The remainder of this section is a discussion of the energetics of the stable and unstable waves which are supported by an arbitrary basic state (U, H, Q). For clarity and generality we make no assumption about the potential vorticity of this flow. Following Ripa (1983) and Marinone & Ripa (1984) we write the solution of the full, nonlinear equations of motion as

$$u_* = U(y) + u(x,y,t), \tag{2.7a}$$

$$v_* = v(x,y,t), \tag{2.7b}$$

$$h_* = H(y) + h(x,y,t), \tag{2.7c}$$

i.e. the basic state plus a nonlinear disturbance. This notation differs from (2.1) but the context will clearly imply what is intended. The disturbance fields (u, v, h) can further be decomposed into a zonal mean (an x-average), denoted by an overbar, and a wave

$$u = \overline{u}(y,t) + u'(x,y,t) \quad \text{etc.} \tag{2.8}$$

For the weakly nonlinear disturbances considered here

$$U \gg u' \gg \overline{u}. \tag{2.9}$$

Note that even though the wave fields, u' etc., are much larger than the mean flow modifications, $\overline{u}$ etc., energetically both of these terms are on the same footing.

The linearized equations of motion are

$$\mathrm{D}u + HQv = -h_x - \theta_x, \tag{2.10a}$$

$$\mathrm{D}v + yu = -h_y - \theta_y, \tag{2.10b}$$

$$\mathrm{D}h + (Hu)_x + (Hv)_y = 0, \tag{2.10c}$$

where

$$\mathrm{D} = \frac{\partial}{\partial t} + U\frac{\partial}{\partial x} \tag{2.11}$$

is the linearized convective derivative. We have suppressed the primes in (2.10). The context will imply our intention. In writing (2.10) the quadratic nonlinearities have been neglected. We have also introduced a forcing term represented by the gradient of an arbitrary, external potential $\theta(x, y, t)$ into the momentum equations. This proves to be a useful device in understanding the properties of the stable modes that exist on the basic state. The essential idea is that these modes can be generated by the force $\boldsymbol{\nabla}\theta$, and we shall be able to calculate how the total (wave plus mean) energy and momentum of the fluid must be changed by this force in order to excite a stable wave. Of course the unstable waves do not require an external force to generate them. They can appear spontaneously because the total energy and momentum of the fluid is unaltered by their presence.

The first consequence of (2.10) is potential-vorticity conservation:

$$\mathrm{D}q+vQ_y = 0, \tag{2.12a}$$

$$q \equiv H^{-1}\{v_x-u_y-Qh\}, \tag{2.12b}$$

and we note that (2.12) is unaffected by the external force $\nabla\theta$. Thus the potential force can change the energy and momentum of the fluid (see the discussion below) but cannot alter the potential vorticity of fluid particles. Equation (2.12) can be integrated in the usual way by introducing the particle displacement η:

$$\mathrm{D}\eta = v \tag{2.13}$$

with the consequence

$$q = -\eta Q_y. \tag{2.14}$$

In obtaining (2.14) from (2.12*a*) it has been assumed that q is initially zero. Thus we are confining attention to normal mode perturbations which can arise without an external alteration of the initial potential vorticity of the fluid. A very important consequence of (2.13) and (2.14) is

$$\overline{vq} = -\tfrac{1}{2}(\overline{\eta^2})_t\, Q_y, \tag{2.15}$$

which shows that there are no potential-vorticity fluxes if the basic state has uniform potential vorticity (indeed there is then no perturbation potential vorticity in a normal mode).

2.3. *A strategy for finding conservation laws of the linearized equations illustrated using mass conservation*

A theme of this article is the relation between conservation laws of the exact nonlinear dynamics and their expression in the linearized stability problem. In principle one can ignore the nonlinear problem and discover these linearized conservation laws solely by manipulating the linearized system. But this is unsystematic and often requires ingenuity and patience. A better approach is to return to the nonlinear dynamics and use it to guide ones manipulation of the linearized system. We illustrate this procedure by discussing the linearized conservation law which corresponds to mass conservation.

This is not entirely straightforward because there is a free boundary (a 'front') at which $h_* = 0$. Thus the area of the (x,y)-plane occupied by the disturbed fluid differs from the area occupied by the basic state. In analogy with (2.7) we write

$$a_* = A+a, \tag{2.16}$$

where a_* is the area occupied by the disturbed fluid.

It must be admitted that this distortion has already introduced some difficulties which have been glossed over. Specifically, is it sensible to use a zonal average as in (2.8) and (2.15) when only part of a latitude line is covered with fluid? An entirely satisfactory answer to this question requires a redefinition of 'average' and probably the introduction of Lagrangian variables (e.g. Andrews & McIntyre 1978). However, here we confine attention to weakly nonlinear disturbances and neglect terms of order amplitude cubed. Thus we deal mostly with (2.10) which is defined on the undistorted band of latitudes. We find consistent (to order amplitude squared) conservation laws by multiplying (2.10) by u, v or h and zonally averaging over this undistorted band.

An example will make this clear. Because the disturbance does not alter the mass of fluid

$$\int H\,\mathrm{d}A = \int h_*\,\mathrm{d}a_*,$$

and using (2.7*c*) and (2.16) this is equivalent to

$$0 = \int h \, \mathrm{d}a + \int H \, \mathrm{d}a + \int h \, \mathrm{d}A. \tag{2.17}$$

The above is an exact nonlinear expression for the 'disturbance mass', i.e. the difference between the mass of the fluid with and without the disturbances. In the weakly nonlinear approximation only first term on the right-hand side of (2.17) can be calculated directly with the solution of (2.10). In fact a little thought shows that

$$\int h \, \mathrm{d}a = [\overline{h\eta}] + O(\text{amplitude})^3, \tag{2.18}$$

where the overbar denotes a zonal average and the square bracket denotes the difference between its contents evaluated at the two bounding, undisturbed fronts. One might refer to $\int h \, \mathrm{d}a$ as the 'wave mass', i.e. this is the contribution to the disturbance mass in (2.17) that is explicitly $O(\text{amplitude})^2$ and whose time derivative can be calculated from the linearized system.

Now because mass is conserved we anticipate that if we calculate the rate of change of $\int h \, \mathrm{d}a$ from (2.10) we shall find a conservation law. To do this we evaluate (2.10*c*) at the front, where $H = 0$, and use (2.13) at the front to integrate this relation. One obtains

$$h = -\eta H_y, \tag{2.19}$$

which we also recognize as the Taylor series expansion of the exact nonlinear result

$$H(y+\eta) + h(x, y, t) = 0. \tag{2.20}$$

From (2.19) we see directly that

$$\overline{h\eta} + \overline{\eta^2} H_y = 0. \tag{2.21}$$

This is a conservation law whose physical content is mass conservation. From (2.17) and (2.18) its implication is that

$$-[\overline{h\eta}] = [\overline{\eta^2} H_y] = \int H \, \mathrm{d}a + \int h \, \mathrm{d}A, \tag{2.22}$$

i.e. the nonlinear contributions to the disturbance mass can be expressed in terms of the solutions of the linearized problem.

The remainder of this section concerns energy and momentum conservation and the strategy is similar to the preceding calculation. One first expands the nonlinear conservation law as in (2.17), and then identifies the term, such as (2.18), whose rate of change can be calculated with the linearized equations of motion. This calculation is guaranteed to reveal a linearized conservation law such as (2.21). This is essentially the procedure used by Ripa (1983) and the importance of these results justifies their recapitulation here. Additionally, the free boundary introduces complications not considered by Ripa.

2.4. *Energy and momentum conservation*

In an unforced problem both the energy

$$E_* = \frac{1}{2} \int h_*(u_*^2 + v_*^2) + h_*^2 \, \mathrm{d}a_* \tag{2.23}$$

and momentum

$$M_* = \int h_*(u_* - \tfrac{1}{2} y^2) \, \mathrm{d}a_* \tag{2.24}$$

are conserved. Now expand (2.23) using (2.7) and (2.16). Neglecting the cubic terms leaves

$$E_* = E_0 + E_a + E_1 + E_2, \tag{2.25a}$$

$$E_0 = \frac{1}{2}\int HU^2 + H^2 \, \mathrm{d}A, \tag{2.25b}$$

$$E_a = \frac{1}{2}\int HU^2 + hU^2 \, \mathrm{d}a, \tag{2.25c}$$

$$E_1 = \int HUu + (H + \tfrac{1}{2}U^2)h \, \mathrm{d}A, \tag{2.25d}$$

$$E_2 = \frac{1}{2}\int H(u^2+v^2) + 2Uhu + h^2 \, \mathrm{d}A, \tag{2.25e}$$

which, apart from the extra term E_a, is the decomposition introduced by Marinone & Ripa (1984). Because H vanishes linearly with distance from the front, terms such as $H^2 \, \mathrm{d}a$ are negligible. The term E_2 is the wave energy and its rate of change can be calculated from (2.10). Because $E_* - E_0$ is a constant, changes in E_2 must be balanced by changes in E_1 and E_a. $E_1 + E_a$ is referred to as the mean energy. In some circumstances it might be informative to distinguish between E_1 and E_a, but here we lump them into one category. Finally the disturbance energy (or pseudoenergy) is the sum of all the terms quadratic in amplitude:

disturbance energy (E) = wave energy (E_2) + mean energy ($\hat{E}_1 = E_1 + E_a$).

Thus the disturbance energy is the difference between the energy of the fluid with the disturbance (E_*) and without the disturbance (E_0).

In general E, $\hat{E}_1$ and E_2 all have indefinite sign. It is perhaps counterintuitive that E is sometimes negative. One somehow feels that creating a disturbance in a parallel flow increases the total energy of the fluid. But this is not the case and the earliest references to this that we are aware of are Landahl (1962) and Benjamin (1963). (In these articles E is called the 'activation energy'.) Waves with $E < 0$ will be referred to as negative-energy disturbances and those with $E > 0$ as positive-energy disturbances. In the absence of forcing and dissipation both of these are stable because exciting them changes the total energy of the system.

An analogous procedure can be applied to (2.24):

$$M_* = M_0 + M_a + M_1 + M_2, \tag{2.26a}$$

$$M_0 = \int H(U - \tfrac{1}{2}y^2) \, \mathrm{d}A, \tag{2.26b}$$

$$M_a = \int H(U - \tfrac{1}{2}y^2) + (U - \tfrac{1}{2}y^2)h \, \mathrm{d}a, \tag{2.26c}$$

$$M_1 = \int Hu + (U - \tfrac{1}{2}y^2)h \, \mathrm{d}A, \tag{2.26d}$$

$$M_2 = \int hu \, \mathrm{d}A, \tag{2.26e}$$

where again cubic terms are neglected. M_2 is the wave momentum and the remaining quadratic terms, $M_1 + M_a$, define the mean momentum. The disturbance momentum (or pseudomomentum) is the sum of all the quadratic terms:

$$\text{disturbance momentum } (M) = \text{wave momentum } (M_2) + \text{mean momentum } (\hat{M}_1 = M_1 + M_a).$$

The disturbance momentum is the difference between the momentum of the fluid with the disturbance (M_*) and without the disturbance (M_0).

In Andrews & McIntyre's (1978) Generalized Lagrangian Mean (GLM) theory, M is pseudomomentum and E the pseudoenergy. We have preferred to refer to them as the disturbance momentum and disturbance energy because this emphasizes that they are equal to the amount of energy and momentum in the disturbed fluid minus that in the undisturbed. Notice that this definition does not rely on a small-amplitude expansion and indeed the GLM theory could be used to generalize the preceding results to finite amplitude. It is also likely that a Lagrangian theory would more elegantly accommodate the complications of a moving, free boundary.

For an unstable mode it is clear that both E and M must be zero, i.e. the disturbance amplifies spontaneously without an external source of energy or momentum. Stated differently, the total energy and momentum of an ideal fluid is unchanged by an unstable wave. On the other hand stable modes have non-zero disturbance energy or momentum. This is why they are stable – exciting them changes the total energy or momentum of the fluid and requires an external source. In fact we show below that for a stable mode E and M are connected by

$$E = cM, \tag{2.27}$$

where c is the phase speed of the mode.

Because $E = \hat{E}_1 + E_2$ is zero for an unstable mode, three different types of instability can be distinguished on the basis of energetics:

(*a*) $\hat{E}_1 \to -\infty$, $E_2 \to \infty$ (positive-wave-energy instability),
(*b*) $\hat{E}_1 = 0$, $E_2 = 0$ (zero-wave-energy instability),
(*c*) $\hat{E}_1 \to \infty$, $E_2 \to -\infty$ (negative-wave-energy instability).

In all three cases $E = \hat{E}_1 + E_2 = 0$ so that energy is conserved as the wave grows. Of course it is possible to refine the tripartite scheme above. For instance in (*b*) it might be interesting to calculate the transfers between E_1 and E_a. However for most purposes the above is probably adequate especially because we can now show that transfers between $\hat{E}_1$ and E_2 are very simply expressed in terms of potential-vorticity fluxes.

2.5. *The connection between energy and momentum transfers and particle displacements*

By direct calculation from (2.10) one can show that

$$m_{2t} + H^2\overline{vq} = -(H\overline{uv})_y - \overline{h\theta}_x, \tag{2.28}$$

where $m_2 = \overline{hu}$ is the density of M_2. Likewise for the energy

$$e_{2t} + UH^2\overline{vq} = -(UH\overline{uv} + H\overline{vh} + H\overline{v\theta})_y - \overline{h_t\theta}, \tag{2.29}$$

where $e_2 = \frac{1}{2}H(\overline{u^2} + \overline{v^2}) + (U\overline{hu} + \frac{1}{2}\overline{h^2})$ is the density of E_2. Apart from the terms involving the external forcing, these results, and their consequences (2.30) and (2.31) below, were given previously by Ripa (1983). Their importance justifies their recapitulation here. They show that the wave energy and momentum can change because of either external forcing (θ) or transfers with the mean flow. This second process cannot change the total energy or momentum of the fluid and indeed using (2.15) we can rewrite (2.28) and (2.29) as

$$m_t = -(H\overline{uv})_y - \overline{h\theta_x}, \tag{2.30a}$$

$$e_t = -(UH\overline{uv} + H\overline{vh} + H\overline{v\theta})_y - \overline{h_t\theta}, \tag{2.30b}$$

where

$$e = \tfrac{1}{2}H(\overline{u}^2+\overline{v}^2) + U\overline{hu} + \tfrac{1}{2}\overline{h}^2 - \tfrac{1}{2}H^2UQ_y\,\overline{\eta^2}, \tag{2.31a}$$

$$m = \overline{hu} - \tfrac{1}{2}H^2Q_y\,\overline{\eta^2}, \tag{2.31b}$$

are the densities of disturbance energy and disturbance momentum. Finally, integrating (2.30) over the area occupied by the basic state leaves the conservation laws

$$M_t = -\int \overline{h\theta_x}\,\mathrm{d}A, \quad E_t = -\int \overline{h_t\theta}\,\mathrm{d}A. \tag{2.32a, b}$$

This result, without the right-hand side, was anticipated by the arguments in the previous section. The direct calculation above has provided expressions for $\hat{M}_1$ and $\hat{E}_1$ in terms of particle displacements, e.g. the last terms in (2.31). This result is well known in the quasi-geostrophic context (e.g. Pedlosky 1979) and as one anticipates from Andrews & McIntyre's GLM formulation it is much more general. Indeed this provides further justification for referring to E_2 as the wave energy. That is, the rate of change of the positive definite quantity

$$\frac{1}{2}\int H(u^2+v^2)+h^2\,\mathrm{d}A \tag{2.33}$$

is not proportional to a weighted integral of the potential-vorticity flux and cannot be related to particle displacements in such a way that conservation of disturbance energy is transparent as in (2.32).

In a recent, very interesting discussion of the quasi-geostrophic initial-value problem, Held (1985) showed that disturbance momentum and disturbance energy are of fundamental importance because linear shear modes are orthogonal with respect to the norm defined by these quantities. Thus they can be used to decide whether an initial condition projects onto the discrete or continuum modes. Held also noted that in the quasi-geostrophic case the disturbance energy and disturbance momentum of a shear mode are related simply as in (2.27) where c is the phase speed of the mode. We now show this result is true for ageostrophic modes.

Imagine exciting a stable mode with phase speed c by gradually switching on external forcing of the form $\theta(x-ct, y, \epsilon t)$. It follows from (2.32) that $E-cM$ is constant, and zero before the forcing acted, so (2.27) is a consequence.

The results in this section are not dependent on any particular assumption about the potential vorticity of the basic state, but we now note that if $Q = (y-U_y)/H$ is uniform then q and $\overline{vq}$ are zero. This follows from (2.14) and (2.15). The implications for energy and momentum transfers between the wave and the mean flow modifications are startling. The last terms in (2.31a, b) are zero and so

$$M = M_2, \quad \hat{M}_1 = 0, \qquad E = E_2, \quad \hat{E}_1 = 0, \tag{2.34}$$

if $Q_y = 0$. Thus for a stable mode all of the disturbance energy and momentum are in the wave and none is in the mean flow modifications. For an unstable mode M, M_2, $\hat{M}_1$, E, E_2 and $\hat{E}_1$ must all be zero. There is no transfer of energy and momentum between the growing wave and the mean flow. It is this counterintuitive case that is analysed in the next two sections.

3. The linearized stability problem for flows with zero potential vorticity

3.1. *The basic state: zero-potential-vorticity zonal flow*

Suppose that the unperturbed flow (the basic state) is zonal (i.e. $\partial/\partial x = 0$) and has zero potential vorticity. Then from (2.4)

$$U_y = y, \tag{3.1}$$

where capital letters denote basic-state variables. Integrating (3.1) and then calculating H from the geostrophic balance gives

$$U = \tfrac{1}{2}(y^2-1)+\phi, \tag{3.2a}$$

$$H = \tfrac{1}{8}(y^2-1)(y^2-1+4\phi), \tag{3.2b}$$

$$\tfrac{1}{2}U^2+H = \tfrac{1}{2}\phi^2, \tag{3.2c}$$

where one of the constants of integration is ϕ and the other is determined by requiring that the front is at $y = 1$ (Salmon 1982). (This amounts to a specification of L in (2.3).) In terms of dimensional variables

$$\phi = \frac{U_f}{\beta L^2},$$

where U_f is the value of U at the front ($y = L$).

As the parameter ϕ is varied there are marked changes in the structure of the basic state: see figure 1. When $\phi > \frac{1}{4}$ the flows span the equator. If $\phi < \frac{1}{4}$ the current separates from the equator and extends from $y = \pm(1-4\phi)^{\frac{1}{2}}$ to $y = \pm 1$. Thus in this case the width of each branch is

$$\mu \equiv 1-(1-4\phi)^{\frac{1}{2}}. \tag{3.3}$$

Of course the dynamics of these two branches are uncoupled and by considering each in isolation we have a basic state that is asymmetric about the equator. As ϕ is reduced the width of the current (i.e. $\mu(\phi)$ in (3.3)) becomes smaller. Thus the f-plane limit discussed by GKS is recovered after rescaling the y-axis as in (3.4) below.

When $\phi < \frac{1}{4}$ it is often convenient to rescale and translate y so that the current spans the interval (0, 1). (Note how the current width μ depends on ϕ in (3.3).) Thus if

$$Y = \frac{y}{\mu}+1-\frac{1}{\mu}, \tag{3.4}$$

where μ is defined in (3.3), then the basic state is

$$U = \tfrac{1}{2}\mu^2(Y-Y_+)(Y-Y_-), \tag{3.5a}$$

$$H = -\tfrac{1}{8}\mu^4 Y(Y-1)\left(Y-2+\frac{2}{\mu}\right)\left(Y-1+\frac{2}{\mu}\right), \tag{3.5b}$$

$$Y_\pm = 1-\frac{1\pm(1-\mu+\mu^2/2)^{\frac{1}{2}}}{\mu}. \tag{3.5c}$$

This particular representation of the basic state makes it easy to take the singular limit $\phi \rightarrow 0$ and recover the f-plane problem discussed by GKS.

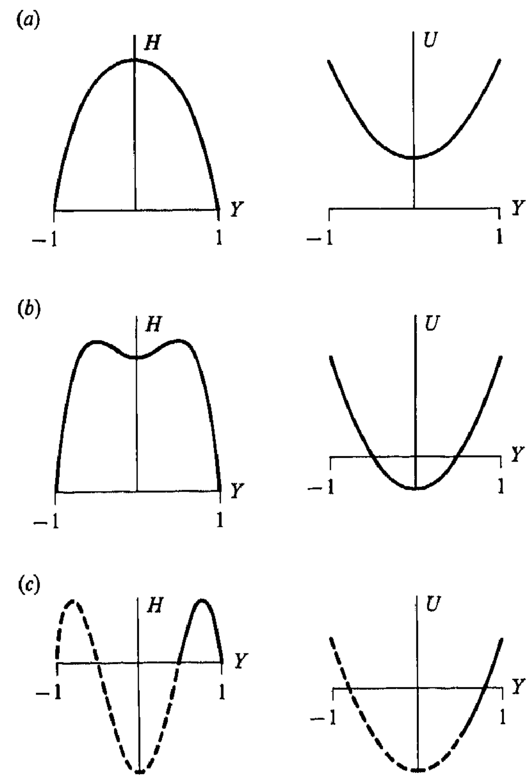

FIGURE 1. The basic state for various values of ϕ. (a) $\phi > \frac{1}{2}$, (b) $\frac{1}{2} > \phi > \frac{1}{4}$, (c) $\frac{1}{4} > \phi$. In (c) the flow is separated from the equator and the fronts are at $y = (1-4\phi)^{\frac{1}{2}}$ and 1.

3.2. *The linearized perturbation problem*

Now suppose that a perturbation (u, v, h) is superimposed on the basic state $(U, 0, H)$. Thus

$$u_*(x,y,t) = U(y) + u(x,y,t), \tag{3.6a}$$

$$v_*(x,y,t) = v(x,y,t), \tag{3.6b}$$

$$h_*(x,y,t) = H(y) + h(x,y,t), \tag{3.6c}$$

where $*$ indicates a total field. Substituting (3.16) into the equations of motion gives:

$$u_t + Uu_x + h_x = \text{quadratic terms}, \tag{3.7a}$$

$$v_t + Uv_x + yu + h_y = \text{quadratic terms}, \tag{3.7b}$$

$$h_t + Uh_x + (vH)_y + Hu_x = \text{quadratic terms}. \tag{3.7c}$$

Note that in (3.7a), the term $(U_y - y)\,v$ is absent because the basic state has zero potential vorticity. Assuming that the perturbation has normal mode form:

$$(u, v, h) = (u, v, h)\exp[\mathrm{i}k(x-ct)]$$

and neglecting the quadratic terms, reduces (3.7) to

$$(U-c)\,u + h = 0, \tag{3.8a}$$

$$yu + \mathrm{i}k(U-c)\,v + h_y = 0, \tag{3.8b}$$

$$Hu - \mathrm{i}k^{-1}(Hv)_y + (U-c)\,h = 0. \tag{3.8c}$$

Elimination of h between (3.8*a*) and (3.8*b*) gives an important intermediate result:

$$\mathrm{i}kv = u_y, \tag{3.9}$$

i.e. normal mode perturbations have zero potential vorticity. This is a special case of a more general, finite-amplitude result: because the basic state has constant potential vorticity, perturbation potential vorticity cannot be created by displacing particles. Equation (3.9) also implies that u_y is non-singular everywhere (even at the fronts where $H = 0$) and this is the boundary condition that is applied to the eigenvalue problem below.

Using (3.9), (3.8) can be put in the form

$$(Hu_y)_y - k^2[H-(U-c)^2]\,u = 0. \tag{3.10}$$

Together with the boundary condition that u is non-singular where $H = 0$, this is the linearized stability problem. Without loss of generality we can normalize the solution so that

$$u = 1 \quad \text{at } y = 1. \tag{3.11}$$

Before turning to the solution of (3.10) we first derive some results that restrict the distribution of eigenvalues (a weak version of Ripa's theorem and a semicircle theorem).

We should note that the eigenvalue problem (3.10) and (3.11) is formally very similar to that considered by Blumen (1970) and Blumen, Drazin & Billings (1975) in the context of shear instability in an inviscid, compressible fluid. Also formally similar, and closely related physically, is Satomura's (1981*a*, *b*) study of shear instability in non-rotating shallow water. Perhaps, however, the most closely related predecessor is the discussion in GKS of the ageostrophic instability of gravity currents on an f-plane. This is an important special case ($\phi = 0$) of the present investigation.

3.3. *A necessary condition for instability*

The calculations below will show that the stability boundary (i.e. the shape of the various neutral curves) has an intricate dependence on ϕ and k. Thus at the outset it is important to note any general results that restrict the region of parameter space in which instability is possible. As we mentioned in §1 one such is Ripa's (1983) theorem. For the basic flow in (3.2), (1.1*b*) is satisfied for any α. Condition (1.1*a*) requires that $\alpha = \phi$ because $H = 0$ at $y = \pm 1$. It is then easy to see that if the inequality is to obtain then $\phi > \frac{3}{4}$. Thus the zero-potential-vorticity profile in (3.2) is stable if $\phi > \frac{3}{4}$ and unstable if $\phi = 0$ (GKS).

For completeness we now derive this result directly from (3.10). We emphasize that this is a weaker version of Ripa's theorem both because normal mode perturbations are assumed and Lyapunov stability is not assured.

We follow Paldor (1983) and multiply (3.10) by the complex conjugate of u. We then integrate across the current and separate real and imaginary parts. Introducing

$$\tilde{U} = U - \phi = y^2 - 1, \tag{3.12a}$$

$$\tilde{c} = c - \phi = \tilde{c}_r + \mathrm{i}\tilde{c}_i \tag{3.12b}$$

one can put these in the form

$$\int \mathrm{d}y H|u_y|^2 + [H - \tilde{U}^2 + 2\tilde{U}\tilde{c}_r - \tilde{c}_r^2 + \tilde{c}_i^2]\,k^2|u|^2 = 0, \tag{3.13a}$$

$$k^2\tilde{c}_i \int \mathrm{d}y [\tilde{U} - \tilde{c}_r]\,|u|^2 = 0. \tag{3.13b}$$

If $\tilde{c}_i \neq 0$ then (3.13) becomes

$$\int \mathrm{d}y\, H|u_y|^2 + [H - \tilde{U}^2 + \tilde{c}_r^2 + \tilde{c}_i^2]\, k^2|u|^2 = 0. \tag{3.14}$$

But because

$$H - \tilde{U}^2 = \tfrac{1}{8}(1-y^2)(3y^2+4\phi-3) \tag{3.15}$$

the left-hand side of (3.14) is positive definite if $\phi > \frac{3}{4}$. Thus in this case $\tilde{c}_i$ must be zero.

3.4. *A semicircle theorem*

To obtain bounds on the size of the real and imaginary parts of c or $\tilde{c}$ we follow Howard (1961). Noting that

$$0 > \tilde{U} > a, \tag{3.16}$$

where

$$a = -\tfrac{1}{2} \quad \text{if } \phi > \tfrac{1}{4},$$
$$a = -2\phi \quad \text{if } \phi < \tfrac{1}{4},$$

one has

$$0 \geqslant \int (\tilde{U} - a)\, \tilde{U}|u|^2\, \mathrm{d}y = \int \tilde{U}^2|u|^2 - a\tilde{U}|u|^2\, \mathrm{d}y, \tag{3.17}$$

and then from (3.13) and (3.14)

$$0 \geqslant \int H\{|u_y|^2 + k|u|^2\}\, \mathrm{d}y + \{\tilde{c}_r^2 + \tilde{c}_i^2 - a\tilde{c}_r\} \int k^2|u|^2\, \mathrm{d}y. \tag{3.18}$$

Because the first term on the right-hand side is positive definite, this implies

$$(\tilde{c}_r - \tfrac{1}{2}a)^2 + \tilde{c}_i^2 \leqslant (\tfrac{1}{2}a)^2 \tag{3.19}$$

so that the complex wave speed for any unstable mode must lie inside the semicircle in the upper half-plane which has the range of U for diameter.

4. Numerical solution of the eigenvalue problem

4.1. *The case $\phi = 0$*

We begin discussion of the numerical results with the case $\phi = 0$. The numerical method is described in Appendix A. With the representation (3.2) the width of the current vanishes as $\phi \to 0$ and so it is convenient to use the rescaled basic state in (3.4) and (3.5). In fact this was done whenever ϕ was less than $\frac{1}{4}$ to ensure that all cases had identical numerical resolution, i.e. the basic flow extended from $Y = 0$ to $Y = 1$. Because the length has been rescaled by the factor μ it is also necessary to rescale c and k according to

$$\hat{k} \equiv \mu k, \quad \hat{c} \equiv \frac{c}{\mu},$$

and define

$$\hat{H} \equiv \frac{H}{\mu^2}, \quad \hat{U} \equiv \frac{U}{\mu},$$

so that the only changes in (3.10) are notational.

Figure 2 shows the dispersion curves, $\hat{c}$ as a function of $\hat{k}$ when $\phi = 0$. Apart from a factor of $\surd 2$ in the definition of wavenumber these results are identical with those in GKS. However besides these unstable Kelvin waves there are additional gravity-wave instabilities at higher wavenumbers, e.g. $k \approx 5.6$. These occur whenever the dispersion curves in figure 2(*a*) intersect. This pattern of unstable 'resonances' at

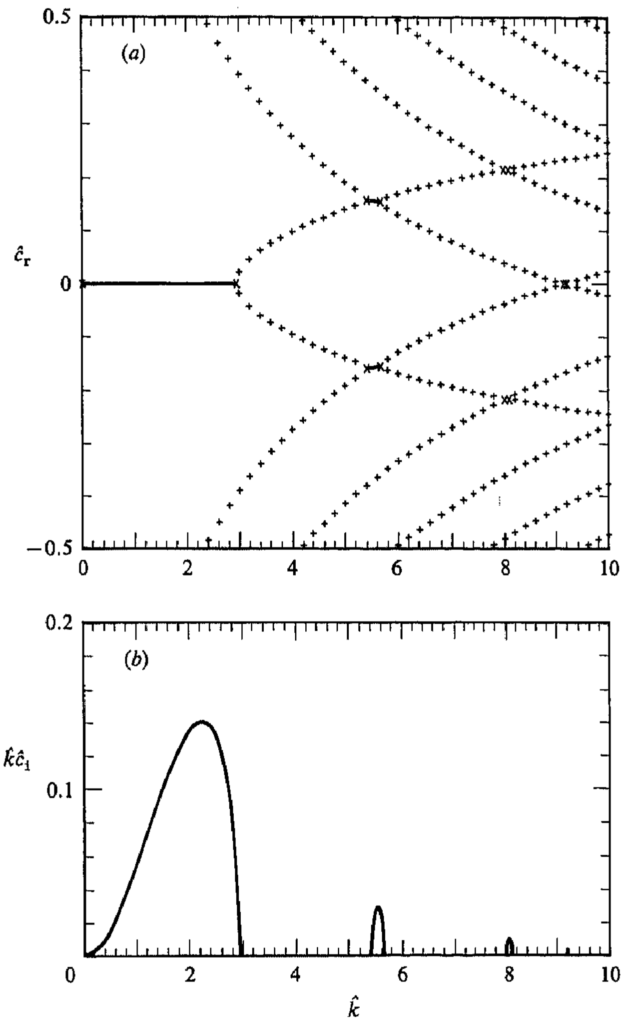

FIGURE 2. Dispersion curves when $\phi = 0$. (*a*) The real part of the phase speed as a function of wavenumber. (*b*) Growth rate as a function of wave number. In (*a*) intervals of instability are indicated by solid lines connecting ×'s. The instability with $c_r = 0$ is a resonance between two Kelvin waves and these modes are distinct and stable when $k > 2.95$. In addition to the Kelvin modes there are gravity-wave modes whose phase speeds are proportional to k^{-1} as $k \to 0$.

wavenumbers where the dispersion curves cross is very similar to that presented by Satomura (1981*a, b*). In §5 we advance a physical 'explanation' of this instability. In anticipation of the more detailed discussion in that section we remark that half of the stable modes in figure 2(*a*), have positive disturbance energy E and the other half have negative disturbance energy. The instabilities occur when a mode with $E > 0$ resonates (i.e. has the same phase speed and wavenumber) as a mode with $E < 0$. The unstable, resonant structure has $E = 0$ and can be thought of as transferring disturbance energy from the constituent with $E < 0$ to the one with $E > 0$. In this way both partners can spontaneously grow without an external source of energy. An analogous interpretation in terms of disturbance momentum is straightforward because of (2.27). This result shows that stable modes with opposite-

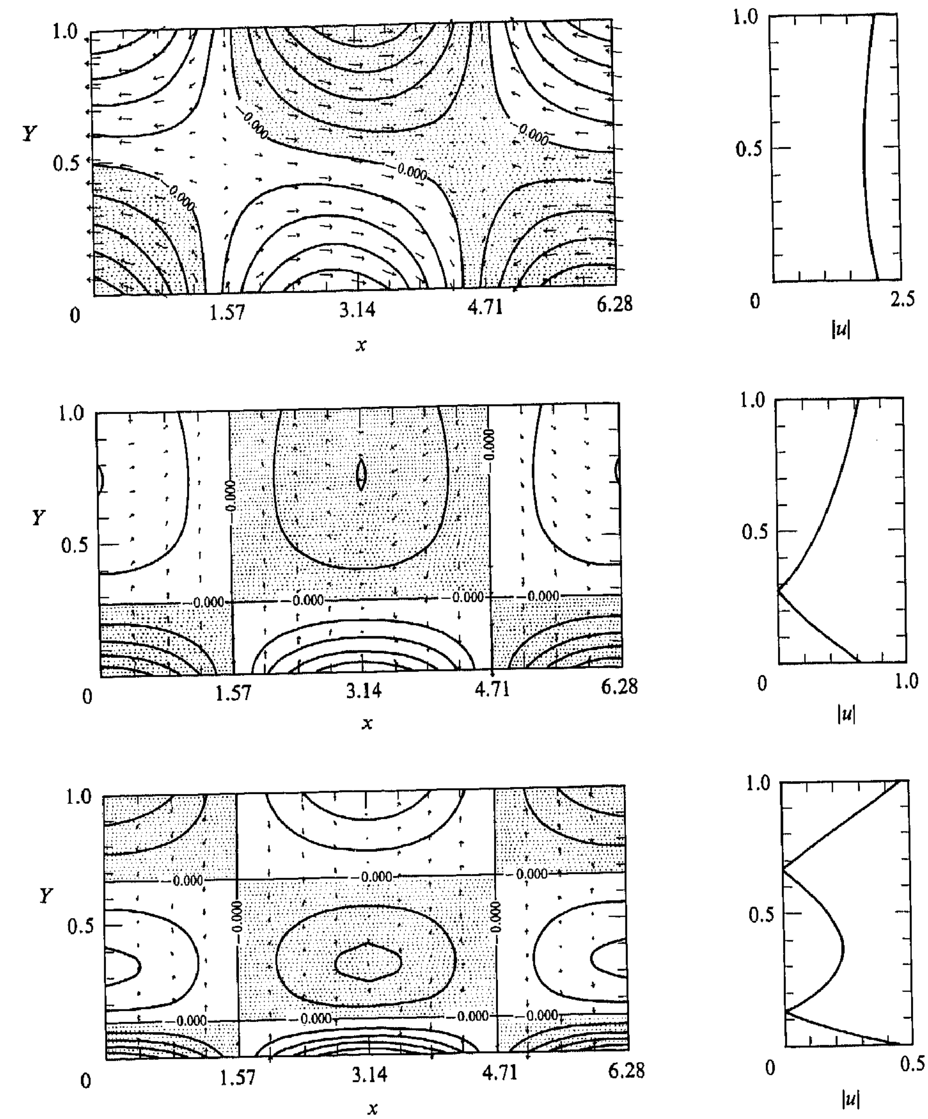

FIGURE 3. Horizontal structure of the waves when $\phi = 0$. (*a*) An unstable Kelvin mode with $\hat{k} = 1$, $\hat{c}_r = 0$, $\hat{k}\hat{c}_i = 5.818 \times 10^{-2}$. (*b*) A stable gravity mode with $\hat{k} = 1$ and $\hat{c}_r = 1.047$. (*c*) A stable gravity mode with $\hat{k} = 1$ and $\hat{c}_r = 1.754$.

signed energies, and the same phase speeds, also have opposite-signed disturbance momentum. Also it is clear that a linear combination that has $E = 0$ also has $M = 0$. Thus the energy and momentum constraints are simultaneously satisfied.

Figure 3 shows the horizontal structure of stable and unstable waves when $\phi = 0$. The shaded areas indicate regions where h is less than zero. In figure 3(*a*) we show the unstable Kelvin modes. The right-hand panels show $|u| = [u_r^2 + u_i^2]^{\frac{1}{2}}$. The waves in figure 3(*b*, *c*) are the first two stable gravity modes in figure 2(*a*). These higher modes are very similar to those discussed by Satomura (1981*a*, *b*). Because he neglected rotation, there was no Kelvin-like mode in his study.

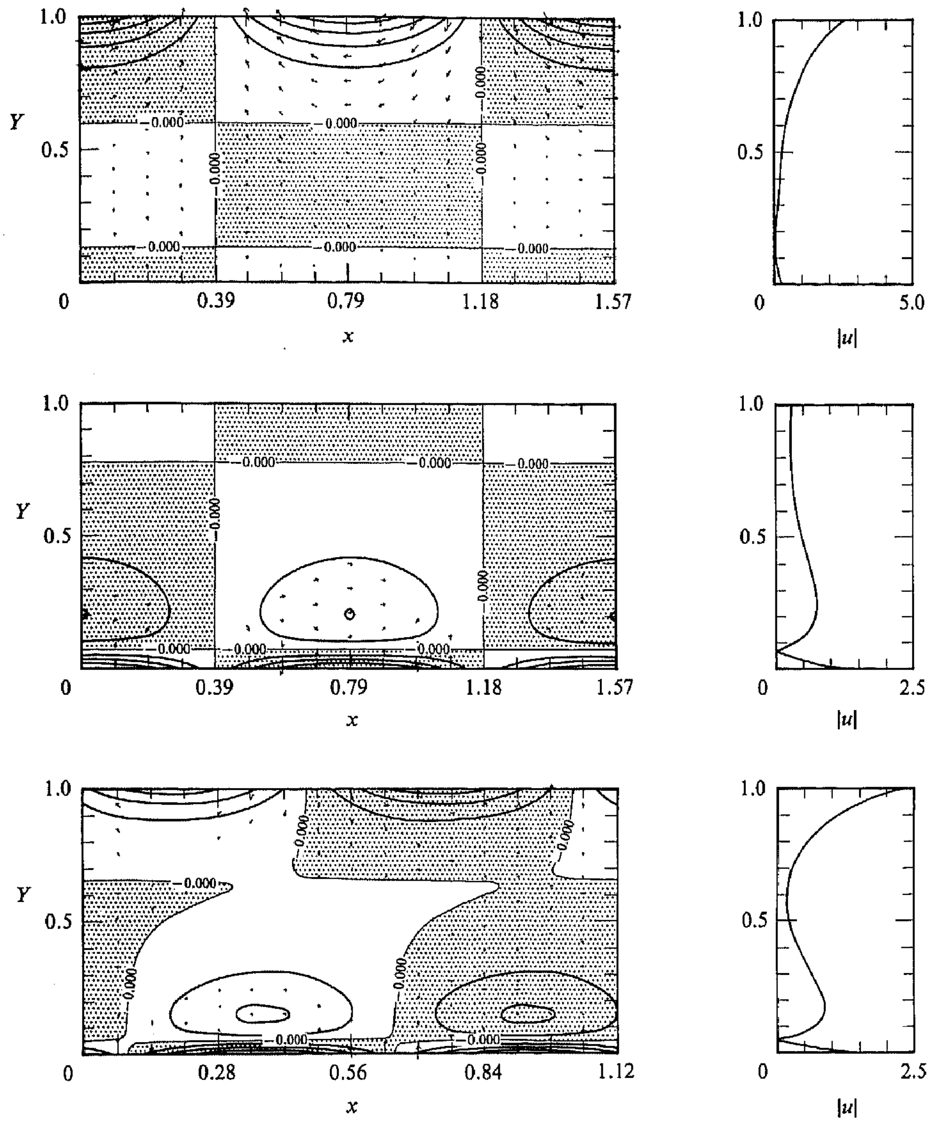

FIGURE 4. An unstable resonance between a Kelvin wave and a gravity wave with $\phi = 0$. (*a*) $\hat{k} = 4.0$ and $\hat{c}_r = 0.0964$. This is a stable Kelvin wave and at this value of $\hat{k}$ it is concentrated at the front $Y = 1$. (*b*) $\hat{k} = 4.0$ and $\hat{c}_r = 0.276$. This is a stable gravity wave which is trapped at the $Y = 0$ front. (*c*) An unstable superposition of the two modes above. In this case $k = 5.6$, $\hat{c}_r = 0.156$ and $\hat{k}\hat{c}_i = 0.0249$. This instability occurs when the dispersion curves intersect – see figure 2(*a*).

Figure 4 illustrates the unstable resonance between the Kelvin mode and first gravity mode in figure 2(*a*). In (*a*) and (*b*) $\hat{k} = 4$ and both modes are stable, boundary trapped, and have different phase speeds. One has $E > 0$ and the other has $E < 0$. In (*c*) $\hat{k}$ has increased to 5.6 and the phase speeds are equal – see figure 2(*a*). The ensuing resonance is unstable, with $E = 0$, but the growth rate is rather less than that of the Kelvin resonance in figure 3(*a*). Note that the structure of this unstable wave is roughly a linear superposition of its two resonating constituents.

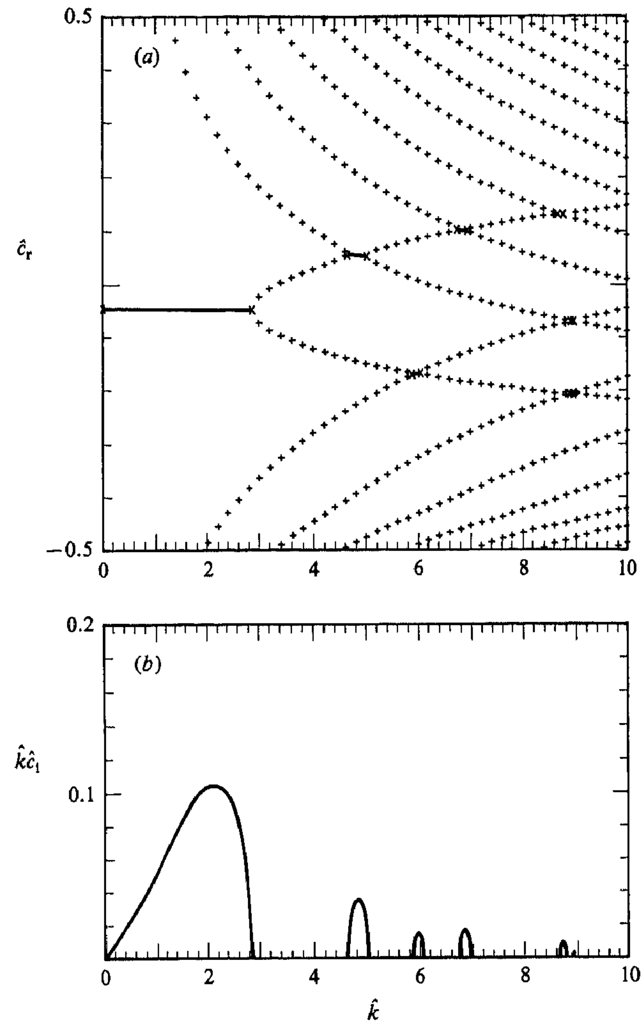

FIGURE 5. Dispersion curves when $\phi = 0.2$. (*a*) The real part of the phase speed as a function of wavenumber. The intervals of instability are indicated by the solid lines connecting the ×'s. (*b*) Growth rate as a function of wavenumber. Note that the unstable Kelvin mode has a westward phase speed.

4.2. *The case* $\phi = 0.2$

Figure 5 shows the effects of planetary vorticity gradient, $\phi = 0.2$. The basic state is now asymmetric about $Y = \frac{1}{2}$ and consequently the dispersion curves in figure 8(*a*) are now asymmetric about $\hat{c}_r = 0$. In particular the most unstable wave, which is again a resonance between the two frontal Kelvin waves, now has a westward phase speed. On the whole the global structure is quite similar to $\phi = 0$ but the growth rates are reduced.

4.3. *The cases* $\phi = 0.26$ *and* 0.28: *a transequatorial basic state*

Figure 6 shows the case $\phi = 0.26$. The flow spans the equator. There are now two types of modes: symmetric $(u(y) = u(-y))$ indicated by + and antisymmetric $(u(y) = -u(-y))$ indicated by −.

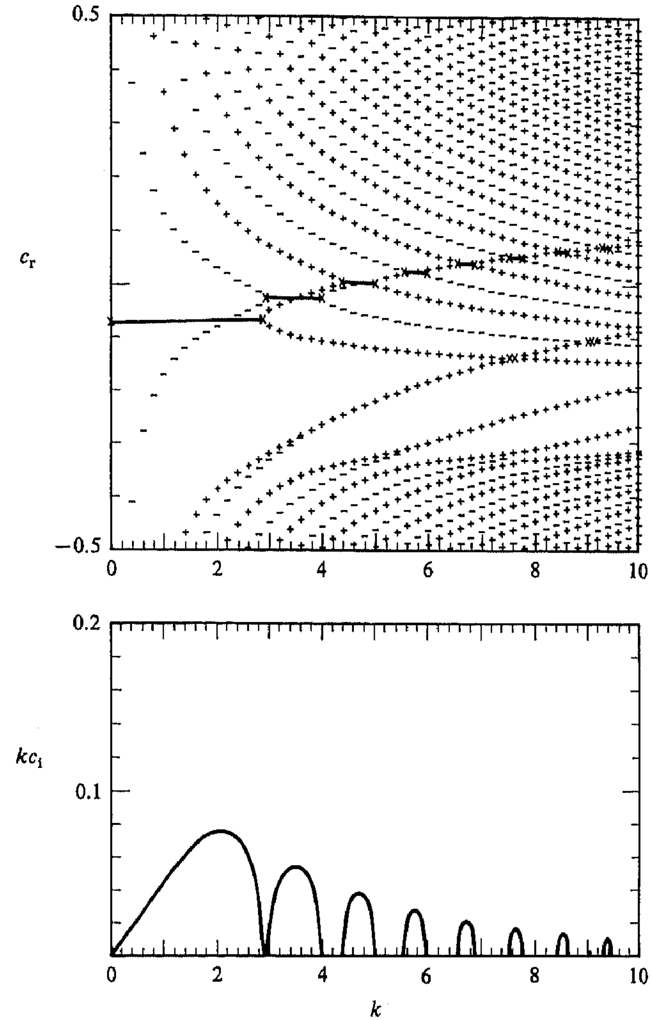

FIGURE 6. Dispersion curves for $\phi = 0.26$. Symmetric modes are indicated by + and antisymmetric by −. At this value of ϕ the current spans the equator and the intervals of instability are dense as the frontally trapped Kelvin wave (i.e. the one whose phase speed becomes positive as k increases) resonates with various gravity modes.

Figure 7 shows the horizontal structure of the Kelvin waves in both the stable and unstable regimes. In figure 7 (a), $k = 1$ and the modes are unstably coupled. The zonal velocity is virtually uniform across the channel and is roughly in geostrophic balance. In figure 7 (b, c), $k = 3$ and the unstable wave in 7 (a) has divided into two stable, symmetric waves with different phase speeds – see figure 6 (a). The wave in figure 7 (b) is concentrated at the fronts ($y = \pm 1$) and has $E > 0$ while in figure 7 (c) the disturbance is largest at the equator and has $E < 0$.

Figure 8 shows the dispersion curves when $\phi = 0.28$. At this value long waves are stable and instead the unstable resonance occurs between roughly $k = 2.4$ and 3.8.

Figure 9 shows the horizontal structure of these Kelvin waves. (a) and (b) display the stable long waves. Once again at these relatively small wavenumbers the velocity is approximately uniform across the current and the zonal velocity is approximately

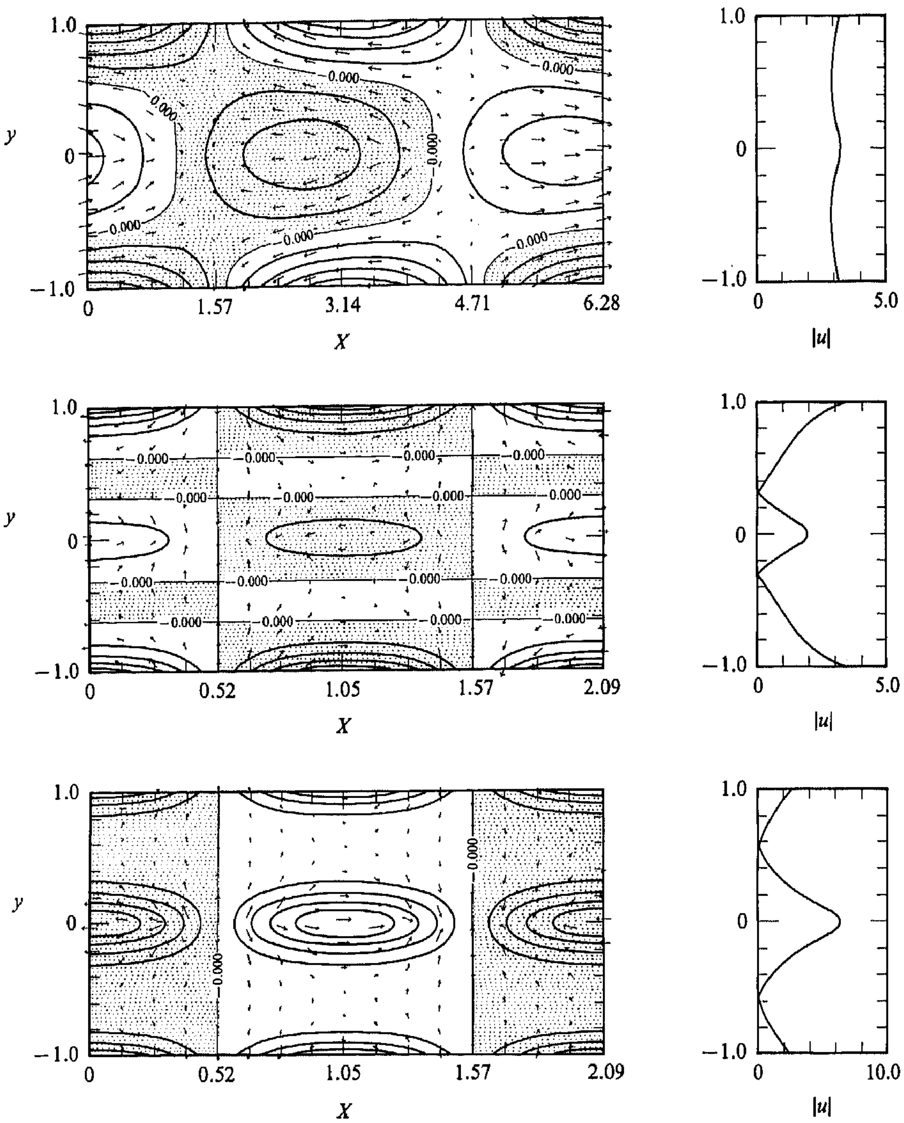

FIGURE 7. Horizontal structure of the Kelvin modes in stable and unstable cases when $\phi = 0.26$. (*a*) An unstable mode with $k = 1.00$, $c_r = -0.0729$, $kc_i = 0.0466$. As k increases this unstable mode splits into two stable modes (see figure 9*a*). (*b*) A stable Kelvin mode with $k = 3.00$ and $c_r = -0.0511$. (*c*) The other stable mode with $k = 3.00$ and $c_r = -0.0794$. The stable modes in (*b*) and (*c*) bifurcate from the unstable mode in (*a*) when k exceeds about 2.8. The first is concentrated at the fronts, the second at the equator.

geostrophic. (*c*) shows the unstable wave resulting from the coalescence of these two modes.

An overview of changes in the geometry of the dispersion curves as ϕ is altered is given in figure 10, where c_r is plotted as a function of k. One has the impression that intervals of instability are pushed out to higher wavenumbers as ϕ increases.

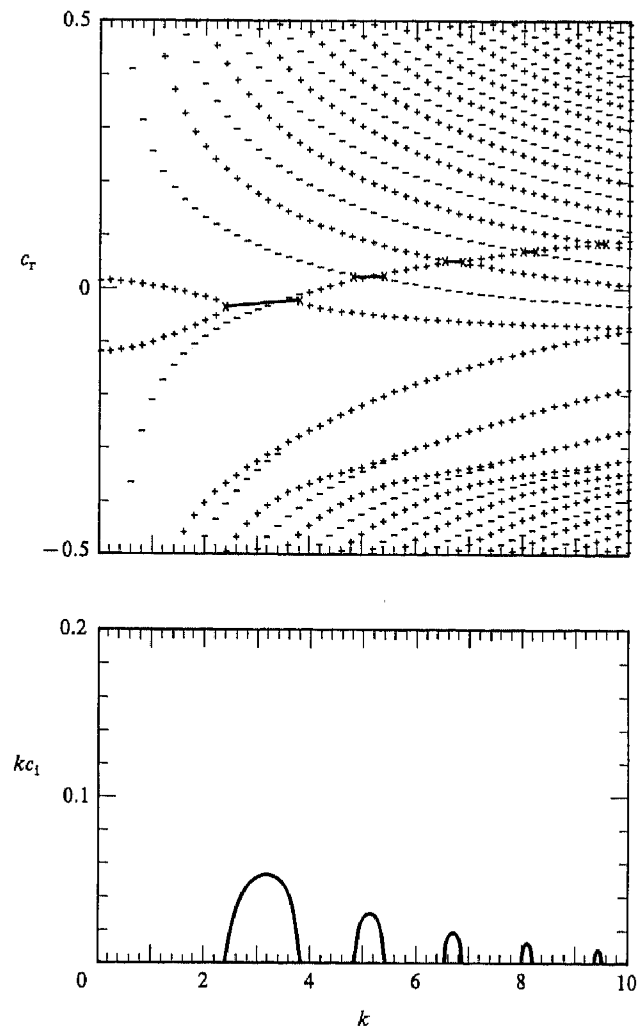

FIGURE 8. Dispersion curves when $\phi = 0.28$. At this value of ϕ there are two stable Kelvin waves at small values of k which combine at about $k = 2.2$ to form an unstable mode with westward phase speed.

5. Unstable resonances between shear modes

Instabilities are often 'explained' physically using energy arguments that distinguish the wave energy from the mean energy. In the non-divergent limit the wave energy is positive definite because $h = 0$ (see (2.25*e*)) and so instability requires the transfer of energy from the mean to the wave. Based on this special case it is often concluded that the mean flow is a source of energy for instability or that instability is characterized by the preferential transfer of energy to the growing wave. But Marinone & Ripa (1984) have shown that this interpretation is inadequate once the idealization of non-divergent flow is removed. A striking example of this is the instability illustrated in §4 in which there is no transfer of energy between the growing wave and the mean flow.

Is there a physical mechanism that is common to all cases and that characterizes instability? Following Cairns (1979) we suggest that it is illuminating to examine the

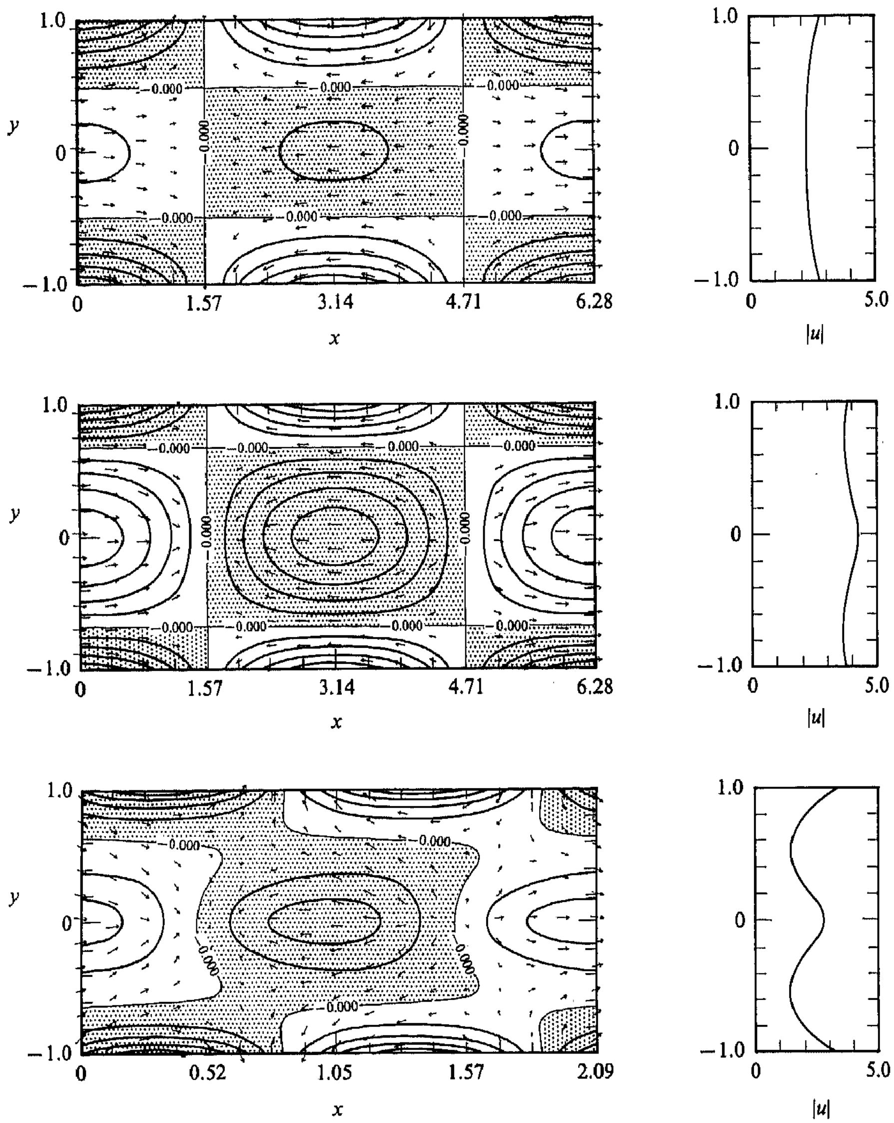

FIGURE 9. Horizontal structure of the Kelvin waves when $\phi = 0.28$. At this value of ϕ long waves are stable. (*a*) A stable Kelvin wave with $k = 1.00$ and $c_r = -0.102$. (*b*) The other Kelvin mode with $k = 1.00$ and $c_r = 0.00447$. (*c*) An unstable combination of the modes in (*a*) and (*b*) with $k = 3.00$, $c_r = -0.0292$, $kc_i = 0.0511$.

energy and momentum of the stable modes which combine to produce instability when their dispersion curves cross (e.g. figure 4). What we find is that instability occurs when a wave with negative disturbance energy ($E < 0$) 'resonates' (i.e. has the same wavenumber and phase speed) with a wave that has positive disturbance energy ($E > 0$). The unstable mode, which is roughly a linear combination of the two modes that would be stable in isolation (figures 4, 7 and 9), has $E = M = 0$. The instability occurs because the disturbance with positive energy gains energy from the disturbance with negative energy. Both disturbances amplify.

A useful rule for determining the sign of E for the stable modes of (3.10) is that

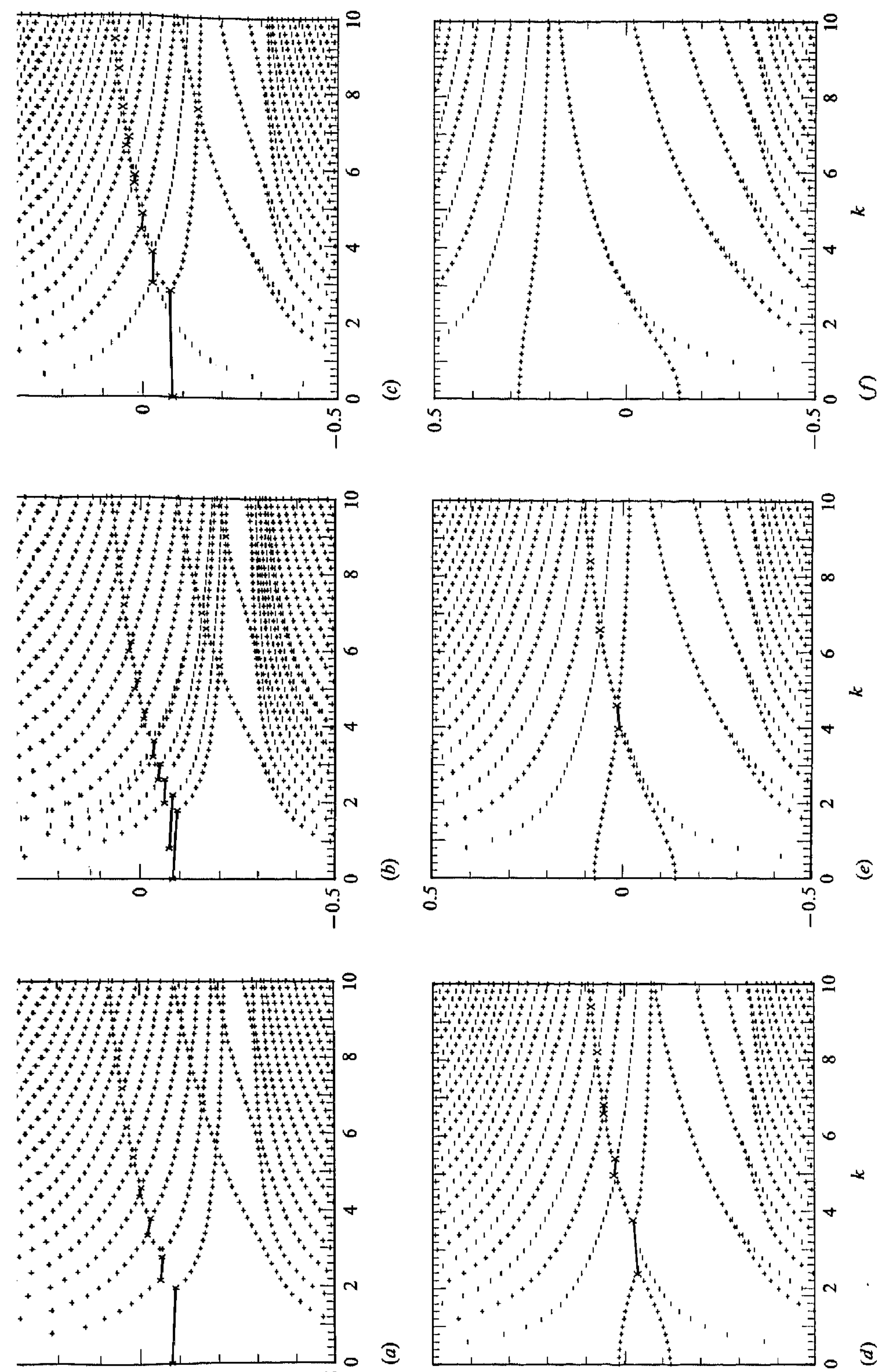

FIGURE 10. An overview of changes in the geometry of the dispersion curves as ϕ is altered. In all cases c_r is plotted as a function of k. Unstable solutions are indicated by '×'. (*a*) $\phi = 0.2499$, (*b*) 0.2501, (*c*) 0.26, (*d*) 0.28, (*e*) 0.3, (*f*) 0.4.

E is opposite in sign to $c^{-1}\,dc/dk$.† Thus from the dispersion curves in figures 2, 5, 6 or 8 one can easily infer the sign of the disturbance energy and verify that instabilities always occur when waves with different-signed disturbance energies interact.

Is this physical explanation of instability peculiar to the special basic state analysed in §4? In that example, and also in those discussed by Cairns (1979), there are no potential vorticity gradients so that the disturbance energy E is equal to the wave energy E_2. However we now argue that this is not essential and that even familiar, non-divergent shear-flow instabilities are subject to an analogous explanation. The key is that one must examine the disturbance energy (rather than the wave energy) of a mode.

An example will make this clear. Consider the 'triangular jet'

$$\left.\begin{aligned} U &= 0 && \text{if } y > b, \\ U &= U_0\left\{1-\frac{|y|}{b}\right\} && \text{if } |y| < b, \\ U &= 0 && \text{if } y < -b, \end{aligned}\right\} \tag{5.1}$$

whose instability was analysed by Rayleigh (1896). In the non-divergent limit

$$u = -\psi_y, \quad v = \psi_x,$$

and the stream function ψ is a solution of the Rayleigh equation

$$\psi_{yy} - \left\{k^2 + \frac{Q_y}{c-U}\right\}\psi = 0. \tag{5.2}$$

For the basic state in (5.1), Q_y is a set of delta functions at the kinks in the velocity profile. When k is very large, (5.2) has three stable modes, each localized at one of the shear discontinuities.

For instance if $|kb| \gg 1$ then there are stable modes trapped near $|y| = b$. Taking $y = b$ and solving (5.2) gives

$$\left.\begin{aligned} \psi &= a\,e^{-k(y-b)} && y > b, \\ \psi &= a\,e^{k(y-b)} && y < b, \\ c(k) &= \frac{U_0}{2kb}, \end{aligned}\right\} \tag{5.3}$$

where a is the amplitude. Using (2.31 b), and noting that $\hat{M}_2 = 0$ in the non-divergent limit ($h = 0$), one calculates the disturbance momentum of this mode

$$M = \frac{a^2k^2b}{U_0} > 0 \tag{5.4}$$

and the disturbance energy follows easily from (2.27).

In addition to the modes at $|y| = b$ there is a mode at $y = 0$. Again when $kb \gg 1$

$$\left.\begin{aligned} \psi &= a\,e^{-ky} && y > 0, \\ \psi &= a\,e^{ky} && y < 0, \\ c &= U_0\{1-(bk)^{-1}\}. \end{aligned}\right\} \tag{5.5}$$

† The complete result, which follows from (3.10), is that $2k^3E(dc/dk) = -c\int Hu'^2\,dy$.

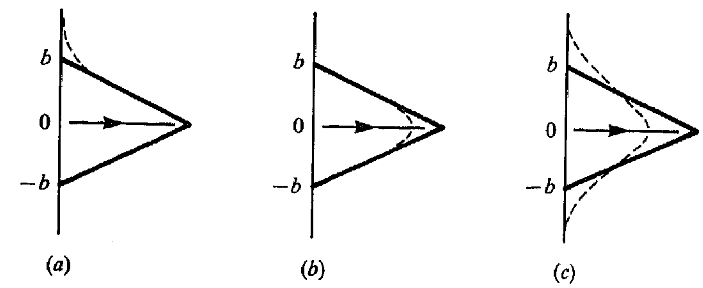

FIGURE 11. Sketch of how the mean velocity profile is modified (dashed line) by the excitation of the stable and unstable shear modes of a triangular jet. (*a*) The localized ($kb \gg 1$), stable mode at $y = b$. Exciting this mode increases both the energy and momentum of the fluid. (*b*) The localized ($kb \gg 1$), stable mode at $y = 0$. Exciting this mode decreases both the energy and momentum of the fluid. The decrease in x-momentum is obvious from the sketch. But the disturbance energy is also negative because the loss of mean energy overwhelms the positive wave energy. (*c*) An unstable mode with $kb = O(1)$. This instability appears when kb is reduced and the modes in (*a*) and (*b*) resonate. The instability 'diffuses' the velocity profile.

The disturbance momentum of this central mode is

$$M = -\frac{a^2k^2b}{2U_0}. \tag{5.6}$$

The disturbance energy follows from (2.27) and is negative when $kb \gg 1$. Thus to excite this stable mode, energy and momentum must be removed from the fluid. Even though the wave energy is positive, to excite the mode a larger amount of energy must be removed from the mean flow. These results are summarized in figure 11.

As kb is reduced these shear modes become less localized. Also their phase speeds become less disparate. In fact, according to (5.3) and (5.5) when

$$bk = \tfrac{3}{2} \tag{5.7}$$

all three modes have the same phase speed viz. $\frac{1}{3}U_0$. Of course this is only a rough estimate based on the limit $kb \gg 1$. But our previous remarks suggest that at about this wavenumber instability will occur because modes with opposite-signed disturbance energy and momentum have the same wavenumber and phase speed. In fact Rayleigh shows that instability occurs at the critical wavenumber $kb = 1.75$ and maximum growth rate is at $kb = 1.25$. The rough estimate in (5.7) lies between these two. Encouraged by this, one can show that the instabilities of all the piecewise-linear velocity profiles discussed by Rayleigh occur when modes of opposite-signed disturbance energy and momentum coalesce. Further, one can roughly predict the critical wavenumber by equating the phase speeds of the stable constituents in the limit $kb \gg 1$.

One unresolved difficulty with the preceding 'broken-line' model is how the scenario it suggests generalizes to smoothly varying velocity profiles with critical-layer singularities. If this smooth profile has an inflexion point, one establishes the existence of unstable modes by first finding a regular neutral mode (i.e. one with a critical layer at the inflexion point) at a certain wavenumber k_c. This is a stability boundary and on one side of it, say $k < k_c$, the Tollmien–Lin perturbation procedure constructs unstable modes. The previous argument suggests that on the other side of this boundary, $k > k_c$, we should find two neutral modes which resonantly coalesce at k_c. Instead there is a continuum of singular neutral modes. The unresolved question which concludes this section is whether the regular neutral mode at k_c is a resonance between two members of the continuum?

6. Conclusion

Energy and momentum transfer between a growing unstable wave and the mean flow is not essential for instability. This counterintuitive result is implicit in previous stability studies of the shallow-water equations but it has been either unnoticed (GKS) or disguised by defining the 'wave energy' as (2.33) and introducing $\int Uhu\,\mathrm{d}A$ as an 'interaction energy' which is distinct from the 'wave energy' and the mean energy (Satomura 1981*a, b*). We would argue that not much physical insight is gained by this and in fact basic results such as (2.29) and (2.30) are obscured. The most straightforward conclusion is that there is nothing special about the direction of transfer of energy in instabilities.

Two questions are suggested by these results. How does one categorize non-dissipative instabilities using the direction of energy transfer? And are there common features that characterize all non-dissipative instabilities of the shallow-water equations?

In answer to the first question, we propose the tripartite scheme in §2.4. In the non-divergent limit, where the Rayleigh equation is recovered, all instabilities have positive wave energy and are in the first category. The second category is exemplified by the analysis in §4 and in fact we showed that if the basic state has uniform potential vorticity then its instabilities are in this category. Thus the instabilities discussed by GKS and Satomura have zero wave energy and are in the second category.† As for the third category, the only *bona fide* examples we know of are those exhibited by Marinone & Ripa (1984). But we expect that many of the ageostrophic, frontal instabilities discussed in the literature are, in fact, in this third category.

The second question posed above is whether these diverse instabilities have any features in common. One obvious answer to this question is that in every case energy and momentum are conserved so that the disturbance energy and disturbance momentum of an unstable mode must be zero. In §§4 and 5 we suggested that this constraint can be satisfied when two otherwise stable modes, with opposite-signed disturbance energy and momentum, resonate. The combination grows because the partner with negative energy transfers energy to the one with positive energy.

We thank the 1984 Woods Hole Geophysical Fluid Dynamics program for its support and hospitality. Support for the 1984 Study Program in Geophysical Fluid Dynamics was provided by the Office of Naval Research under contract N00014-82-G-0079 and the National Science Foundation under Grant MCS-82-000450. W.R.Y. also acknowledges support from the Office of Naval Research under contract N-00014-86-K-0325.

During the course of this work we enjoyed conversations with Andrew Fowler, Richard Lindzen and Glenn Flierl.

† In the context of stratified shear flow, the piecewise-constant profiles of velocity and density studied by Cairns (1979) are also in this second category. In these Kelvin–Helmholtz instabilities there is no disturbance energy in the mean flow, i.e. $E = E_2$ and $\hat{E}_1 = 0$.

Appendix A. The numerical solution of the eigenvalue problem

Rather than deal with (2.19), in which the eigenvalue appears quadratically, it is best to use (2.18) to eliminate v from (2.17). The result is two coupled equations for h and u:

$$\left.\begin{aligned} Uu+h &= cu, \\ \{-k^{-2}(d/dy)H(d/dy)+H\}u+Uh &= ch, \end{aligned}\right\} \tag{A 1}$$

in which the eigenvalue c is conveniently placed. If we now discretize the y-interval, (A 1) is equivalent to an eigenvalue problem

$$\mathrm{A}x = cx, \tag{A 2}$$

where x is the eigenvector and A is the finite-difference form of the various differential operators and y-dependent coefficients on the left-hand side of (A 1). The eigenvalues are computed by first transforming A into Hessenberg form and then using the double QR method. The eigenvectors were then computed using inverse iteration.

This method was checked against a shooting method which used a fourth-order Runge–Kutta scheme and the agreement was satisfactory. We also checked the integral relationship:

$$\int_{-1}^{1} H\left|\frac{\mathrm{d}u}{\mathrm{d}y}\right|^2 + k^2\{H-(U-c)^2\}|u|^2\,\mathrm{d}y = 0. \tag{A 3}$$

In fact with 100 subintervals (i.e. $\delta y = 0.02$) the matrix method gives eigenfunctions for which this integral is about 10^{-12}. Shooting with a step size of $\delta y = 10^{-4}$ (i.e. 20000 intervals) produces eigenfunctions for which the integral is only 10^{-4}. From a number of comparisons our conclusion is that for this particular eigenvalue problem the matrix method is more accurate and systematic than shooting.

REFERENCES

Andrews, D. G. & McIntyre, M. E. 1978 An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.* **89**, 609–646.

Benjamin, T. B. 1963 The threefold classification of unstable disturbances in flexible surfaces bounding inviscid fluids. *J. Fluid Mech.* **16**, 436–450.

Blumen, W. 1970 Shear layer instability of an inviscid compressible fluid. *J. Fluid Mech.* **40**, 769–781.

Blumen, W., Drazin, P. G. & Billings, D. F. 1975 Shear layer instability of a compressible fluid. Part 2. *J. Fluid Mech.* **71**, 305–316.

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

Griffiths, R. W., Killworth, P. D. & Stern, M. E. 1982 Ageostrophic instability of ocean currents. *J. Fluid Mech.* **117**, 343–377.

Held, I. M. 1985 Pseudomomentum and orthogonality of modes in shear flows. *J. Atmos. Sci.* **42**, 2280–2288.

Hoskins, B. J., McIntyre, M. E. & Robertson, A. W. 1985 On the use and significance of isentropic potential vorticity maps. *Q. J. R. Met. Soc.* **111**, 877–946.

Howard, L. N. 1961 Note on a paper of John W. Miles. *J. Fluid Mech.* **10**, 509–512.

Landahl, M. T. 1962 On the stability of a laminar incompressible boundary layer over a flexible surface. *J. Fluid Mech.* **13**, 609–632.

Marinone, S. G. & Ripa, P. 1984 Energetics and instability of a depth independent Equatorial jet. *Geophys. Astrophys. Fluid Dyn.* **30**, 105–130.

Paldor, N. 1983 Linear stability and stable modes of geostrophic fronts. *Geophys. Astrophys. Fluid Dyn.* **24**, 299–326.

PEDLOSKY, J. 1979 *Geophysical Fluid Dynamics*. Springer. 624 pp.

RAYLEIGH, LORD 1896 *The Theory of Sound*, vol. II. Dover.

RIPA, P. 1983 General stability conditions for zonal flows in a one layer model on the beta-plane or the sphere. *J. Fluid Mech.* **126**, 463–487.

SALMON, R. 1982 The shape of the main thermocline. *J. Phys. Oceanogr.* **12**, 1458–1479.

SARDESHMUKH, P. D. & HOSKINS, B. J. 1985 Vorticity balances in the tropics during the 1982–83 El Nino–Southern Oscillation event. *Q. J. R. Met. Soc.* **111**, 261–278.

SATOMURA, T. 1981*a* An investigation of shear instability in a shallow water. *J. Met. Soc. Japan* **59**, 148–167.

SATOMURA, T. 1981*b* Supplementary note on shear instability in a shallow water. *J. Met. Soc. Japan* **59**, 168–171.

TAYLOR, G. I. 1915 Eddy motion in the atmosphere. *Phil. Trans. R. Soc. Lond.* A **215**, 1–26.