J. Fluid Mech. (1985), *vol.* 161, *pp.* 449–492
Printed in Great Britain

Do Rossby-wave critical layers absorb, reflect, or over-reflect?

By PETER D. KILLWORTH† AND MICHAEL E. McINTYRE
Department of Applied Mathematics & Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, U.K.

(Received 4 April 1985 and in revised form 6 August 1985)

The Stewartson–Warn–Warn (SWW) solution for the time evolution of an inviscid, nonlinear Rossby-wave critical layer, which predicts that the critical layer will alternate between absorbing and over-reflecting states as time goes on, is shown to be hydrodynamically unstable. The instability is a two-dimensional shear instability, owing its existence to a local reversal of the cross-stream absolute vorticity gradient within the long, thin Kelvin cat's eyes of the SWW streamline pattern. The unstable condition first develops while the critical layer is still an absorber, well before the first over-reflecting stage is reached. The exponentially growing modes have a two-scale cross-stream structure like that of the basic SWW solution. They are found analytically using the method of matched asymptotic expansions, enabling the problem to be reduced to a transcendental equation for the complex eigenvalue. Growth rates are of the order of the inner vorticity scale δq, i.e. the initial absolute vorticity gradient $\mathrm{d}q_0/\mathrm{d}y$ times the critical-layer width scale. This is much faster than the time evolution of the SWW solution itself, albeit much slower than the shear rate $\mathrm{d}u_0/\mathrm{d}y$ of the basic flow. Nonlinear saturation of the growing instability is expected to take place in a central region of width comparable to the width of the SWW cat's-eye pattern, probably leading to chaotic motion there, with very large 'eddy-viscosity' values. Those values correspond to critical-layer Reynolds numbers $\lambda^{-1} \ll 1$, suggesting that for most initial conditions the time evolution of the critical layer will depart drastically from that predicted by the SWW solution. A companion paper (Haynes 1985) establishes that the instability can, indeed, grow to large enough amplitudes for this to happen.

The simplest way in which the instability could affect the time evolution of the critical layer would be to prevent or reduce the oscillations between over-reflecting and absorbing states which, according to the SWW solution, follow the first onset of perfect reflection. The possibility that absorption (or over-reflection) might be prolonged indefinitely is ruled out, in many cases of interest (even if the 'eddy viscosity' is large), by the existence of a rigorous, general upper bound on the magnitude of the time-integrated absorptivity $\alpha(t)$. The bound is uniformly valid for all time t. The absorptivity $\alpha(t)$ is defined as the integral over all past t of the jump in the wave-induced Reynolds stress across the critical layer. In typical cases the bound implies that, no matter how large t may become, $|\alpha(t)|$ cannot greatly exceed the rate of absorption predicted by linear theory multiplied by the timescale on which linear theory breaks down, say the time for the cat's-eye flow to twist up the absolute vorticity contours by about half a turn. An alternative statement is that $|\alpha(t)|$ cannot greatly exceed the initial absolute vorticity gradient $\mathrm{d}q_0/\mathrm{d}y$ times the cube of the

† Present affiliation: Robert Hooke Institute for Atmospheric Research, Dept. of Atmospheric Physics, Clarendon Laboratory, Oxford OX1 3PU.

widthscale of the critical layer. In typical cases, therefore, a brief answer to the question posed in the title is that the critical layer absorbs at first, at a rate $\propto \mathrm{d}q_0/\mathrm{d}y$, whereas after linear theory breaks down the critical layer becomes a perfect reflector in the long-time average. If absolute vorticity gradients vanish throughout the critical layer then the bound is zero, implying perfect reflection for *all* t.

The general conditions for the bound to apply are that the wave amplitude and critical-layer width are uniformly bounded for all t, the motion is two-dimensional, and vorticity is neither created nor destroyed within the critical layer, nor transported into or out of it by diffusion, by advection, or by other means. Vorticity may, however, be diffused or turbulently transported within the critical layer, provided that the region within which the transport acts is of bounded width and the range of values of vorticity within that region remains bounded. There are no other restrictions on wave amplitude, none on wavelength, and no assumptions about flow details within the critical layer nor about the initial vorticity profile $q_0(y)$, apart from an assumption that $q_0(y)$ has singularities no worse than a finite number of jump discontinuities. The proof, in its most general form, makes use of a new finite-amplitude conservation theorem for disturbances to parallel shear flows, generalizing the classical results of Taylor, Eliassen & Palm, and others.

1. Introduction and résumé

1.1. *Background*

There has been a continuing debate in recent years about what happens when monochromatic Rossby waves are incident on a critical line in a shear flow, where phase speed matches mean-flow speed. A basic question, posed epigrammatically, is whether and when the nonlinear 'critical layer' surrounding the critical line acts as an absorber, a reflector, or an over-reflector.

The problem is important in a wider context than the meteorological one which directly motivates it (e.g. Tung & Lindzen 1979*a*, *b*; Tung 1979; Nigam & Held 1983 and references), since closely related questions arise whenever stable or unstable disturbances to shear flows are of interest, these disturbances having coherent structures and definite phase speeds. Disturbances to the jets and shear layers encountered in aerodynamics are cases in point. As was noted by Stewartson (1978), the case of two-dimensional Rossby waves in constant shear provides the simplest context in which one of the central theoretical problems can be studied in detail, namely the onset of nonlinear effects in a time-dependent critical layer and the consequent changes in the matching conditions across the layer. For surveys of the topic in general the reader may consult the reviews by Maslowe (1981) and Stewartson (1981), and for some very recent meteorological developments the papers by Al-Ajmi *et al.* (1985), Clough *et al.* (1985), Leovy *et al.* (1985), Hoskins *et al.* (1985), and McIntyre & Palmer (1983, 1984). The last-named group of papers presents direct evidence, from satellite and other observations, indicating that effects of the general kind modelled *inter alia* by the time-dependent theory of nonlinear critical layers play an essential role in the large-scale dynamics of the Earth's atmosphere.

The linear, time-dependent problem, for monochromatic Rossby waves incident upon an initially undisturbed critical line in constant shear, was originally solved by Dickinson (1970), and further elucidated by Warn & Warn (1976), for the case of inviscid flow. It predicts that the critical layer acts as a perfect absorber, as long as the linearization remains valid, but that nonlinear effects become important after

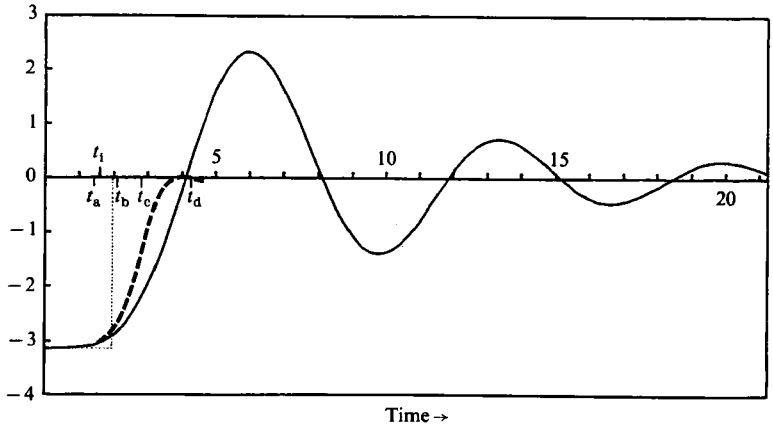

FIGURE 1. The *continuous curve* represents the time evolution of the absorptivity of a Rossby-wave critical layer according to the SWW solution. The absorptivity is expressed as a normalized Reynolds-stress or momentum-flux jump across the critical layer, equivalent in this problem to the phase-slope jump or 'logarithmic phase jump' of the fundamental harmonic. The initial value is $-\pi$ and represents the absorbing state predicted by linear critical-layer theory; zero represents perfect reflection, and positive values over-reflection; see (2.36) ff. The time is in units of $\Lambda^{-1}\hat{k}^{-1}\Delta y^{-1}$, where Λ is the shear of the basic flow, $\hat{k}$ the x-wavenumber of the Rossby wave, and Δy the critical-layer width scale (the units in which the y-coordinate is expressed in figure 2). Δy is equal to $\epsilon^{\frac{1}{2}}\Lambda/\beta$, where $\beta = \mathrm{d}q_0/\mathrm{d}y$ is the initial absolute (potential) vorticity gradient, assumed constant in this case, and ϵ the disturbance amplitude defined in §2. The area under the entire curve, between time zero and infinity, is equal to the area enclosed by the finely dotted rectangle (see §1.4). The *dashed curve* schematically represents a result from the pioneering numerical simulation by Béland (1976, figure 5); see also figure 5 of this paper. Δy varies with time in the problem solved by Béland, so that a quantitatively meaningful comparison in the same time units is not possible. The dashed curve has been drawn on the assumption that the effective value of Δy was increased (by about 10%) by the effects of instability in Béland's simulation; cf. (1.11) with $b \propto \Delta y$.

a sufficient time no matter how small the amplitude of the incident wave. Our understanding of the subsequent nonlinear time evolution has been greatly advanced in recent years, particularly through the numerical work of Béland (1976, 1978) and the analytical and numerical work of Warn & Warn (1978), Stewartson (1978), Brown & Stewartson (1978), Smith & Bodonyi (1982), and Ritchie (1985). The most detailed formal analysis of the nonlinear time evolution is that of Warn & Warn (1978), who by making simultaneous use of matched-asymptotic and multiple-scale techniques clarified the general structure of the inviscid initial-value problem. In addition, the important discovery was made that special cases exist (Warn & Warn 1978) in which a complete analytical solution can be found to leading order (Stewartson 1978). We thus have examples in which a self-consistent, quantitative description of the inviscid evolution of the critical layer, or rather a description of one possible such evolution, is available out to times t very much larger than the time at which linear theory breaks down. That evolution is inaccessible to almost any kind of numerical method, because of the exponentially large vorticity gradients which develop on exponentially small spatial scales in certain parts of the critical layer.

The Stewartson–Warn–Warn solution, as we shall call this analytical solution for convenience (hereinafter referred to as the 'SWW solution'), predicts that nonlinearity turns the critical layer from a wave absorber into a wave reflector, then immediately into an over-reflector (reversing the wave-induced momentum flux outside the critical

layer), then back into a weaker absorber, and so on. The state of the critical layer continues to oscillate about perfect reflection in this manner, as shown by the continuous curve in figure 1. The oscillations die away like t^{-2}. As far as the wave motion outside the critical layer is concerned, the perfectly reflecting state which is approached at $t \to \infty$ is nearly equivalent to that predicted by the earlier, steady-state theories of Benney & Bergeron (1969), Davis (1969), and Haberman (1972) (hereinafter referred to as BBDH), in the limit of small viscosity. During the long evolution towards perfect reflection, the critical layer causes higher harmonics of the incident wave to appear in the flow outside it, as Warn & Warn showed would generally occur (figure 3 below). However, at least in the case of the SWW solution itself, the contribution from these higher harmonics to the stream function in the outer flow decays to zero in the long-time limit. The SWW solution and its physical meaning are reviewed in §2 of this paper.

1.2. *Instability*

The present work was motivated by the realization that the SWW solution represents an unstable flow, with growth rates generally much faster than the timescale of the evolution just described.† The unstable condition first develops when t exceeds the value t_1 marked in figure 1. This raised the possibility that the long-time evolution might be quite different in reality. Indeed Béland's numerical work did, originally, predict a qualitatively different time evolution, a fact which to our knowledge has never been satisfactorily explained. We now believe that the instability is the most likely explanation, even though the numerical scheme used by Béland may not have represented the unstable disturbances with great accuracy. In his best-resolved simulation, which predicted a time evolution like that indicated schematically by the dashed curve in figure 1, with little indication of over-reflection (Béland 1976, figure 5), disturbances appeared at the expected stage in the evolution and showed a spatial structure very like that predicted by the instability theory (figures 5*b*, *c* below).‡

The instability theory is presented in §§3 and 4. The instability is an ordinary two-dimensional inviscid shear instability, in the sense that it depends in the usual way upon a reversal of the absolute vorticity gradient in the y- or cross-stream direction. The unstable modes have a two-scale spatial structure in the y-direction paralleling that of the SWW solution itself. This is related to the fact that the shear of the basic flow,

$$\Lambda = \frac{\mathrm{d}u_0(y)}{\mathrm{d}y},$$

is large in comparison with the absolute vorticity contrast δq across the narrow region of reversed vorticity gradient to which the instability owes its existence. The maximum growth rate turns out to be of the order of δq. Even though this is much slower than the basic shear rate Λ, it is much faster than the rate at which the SWW solution itself evolves. Consequently there is a separation of timescales which permits the time evolution of the SWW flow to be ignored in the instability analysis.

† This result was first announced in December 1979 at the 17th IUGG General Assembly; see Ruttenberg (1980).

‡ It is also possible that the disturbances appearing in Béland's simulation had a different cause, namely the resonant amplification of stable harmonics recently studied by Ritchie (1985). But for reasons to be given in §4 we think that the explanation in terms of instability is the more likely alternative, on present evidence.

Moreover, there is a corresponding separation of streamwise or x-scales, which permits the use of a local parallel-flow approximation. The disturbance x-scale is much smaller than that of the SWW solution, albeit much larger than the width of the critical layer. These scaling relations imply, incidentally, that the unstable disturbances can always be treated as inviscid. If viscosity is small enough for the SWW solution to be applicable in the first place, then it is *a fortiori* negligible for the faster-evolving instabilities, since the same minimum spatial scale, the critical-layer width, is relevant to both.

The scale separation in y between the inner, critical-layer region and its surroundings gives rise to a further parallel between the instability and the basic SWW solution. This will prove to be significant for understanding the long-time evolution of the critical layer. The matched-asymptotic theory, which is used to compute the structure of the unstable disturbances to leading order, predicts that the nonlinear saturation of the instability, and the concomitant rearrangement of vorticity, will take place in the inner, critical-layer region only. Although the outer region plays an essential role in the nonlinear evolution of the instability, through the coupling to the inner region expressed by the matching conditions, the dynamical behaviour of the outer region itself remains linear to a first approximation, just as in the SWW solution itself.†

1.3. *Could the instability give rise to prolonged absorption?*

Accurate calculations of the nonlinear development of the instability are given in the following paper (Haynes 1985). As anticipated from order-of-magnitude estimates (§3 below), the amplitudes at which the instability typically saturates turn out to be more than enough to cause a substantial two-dimensional rearrangement of vorticity within a finite central region of the critical layer. The width of this region is numerically of the same order as that of the cat's-eye pattern in the SWW stream function. Such a rearrangement of vorticity may be expected to lead to a time evolution which differs substantially from that indicated by the continuous curve in figure 1 since, as will be recalled in §2, the absorptivity of the critical layer is a functional of the vorticity distribution within it.

At this point, one might be tempted to speculate (as in fact we did in the early stages of this work) that the instability might lead to prolonged absorption, and that this might explain the absorptivity seen in the Earth's atmosphere over timescales of the order of a month or more, according to certain well-established meteorological statistics (e.g. Starr 1968; Edmon, Hoskins & McIntyre 1980; Karoly 1982; Hamilton 1982; Held 1983, p. 145 and references). Although Béland's results did not suggest this (cf. dashed curve in figure 1), it seemed reasonable to suppose that they might have done so had the numerical resolution been higher, and the instability better resolved. The idea appeared to be supported by an intuitive appeal to the notions of 'mixing length' and 'eddy viscosity', since the disturbance velocity and mixing-length scales predicted by the instability theory and its nonlinear extension can be shown to correspond to an exceedingly viscous critical layer and therefore, it might

† This aspect of the situation finds an illuminating analogy in what, at first sight, might appear to be an unrelated problem, namely that of the action of musical oscillators such as the clarinet or bowed string. The nonlinearity controlling the saturation amplitude is largely concentrated in one place, at the reed or bow, whereas the time evolution of the system as a whole depends on the coupling between the nonlinear element and the rest of the system, exactly as it does in the critical-layer problem. (For recent reviews see Fletcher 1979, and McIntyre, Schumacher & Woodhouse 1983.)

be thought, to an absorber (Lin 1955; Haberman 1972; Béland 1978; Brown & Stewartson 1978; Tung 1979; Smith & Bodonyi 1982).

However, it will be shown in §§5–7 that, in a very wide set of circumstances, the opposite is the case. This well illustrates the dangers inherent in too careless a use of the notion of 'eddy viscosity', even for the purpose of an order-of-magnitude argument. We shall see that over long periods of time the critical layer must, on average, be a *perfect reflector*, in a sense to be defined in §1.4 below, no matter how chaotic the motion may be in any finite central region. More precisely, a critical layer of bounded width, such as might be set up by a Rossby wave of bounded amplitude and constant streamwise phase speed, will be a perfect reflector in the long-time average – even though the time evolution of its absorptivity may well differ, in other respects, from that suggested by the SWW solution – whenever the fluid motion is

(i) two dimensional,
(ii) free of sources and sinks of vorticity due to external forcing,
(iii) unable to advect vorticity into or out of the critical layer (implying that the critical layer always consists of the same material fluid elements,† and
(iv) unable to diffuse vorticity towards or away from the critical layer.

The vorticity may be subject to diffusion or to any other laminar or 'turbulent' transport process *within* the critical layer, provided that the transport is confined to some central region having finite width, so that (iv) is satisfied and provided also that

(v) the transport process just referred to is such that the range of values of absolute vorticity stays bounded in the region affected.

Even up-gradient vorticity transport is allowed, therefore, as long as it is not so persistently up-gradient as to cause the range of values to increase without bound in the region affected by the transport. Note that, in the assumed circumstances,

(vi) the critical layer is free of external sources and sinks of mean momentum,

where 'mean' refers to the usual streamwise Eulerian mean. Otherwise, (ii) would be contradicted for a critical layer of finite width. Finally, the motion is assumed incompressible to a sufficient approximation, as is usual when talking about Rossby waves.

The foregoing statements imply, *inter alia*, that the sustained absorption exhibited by the viscous critical-layer models usually studied cannot be attributed solely to the diffusion of vorticity within a given central region. The fact that these models violate condition (iv) is crucial. The work of Haberman (1972), Brown & Stewartson (1978) and Smith & Bodonyi (1982) has shown clearly that, in such models, which assume spatially uniform viscosity, x-averaged vorticity does indeed diffuse further and further away from the centre of the critical layer as time goes on, occupying regions of width $O(t^{\frac{1}{2}})$ on either side as $t \to \infty$.

The sustained absorption seen in long-term meteorological statistics seems likely, by contrast, to be due mainly to violation of conditions (ii) and (iii), reading 'potential vorticity on an isentropic surface' for 'vorticity' (e.g. Charney & Stern 1962; Hoskins *et al.* 1985). If eddy viscosity values are much smaller outside the central critical-layer region than within it, as the instability theory suggests might typically be the case in the real atmosphere, then condition (iv) could be relatively well satisfied and the usual concept of a viscous critical layer irrelevant.

† Note that condition (iii) is violated if fresh fluid elements are carried into the critical layer by mean circulations, or if the critical layer moves through the fluid.

1.4. *A bound on the time-integrated absorptivity*

The general measure of absorptivity to which the foregoing statements will be shown to apply is the jump $[\overline{u'v'}]$ in the wave-induced Reynolds stress or Eulerian momentum flux $\overline{u'v'}$ across the critical layer. (In the meteorological case, read 'convergence of the Eliassen–Palm flux onto the critical layer'.) The overbar represents the Eulerian mean with respect to the streamwise coordinate x, and (u', v') are the (x, y) components of the departure of the velocity field from its Eulerian mean. In the case of the SWW solution, the continuous curve in figure 1 can be regarded as a graph of $-[\overline{u'v'}]$ against time, the normalization and sign convention being chosen such that the value at $t = 0$ corresponds to the rate of absorption predicted by linear theory. The value zero corresponds to perfect reflection, as already implied.

We define the time-integrated absorptivity as

$$\alpha(t) = \int_0^t [\overline{u'v'}]\,\mathrm{d}t \tag{1.1}$$

(with the sign conventions used by SWW). In the case of the SWW solution, it is proportional to minus the area under the continuous curve in figure 1 up to time t, and it is shown in the appendix to take the finite limiting value

$$\alpha(\infty) = 3.0858\,\frac{\mathrm{d}q_0}{\mathrm{d}y}\left(\frac{\hat{v}}{\hat{k}A}\right)^{\frac{3}{2}} \quad \text{(SWW case)}, \tag{1.2}$$

where $\hat{k}$ is the x-wavenumber of the incident Rossby wave, and $\hat{v}$ the amplitude of the y-component of the disturbance velocity at the critical layer, which is constant with time in the SWW solution. The magnitude of $\alpha(\infty)$ is shown in figure 1 as the area of the finely dotted rectangle.

The general result to be proved in §§5–7 may be stated as follows. If conditions (i)–(vi) of §1.3 are satisfied, a critical layer of bounded width set up by a Rossby wave of bounded amplitude (the bounds on the width and amplitude being independent of t) must satisfy a relation of the form

$$|\alpha(t)| \leqslant \alpha_{\max} \quad \text{(for all } t\text{)}, \tag{1.3}$$

where the bound $\alpha_{\max}$ depends on the width and amplitude but is independent of t. It is in this, quite strong, sense that the critical layer must be a perfect reflector in the long-time average, regardless of the details of its time evolution. In particular, if a steady state should be reached it must be one of perfect reflection, just as was originally predicted by the BBDH model, where 'reflection' would in general include contributions to $\overline{u'v'}$ due to radiation of higher harmonics from the critical layer, if any.† Explicit expressions for $\alpha_{\max}$ are presented in §§6, 7, after giving appropriate definitions of 'bounded wave amplitude' and 'bounded critical-layer width'. There are different versions representing different tradeoffs between simplicity and sharpness of the bound. The expressions for $\alpha_{\max}$ have the order of magnitude suggested by the right-hand side of (1.2), or more generally by (1.10)–(1.12) below.

One implication of these results is that in cases where the absolute vorticity is uniform throughout the critical layer $\alpha_{\max}$ is zero. Therefore $\alpha(t)$ itself must be zero. It follows that in such cases the critical layer is a perfect reflector for all time t, whether in the linear or nonlinear stages of its development. In other cases where

† It happens that there are no such contributions in the BBDH and SWW models; in the latter case this is a consequence of the boundary conditions imposed, which prevent the higher harmonics from radiating away.

(potential) vorticity gradients are sufficiently weak within the critical layer, any absorption or over-reflection that takes place must be correspondingly weak for all t, a fact which may be significant for certain problems in large-scale atmospheric dynamics (e.g. Tung & Lindzen 1979*a*, *b*; Tung 1979; McIntyre 1982; McIntyre & Palmer 1983).

Although the wave amplitude must be assumed to be bounded for all t, in order for the result (1.3) to apply, it need not be small. Included is the meteorologically important case of large wave amplitude, in which 'critical layers' are certainly nonlinear, but certainly not thin. Thus the results of §§5–7 represent a significant extension of the scope of critical-layer theory. In particular, the method of matched asymptotic expansions is not used in the proof of (1.3). Nor is any particular cat's-eye structure assumed.

The result (1.3) is a direct consequence of the conservative behaviour of (potential) vorticity, which is neither created nor destroyed in the circumstances considered (although it may, for instance, be diffused). For the purpose of proving (1.3), the 'critical layer' is defined as the material region within which significant rearrangement, mixing, diffusion, or turbulent transport of vorticity takes place. In the assumed circumstances, the width of this region will be bounded provided that condition (iv) holds. It turns out to be a matter of some delicacy to decide precisely how much vorticity rearrangement is 'significant'. In the case of the SWW solution, for instance, most of the vorticity rearrangement does take place in and near the cat's eyes of the central region, as illustrated in figure 2 below, but a certain amount takes place also in a much wider region including the outer flow. This is due to the presence of starting transients which render the disturbance not quite monochromatic with respect to streamwise phase speed (Dickinson 1970). They take the form of 'sheared disturbances', which are equivalent to a continuous spectrum of phase speeds and are described in the well-known way by disturbance fields having the form of a function of t times a sinusoidal function of $(x-\Lambda yt)$ (e.g. Kelvin 1887; Yamagata 1976; Brown & Stewartson 1980; Rhines & Young 1983; Shepherd 1985 and references). Their presence in the outer flow is the reason why a full asymptotic solution of the problem for large t requires the use of multiple-scale as well as matched-asymptotic techniques, beyond a certain level of approximation, as was shown by Warn & Warn (1978).

If we ignore this point for the present – it is dealt with rigorously in §§5–7 – then a simple 'mixing argument' can be used to make the result (1.3) immediately plausible, and to give a rough idea of likely values for $\alpha(\infty)$ and $\alpha_{\max}$. Consider a thought experiment in which the source of Rossby waves is gradually turned on and, after an arbitrarily long time interval, gradually turned off again. Assume that the initial and final states are both parallel shear flows, independent of x, with absolute vorticity profiles $q_0(y)$ and $q_\infty(y)$ respectively.† Assume further that there is no permanent rearrangement of vorticity (including that due to diffusion or other transport processes) anywhere except within a 'mixing region'

$$|y| < \tfrac{1}{2}b, \tag{1.4}$$

† The notation q for absolute vorticity is chosen for later convenience and also to suggest the meteorological generalization to stratified, rotating flow, in which the symbol q is often used to denote a quantity that is then relevant, the quasi-geostrophic potential vorticity. The horizontal distribution of q approximately corresponds to the (more fundamentally significant) *isentropic* distribution of Rossby–Ertel potential vorticity. See, for example, the important paper by Charney & Stern (1962), and the recent review by Hoskins *et al.* (1985).

of constant width, b, which for present purposes we identify with the critical layer. The difference

$$\Delta q(y) = q_\infty(y) - q_0(y)$$

between the initial and final absolute vorticity profiles then vanishes outside the mixing region, and vorticity conservation dictates that the integral of Δq across the mixing region must be zero:

$$\int_{-\frac{1}{2}b}^{\frac{1}{2}b} \Delta q(y)\,\mathrm{d}y = 0. \tag{1.5}$$

$\Delta q(y)$ is related to the corresponding change $\Delta u(y)$ in the velocity profile by $\Delta q(y) = -\partial \Delta u(y)/\partial y$, so that

$$\Delta u(y) = -\int_{-\frac{1}{2}b}^{y} \Delta q(y)\,\mathrm{d}y. \tag{1.6}$$

$\Delta u(y)$ vanishes outside the mixing region, by (1.5). It is related to the time-integrated absorptivity by momentum conservation, in virtue of condition (vi) above. We have

$$\alpha(\infty) = \int_0^\infty [\overline{u'v'}]_{-\frac{1}{2}b}^{\frac{1}{2}b}\,\mathrm{d}t = -\int_{-\frac{1}{2}b}^{\frac{1}{2}b} \Delta u(y)\,\mathrm{d}y = -\int_{-\frac{1}{2}b}^{\frac{1}{2}b} y\,\Delta q(y)\,\mathrm{d}y, \tag{1.7}$$

where the last step uses integration by parts and depends on the vanishing of $\Delta u(y)$ outside the mixing region. Suppose now that the range of values of absolute vorticity q within the mixing region satisfies

$$q_{\min} \leqslant q \leqslant q_{\max}. \tag{1.8}$$

Here $q_{\min}$ and $q_{\max}$ can be taken as finite constants for all t, by condition (v). There are two functions $\Delta q(y)$ that maximize the absolute value of the last integral in (1.7) under the constraint (1.8), namely the step functions

$$\Delta q(y) = \pm (q_{\max} - q_{\min}) \operatorname{sgn} y. \tag{1.9}$$

Evaluating the integral, we deduce that (1.3) holds with $t = \infty$ and

$$\alpha_{\max} = \tfrac{1}{4}(q_{\max} - q_{\min})\, b^2. \tag{1.10}$$

Such a $\Delta q(y)$ is of course highly improbable – indeed it may not even represent a q-conserving rearrangement, for general $q_0(y)$ – and so we do not expect this bound to be at all sharp. For instance, suppose that the initial absolute vorticity profile has constant gradient $\mathrm{d}q_0/\mathrm{d}y$, and assume perfect mixing. Then the final value of the absolute vorticity gradient is zero in the mixing region, and evaluation of the right-hand side of (1.7) gives

$$\alpha(\infty) = \frac{1}{12}\frac{\mathrm{d}q_0}{\mathrm{d}y}\, b^3. \tag{1.11}$$

This is one third of the bound given by substituting $q_{\max} - q_{\min} = b\,\mathrm{d}q_0/\mathrm{d}y$ into (1.10), namely

$$\alpha_{\max} = \frac{1}{4}\frac{\mathrm{d}q_0}{\mathrm{d}y}\, b^3. \tag{1.12}$$

The assumption of perfect mixing may not be very realistic (e.g. Clough *et al.* 1985; McIntyre & Palmer 1984; Haynes 1985), but the expression (1.11) is probably closer, nevertheless, to actual large-time values of $\alpha(t)$ than (1.12) would be.

The estimate (1.11) suggests an intuitively appealing way of characterizing the time-integrated absorptivity of a Rossby-wave critical layer, in terms of a 'mixing

width' b_m. For any Rossby-wave critical layer whose absolute vorticity gradient has a constant initial value dq_0/dy and for which the limit $\alpha(\infty)$ is attained, one could define its 'mixing width' b_m as the value of b which makes the right-hand side of (1.11) equal to the actual time-integrated absorptivity, i.e.

$$b_m = \left| 12 \frac{\alpha(\infty)}{dq_0/dy} \right|^{\frac{1}{3}}. \tag{1.13}$$

In the case of the SWW solution, (1.2) and (1.13) give

$$b_m = 3.3331 \left| \frac{\hat{v}}{kA} \right|^{\frac{1}{2}}, \tag{1.14}$$

which is about five-sixths of the width of the cat's eyes in the stream-function pattern. This is indicated by the bar at the centre of figure 2(*d*) below.

The analysis to be given in §§5–7 improves on the simple mixing argument just given, by replacing the unrealistic assumption (1.4), that of a finite mixing region, with a realistic assumption about the rate at which disturbances to the vorticity field fall off with distance from the central region. The resulting proof of (1.3) is rigorous and quite generally applicable. It applies for all t, does not require the incident wave to die away after a certain time, and does not require vorticity rearrangement to be wholly confined to a region of finite width. The mathematical questions involved are not entirely trivial. This can be seen at once from the presence of the factor y in the integral on the right of (1.7) together with the fact that in some cases, for instance that described by the SWW solution, the magnitude of the vorticity disturbance falls off as slowly as $|y|^{-1}$ as $|y| \to \infty$.

The general proof also avoids reference to the actual momentum change $\int \Delta u(y)\, dy$. Rather, it is expressed directly in terms of disturbance correlations like $\overline{u'v'}$ and the corresponding vorticity flux $\overline{v'q'}$, which as is well known are related by the Taylor identity

$$\frac{\partial}{\partial y}(\overline{u'v'}) = -\overline{v'q'}. \tag{1.15}$$

This route to (1.3) is a better preparation for extending the analysis to the three-dimensional, meteorological case. In that case the actual momentum change does not satisfy any relation as simple as (1.7), because of the existence of secondary circulations in the (y, z)-plane which redistribute absolute angular momentum. However, (1.15) generalizes immediately, the left-hand side being replaced by the convergence of the Eliassen–Palm flux, as is well known, and the right-hand side by minus the flux of quasi-geostrophic potential vorticity.

The mathematical tool on which the general proof depends is a finite-amplitude conservation theorem for time-dependent disturbances to parallel shear flows, which appears to be new and which generalizes the well-known results of Taylor (1915) and (for the three-dimensional case) of Eliassen & Palm (1961). The two-dimensional version of this theorem is proved in §§5 and 7. The 'diffusive' case is included by means of a formalism which expresses condition (v) of §1.3 in a somewhat abstract but very general way, thus allowing for a very wide class of vorticity-transport processes of which ordinary downgradient diffusion is a special case. The only restrictions are those dictated by condition (v), and implicitly by condition (vi): the transport process must not, of course, violate momentum conservation. The conservation theorem may have applications other than the present one. In the non-diffusive case the theorem can be inferred from a mathematical analogy between the present problem and the notion of 'available potential energy' for a stratified

fluid, as suggested by the forms of the right-hand sides of (1.7) and (1.15). The result (1.3) corresponds, in the analogy, to the statement that the potential-energy change due to rearrangement of a stratified layer of finite depth is bounded.

We now turn to the SWW solution and its instability.

2. The SWW solution

The review to be given in this section will introduce some essential notation, and will also remind us of the fact that the fluid-dynamical situation described by the SWW solution has two cardinal features in common with other, more complicated cases, including cases in which the instability is excited and modifies the properties of the critical layer. First, the absorbing or reflecting properties of the critical layer depend mainly on the distribution of vorticity within and near the Kelvin 'cat's eyes' of the stream-function pattern. This vorticity field induces a velocity field which extends outside the critical layer and thereby influences the outer flow. Secondly, the time evolution of the vorticity distribution within the critical layer is given by a conceptually simple rule: take the *leading approximation* to the cat's-eye streamline pattern, which is controlled by the linear, wavelike flow outside the critical layer through the leading-order matching condition, and use the corresponding approximate velocity field to advect the absolute vorticity distribution within the critical layer. For this purpose one can neglect the correction to the velocity field induced by the changing vorticity distribution. Within the critical layer, it is a higher-order contribution. In summary,

(*a*) as far as the inner, critical-layer problem is concerned the absolute vorticity behaves like a passive tracer to leading order, but

(*b*) its induced velocity field is important to leading order *outside* the critical layer, being the means whereby the critical layer exerts its influence upon the outer flow and acts as an absorber, reflector, or over-reflector.

In the case of the SWW solution itself, there is a further simplification. Conditions in the outer flow, including the boundary conditions, can be chosen in a special way (Warn & Warn 1978, §6; (2.24) below) such that the changes induced in the outer flow by the evolving vorticity distribution within the critical layer do not react back upon the leading-order cat's-eye streamline pattern. The cat's-eye flow can therefore be taken as steady, to leading order, throughout the nonlinear evolution. The resulting absolute vorticity distribution within the critical layer is particularly simple. It is illustrated in figures 2(*a*), (*b*), (*c*), (*d*), at four successive instants corresponding respectively to the times marked t_a, t_b, t_c, and t_d on the abscissa in figure 1. The cat's-eye flow is twisting up the contours of constant absolute vorticity like spaghetti on a fork. The possibility of shear instability is at once apparent from the fact that regions of reversed absolute vorticity gradient exist after a certain time t_i, lying between t_a and t_b and also marked on the abscissa in figure 1. The time t_i is the time for the centre of the cat's eye to rotate through one right angle.

The leading-order analytical details of the SWW solution will now be summarized, together with the scaling assumptions which define its parameter regime. As before, (x, y) are streamwise and spanwise Cartesian coordinates and (u, v) the corresponding velocity components. The starting point is the barotropic vorticity equation for inviscid, two-dimensional, incompressible motion on a beta-plane, namely

$$\frac{\mathrm{D}q}{\mathrm{D}t} = 0, \tag{2.1}$$

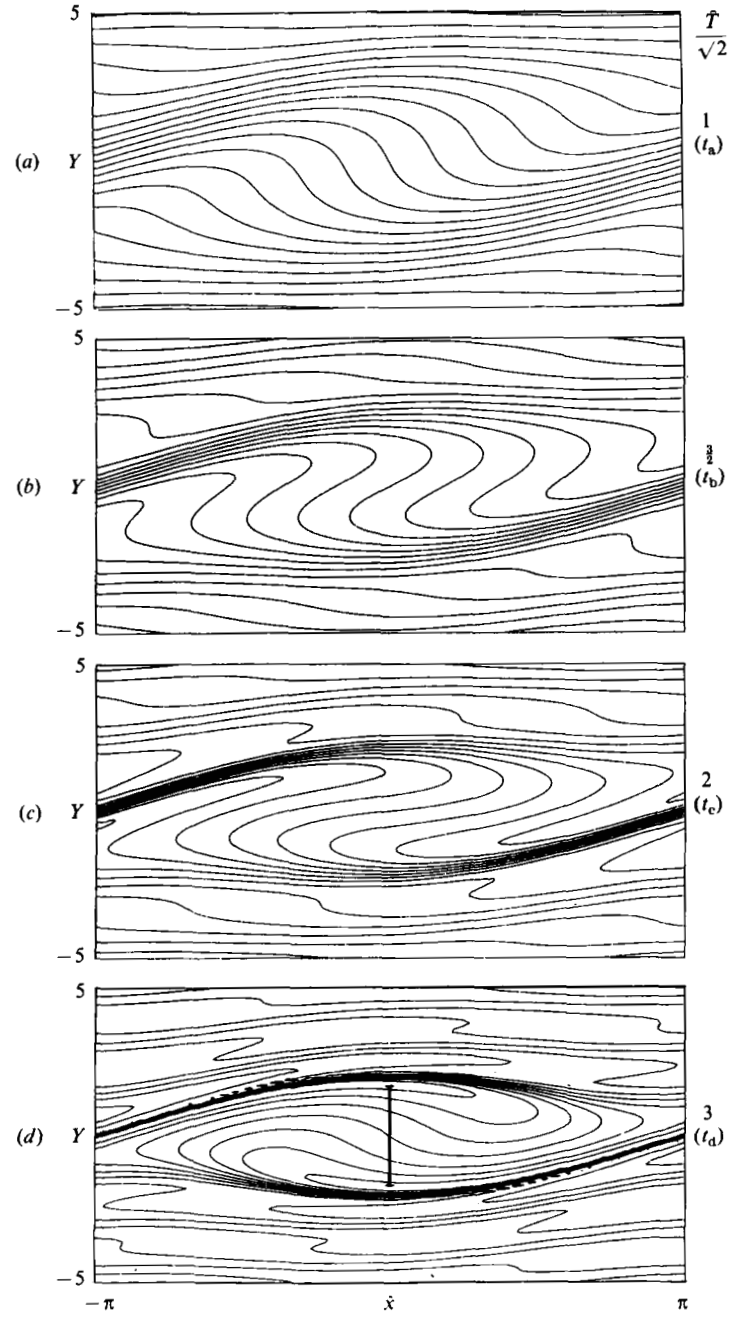

FIGURE 2. Contours of the absolute vorticity field predicted by the SWW analytical solution, at four times t_a, t_b, t_c and t_d. The dimensionless time units are the same as in figure 1; see also (2.21). Linear theory (Dickinson 1970) is beginning to break down at time t_a, panel (*a*), and perfect reflection is attained just before time t_d, panel (*d*). The *y*-scale is exaggerated for clarity. The abscissa Y is expressed in units of Δy, where Δy is equal to $\epsilon^{\frac{1}{2}}\Lambda/\beta$, where $\beta = \mathrm{d}q_0/\mathrm{d}y$ is the basic absolute (potential) vorticity gradient, and ϵ the disturbance amplitude defined in §2. The bar at the centre of (*d*) represents the 'mixing width' b_m defined by equation (1.14).

where the material derivative

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}, \tag{2.2}$$

and where the materially conserved quantity q is the absolute vorticity, as before. It is defined by

$$q = \beta y + \zeta + \text{constant}, \tag{2.3}$$

where

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \tag{2.4}$$

the relative vorticity, and where β is a positive constant, representing the planetary vorticity gradient or northward gradient of the Coriolis parameter. The incompressibility condition

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{2.5}$$

permits the introduction of a stream function $\psi(x, y, t)$, for which we adopt the sign convention

$$u = -\frac{\partial \psi}{\partial y}, \quad v = +\frac{\partial \psi}{\partial x}, \tag{2.6}$$

giving $\zeta = +\nabla^2\psi$ so that

$$q = \beta y + \nabla^2\psi + \text{constant}. \tag{2.7}$$

These equations and definitions comprise the basic mathematical model to be used until §7, where we generalize (2.1) to include the possibility of a non-advective transport consistent with condition (v) of §1.3.

SWW examined weak, x-periodic disturbances, of wavelength $2\pi/\hat{k}$, say, to an initial state with constant shear Λ:

$$u = u_0(y) = \Lambda y, \quad q = q_0(y) = \beta y + \text{constant}. \tag{2.8}$$

Thus

$$\psi = -\tfrac{1}{2}\Lambda y^2 + \epsilon\phi(x, y, t), \tag{2.9}$$

say, where ϵ is a small, dimensionless parameter. It is convenient at this point to redefine the symbols x, y, t, q, ψ and ϕ so that they become dimensionless quantities, using the natural scales

$$\left.\begin{array}{ll} \Lambda/\beta & \text{for } x, y, \\ \Lambda^{-1} & \text{for } t, q^{-1}, \\ \Lambda^3/\beta^2 & \text{for } \psi, \phi. \end{array}\right\} \tag{2.10}$$

When ϕ and y are both of order unity, i.e. away from the critical layer, the small dimensionless parameter ϵ measures the order of magnitude of the particle displacement in the y-direction in units of Λ/β. The dimensionless forms of (2.7) and (2.9) are the same expressions with β and Λ set equal to unity. Substituting these into (2.1) ff., we get

$$\nabla^2\phi_t + y\nabla^2\phi_x + \phi_x + \epsilon(\phi_x\nabla^2\phi_y - \phi_y\nabla^2\phi_x) = 0, \tag{2.11}$$

where the suffixes x, y and t denote partial differentiation.

The SWW solution depends not only on assuming that ϵ is small, but also on considering solutions to (2.11) whose x-wavelengths are *long* in comparison with the natural lengthscale Λ/β (more precisely, $\hat{k} \ll \beta/\Lambda$), so that the operator ∇^2 appearing in (2.7) and (2.11) may be approximated by $\partial^2/\partial y^2$ for all y. (Warn & Warn's formal

analysis covers many other cases as well, but these do not admit analytical solutions and will not be considered here.) Let

$$\mu = \hat{k}\frac{\Lambda}{\beta} \ll 1, \tag{2.12}$$

and introduce the corresponding slow x and t variables

$$\hat{x} = \mu x, \quad \hat{t} = \mu t; \tag{2.13}$$

$\hat{x}$ is the x-distance measured in radian wavelengths, and $\hat{t} = 1$ represents the time for a fluid element at $y = 1$ (dimensionally, at $y = \Lambda/\beta$) to be carried a dimensional distance $\hat{k}^{-1}$ by the basic flow (2.8). These assumptions allowed SWW to simplify (2.11) to

$$\hat{\phi}_{yy\hat{t}} + y\hat{\phi}_{yy\hat{x}} + \hat{\phi}_{\hat{x}} + \epsilon(\hat{\phi}_{\hat{x}}\hat{\phi}_{yyy} - \hat{\phi}_y\hat{\phi}_{yy\hat{x}}) = O(\mu^2), \tag{2.14}$$

where the symbol $\hat{\phi}$ denotes ϕ regarded as a function of $\hat{x}, y$ and $\hat{t}$.

SWW considered flow in a domain $-\infty < y < y_b$, with y_b formally of order unity, and took the initial and boundary conditions to be

$$\hat{\phi} = 0 \quad (y \leqslant y_b, \hat{t} = 0), \tag{2.15a}$$

$$\hat{\phi} \to 0 \quad (y \to -\infty, \hat{t} > 0), \tag{2.15b}$$

$$\hat{\phi} = \mathrm{Re}\,(a\,\mathrm{e}^{\mathrm{i}\hat{x}}) \quad (y = y_b, \hat{t} > 0). \tag{2.15c}$$

These describe the switching on, at $\hat{t} = 0$, of a steady sinusoidal disturbance at $y = y_b$ such as might be generated by the flow past a corrugated boundary. The constant a appearing in (2.15c) is taken to be real and of order unity; its numerical value will be chosen shortly in a way that will prove convenient.

The work of Dickinson (1970), Warn & Warn (1976) and SWW showed that the leading approximation to the linear ($\epsilon = 0$) solution of (2.14) and (2.15) tends to a steady state, for $\hat{t} \gg 1$, everywhere except within a non-dimensional distance $O(t^{-1})$ of the critical line $y = 0$. This steady-state, linear solution is

$$\hat{\phi} = \begin{cases} \mathrm{Re}\,[\{A_0 f(y) + B_0 g(y)\}\,\mathrm{e}^{\mathrm{i}\hat{x}}] & (y > 0), \quad (2.16a) \\ \mathrm{Re}\,[A_0 h(y)\,\mathrm{e}^{\mathrm{i}\hat{x}}] & (y < 0), \quad (2.16b) \end{cases}$$

where the real-valued functions $f(y)$, $g(y)$ and $h(y)$ are defined in terms of the Bessel functions Y_1, J_1 and K_1 by

$$\begin{aligned} f(y) &= -\pi y^{\frac{1}{2}} Y_1(2y^{\frac{1}{2}}) \\ &= 1 - y\log y - (2\gamma - 1)\,y + O(y^2 \log y), \end{aligned} \tag{2.17a}$$

$$g(y) = y^{\frac{1}{2}} J_1(2y^{\frac{1}{2}}) = y + O(y^2), \tag{2.17b}$$

$$\begin{aligned} h(y) &= 2\,|\,y\,|^{\frac{1}{2}} K_1(2\,|\,y\,|^{\frac{1}{2}}) \\ &= 1 + |\,y\,|\log|\,y\,| + (2\gamma - 1)\,|\,y\,| + O(|\,y\,|^2 \log|\,y\,|), \end{aligned} \tag{2.17c}$$

$\gamma = 0.57722$ being Euler's constant and the O symbols referring to the limit $y \to 0$. It is straightforward to check that (2.16) satisfies (2.14) with the terms in $\partial/\partial t$, ϵ and μ neglected. The constants A_0 and B_0 in (2.16) are determined, in this linear problem, by the relations

$$B_0 = -\mathrm{i}\pi A_0 \tag{2.18}$$

and

$$A_0 f(y_b) + B_0 g(y_b) = a. \tag{2.19}$$

The last relation expresses the boundary condition (2.15c).

According to (2.16) and (2.17), $\hat{\phi}$ is continuous across the critical line $y = 0$. The continuity of $\hat{\phi}$, together with the relation (2.18), can be deduced either from the exact solution of the initial-value problem (Dickinson 1970; Warn & Warn 1976) or alternatively from an application of the method of matched asymptotic expansions (SWW). We note that the dimensionless Reynolds stress $-\overline{\hat{\phi}_{\hat{x}}\hat{\phi}_y}$ implied by (2.16) is constant in $y > 0$ (Taylor 1915; Foote & Lin 1950; Eliassen & Palm 1961), being proportional to the Wronskian $fg' - f'g = 1$. The Reynolds stress is zero in $y < 0$, and the resulting jump discontinuity across $y = 0$ is given by

$$[-\overline{\hat{\phi}_{\hat{x}}\hat{\phi}_y}] = \tfrac{1}{2}\operatorname{Im}(A_0 B_0^*), \tag{2.20}$$

where the asterisk denotes the complex conjugate.

The validity of linear theory ceases when $\hat{t}$ becomes large, of order $\epsilon^{-\frac{1}{2}}$ (Warn & Warn 1976), at which time the ϵ-term in (2.14) becomes significant. It does so within the critical-layer region $y = O(\epsilon^{\frac{1}{3}})$. From then on a still slower time variable

$$\hat{T} = \epsilon^{\frac{1}{2}}\hat{t} = \epsilon^{\frac{1}{2}}\mu t \tag{2.21}$$

is appropriate; it is this that corresponds to the timescale of the nonlinear evolution illustrated in figures 1 and 2. SWW showed that the steady-state y-structure expressed by (2.16) and (2.17) is still relevant outside the critical layer on this timescale, but that because of the nonlinear evolution within the critical layer the constants A_0, B_0 in (2.16) become functions of $\hat{T}$, with (2.18) ceasing to hold, while extra terms representing the higher harmonics of $\exp(\mathrm{i}\hat{x})$ appear in the solution. Following Stewartson, we express this state of affairs succinctly by writing

$$\left.\begin{array}{c} A(\hat{x},\hat{T}),\ B(\hat{x},\hat{T}) \\ \text{in place of} \\ \operatorname{Re}[A_0 \mathrm{e}^{\mathrm{i}\hat{x}}],\ \operatorname{Re}[B_0 \mathrm{e}^{\mathrm{i}\hat{x}}], \end{array}\right\} \tag{2.22}$$

in (2.16), with $A(\hat{x},\hat{T}), B(\hat{x},\hat{T})$ real, so that, in particular, the boundary condition (2.19), with a real, is replaced by

$$A(\hat{x},\hat{T})\,f(y_b) + B(\hat{x},\hat{T})\,g(y_b) = a\cos\hat{x}. \tag{2.23}$$

The SWW solution describes the nonlinear behaviour, for order-unity values of $\hat{T}$, in the set of special cases for which $g(y_b) = 0$ (and $f(y_b) \neq 0$) so that (2.23) constrains A to be independent of $\hat{T}$. The boundary positions for which $g(y_b) = 0$ are

$$y_b = (\tfrac{1}{2}j_{1,n})^2 = 3.671, 12.305, 25.875, \ldots, \tag{2.24}$$

$j_{1,n}$ being the nth zero of the Bessel function J_1. In the remainder of this section, and in §§3 and 4, we restrict attention to these cases, and without loss of generality choose a in (2.15c) and (2.23) to be such that

$$A(\hat{x},\hat{T}) = A(\hat{x}) = \cos\hat{x}. \tag{2.25}$$

Now if we substitute into (2.16) and the dimensionless form of (2.9) the leading-order, $O(1)$ terms of (2.17), we obtain the following small-y approximation to $\hat{\psi}$, the dimensionless total stream function ψ regarded as a function of $\hat{x}$, y and $\hat{t}$:

$$\hat{\psi} \sim -\tfrac{1}{2}y^2 + \epsilon\cos\hat{x}. \tag{2.26}$$

SWW showed that this gives the leading approximation to the total stream function even *inside* the critical layer, $y = O(\epsilon^{\frac{1}{3}})$, and not only at the end of the linear stage (as already shown by Dickinson's solution) but also throughout the nonlinear stage

of evolution, $\hat{T} = O(1)$. The velocity field whose stream function is (2.26) is a steady Kelvin cat's-eye flow, the bounding streamline through the saddle points of $\hat{\psi}$ being given by $\hat{\psi} = -\epsilon$, i.e. by

$$y = \pm 2\epsilon^{\frac{1}{2}} |\cos(\tfrac{1}{2}\hat{x})|, \tag{2.27}$$

showing that the maximum y-extent of the cat's eyes is just $\pm 2\epsilon^{\frac{1}{2}}$ in units of A/β, or ± 2 in terms of the scaled y-coordinate

$$Y = \epsilon^{-\frac{1}{2}} y \tag{2.28}$$

for the critical layer. The timescale for the nonlinear evolution, corresponding to $\hat{T} \sim 1$, can be thought of as the time for a fluid element to travel a substantial portion of the way around one of the cat's eyes. For instance a fluid element near the centre of a cat's eye, e.g. an element near the origin, makes a single round trip when $\hat{T}$ increases by 2π.

SWW's analysis showed that the absolute vorticity q in the critical layer takes the form

$$q = \epsilon^{\frac{1}{2}} \hat{Q}(\hat{x}, Y, \hat{T}) + \text{constant}, \tag{2.29}$$

where

$$\hat{Q}(\hat{x}, Y, \hat{T}) = Y + \hat{Q}_1(\hat{x}, Y, \hat{T}) + O(\epsilon^{\frac{1}{2}} \log \epsilon), \tag{2.30}$$

$\hat{Q}, \hat{Q}_1$ as well as Y being order-unity quantities, and where $Y + \hat{Q}_1$, the first approximation to the scaled absolute vorticity, satisfies

$$\left(\frac{\partial}{\partial \hat{T}} + \frac{\partial \hat{\Psi}_0}{\partial \hat{x}} \frac{\partial}{\partial Y} - \frac{\partial \hat{\Psi}_0}{\partial Y} \frac{\partial}{\partial \hat{x}}\right)(Y + \hat{Q}_1) = 0. \tag{2.31}$$

Here $\hat{\Psi}_0$ is the appropriately scaled form of (2.26) (scale ϵ) written as an order-unity function of $(\hat{x}, Y)$:

$$\hat{\Psi}_0(\hat{x}, Y) = -\tfrac{1}{2} Y^2 + \cos \hat{x}. \tag{2.32}$$

Since $\hat{\Psi}_0$ is a known function, $\hat{Q}_1$ may be computed to leading order simply by integrating the first-order hyperbolic equation (2.31) along each steady streamline of (2.32), either numerically, or analytically using Stewartson's implicit solution in terms of elliptic functions.† Either procedure, as we ourselves have verified, gives the results shown in figure 2, where $\hat{T}/\sqrt{2} = 1$, 1.5, 2 and 3 in panels (*a*), (*b*), (*c*) and (*d*) respectively. The situation described by (2.31) and figure 2 is a special case of the more general statement (*a*) made at the beginning of this section.

For later reference we note the asymptotic approximation

$$\hat{Q}_1 \simeq \frac{1}{Y}\{\cos(\hat{x} - Y\hat{T}) - \cos \hat{x}\} \quad (Y \gg \hat{T}) \tag{2.33}$$

(Warn & Warn 1976, 1978; Stewartson 1978), which satisfies (2.31) with zero initial disturbance if we make the approximation corresponding to linear critical-layer theory, viz. neglect the term $-\sin \hat{x}\, \partial \hat{Q}_1/\partial y$ in (2.31) with (2.32) substituted. It is the large-Y behaviour indicated by (2.33) that gives rise to the difficulty in using (1.7) mentioned earlier; (2.33) also illustrates the development outside the cat's eyes of the 'sheared disturbances' mentioned in §1.4. These are visible near the periphery of figure 2(*d*).

The vorticity anomaly $\hat{Q}_1$ induces a jump $[-\hat{\phi}_y]$ in the dimensionless x-velocity

† The fluid flow (2.32) is the same as the (incompressible) phase-space flow for a simple pendulum (Whittaker 1937, p. 72–74).

component $-\hat{\phi}_y$, as we move across the critical layer, so that (2.18) is replaced by

$$B(\hat{x}, \hat{T}) = \oint_{-\infty}^{\infty} \hat{Q}_1 \, \mathrm{d}Y \quad (= [\hat{\phi}_y]). \tag{2.34}$$

Here $\oint$ and [ ] denote the Cauchy principal values of the respective limits $|Y| \to \infty$ and $|y| \to 0$. $B(\hat{x}, \hat{T})$ is an odd, periodic function of $\hat{x}$ whose nth harmonic is

$$B_n(\hat{T}) \sin(n\hat{x}),$$

where

$$B_n(\hat{T}) = \frac{1}{\pi} \int_{-\pi}^{\pi} B(\hat{x}, \hat{T}) \sin(n\hat{x}) \, \mathrm{d}\hat{x}. \tag{2.35}$$

Thus, recalling that $A(\hat{x}, \hat{T}) = \cos\hat{x}$, we see that (2.20) is replaced by

$$[-\overline{\hat{\phi}_{\hat{x}} \hat{\phi}_y}] = \tfrac{1}{2} B_1(\hat{T}), \tag{2.36}$$

giving the time evolution of the dimensionless Reynolds-stress jump according to the SWW solution. The quantity plotted in figure 1 is actually $-B_1(\hat{T})$, which is equal to the traditionally defined 'logarithmic phase jump'. As figure 1 shows, it takes the value $-\pi$ for small $\hat{T}$. The value $-\pi$ can be directly verified from the asymptotic solution (2.33) using contour integration in (2.34), or alternatively from compatibility with (2.18), or yet again from the fact that, because of the well-known properties of linear differential equations, (2.16*b*) must be the analytic continuation of (2.16*a*), via the lower half-plane, at the end of the linear stage. The value $-B_1 = -\pi$ is often regarded as signifying perfect absorption, since this is true in a certain class of incident-wave problems, including the one originally solved by Dickinson, in which the basic shear and absolute vorticity gradient vary on a much larger y-scale than A/β. In other cases the absorption may be only partial even when $-B_1 = -\pi$ (e.g. Tung 1979, equation (46)), depending very much on the relative values of the basic absolute vorticity gradient near and away from the critical layer; recall the remarks in the paragraph below (1.3).

In connection with the terminology 'phase jump' it should perhaps be recalled that $-B_1$ does not represent any actual phase discontinuity in the spatial structure of the disturbance itself, as is clear for instance from the fact that $-B_1 = -\pi$ does not signify the same state of things as $-B_1 = +\pi$. Rather, as Warn & Warn show (1978, p. 43, *q.v.* for further discussion), $-B_1$ can be regarded as the jump in the dimensionless phase *slope* $\partial\theta/\partial y$, where $\theta(y, \hat{T})$ is the phase of the fundamental complex Fourier component of the $\hat{x}$-dependence of the disturbance stream function. A better terminology might therefore be to call $-B_1$ the 'phase-slope jump' (of the fundamental harmonic).

Figure 3 compares the time dependence of the second and third harmonics $-B_2(\hat{T})$ and $-B_3(\hat{T})$ with that of $-B_1(\hat{T})$.

To complete our description of the SWW solution, we note the approximation to the total stream function within the critical layer that corresponds to the leading-order relative-vorticity distribution $\hat{Q}_1$. This may be written in terms of the scaled critical-layer stream function $\hat{\Psi}$ defined by

$$\hat{\Psi}(\hat{x}, Y, \hat{T}) = \epsilon^{-1} \hat{\psi}(\hat{x}, y, \hat{T}), \tag{2.37}$$

to which $\hat{\Psi}_0$ in (2.32) is the leading approximation. We have

$$\hat{\Psi}(\hat{x}, Y, \hat{T}) = \hat{\Psi}_0(\hat{x}, Y) - \tfrac{1}{2}(\epsilon^{\frac{1}{2}} \log\epsilon) \, Y \cos\hat{x} - \epsilon^{\frac{1}{2}}\{(2\gamma - 1) \, Y \cos\hat{x} + \hat{\Psi}_1\} + O(\epsilon \log\epsilon), \tag{2.38}$$

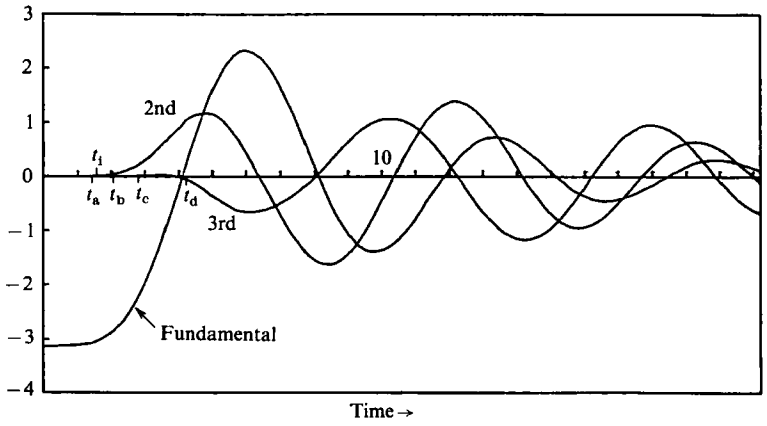

FIGURE 3. Time evolution of the first three harmonics $-B_1$, $-B_2$, $-B_3$, defined in (2.35), of the Cauchy-principal-value velocity jump across the critical layer, according to the SWW solution. The first harmonic $-B_1$ gives the same curve as in figure 1; see (2.36) ff.

where $\hat{\Psi}_1(\hat{x}, Y, \hat{T})$ is related to $\hat{Q}_1(\hat{x}, Y, \hat{T})$ by

$$\frac{\partial^2}{\partial Y^2}\hat{\Psi}_1(\hat{x}, Y, \hat{T}) = \hat{Q}_1(\hat{x}, Y, \hat{T}). \tag{2.39}$$

The remaining terms in (2.38) all correspond to irrotational motions, to this order. It is straightforward to verify, using the matching rules given, for instance, by Lesser & Crighton (1975) and Van Dyke (1975), and recalling the asymptotic result (2.33), that (2.38) represents a three-term inner expansion to the first three orders of the asymptotic sequence

$$(1, \epsilon^{\frac{1}{2}} \log \epsilon, \epsilon^{\frac{1}{2}}, \epsilon \log \epsilon, \ldots), \tag{2.40}$$

which matches the two-term outer expansion represented by (2.16) and (2.17) together with the dimensionless form of (2.9). We note, in particular, that (2.39) is consistent with (2.34) ff. We note also that, at the first two orders, the inner problem permits $\hat{\Psi}$ to be any function of the (approximately irrotational) form $F(\hat{x}, \hat{T})\, Y + G(\hat{x}, \hat{T})$, but that at the first order this cannot match the outer solutions (2.16a, b) unless $F = 0$. Together with (2.23) and (2.24) this is what fixes the form of the first approximation $\hat{\Psi}_0$ to be the form given in (2.32). The same principle will apply in the next section when finding the stream function describing the unstable disturbance.

In preparation for the instability analysis it may also be useful to re-emphasize what has already been said in several ways, namely that the higher-order terms in (2.38) are not needed to describe the velocity field responsible for advecting vorticity within the critical layer (recall (2.31)). At the order of accuracy to which the problem has been solved, the higher-order terms in (2.38) arise solely as part of the machinery whereby the matched asymptotics express inversion of the relation (2.7), regarded as an equation for ψ with q given. This of course is what is meant by the velocity field 'induced' by a given vorticity field. The inversion is an elliptic boundary-value problem involving the entire flow domain, both inner and outer regions simultaneously – albeit slightly disguised by the fact that, under the present assumption (2.12), the operator ∇^2 is everywhere replaced by $\partial^2/\partial y^2$. It is only in solving that boundary-value problem that matching between the two regions is involved. By

contrast, solution of the leading-order vorticity equation (2.31), for given $\hat{\Psi}_0$, does not depend on any matching condition. This is evident from the hyperbolic character of that equation.

If one were to extend the solution to higher orders of accuracy, one would find that the matching would hold in a coarse-grained sense only, as Warn & Warn (1978) made clear. This is because (2.16) and (2.17) neglect the sheared disturbances whose presence is indicated by the large-Y behaviour of (2.33), and which extend into the outer region where they give rise to a fine-grained structure with typical y-scale $\epsilon^{\frac{1}{2}}$, vorticity amplitude ϵ, and stream-function amplitude ϵ^2, in the units of (2.10). Warn & Warn handle this explicitly by using a combination of matched-asymptotic and multiple-scale techniques. In particular, the analogue of (2.29) for the outer solution is

$$q = y + \epsilon\hat{q}(\hat{x}, y, Y, \hat{T}) + \text{constant}. \tag{2.41}$$

Strictly speaking, the y-dependence of the stream function (2.9) should be rewritten in the corresponding way, but to the order considered this is inconsequential because of the fact that the leading-order fine-grained contribution to the stream function ψ is $O(\epsilon^2)$, $\ll \epsilon$, because of (2.7) and (2.28).

3. The instability problem

We now consider disturbances to the slowly evolving flow discussed in §2. In place of (2.29), the dimensionless absolute vorticity is assumed to take the form

$$q = \epsilon^{\frac{1}{2}}\{\hat{Q}(\hat{x}, Y, \hat{T}) + \tilde{Q}(x, Y, T)\} + \text{constant} \tag{3.1a}$$

in the inner region. $\hat{Q}$ is the undisturbed flow (to which figure 2 represents the leading approximation) and $\tilde{Q}$ is the disturbance. Similarly, in place of (2.41), we assume

$$q = y + \epsilon\{\hat{q}(\hat{x}, y, Y, \hat{T}) + \tilde{q}(x, y, T)\} + \text{constant} \tag{3.1b}$$

in the outer region. T is a new time variable whose definition is

$$T = \epsilon^{\frac{1}{2}}t, \tag{3.2}$$

and x is the dimensionless coordinate introduced in (2.10). The corresponding stream functions are written as

$$\psi = \epsilon\{\hat{\Psi}(\hat{x}, Y, \hat{T}) + \tilde{\Psi}(x, Y, T)\} \tag{3.3a}$$

and

$$\psi = -\tfrac{1}{2}y^2 + \epsilon(\hat{\phi}(\hat{x}, y, \hat{T}) + \tilde{\phi}(x, y, T)\}, \tag{3.3b}$$

for the inner and outer regions respectively. Strictly speaking, the y-dependences of $\tilde{q}$ and $\tilde{\phi}$, as well as that of $\hat{\phi}$, should all be written in multiple-scale notation like the y-dependence of $\hat{q}$ because of the presence of the fine-grained structure in $\hat{q}$ mentioned at the end of §2. But it will be found that the fine-grained structure does not enter the instability problem to leading order. Consequently, to that order, $\tilde{q}$ and $\tilde{\phi}$ have the simple outer structure just indicated in (3.1*b*) and (3.3*b*). The whole picture will be seen to be justified *post hoc* when we find growth-rate maxima within the assumed class of disturbances, with modal structures consistent with the forms of (3.1)–(3.3).

Before proceeding, it may be useful to summarize all the time variables that have been introduced. They are

$$t, \quad \hat{t} = \mu t, \quad T = \epsilon^{\frac{1}{2}}t, \quad \hat{T} = \epsilon^{\frac{1}{2}}\mu t. \tag{3.4}$$

It is immaterial to our results which of the middle two is the slower time variable, but it is essential that both are much slower than the left-hand variable t, and that

both are much faster than the right-hand variable $\hat{T}$. We note that the dimensional timescale corresponding to T is the reciprocal of

$$\delta q = \epsilon^{\frac{1}{2}} \Lambda, \tag{3.5}$$

which is therefore the dimensional scale for disturbance growth rates. Maximum growth rates will in fact turn out to be equal to δq times dimensionless numbers ranging up to 1.24, in the cases considered. The same quantity δq measures the typical strength of vorticity variations within the critical layer, being β times the critical-layer thickness $\epsilon^{\frac{1}{2}} \Lambda/\beta$. Moreover, the corresponding timescale δq^{-1} is equal to the time for a fluid element in the critical layer to move a streamwise distance of order Λ/β, (non-dimensionally, $x \sim 1$), the streamwise scale of the disturbance. It will be recalled that, by our earlier assumption (2.12), Λ/β is much smaller than the scale $\hat{k}^{-1}$ (non-dimensionally, $\hat{x} \sim 1$), characteristic of the SWW solution.

The foregoing scale relations have two important consequences. First, the disturbances can be analysed locally in x (although not in y); we can speak of growth rates and disturbance structures 'at' different locations along the length of the SWW cat's eye. Secondly, growth will be effectively instantaneous when viewed on the longest of the four timescales, that of the nonlinear evolution of the SWW solution. In particular, the instability will have taken effect shortly after the time t_{i}, and long before the time t_{b}, marked in figure 1.

To derive the equation satisfied by the disturbance in the inner region, we substitute the expressions (3.1a) and (3.3a) into the dimensionless forms of (2.1) ff., subtract out the equation satisfied by the undisturbed flow, and replace t by $\epsilon^{-\frac{1}{2}} T$ and y by $\epsilon^{-\frac{1}{2}} Y$, etc. The result is

$$\left(\frac{\partial}{\partial T} + \mu \hat{\Psi}_{\hat{x}} \frac{\partial}{\partial Y} - \hat{\Psi}_Y \frac{\partial}{\partial x}\right) \tilde{Q} + \left\{ \tilde{\Psi}_x \frac{\partial}{\partial Y} - \tilde{\Psi}_Y \left(\frac{\partial}{\partial x} + \mu \frac{\partial}{\partial \hat{x}}\right) \right\} (\hat{Q} + \tilde{Q}) = 0. \tag{3.6}$$

We may now introduce the approximate expressions (2.30) and (2.38) to represent the undisturbed (SWW) flow, and drop all but the leading terms under our scaling assumptions. The inner disturbance equation simplifies to

$$\left(\frac{\partial}{\partial T} - \hat{\Psi}_{0Y} \frac{\partial}{\partial x}\right) \tilde{Q} + \tilde{\Psi}_x \frac{\partial}{\partial Y} (Y + \hat{Q}_1) = -\tilde{\Psi}_x \tilde{Q}_Y + \tilde{\Psi}_Y \tilde{Q}_x, \tag{3.7}$$

with relative errors $O(\mu)$ and $O(\epsilon^{\frac{1}{2}} \log \epsilon)$. The only surviving factor involving $\hat{\Psi}$ is $\partial \hat{\Psi}_0 / \partial Y$, which by (2.32) is just $-Y$. The remaining contributions from (2.38) are still negligible as far as their advective effects are concerned. Going through the corresponding procedure for the outer region, we find simply

$$y \tilde{q}_x + \tilde{\phi}_x = 0, \quad \text{where} \quad \tilde{q} = \nabla^2 \tilde{\phi}. \tag{3.8a,b}$$

The dominant relative error is $O(\epsilon^{\frac{1}{2}})$, from neglecting the $\partial \tilde{q}/\partial T$ and $\tilde{\phi}_x \partial \hat{q}/\partial Y$ terms.† The full two-dimensional Laplacian appears in (3.8b) because of the fact that the x- and y-scales for the instability are both Λ/β.

It is noteworthy that $\hat{q}$ does not appear in (3.8) at all. The outer problem is linear, under our scaling assumptions, and the principle of superposition applies to the outer

† This is the dominant formal error from the process of approximating the outer equation, but it should be added that we expect a very slightly larger relative error $O(\epsilon^{\frac{1}{2}} \log \epsilon)$ in the leading-order outer solution to be obtained. This is not due to any term neglected in the outer equation. Rather, it is induced by the inner vorticity distribution, whose relative error will be $O(\epsilon^{\frac{1}{2}} \log \epsilon)$ just as it is in (2.30).

portions of the unstable disturbance and the SWW solution when both are regarded as disturbances to the original parallel shear flow. (This is why the fine-grained outer structure does not enter the instability problem to this order.) By contrast, the inner equation (3.7) has nonlinear terms on its right-hand side, a consequence of the fact that so far $\tilde{Q}$ has been allowed to be comparable in magnitude to $\hat{Q}$, both being dimensionless quantities of order unity. Although, as it happens, the second term on the right of (3.7) is negligible to leading order, for reasons about to be explained, the first term is by no means negligible except in the early exponentially growing stages of the disturbance evolution.

The second term on the right of (3.7) is negligible, even when the first term becomes important, for the same reason as was indicated earlier for the SWW solution. The expansion for $\tilde{\Psi}$ has the same structure as the expansion (2.38) for $\hat{\Psi}$, and the matching condition on the leading term $\tilde{\Psi}_0(x, Y, T)$ dictates, in essentially the same way as before, that this term is a function of x and T only:

$$\tilde{\Psi}_0(x, Y, T) = \tilde{\Psi}_0(x, T). \tag{3.9}$$

This is the only significant contribution to (the advective effects of) the disturbance stream function $\tilde{\Psi}$, both on the left and on the right of (3.7), in precise analogy to the situation with $\hat{\Psi}$ already explained.

The forms of (3.7) and (3.8) (together with the fact that growing solutions consistent with the implied scalings will indeed be found) imply that the nonlinear saturation of growing disturbances will take place in the inner region, as was asserted in §1.2, and that the disturbance will indeed reach amplitudes such that $\tilde{Q}$ is comparable to unity, at least in an order-of-magnitude sense. Moreover, numerical evaluation of the first nonlinear term on the right of (3.7), using the fastest-growing disturbances found from linear theory (§4), strongly indicates that nonlinearity cannot stop $\tilde{Q}$ from growing to amplitudes numerically $\gtrsim 1$, at least in some cases. This implies that substantial rearrangement of the SWW vorticity patterns shown in figure 2 must take place, on the timescale δq^{-1}. Details of these estimates are omitted, since they have been superseded (and their implications confirmed) by the quantitative calculations of nonlinear saturation amplitudes given in Haynes (1985). Note that, if we were to assume (naively) that the effect of the rearrangement is equivalent to that of an eddy viscosity compatible with the foregoing scaling relations, then its value would greatly exceed the values required to make the SWW solution into a viscous critical layer. This follows from the fact that fluid elements travel across the critical layer in a time of order δq^{-1}, which as already emphasized, is much faster than the time on which the SWW solution evolves.

We now turn from order-of-magnitude considerations to an explicit consideration of the disturbance structure.

Since (3.7) and (3.8) have x-independent (albeit $\hat{x}$-dependent) coefficients, the problem is locally (in x) a parallel-flow instability problem. It is therefore natural to represent the local disturbance structure by means of Fourier integrals. Thus, in the outer region, we assume that the leading-order disturbance is locally represented by

$$\tilde{q}(x, y, T) = \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{\mathrm{i}kx}\, \hat{q}(k, y, T), \tag{3.10a}$$

and

$$\tilde{\phi}(x, y, T) = \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{\mathrm{i}kx}\, \hat{\phi}(k, y, T). \tag{3.10b}$$

The corresponding transforms $\hat{Q}(k, y, T)$ and $\hat{\Psi}_0(k, T)$ will be used to represent $\tilde{Q}(x, Y, T)$ and $\tilde{\Psi}_0(x, T)$ in the inner region. From (3.8), $\hat{\phi}$ satisfies

$$\hat{\phi}_{yy} + \left(\frac{1}{y} - k^2\right)\hat{\phi} = 0. \tag{3.11}$$

In the absence of the boundary at $y = y_b$, an appropriate ($|y|$-evanescent) solution to (3.11) would be

$$\hat{\phi} = \begin{cases} \hat{C}(k,T)\,\Gamma\left(\frac{-1}{2|k|}+1\right) U\left(\frac{-1}{2|k|}, 0, 2|ky|\right) e^{-|ky|} & (y > 0), \qquad (3.12a) \\ \hat{C}(k,T)\,\Gamma\left(\frac{1}{2|k|}+1\right) U\left(\frac{1}{2|k|}, 0, 2|ky|\right) e^{-|ky|} & (y < 0), \qquad (3.12b) \end{cases}$$

where $\Gamma(\,.\,)$ denotes the gamma function and $U(\,.\,,\,.\,,\,.\,)$ the second confluent hypergeometric function, in the notation of Abramowitz & Stegun (1965). The expressions (3.12a, b) decrease exponentially towards zero as $y \to \pm\infty$. To satisfy the boundary condition $\hat{\phi} = 0$ at $y = y_b < \infty$ we must add to (3.12a) a term proportional to

$$\lim_{\sigma \to 0} \{\Gamma(\sigma)\}^{-1} M\{-(2|k|)^{-1}, \sigma, 2|ky|\},$$

which cancels the contribution from U at $y = y_b$. Here M denotes the first confluent hypergeometric function. However, we may neglect the M-term for present purposes since it was found to change growth rates for the fastest-growing disturbances by only a few percent at most, in the cases considered, even when y_b takes its smallest possible value in (2.24). For the more distant boundary positions the effect is far smaller still, indeed utterly negligible *vis-à-vis* the numerical accuracy of the growth-rate computations to be presented in the next section. The reason is the exponentially decreasing behaviour of (3.12a) as y increases towards y_b, and the exponentially decreasing behaviour of M as y decreases back towards 0.

The coefficients in (3.12a, b) have been chosen to make $\hat{\phi}$ continuous across $y = 0$, as was done in (2.16). This is necessary to match with (3.9).

For small y, (3.12) reduces to

$$\hat{\phi} = \begin{cases} \hat{C}(k,T)\,(1 - y\log y + a_> y + O(y^2 \log y)) & (y > 0), \qquad (3.13a) \\ \hat{C}(k,T)\,(1 + |y|\log|y| + a_< |y| + O(|y|^2 \log|y|)) & (y < 0), \qquad (3.13b) \end{cases}$$

where the coefficients

$$a_> = |k| - \log(2|k|) - \left\{\nu\left(\frac{-1}{2|k|}\right) + 2\gamma - 1\right\},$$

$$a_< = |k| + \log(2|k|) + \left\{\nu\left(\frac{1}{2|k|}\right) + 2\gamma - 1\right\}.$$

See e.g. (13.1.6) and (13.4.17) of Abramowitz & Stegun (1965). Here γ is Euler's constant as before, and

$$\nu(\ ) = \Gamma'(\ )/\Gamma(\ ), \tag{3.14}$$

the logarithmic derivative of the gamma function, more usually denoted by ψ. Note that to match with (3.9) we have

$$\hat{C}(k, T) = \hat{\Psi}_0(k, T), \tag{3.15}$$

where $\hat{\Psi}_0(k, T)$ is the Fourier transform of $\Psi_0(x, T)$. Furthermore, (3.13) shows that the Cauchy principal value of the jump in the y-derivative $\hat{\phi}_y$ across $y = 0$ is

$$[\hat{\phi}_y] = \hat{C}(k,T)\,(a_> + a_<) = \hat{C}(k,T)\left\{2|k| - \nu\left(\frac{-1}{2|k|}\right) + \nu\left(\frac{1}{2|k|}\right)\right\},$$

$$= -\pi\hat{C}(k,T)\cot\left(\frac{\pi}{2|k|}\right). \tag{3.16}$$

The last step uses the fact that

$$\nu(p) - \nu(-p) + \frac{1}{p} = -\pi\cot(p\pi),$$

from (6.3.5) and (6.3.7) of Abramowitz & Stegun (1965), or from logarithmic differentiation of the well-known formula $\Gamma(p)\,\Gamma(-p) = -p^{-1}\pi\,\mathrm{cosec}\,(p\pi)$. It follows that, in order for $\hat{\phi}$ to match the corresponding Fourier component of the inner solution, the latter must satisfy

$$\int_{-\infty}^{\infty} \hat{Q}(k, Y, T)\,\mathrm{d}Y = -\pi\hat{C}(k,T)\cot\left(\frac{\pi}{2|k|}\right). \tag{3.17}$$

As before, this result can be verified from the matching rules, in the same way as (2.34).

The problem has now been reduced to solving the Fourier transform of

$$\left(\frac{\partial}{\partial T} + Y\frac{\partial}{\partial x}\right)\tilde{Q} + \tilde{\Psi}_{0x}\frac{\partial}{\partial Y}(Y + \tilde{Q}_1) = -\tilde{\Psi}_{0x}\tilde{Q}_Y, \tag{3.18}$$

the relevant simplification of (3.7), with $\hat{\Psi}_0(k, T)$ given by (3.15) and (3.17). In the next section we establish that this problem admits exponentially growing solutions when the right-hand side of (3.18) is neglected, and calculate some representative growth rates. The work of Haynes (1985) extends these results to the fully nonlinear problem in which the right-hand side of (3.18) is not neglected. For the occurrence of instability it is necessary that

$$\hat{T} > \hat{T}_1 = \tfrac{1}{2}\pi \tag{3.19}$$

(the right-hand side being the dimensionless counterpart of t_1 in figure 1), since otherwise the basic (SWW) absolute vorticity gradient $(Y + \hat{Q}_1)_Y$ appearing in (3.18) is monotonic. In that case the analogue of Rayleigh's inflection-point theorem rules out unstable solutions of the type considered here, the situation shown in figure 2(*a*) being a case in point. We note in addition that Haynes (1985) uses a variant of the Tollmien–Lin argument to prove that (3.19) is also a *sufficient* condition for instability when the boundary is at a sufficiently large distance y_b.

4. Linear instability and maximum growth rate

Neglecting the right-hand side of (3.18) we obtain a linear problem, so that attention can be restricted to a single Fourier component. Taking $\hat{Q}(k, T)$ and $\hat{C}(k, T)$ both proportional to $\exp(-ikcT)$, substituting this into (3.15) and into the Fourier transform of (3.18) with the right-hand side neglected, and using (3.17), we get

$$\int_{-\infty}^{\infty} \frac{\partial(Y + \hat{Q}_1)}{\partial Y}\frac{\mathrm{d}Y}{Y - c} = \pi\cot\left(\frac{\pi}{2|k|}\right). \tag{4.1}$$

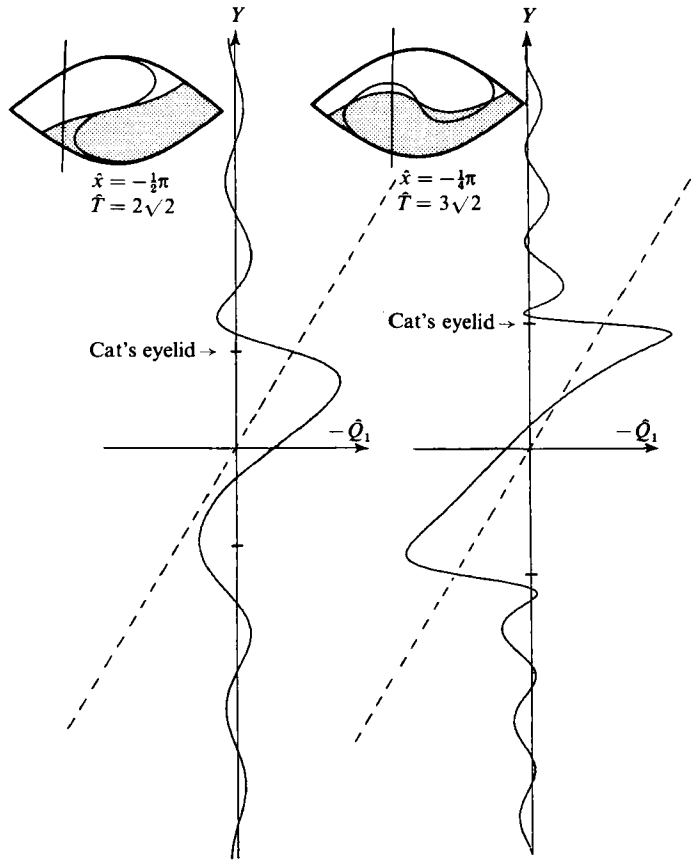

FIGURE 4. Two sample profiles across the critical layer (solid curves) of minus the basic (SWW) relative vorticity, $-\hat{Q}_1$, calculated from Stewartson's elliptic-function solution, and plotted on the same scales as the undisturbed absolute-vorticity profile (dashed lines). The accompanying sketches of the overall cat's-eye configuration show where the cross-sections are taken. The tick marks on the Y-axes show the positions of the bounding streamline or 'cat's eyelid' on each section. Note the reversed gradients of $Y+\hat{Q}_1$ within the cat's eyes. The times correspond to those of figures 2(*c*), (*d*). The accompanying sketches also show schematically the absolute vorticity contour through the centre of the cat's eye (for the precise shapes see figures 1*c*, *d*), while the shading indicates the main regions of positive $\hat{Q}_1$ contributing to the Cauchy-principal-value integral in (2.34). The right-hand sketch corresponds to near-perfect cancellation in the first Fourier component, (2.35) with $n = 1$, of that integral, which in this problem implies a state of near-perfect reflection.

The mathematical problem posed by (3.15), (3.17) and the linearized version of (3.18) has thus been reduced to solving the transcendental equation (4.1) for the unknown eigenvalue

$$c = \mathrm{Re}\, c + \mathrm{i}\, \mathrm{Im}\, c = c_\mathrm{r} + \mathrm{i}c_\mathrm{i},$$

say.

Typical examples of the Y-dependence of $-\hat{Q}_1$ are shown as the continuous curves in figure 4. The sloping dashed lines show the undisturbed absolute vorticity gradient, whose dimensionless value is unity; the existence of regions where $-\partial\hat{Q}_1/\partial Y > 1$ implies changes in the sign of the absolute gradient $(Y+Q_1)_Y$. It is easy to see graphically that such profiles admit eigensolutions with complex values of the phase velocity c, implying exponential instability. For instance one may note the qualitative

$\hat{T} = 1.5\sqrt{2}$:							
$\hat{x}$	$-\frac{3}{4}\pi$	$-\frac{1}{2}\pi$	$-\frac{1}{4}\pi$	0	$\frac{1}{4}\pi$	$\frac{1}{2}\pi$	$\frac{3}{4}\pi$
k	1.61	1.53	1.40	1.25	1.09	0.95	0.84
c_r	−0.67	−0.13	0.20	0.35	0.43	0.58	0.91
c_i	0.154	0.273	0.382	0.435	0.398	0.290	0.163
kc_i	0.247	0.418	0.536	0.542	0.434	0.275	0.137
$\hat{T} = 2\sqrt{2}$:							
$\hat{x}$							
k	2.52	2.20	1.87	1.59	1.34	1.11	0.92
c_r	−0.06	0.43	0.66	0.63	0.48	0.41	0.60
c_i	0.245	0.379	0.525	0.610	0.586	0.459	0.305
kc_i	0.618	0.835	0.982	0.967	0.783	0.508	0.282
$\hat{T} = 3\sqrt{2}$:							
$\hat{x}$							
k	6.43	4.07	3.18	2.78	1.62	1.54	1.16
c_r	0.50	1.02	1.30	1.27	0.45	0.21	0.18
c_i	0.139	0.280	0.392	0.400	0.576	0.518	0.293
kc_i	0.891	1.139	1.244	1.112	0.931	0.798	0.339

TABLE 1. Maximum growth rates kc_i at various stages of development of the SWW solution ($\hat{T}/\sqrt{2} = 1.5, 2, 3$, corresponding to figures 2*b*, *c*, *d* respectively), and at various positions along the SWW cat's eye. Note that $\hat{x} = 0$ is at the centre of the cat's eye and that $\hat{x} = \pm\pi$ are at its corners.

shapes of the graphs against Y of the real and imaginary parts of the factor $(Y-c)^{-1}$ in the integrand of (4.1). The real and imaginary parts are respectively odd and even functions of $Y-c_r$, whose shapes are invariant apart from shifts of origin (as c_r is varied) and of scale (as c_i is varied). If the Y-scale is narrow enough (c_i not too large), then it is easy to see that the origin may be shifted so as to make the imaginary part of the integral in (4.1) vanish, by taking the point $Y = c_r$ somewhere near the location of a sign change in $(Y+\hat{Q}_1)_Y$. The real part of the integral will then agree with the real-valued function on the right-hand side of (4.1) for some $|k|$, since the right-hand side runs through all real values as $|k|$ varies. Indeed, the form of the right-hand side shows that there are an infinite number of such $|k|$'s; the largest will give the fastest growth rate $|k|c_i$. See also Haynes (1985, §2). Arguments of this kind can be developed to predict a number of interesting properties of the spectrum of instabilities, but we omit them for brevity since our main purpose is simply to establish the existence of substantial growth rates, and then to return to the question of possible implications for the evolution of the basic critical layer.

To find quantitative solutions, Stewartson's elliptic-function solution was used to generate numerical tables of $\hat{Q}_1$ for use in the left-hand side of (4.1), while the contributions to the integral for large Y were handled analytically using the asymptotic solution (2.33). Table 1 presents some typical results, confining attention to the fastest growth rates kc_i found for a given $\hat{Q}_1$ configuration. The program which produced these results was spotchecked against some 'brute force' numerical solutions of the linear instability problem, using a Taylor-series ordinary differential equation solver (and without using matched asymptotic expansions). The results have been checked yet again by the independently programmed calculations of Haynes (1985).

It is interesting that many of the dimensionless growth-rate values shown in table 1

approach or exceed unity. This implies that the quantity δq defined in (3.5) is not only a nominal scale for growth rates, but is also a reasonable guide numerically, unlike what is often found in other shear-instability problems.

Many of the associated k-values are considerably in excess of unity, showing that the fastest-growing instabilities often have a somewhat smaller radian wavelength than the nominal scale Λ/β (itself much smaller than the SWW lengthscale $\hat{k}^{-1}$). The fact that these fastest-growing k's are not too small numerically, together with the known asymptotic behaviour of the second confluent hypergeometric function, helps to account for the fact, mentioned earlier, that the results are hardly changed by the presence of a boundary even at the closest possible position allowed by (2.24). Even in the latter case, it was found that most of the numbers in table 1 do not change at all, the only exceptions being the smallest values of k, for which the disturbance structure reaches furthest into the outer flow. In those cases, growth rates go down slightly (in a few entries at the top right of table 1), by amounts of the order of a few percent.

The demonstration that maximum growth rates have substantial values is sufficient for our purposes. However, this does not mean that modes having different values of k could not be important physically. For example, smaller values of k may have slower growth rates but may, on the other hand, reach larger saturation amplitudes, or lead to substantial vorticity rearrangement in a different Y-interval. They could thus be important in modifying the evolution of the critical layer. Evidence that this may indeed be the case is provided by the fully nonlinear calculations of Haynes (1985). One of the interesting properties of (4.1) is the form of the right-hand side, which implies, as already mentioned, that for any given eigensolution there will be an infinite sequence of further eigensolutions with smaller values of $|k|$ and the same value of c. A finite number of these may be consistent with the scale assumptions and boundary constraints. (The smaller the value of $|k|$, the more important the boundary constraint and the less justified the neglect of the M contribution to (3.12a).) The eigensolutions in question have successively smaller growth rates, since $|k|$ is smaller and c the same. But, again, some of them might grow to larger amplitudes if boundary constraints permit.

Finally, we note that the disturbance structures implied by these unstable eigensolutions are, indeed, qualitatively like those occurring in Béland's numerical simulations of Rossby-wave critical layers, particularly the simulation reported in Béland (1976), which had the best resolution in the y-dimension. Figure 4 suggests that this would have been the most difficult dimension to resolve numerically. Béland presented a series of simulated streamline patterns, of which three are reproduced in figure 5. Note that the short-wave disturbances in Béland's figure extend well outside the cat's eyes, with wavy streamlines and imperceptible phase tilts, just as predicted by the real-valued outer solution (3.12). The waviness penetrates further out to the positive-y side (positive basic-flow velocity) than to the negative side. Our solutions have the same property, as can be seen for instance by inspection of (3.11). In Béland's cat's-eye region there are clear indications of the presence of closed streamlines on the disturbance x-scale, implying local nonlinearity. Again, this is precisely as expected from (3.18).

It should be cautioned that disturbances with this structure could conceivably arise from a different mechanism, namely resonant amplification, if boundary conditions and parameter values happened to be tuned in such a way that the appropriate harmonics of the basic Rossby wave were close to resonance (Warn & Warn 1978; Ritchie 1985). Inspection of figure 5 suggests that more than one harmonic would

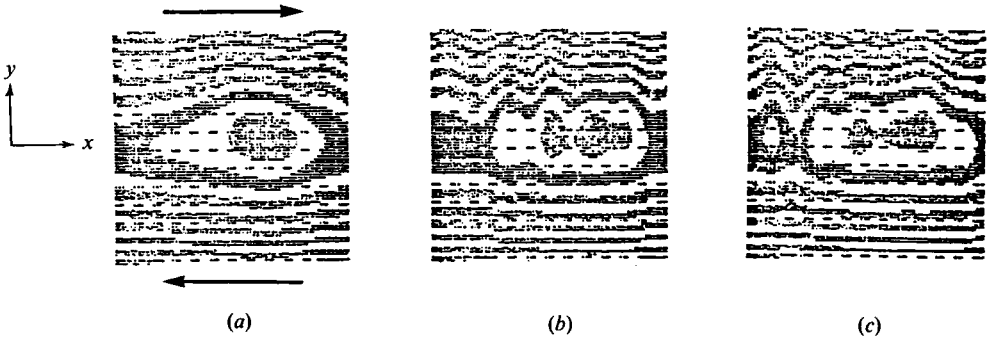

FIGURE 5. The last three panels from figure 4(*c*) of Béland (1976), showing the stream function at three successive times for a fully nonlinear numerical simulation of a Rossby-wave critical layer, the same simulation as the one referred to in figure 1. Perfect reflection was first achieved at around the time corresponding to the middle panel.

have had to be close to resonance, in this case, in order to account for the short-wave disturbance seen there. This is in contrast with the cases studied by Ritchie (1985), in which only one harmonic was amplified significantly. It would be interesting to repeat Béland's simulation with more detailed diagnostics in order to check the point. We note also that the value of μ in the simulation was 0.25 in our units. A question might therefore be raised as to whether the assumption $\mu \ll 1$ was sufficiently well satisfied and in particular whether there was sufficient time for the instability mechanism to operate in the simulation. It is not known at what finite values of μ, if any, the instability mechanism might be suppressed, but the large dimensionless growth rates found in table 1 suggest to us that such suppression would not be very likely at $\mu = 0.25$ although, again, the question merits further investigation.

5. A finite-amplitude conservation theorem for Rossby waves and other disturbances to a parallel shear flow

We now return to the question of what can be said about the consequences of the instability, recalling that in many cases where it arises, and in many other cases of interest, the possibilities are restricted by the bound (1.3) on the time-integrated absorptivity. The general proof of (1.3), to be given in §§6 and 7, depends on a finite-amplitude conservation theorem, which we establish first, in this section, on the assumption that the model equations (2.1)–(2.5) continue to hold, and then, in §7, in a more general form allowing for diffusion and other forms of vorticity transport.

There are two mathematical obstacles to a proof of (1.3) more general than that given in §1.4. The first is the one already mentioned there, and illustrated by the large-Y behaviour of (2.33). This is the fact that vorticity rearrangement is not generally confined to a central region of width comparable to the true 'mixing width' b_{m} defined by (1.13), and exemplified by the size of the bar at the centre of figure 2(*d*). Some vorticity rearrangement may go on in a far wider region, which may include the entire *outer* region, as implied at the end of §2. The typical magnitude of the vorticity fluctuations may die off as slowly as $O(y^{-1})$, with disastrous results for simple estimates of the right-hand side of (1.7). The second and related difficulty is that, even if there were no permanent rearrangement of vorticity outside the central

region, we would still want to be able to deal with an x-dependent state in which the contours of constant absolute vorticity q, i.e. material contours, undulate in some manner throughout a relatively large region. The simple argument of §1.4 is applicable only if such undulations have died away altogether outside the mixing region, and so can at best apply only at $t = \infty$, and then only in the case where the incident Rossby wave has died away to zero after a certain time.

A way round these obstacles is provided by the use of the conservation theorem to be proved here. The essential idea can be quickly appreciated by recalling some well-known properties of the model equations (2.1)–(2.5). We revert to dimensional variables since the results to be obtained do not depend on taking any particular parameter limits, such as the small-ϵ limit involved in the matched-asymptotic theory. As in §1.4, an overbar will denote the Eulerian mean in the x-direction. This is assumed to be well defined so that

$$\frac{\partial}{\partial x}\overline{(\ )} = \overline{\frac{\partial}{\partial x}(\ )} = 0 \tag{5.1}$$

for all quantities ( ) of interest. Subscript zero will denote the initial fields, which are functions of y alone. Thus $q_0(y)$ will denote, as before, the initial absolute vorticity field, and $u_0(y)$ the initial shear flow. Departures from the Eulerian mean will be denoted by primes, as before, and departures from the initial fields ('excess values') will be denoted by subscripts 'e'. Thus, for instance,

$$q(x,y,t) = \bar{q}(y,t) + q'(x,y,t) \tag{5.2}$$

$$= q_0(y) + q_e(x,y,t). \tag{5.3}$$

Now consider the following two well-known consequences of (2.1)–(2.5). First, equations (2.3)–(2.5) imply the exact identity

$$\frac{\partial}{\partial y}(\overline{u'v'}) = -\overline{v'q'}, \tag{5.4}$$

as is easily verified using (5.1). Second, (2.1), $\mathrm{D}q/\mathrm{D}t = 0$, after linearization about either the initial or the Eulerian-mean state, implies the small-amplitude relation

$$\tfrac{1}{2}q_{0y}\frac{\partial}{\partial t}\overline{\eta^2} = -\overline{v'q'}. \tag{5.5}$$

Both results were noted and used by Taylor (1915). In (5.5), η is the y-displacement of a fluid element from its initial position, and q_{0y} denotes the gradient $\mathrm{d}q_0/\mathrm{d}y$ of the initial vorticity profile $q_0(y)$. Equations (5.4) and (5.5) imply the x-averaged conservation relation

$$\frac{\partial}{\partial t}(\tfrac{1}{2}q_{0y}\overline{\eta^2}) - \frac{\partial}{\partial y}(\overline{u'v'}) = 0 \tag{5.6}$$

for inviscid, small-amplitude disturbances. Terms of the third order or higher in disturbance amplitude have been neglected. This conservation relation is a special case of what may be called a 'generalized Eliassen–Palm relation' (e.g. Edmon *et al.* 1980, and references). If we integrate it over any prescribed region

$$y_1 \leqslant y \leqslant y_2, \tag{5.7}$$

and over time from 0 to t, we obtain the following (approximate) expression for the time-integrated absorptivity $\alpha(t)$:

$$\alpha(t) = \int_0^t [\overline{u'v'}]_{y_1}^{y_2} \,\mathrm{d}t = \int_{y_1}^{y_2} \tfrac{1}{2} q_{0y} \overline{\eta^2} \,\mathrm{d}y. \tag{5.8}$$

The right-hand side is bounded whenever η and q_{0y} are bounded in the region $y_1 < y < y_2$. Moreover, it is convergent for large y_1, y_2. For the SWW solution, η^2 diminishes away from the central region like $|y|^{-2}$, and $q_{0y} = \beta =$ constant, which is bounded. For any similar problem in which the incident Rossby wave is switched on less abruptly we expect η^2 to be smaller still, if anything, for large $|y|$. So we expect no trouble at the periphery of the y-integration, no matter how large the interval (y_1, y_2). There is no such simple way of bounding the integral on the right of (1.7) when its limits of integration are large.

To turn this idea into a rigorous proof, we need an exact, finite-amplitude version of the generalized Eliassen–Palm relation (5.6). This can be obtained, for the non-diffusive case, if we define η appropriately and work at first with q_e, u_e, etc. instead of q', u', etc. The mathematical device used is the same as that used by Holliday & McIntyre (1981) to obtain an exact, positive-definite expression for 'available potential energy' in an incompressible, stably stratified fluid, but its application to Rossby waves appears to be new. Its use was suggested by the form of the expression $\frac{1}{2} q_{0y} \overline{\eta^2}$, which is the same as the small-amplitude expression for available potential energy when y is vertical and $q_0(y)$ represents the basic density stratification.

Consider the fluid element which finds itself at (x, y) at time t. Let the y-coordinate of this element in the original, undisturbed shear flow be denoted by $y_0(x, y, t)$. Then η is defined by

$$y = y_0(x, y, t) + \eta(x, y, t). \tag{5.9}$$

That is, η is the y-displacement of a fluid element in the usual sense, except that it is expressed as a function of the present rather than the initial position of the fluid element.† We then have

$$v = v_e = \frac{\mathrm{D}y}{\mathrm{D}t} = \frac{\mathrm{D}(y - y_0)}{\mathrm{D}t} = \frac{\mathrm{D}\eta}{\mathrm{D}t}, \tag{5.10}$$

where as before $\mathrm{D}/\mathrm{D}t = \partial/\partial t + u\,\partial/\partial t + v\,\partial/\partial y$. Also

$$q(x, y, t) = q_0\{y_0(x, y, t)\} = q_0\{y - \eta(x, y, t)\}, \tag{5.11}$$

because of the fact that $\mathrm{D}q/\mathrm{D}t = 0$. Now define

$$A(y, \eta) = \int_0^\eta \hat{\eta} q_{0y}(y - \hat{\eta}) \,\mathrm{d}\hat{\eta}. \tag{5.12}$$

$A(y, \eta)$ is a known function of y and η, for any given initial vorticity profile $q_0(y)$. It reduces to $\frac{1}{2} q_{0y} \eta^2$ when η is sufficiently small and $q_0(y)$ is a smooth function. It has the property that

$$\left(\frac{\partial A(y, y - y_0)}{\partial y}\right)_{y_0 = \text{constant}} = \left(\frac{\partial A(y, \eta)}{\partial y}\right)_\eta + \left(\frac{\partial A(y, \eta)}{\partial \eta}\right)_y = -q_0(y_0) + q_0(y), \tag{5.13}$$

† P. B. Rhines has suggested to us that the two functional forms be distinguished from each other by calling them 'arrival' and 'departure' displacements respectively. Thus the present $\eta(x, y, t)$ is an 'arrival displacement'. It is related to the idea of 'arrival diffusivity' used for instance by Haidvogel & Rhines (1983).

as is easily verified. The right-hand side is equal to $-q_e$, by (5.3) and (5.11). Hence, from this and (5.10),

$$\frac{D}{Dt} A(y, \eta) = -v_e q_e, \tag{5.14}$$

the x-average of which may be compared to (5.5). Although it will not be used here, we note for completeness the alternative functional form

$$A = A(q_0, q_e) = \int_0^{q_e} \{\mathscr{Y}(q_0 + \hat{q}_e) - \mathscr{Y}(q_0)\} \, d\hat{q}_e \tag{5.15}$$

expressing A in terms of the values of q_e and q_0 at a given place and time, where $\mathscr{Y}(q_0)$ is the function inverse to $q_0(y)$, assumed monotonic. This purely Eulerian form is possible, as Holliday & McIntyre point out, because by (5.11) q contains all the Lagrangian information relevant to (5.12), when $q_0(y)$ is monotonic and $Dq/Dt = 0$. This provides an alternative route to (5.14), as is easily checked. It is emphasized, however, that since we shall use the form (5.12) we shall not need to assume that $q_0(y)$ is monotonic.

The right-hand side of (5.14) suggests that we use the following analogue of (5.4):

$$\frac{\partial}{\partial y}(u_e v_e) + \frac{\partial}{\partial x}(\tfrac{1}{2}u_e^2 - \tfrac{1}{2}v_e^2) = -v_e q_e. \tag{5.16}$$

This follows from equations (2.3)–(2.5) as before, but this time with no averaging in x. Putting (5.14) and (5.16) together, expanding D/Dt, and again using (2.5), $\partial u/\partial x + \partial v/\partial y = 0$, we get the exact generalized Eliassen–Palm relation in the form

$$\boxed{\frac{\partial}{\partial t} A(y, \eta) + \frac{\partial}{\partial x}\{-\tfrac{1}{2}u_e^2 + \tfrac{1}{2}v_e^2 + (u_0 + u_e) A\} + \frac{\partial}{\partial y}\{-u_e v_e + v_e A\} = 0.} \tag{5.17}$$

In the following section we show how this conservation relation can be used to bound $\alpha(t)$ under the conditions outlined in §1.3. Only the x-averaged form will actually be needed; this is simply

$$\frac{\partial}{\partial t} \overline{A(y, \eta)} + \frac{\partial}{\partial y}\{-\overline{u_e v_e} + \overline{v_e A}\} = 0. \tag{5.18}$$

We note incidentally from (5.12) and (5.15) that the small-amplitude formulae $A \simeq \frac{1}{2} q_{0y} \eta^2 \simeq \frac{1}{2} q_{0y}^{-1} q_e^2$ are exact in the special case $q_{0y} = \beta =$ constant.†

6. Proof of the bound (1.3) in the non-diffusive case

We now choose, as before, some fixed interval

$$y_1 \leqslant y \leqslant y_2 \tag{6.1}$$

in terms of which to define the time-integrated absorptivity

$$\alpha(t) = \int_0^t [\overline{u'v'}]_{y_1}^{y_2} \, dt. \tag{6.2}$$

This will usually be taken to enclose the central critical-layer region by some suitable margin. The best choice will depend on the problem of interest. In particular, it will depend on the amplitude of the incident Rossby wave, which determines the width of the central region within which substantial rearrangement of vorticity takes place.

† *Note added in proof:* This case has previously been noted by Rhines (1977). It has also come to light that the general finite-amplitude expression (5.15) is implicit in an important but largely overlooked paper by Arnol'd (1966) on nonlinear hydrodynamical stability. The connexion will be discussed in a forthcoming paper by McIntyre & Shepherd (1986), together with the interpretation of expressions like (5.15) when $\mathscr{Y}(q_0)$ is not monotonic.

As mentioned earlier, that width may be a sizeable fraction of the whole flow domain in some large-amplitude cases of meteorological interest.

The assumptions needed in order to bound (6.2) are that the incident Rossby wave have bounded amplitude, and that conditions (i)–(v) of §1.3 hold good. The meaning of 'bounded amplitude' can now be made precise in terms of the displacement function $\eta(x, y, t)$ defined in (5.9). The essential requirement is that, within the fixed interval (y_1, y_2), the displacement $\eta(x, y, t)$ be uniformly bounded for all t, i.e. that

$$|\eta(x, y, t)| \leqslant B \quad (y_1 \leqslant y \leqslant y_2) \tag{6.3}$$

for some constant B. For amplitudes that are not too large, a stronger assumption of the type

$$|\eta(x, y, t)| \leqslant B \quad (-\tfrac{1}{2}B \leqslant y \leqslant \tfrac{1}{2}B) \tag{6.4a}$$

together with

$$|\eta(x, y, t)| \leqslant \frac{B^2}{4|y|} \quad (y_1 \leqslant y \leqslant -\tfrac{1}{2}B \text{ or } \tfrac{1}{2}B \leqslant y \leqslant y_2) \tag{6.4b}$$

may be appropriate. This will lead to a sharper bound than (6.3). The stronger assumption (6.4) holds true of the SWW solution, for instance, if the boundaries of the central region $-\frac{1}{2}B \leqslant y \leqslant \frac{1}{2}B$ are taken to be tangent to the extremities of the cat's eyes, so that B is identified with the maximum width of the cat's eyes. In general we shall assume that

$$|\eta(x, y, t)| \leqslant \eta_c(y) \quad (y_1 \leqslant y \leqslant y_2) \tag{6.5}$$

for some fixed, bounded function $\eta_c(y)$. We shall also assume that the associated values of q, namely $q_0(y-\eta)$, see (5.11), are themselves bounded, for $y_1 \leqslant y \leqslant y_2$.

This latter assumption, together with (6.5), means that the factor $\mathrm{d}q_0/\mathrm{d}y$ in the integrand of (5.12) is either bounded or contains delta functions of finite strength. We exclude q_{0y} profiles more pathological than that: to be precise, we allow no more than a finite number of delta functions. The function $A(y, \eta)$ defined by (5.12) is then self-evidently bounded, under (6.5).

Conditions (i) and (ii) of §1.3 are already expressed by the model equations and require no comment. Conditions (iv) and (v) are satisfied trivially in the present, non-diffusive case (provided that the initial profile $q_0(y)$ is finite valued, as is implied by the assumption just made). Condition (iii) states that 'the critical layer always consists of the same material fluid elements'. This implies that there is a frame of reference in which the critical layer does not drift sideways and in which there is no net mass flux across a line $y =$ constant. We shall adopt this frame of reference so that, in particular,

$$\bar{v} = v_0 = 0 \tag{6.6}$$

and

$$v' = v_e = v. \tag{6.7}$$

Consequently

$$\overline{u'v'} = \overline{u_e v_e}, \tag{6.8}$$

and we can use (5.18). Note that the assumption that there is no net mass flux across a line $y =$ constant is consistent with the assumption (6.5) that $|\eta(x, y, t)|$ is uniformly bounded for all t. Since η has been defined as the displacement from the initial y-position of a fluid element, and not from any kind of mean position, η would include any mean sideways drift that occurred.

The construction of bounds on $\alpha(t)$ now proceeds in almost the same way as was suggested by (5.6) and (5.8). From (5.18), (6.2), (6.7) and (6.8) we have

$$\alpha(t) = \int_0^t [\overline{u'v'}]_{y_1}^{y_2} \mathrm{d}t = \int_{y_1}^{y_2} \mathrm{d}y\, \overline{A(y, \eta)} + \int_0^t \mathrm{d}t\, [\overline{vA}]_{y_1}^{y_2}. \tag{6.9}$$

This is the exact result corresponding to (5.8). Just as with the right-hand side of (5.8), it is straightforward to bound $\overline{A(y,\eta)}$ given our assumptions, as already indicated. The only remaining problem is the extra term

$$-\int_0^t \mathrm{d}t\,\overline{vA(y,\eta)}\Big|_{y=y_1} = -\lim_{L\to\infty}\frac{1}{2L}\int_{-L}^{L}\mathrm{d}x\int_0^t \mathrm{d}t\,vA(y,\eta)\Big|_{y=y_1}, \tag{6.10}$$

and its counterpart for $y = y_2$, appearing on the right of (6.9). It is crucial to be able to bound these terms uniformly for all t, despite the arbitrarily long integration time.

We can do this by using the fact that A, by definition, is a function of y and η alone, together with incompressibility and the boundedness of $\eta(x, y_1, t)$ and $\eta(x, y_2, t)$ implied by (6.3), (6.4) or (6.5). We have

$$|\eta(x,y_1,t)| \leqslant \eta_1, \quad |\eta(x,y_2,t)| \leqslant \eta_2, \tag{6.11}$$

say, where η_1 and η_2 are constants, respectively $\eta_c(y_1)$ and $\eta_c(y_2)$ if (6.5) is assumed, $B^2/4|y_1|$ and $B^2/4y_2$ if (6.4) is assumed, or both equal to B if (6.3) is assumed. In view of the definition of η, this says that any fluid element crossing the line $y = y_1$, say, must have originated from somewhere in the interval

$$y_1 - \eta_1 \leqslant y \leqslant y_1 + \eta_1, \tag{6.12}$$

and similarly for an element crossing the line $y = y_2$. Now, if a given fluid element having area $\mathrm{d}m$ and initial position y_0 crosses the line $y = y_1$, it will make a contribution

$$A(y,\eta)\,\mathrm{d}m \tag{6.13}$$

to the double integral on the right of (6.10), with y evaluated as y_1 and η as $y_1 - y_0$. Moreover, if the same element subsequently re-crosses the line $y = y_1$ in the other direction, there will be a further contribution to the double integral that exactly cancels the previous contribution. Both y and η take exactly the same values, respectively y_1 and $y_1 - y_0$, on each occasion when the fluid element crosses or re-crosses the line $y = y_1$. Thus no matter how large t becomes, each fluid element can never make more than *one* contribution $A(y,\eta)\,\mathrm{d}m$ to the double integral on the right of (6.10).

Now consider the total area $\int \mathrm{d}m$ occupied by those fluid elements which were initially in the interval (6.12) and which might therefore contribute to (6.10). This area is $2\eta_1$ per unit x-distance. It follows that

$$\left|\int_0^t \mathrm{d}t\,\overline{vA(y,\eta)}\right|_{y=y_1} \leqslant 2\eta_1 \sup_{|\eta|\leqslant\eta_1} |A(y_1,\eta)|, \tag{6.14}$$

and similarly for $y = y_2$. The right-hand side represents a finite *a priori* bound. The supremum is taken over values of the function $A(y,\eta)$, and that function is (*a*) known *a priori*, for any given initial profile $q_0(y)$, and (*b*) takes finite values, under the assumptions made earlier in the paragraph containing (6.5).

We can now obtain bounds on $\alpha(t)$ itself, using (6.9) and (6.14) together with (6.3), (6.4) or (6.5). From (6.5), for instance, we immediately obtain a bound in the general form

$$|\alpha(t)| \leqslant \alpha_{\max} \equiv \int_{y_1}^{y_2}\mathrm{d}y \sup_{|\eta|\leqslant\eta_c(y)} |A(y,\eta)| + 2\eta_c(y_1)\sup_{|\eta|\leqslant\eta_c(y_1)}|A(y_1,\eta)| + 2\eta_c(y_2)\sup_{|\eta|\leqslant\eta_c(y_2)}|A(y_2,\eta)|. \tag{6.15}$$

Simpler expressions representing less sharp bounds can easily be deduced at need, depending on the nature of the $q_0(y)$ profile and on how much more information we are prepared to throw away. For example, if the initial gradient

$$|q_{0y}| \leqslant K \quad (y_1 \leqslant y \leqslant y_2), \tag{6.16}$$

where K is a constant, then from (5.12) we have

$$|A(y,\eta)| \leqslant \tfrac{1}{2}K\eta^2. \tag{6.17}$$

If, further, (6.4) is assumed, then substitution into (6.15) gives

$$|\alpha(t)| < \tfrac{1}{2}KB^3\left\{\frac{5}{4}-\frac{1}{16}\frac{B}{|y_1|}-\frac{1}{16}\frac{B}{y_2}+\frac{1}{32}\left(\frac{B}{|y_1|}\right)^3+\frac{1}{32}\left(\frac{B}{y_2}\right)^3\right\}. \tag{6.18}$$

Note in particular that the bound, and therefore $\alpha(t)$ itself, is *zero* if the bound K on q_{0y} is zero. That is, if no (potential) vorticity gradient exists in the critical layer, and if all the relevant conditions of §1.3 are satisfied, then, as already mentioned in §1, the critical layer is a perfect reflector for all time t.

The same conclusion can be reached from (2.34)–(2.36), in cases where the standard (matched-asymptotic) critical-layer theory applies, since $\hat{Q}_1 + Y$ would be independent of x if q_{0y} were zero throughout the inner region.

It will be noticed that the simple bound (6.18) is much less sharp than (1.12). The same is true of (6.15). This is mainly because, for the sake of simplicity, we have thrown away information about the kinematically possible distributions of η within the critical layer. Those distributions are restricted by incompressibility, which implies in particular that not all the fluid elements can have their maximum possible values of η simultaneously. However, although refinements taking this into account are clearly possible we shall not pursue them here, both for the sake of brevity, and also because a truly sharp bound is not to be expected in any case, for the reasons explained below (1.10).

7. The finite-amplitude conservation theorem and the bound (1.3) in the 'diffusive' case

We now take up the wider implications of conditions (iv) and (v) of §1.3. Condition (v) says that q has a bounded range of values, for all t, in the central region of the critical layer where non-advective transport may take place. The way in which this enters the heuristic argument given in §1.4 suggests that the bound (1.3) does not really depend on having no non-advective transport.

The same conclusion can be reached from the following consideration, which also serves to motivate the subsequent rigorous proof. For the purpose of evaluating an integral in which q appears linearly, like the right-hand side of (1.7), downgradient transport processes like diffusion can be thought of as equivalent to rearrangement of q by a hypothetical, 'fine-grain' velocity field varying on very small lengthscales. (In the case of ordinary diffusion, the fine-grain motion may be modelled in terms of a random walk, and there are many other possibilities.) If the y-excursions involved in the hypothesized fine-grain motion are bounded, as they would generally have to be under condition (v), then A can evidently be bounded in just the same way as before. We may simply regard the displacement field η as including the fine-grain motion; it is still bounded and, therefore, so is $|\alpha(t)|$.

Although the foregoing consideration makes the diffusive version of the result

intuitively plausible, and could probably be made into a rigorous proof, it is not straightforward to express mathematically since the functions describing the arbitrarily fine-grained velocity and displacement fields would be of an extremely pathological kind. While this difficulty seems to be no more than a technical one, it would nevertheless seem desirable to possess a proof in which the use of such functions is not resorted to. It is, in fact, possible to recast the problem so that only mathematically well-behaved functions are used. That is the purpose of the present section. A bonus will be that the conditions under which a conservation relation like (5.17) can be obtained are widened still further, so as to include the possibility of an arbitrary time-integrated vorticity transport, whether downgradient, upgradient, or unrelated to the gradient. It will be found, as expected, that the bound (1.3) holds, provided always that the transport is not so persistently upgradient as to violate condition (v) that the range of values of q stays bounded. This is related to the fact that upgradient transport can be thought of, if desired, as due to a hypothetical fine-grain rearrangement process in which a suitably chosen initial fine-grained arrangement is 'undone', so that the fine-grain spatial fluctuations in q are reduced in intensity by the hypothesized fine-grain motion.

The reader should recall at this point that some models of the mean effects of three-dimensional turbulence can, in fact, exhibit upgradient vorticity transport, for instance models in which the associated momentum transport is taken to be equivalent to the effect of a (spatially variable) eddy viscosity.† Whether a given turbulent transport model would satisfy condition (v), or not, can be answered only by a detailed consideration of that model. Regarding eddy-viscosity models, as usually understood, these imply a very special, and possibly unrealistic, relation between momentum flux and strain rate, in which momentum is everywhere transferred strictly down its gradient. As far as we are aware it is still an open question whether such a relation is actually a good model of real three-dimensional turbulence. In reality, the direction of transfer might be at a highly variable angle to the mean gradient. It could also take place in the absence of any mean gradient. There is certainly a strong presumption that downgradient momentum transfer is a bad model of real two-dimensional turbulence, such as is of interest in astrogeophysical applications (e.g. Starr 1968). There, downgradient (potential) vorticity diffusion is expected to be a better model, if anything, than downgradient momentum diffusion (e.g. Green 1970; Rhines 1979), although care must be taken, as always, to see that the model does not tacitly assume the existence of local sources of momentum.

To make sure of this last point while retaining a degree of generality commensurate with conditions (iv) and (v), and avoiding any kind of eddy-viscosity or other special assumption about turbulent stresses, we take as our starting point the momentum equation in its general two-dimensional form

$$\frac{\mathrm{D}u_i}{\mathrm{D}t} - \epsilon_{ij} f u_j = -\frac{\partial p}{\partial x_i} + \frac{\partial \sigma_{ij}}{\partial x_j} \quad (i,j = 1,2), \tag{7.1}$$

where repeated suffixes are summed and where σ_{ij} represents a stress, 'turbulent' or otherwise, whose dependence on other variables will be left arbitrary for the moment. A Coriolis term is included, with the Coriolis parameter f a function of $\boldsymbol{x}$, so that model fluid systems of astrogeophysical interest continue to be included. The two-dimensional

† A simple but sufficient example is that of a straight viscous jet surrounded by inviscid fluid. As time goes on, vorticity tends to become more and more concentrated at the edges of the viscous region.

alternating tensor ϵ_{ij} is defined by $\epsilon_{11} = \epsilon_{22} = 0$, $\epsilon_{12} = -\epsilon_{21} = 1$. The notations $(x, y) = \boldsymbol{x} = (x_1, x_2)$ will be used interchangeably. Incompressibility is assumed, as before, and (2.5) may be written

$$\boldsymbol{\nabla}\cdot\boldsymbol{u} = \frac{\partial u_j}{\partial x_j} = 0. \tag{7.2}$$

If we operate on (7.1) with $\epsilon_{ki}\,\partial/\partial x_k$ and use (7.2) we get the relevant form of the vorticity equation:

$$\frac{\mathrm{D}q}{\mathrm{D}t} + \boldsymbol{\nabla}\cdot\boldsymbol{F} = 0, \tag{7.3}$$

where

$$F_j = -\epsilon_{ki}\,\sigma_{ij,\,k}. \tag{7.4}$$

Here $(\ )_{,\,k}$ denotes $\partial(\ )/\partial x_k$. The definition of q is

$$q = f + \epsilon_{ki}\,u_{i,\,k}.$$

The fact that the second term in (7.3) takes the form of the divergence of a flux $\boldsymbol{F}$ shows that, in any model governed by an equation of the form (7.1), q is transported conservatively. That is, q cannot be created or destroyed by the effects of the stress σ_{ij}, but is always *redistributed* in some way, the integral of q over an appropriate area remaining constant.

To describe that redistribution in a sufficiently general manner, we consider the contribution to $q(\boldsymbol{x}, t)$ which originated from the strip lying between y_* and $y_* + \mathrm{d}y_*$ in the initial state. Denote the contribution in question by

$$P(\boldsymbol{x}, t; y_*)\, q_0(y_*)\, \mathrm{d}y_*. \tag{7.5}$$

Then

$$q(\boldsymbol{x}, t) = \int P(\boldsymbol{x}, t; y_*)\, q_0(y_*)\, \mathrm{d}y_*. \tag{7.6}$$

The range of integration in which $P \neq 0$ represents those parts of the initial domain from which q at $(\boldsymbol{x}, t)$ is supposed to have been transported. They will all be located within a strip of finite width, under conditions (iv) ff. of §1.3. We shall choose P such that

$$\int P(\boldsymbol{x}, t; y_*)\, \mathrm{d}y_* = 1. \tag{7.7}$$

This this is possible will be shown below. Note that in the non-diffusive case, with advective transport only, we can take

$$P(\boldsymbol{x}, t; y_*) = \delta\{y_* - y_0(\boldsymbol{x}, t)\}, \tag{7.8}$$

which reproduces (5.11) when substituted into (7.6).

The corresponding contribution to the non-advective vorticity flux $\boldsymbol{F}$ will be denoted by

$$\boldsymbol{R}(\boldsymbol{x}, t; y_*)\, q_0(y_*)\, \mathrm{d}y_*, \tag{7.9}$$

so that

$$\boldsymbol{F}(\boldsymbol{x}, t) = \int \boldsymbol{R}(\boldsymbol{x}, t; y_*)\, q_0(y_*)\, \mathrm{d}y_*, \tag{7.10}$$

where

$$\frac{\mathrm{D}}{\mathrm{D}t} P(\boldsymbol{x}, t; y_*) + \boldsymbol{\nabla}\cdot\boldsymbol{R}(\boldsymbol{x}, t; y_*) = 0 \tag{7.11}$$

for fixed y_*, and where, consistent with (7.7) and (7.11),

$$\int \boldsymbol{R}(\boldsymbol{x}, t; y_*)\, \mathrm{d}y_* = 0 \tag{7.12}$$

and
$$\boldsymbol{R} = 0 \quad \text{for all } \boldsymbol{x}, t \text{ such that } \boldsymbol{F} = 0. \tag{7.13}$$

The fact that $\boldsymbol{R}$ can be chosen to satisfy (7.12) will also be shown below.

Now consider the quantity

$$\langle A \rangle (\boldsymbol{x}, t) = \int P(\boldsymbol{x}, t; y_*)\, A(y, y - y_*)\, \mathrm{d}y_* \tag{7.14}$$

noting that in the non-diffusive case, (7.8), this reduces to

$$\langle A \rangle (\boldsymbol{x}, t) = A\{y, y - y_0(\boldsymbol{x}, t)\} = A\{y, \eta(\boldsymbol{x}, t)\}. \tag{7.15}$$

Define also

$$\langle \boldsymbol{S} \rangle (\boldsymbol{x}, t) = \int \boldsymbol{R}(\boldsymbol{x}, t; y_*)\, A(y, y - y_*)\, \mathrm{d}y_*. \tag{7.16}$$

Using (7.11), we then have

$$\frac{\mathrm{D}}{\mathrm{D}t} \langle A \rangle + \boldsymbol{\nabla} \cdot \langle \boldsymbol{S} \rangle = \int \mathrm{d}y_*\, P(\boldsymbol{x}, t; y_*) \left. \frac{\mathrm{D}A(y, y - y_*)}{Dt} \right|_{y_* \text{ constant}} + \int \mathrm{d}y_*\, R_2(\boldsymbol{x}, t; y_*) \left. \frac{\partial A(y, y - y_*)}{\partial y} \right|_{y_* \text{ constant}} \tag{7.17}$$

where R_2 is the y-component of $\boldsymbol{R}$. Noting also that, with y_* constant, $\mathrm{D}A(y, y - y_*)/\mathrm{D}t = v\, \partial A(y, y - y_*)/\partial y$, and using (5.10), (5.13) with y_* in place of y_0, (7.4) with $j = 2$, (7.6), (7.7), (7.10) and (7.12), we may rewrite (7.17) as

$$\frac{\mathrm{D}}{\mathrm{D}t} \langle A \rangle + \boldsymbol{\nabla} \cdot \langle \boldsymbol{S} \rangle = -v_\mathrm{e} q_\mathrm{e} + \frac{\partial}{\partial x_k} \{\varepsilon_{ki}\, \sigma_{i2}(\boldsymbol{x}, t)\}. \tag{7.18}$$

Recalling the identity (5.16), we deduce the 'diffusive' generalization of the finite-amplitude conservation relation (5.17):

$$\boxed{\frac{\partial}{\partial t} \langle A \rangle + \frac{\partial}{\partial x} \{-\tfrac{1}{2} u_\mathrm{e}^2 + \tfrac{1}{2} v_\mathrm{e}^2 + (u_0 + u_\mathrm{e}) \langle A \rangle + \langle S \rangle_1 - \sigma_{22}\} + \frac{\partial}{\partial y} \{-u_\mathrm{e} v_\mathrm{e} + v_\mathrm{e} \langle A \rangle + \langle S \rangle_2 + \sigma_{12}\} = 0.} \tag{7.19}$$

Now σ_{ij} and $\langle \boldsymbol{S} \rangle$ vanish outside the central region, by condition (iv) ff. of §1.3. Therefore if $y_1 + \eta_1$ and $y_2 - \eta_2$ are chosen to lie outside that region, where η_1 and η_2 are the amplitude bounds appearing in (6.11) ff., we immediately obtain results analogous to (6.9), (6.15), etc. by integrating (7.19) in place of (5.17). Note that (7.8) and (7.15) may be used when evaluating boundary terms like (6.10), or indeed anywhere outside the central region, and that the boundary-term estimate (6.14) therefore still applies. Thus we obtain (6.15), for instance, with the sole change that $\langle A \rangle (\boldsymbol{x}, t)$ replaces $A\{y, \eta(\boldsymbol{x}, t)\}$ in the first term on the right, over some range of integration covering the central region. This bounds $\alpha(t)$ just as before, provided that the quantity $\langle A \rangle$ defined by (7.14) can be bounded within the central region.

Again, one can do this to varying degrees of refinement, depending upon how much information one is prepared to throw away for the sake of getting simple-looking inequalities. A key step will be to note that any vorticity redistribution satisfying condition (v) of §1.3 can, in general, be represented in the central region by a bounded redistribution function P. Then $\langle A \rangle$ and hence $\alpha(t)$ can again be bounded by simple estimates.

For instance, suppose for the sake of definitiveness that the central region lies within the interval $-\frac{1}{2}B < y < \frac{1}{2}B$, where B is a constant. By definition, this means that all fluid elements affected by the non-advective vorticity transport remain within the interval $(-\frac{1}{2}B, \frac{1}{2}B)$ for all time t. As already indicated, we must then choose y_1 and y_2 such that

$$y_1+\eta_1 < -\tfrac{1}{2}B, \quad \tfrac{1}{2}B < y_2-\eta_2. \tag{7.20}$$

Given conditions (iii) and (iv) of §1.3, and the assumption of bounded wave amplitude (in the particle-displacement sense of §6), one can always find constants $B, y_1, y_2, \eta_1, \eta_2$ such that the foregoing statements are true. Now, for a fluid element in the central region $|y| < \frac{1}{2}B$, the range of integration with respect to y_* in (7.14) may be taken to be $(-\frac{1}{2}B, \frac{1}{2}B)$. Anticipating the boundedness of P, suppose that

$$B|P(\boldsymbol{x}, t; y_*)| \leqslant C \quad \text{for all } |y| < \tfrac{1}{2}B, |y_*| < \tfrac{1}{2}B \tag{7.21}$$

(and for all x, t), where C is a finite constant. C is dimensionless and may generally be expected to be of order unity, for consistency with (7.7). Then, from (7.14),

$$|\langle A\rangle(\boldsymbol{x}, t)| \leqslant C \sup_{|y_*|<\frac{1}{2}B} |A(y, y-y_*)| \quad \text{for all } |y| < \tfrac{1}{2}B \tag{7.22}$$

(and for all x, t). The analogue of (6.15) which now follows from (7.19) can be written

$$\begin{aligned} |\alpha(t)| \leqslant \alpha_{\max} \equiv C \int_{-\frac{1}{2}B}^{\frac{1}{2}B} \mathrm{d}y \sup_{|y_*|<\frac{1}{2}B} |A(y, y-y_*)| \\ + \int_{y_1}^{-\frac{1}{2}B} \mathrm{d}y \sup_{|\eta|\leqslant\eta_c(y)} |A(y, \eta)| + \int_{\frac{1}{2}B}^{y_2} \mathrm{d}y \sup_{|\eta|\leqslant\eta_c(y)} |A(y, \eta)| \\ + 2\eta_1 \sup_{|\eta|\leqslant\eta_1} |A(y_1, \eta)| + 2\eta_2 \sup_{|\eta|\leqslant\eta_2} |A(y_2, \eta)|, \end{aligned} \tag{7.23}$$

where $\eta_c(y)$ is the amplitude bound which applies outside the central region, as in (6.5), and η_1, η_2 may be identified with $\eta_c(y_1)$ and $\eta_c(y_2)$ as before.

It remains to verify that functions $P(\boldsymbol{x}, t; y_*)$ and $\boldsymbol{R}(\boldsymbol{x}, t; y_*)$ can indeed be chosen so as to satisfy the foregoing conditions, particularly (7.7), (7.12), and boundedness of $|P|$, for all $\boldsymbol{x}$ and y_* belonging to the central region. The possibility of being able to choose P and $\boldsymbol{R}$ in this way is immediately plausible from the fact that their functional dependence upon y_* adds an extra dimension to the space on which they are defined, as compared to the space $\{\boldsymbol{x}, t\}$ on which the functions q and $\boldsymbol{F}$ themselves are defined. Clearly this allows an enormous latitude of choice, which, indeed, we have already exercised in using (7.8) and (7.15) outside the central region (in the second and third lines of (7.23)), and (7.6) with $|P|$ bounded inside it (in the first line).† This latitude of choice reflects the fact that there is no way of distinguishing vorticity transported from one place from that transported from another, once the transport has taken place. However, it might be thought that the conditions (7.7) and (7.12), which correspond to incompressibility of the equivalent fine-grain rearrangement,

† We note that P has therefore been taken to behave discontinuously whenever a fluid element crosses $y = \frac{1}{2}B$ or $y = -\frac{1}{2}B$, and that $\boldsymbol{R}$ will have a correspondingly singular behaviour. We are at liberty to use this device, which is adopted purely for mathematical convenience, because of the conservation form of (7.19), which ensures that any such singular behaviour integrates out and does not appear explicitly in (7.23). Other choices of P could be made: for instance one could use a representation based everywhere on (7.26*a*) below, at the cost of complicating the second and third lines of (7.23).

impose a restriction on the nature of the transport, making it less general than what is implied by (7.4). We now show that, on the contrary,

(1) the representations (7.6) and (7.10) are completely general, and can always be chosen to satisfy (7.7), (7.12), and the boundedness of $|P|$, whenever $q_{0y} \neq 0$ somewhere in the central region, and

(2) the case $q_{0y} \equiv 0$ is also covered as a limiting case of (1), with consequences to be noted at the end of the section.

To verify assertion (1) it is enough to exhibit just one example of a pair of functions P and $\boldsymbol{R}$, which represent a given q and $\boldsymbol{F}$ through (7.6) and (7.10), and which satisfy the other requirements. The following example is chosen for mathematical simplicity. More sophisticated choices could be made, depending on the detailed profile of $q_0(y)$; this would be important if it were desired to make the bound (7.21) and hence (7.23) as sharp as possible. For any $q_0(y)$ profile for which $q_{0y} \neq 0$ in the central region, we can always find two disjoint intervals I_a and I_b lying within that region such that

$$\int_{\mathrm{I}_a} q_0(y_*)\,\mathrm{d}y_* = q_a\,\Delta y_*, \quad \int_{\mathrm{I}_b} q_0(y_*)\,\mathrm{d}y_* = q_b\,\Delta y_*, \tag{7.24}$$

where Δy_* is the size of each interval, the sizes being taken equal for convenience, and where the numbers q_a and q_b, representing average values of q_0 over each interval, are unequal:

$$q_a \neq q_b. \tag{7.25}$$

Given any pair of functions $q(\boldsymbol{x}, t)$ and $\boldsymbol{F}(\boldsymbol{x}, t)$ satisfying (7.3) it is clear that (7.6), (7.7), (7.10) and (7.12) will all hold if we choose

$$P(\boldsymbol{x}, t; y_*) = \begin{cases} (q-q_b)/(q_a-q_b)\,\Delta y_* & (y_* \in \mathrm{I}_a), \\ (q-q_a)/(q_b-q_a)\,\Delta y_* & (y_* \in \mathrm{I}_b), \\ 0 \quad \text{otherwise} \end{cases} \tag{7.26a}$$

and

$$\boldsymbol{R}(\boldsymbol{x}, t; y_*) = \begin{cases} \boldsymbol{F}/(q_a-q_b)\,\Delta y_* & (y_* \in \mathrm{I}_a), \\ \boldsymbol{F}/(q_b-q_a)\,\Delta y_* & (y_* \in \mathrm{I}_b), \\ 0 \quad \text{otherwise.} \end{cases} \tag{7.26b}$$

This immediately vindicates assertion (1), since (7.26a) is bounded if condition (v) of §1.3 holds, i.e. if the range of q is bounded.

Now consider assertion (2), regarding the limiting case in which the initial gradient q_{0y} goes to zero throughout the central region. This implies that $(q_a-q_b) \to 0$, so that (7.25) fails. Two cases of interest can be distinguished. First, if $|P|$ can be taken to be bounded in the limit, so that (7.21) still holds, as would be true of ordinary diffusion or any other gradient-dependent transport, then the first term on the right of (7.23) vanishes in the limit, by (5.12), and so $|\alpha(t)|$ is again bounded (and by a smaller amount). We note moreover that if q_{0y} vanishes throughout the entire region $[y_2-\eta_2, y_1+\eta_1]$, then the entire right-hand side of (7.23) is zero, in this case of bounded $|P|$. The critical layer is then a perfect reflector for all time, in the same way as noted below (6.18).

The second case of interest is where $|P|$ is unbounded in the limit. This possibility must be reckoned with because the general form (7.4) of the vorticity transport permits a flux in the absence of any gradient. In fact there is no reason why such a flux should not actually occur in some cases, for instance as a result of transport by inhomogeneous, three-dimensional (but statistically two-dimensional) turbulence,

leading to the creation of two-dimensional gradients in the central region where there were none initially. It still follows, nevertheless, that $|\alpha(t)|$ is bounded, albeit no longer necessarily zero, if condition (v) of §1.3 holds. If we take the limit defined by

$$|q_{0y}| = O(\delta), \quad \delta \to 0 \tag{7.27}$$

so that $(q_a - q_b) = O(\delta)$, then (7.26*a*) is $O(\delta^{-1})$ in the same limit if the range of q is bounded. Thus $C = O(\delta^{-1})$ in (7.21). But (5.12) and (6.5) show that $A = O(\delta)$ so that the first term on the right of (7.23) is still bounded in the limit. The boundedness of $|\alpha(t)|$ again follows.

We thank P. H. Haynes for valuable and generous help with both the analysis and the numerical computations. Further assistance with the latter was kindly provided by J. Smith, J. Wheeler, and J. Venn. We are grateful to D. G. Andrews, P. B. Rhines and an anonymous referee for constructive comments.

Appendix. The time-integrated absorptivity predicted by the SWW solution

We now calculate the limiting value $\hat{\alpha}(\infty)$, say, of the dimensionless time-integrated absorptivity $\hat{\alpha}(\hat{T})$ implied by the SWW vorticity field

$$\hat{Q}_{\mathrm{SWW}}(\hat{x}, Y, \hat{T}) = Y + \hat{Q}_1(\hat{x}, Y, \hat{T}). \tag{A 1}$$

By definition, $\hat{Q}_{\mathrm{SWW}}(\hat{x}, Y, \hat{T})$ satisfies (2.31) under an initial condition of no disturbance, $\hat{Q}_1 = 0$. Equation (2.31), it will be recalled, states that $\hat{Q}_{\mathrm{SWW}}$ is advected by the leading-order velocity field, whose stream function is (2.32), i.e.

$$\Psi = \hat{\Psi}_0(\hat{x}, Y) = -\tfrac{1}{2}Y^2 + \cos\hat{x}. \tag{A 2}$$

We may proceed from Stewartson's elliptic-function solution for $\hat{Q}_{\mathrm{SWW}}$ or directly from first principles; the latter is easier. The quantity of interest is the large-time limit of

$$\hat{\alpha}(\hat{T}) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}Y \int_{-\pi}^{\pi} \mathrm{d}\hat{x}\; Y\hat{Q}_1(\hat{x}, Y, \hat{T}), \tag{A 3}$$

as can be seen for instance by analogy with the right-hand side of (1.7). This form is used rather than the form found in §6 because we wish to turn to computational advantage the fine-grain spatial structure in $\hat{Q}_1(\hat{x}, Y, \hat{T})$, which develops for large $\hat{T}$ (and which becomes infinitely fine in the limit $\hat{T} \to \infty$). Since the form (A 3) is a linear functional of $\hat{Q}_1(\hat{x}, Y, \hat{T})$, the fine-grain structure will average out, more and more accurately as $\hat{T} \to \infty$.

Regarding the use of (A 3), it is important to note that, in order for the double integral to be equal to the time-integrated Reynolds-stress jump (1.1), the $\hat{x}$-integration must be performed before letting the limits of the Y-integration tend to infinity. This can be seen by referring again to the analogy with (1.7). Provided that the $\hat{x}$ integration is performed first, the manipulations go through in exactly the same way as in (1.7), since averaging the integrands with respect to $\hat{x}$ makes them small enough at large $|Y|$, as we shall see shortly, for the Y integration by parts to be valid and the Y integration in (A 3) convergent. To show that the integrated term vanishes when $|Y| \to \infty$ it is convenient to use (2.33), (5.18), (6.7), (6.8) and (6.14) together with the $\hat{x}$-averaged $\hat{x}$-momentum equation and the fact that the conserved density $A \simeq \frac{1}{2}\beta\eta^2 = O(Y^{-2})$ as $|Y| \to \infty$.

Now consider the ribbon-like region $\mathrm{R}(\Psi, \delta\Psi)$ lying between two neighbouring streamlines of (A 2) having the values Ψ and $\Psi + \delta\Psi$, where $\delta\Psi$ is small but fixed

as $\hat{T} \to \infty$. On each streamline within R, the time for a fluid element to travel through one period of the streamline pattern varies continuously across R because of the basic shear. Consequently, after a sufficiently long time – the smaller the value of $\delta\Psi$, the longer the time required – the value of $\hat{Q}_{\mathrm{SWW}}$ averaged across R, at any given $\hat{x}$, will tend to the area average of $\hat{Q}_{\mathrm{SWW}}$ over the whole ribbon R. Call this average $\hat{Q}_{\mathrm{R}}(\Psi)$. As the notation indicates, it is a function of Ψ alone. Once $\hat{Q}_{\mathrm{R}}(\Psi)$ has been calculated, we can evaluate $\hat{\alpha}(\infty)$ by substituting

$$\hat{Q}_{\mathrm{R}}\{\Psi(\hat{x}, Y)\} - Y \tag{A 4}$$

for $\hat{Q}_1$ in (A 3). This amounts to using in place of the actual vorticity field $\hat{Q}_{\mathrm{SWW}}(\hat{x}, Y, \hat{T})$ a vorticity field from which the (infinitely) fine limiting structure has been averaged out. This procedure is justifiable because the other factor Y in the integrand of (A 3) is a smooth function.

We note that the implied limiting process is not uniform as the centre $(\hat{x}, Y) = (0, 0)$ of the cat's eye is approached. For any finite $\hat{T}$, no matter how large, there is always some neighbourhood N_0 of $(0, 0)$ within which the local strain rate is so small that the absolute vorticity contours are rotated but not significantly sheared. However, the area of N_0 shrinks to zero as $\hat{T} \to \infty$. Therefore, since $\hat{Q}_1$ is bounded in magnitude, N_0 contributes nothing to (A 3) in the large-time limit. (N_0 has dimensions $O(\hat{T}^{-\frac{1}{2}})$ in both the $\hat{x}$- and Y-directions. It is this, incidentally, that gives rise to the $O(\hat{T}^{-2})$ behaviour of $[\overline{u'v'}]$ mentioned in §1.1, as can be seen by considering the contribution from N_0 to the integral (2.35).)

Outside the cat's eyes ($-\infty < \Psi < -1$), $\hat{Q}_{\mathrm{R}}(\Psi)$ can be evaluated from the initial condition as

$$\hat{Q}_{\mathrm{R}}(\Psi) = \int_{-\pi}^{\pi} Y \frac{\partial Y}{\partial \Psi} \mathrm{d}\hat{x} \Big/ \int_{-\pi}^{\pi} \frac{\partial Y}{\partial \Psi} \mathrm{d}\hat{x}, \tag{A 5}$$

where Y is regarded as a function of $\hat{x}$ and Ψ as defined by (A 2). Differentiating (A 2) at constant $\hat{x}$, we have

$$\frac{\partial Y}{\partial \Psi} = \left(\frac{\partial \Psi}{\partial Y}\right)^{-1} = -Y^{-1} = -\{2(\cos\hat{x} - \Psi)\}^{-\frac{1}{2}}, \tag{A 6}$$

$$= -\tfrac{1}{2} m^{\frac{1}{2}} \{1 - m \sin^2(\tfrac{1}{2}\hat{x})\}^{-\frac{1}{2}}, \tag{A 7}$$

where

$$m = \frac{2}{1 - \Psi}. \tag{A 8}$$

Therefore the numerator of (A 5) is just -2π, and the denominator is recognizable as $-2m^{\frac{1}{2}}$ times the first complete elliptic integral

$$K(m) = \int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}\theta}{(1 - m \sin^2\theta)^{\frac{1}{2}}} \tag{A 9}$$

in the notation of Abramowitz & Stegun (1965, 17.3.1). Hence, outside the cat's eyes,

$$\hat{Q}_{\mathrm{R}}(\Psi) = \pm \pi m^{-\frac{1}{2}} \{K(m)\}^{-1}, \tag{A 10}$$

where the sign is chosen to make $\hat{Q}_{\mathrm{R}}\{\Psi(\hat{x}, Y)\}$ an odd function of Y, positive when Y is positive. Inside the cat's eyes ($-1 < \Psi < 1$), $\hat{Q}_{\mathrm{R}}(\Psi)$ is zero, by symmetry, because each ribbon $R(\Psi, \delta\Psi)$ is a closed loop symmetrical about the $\hat{x}$-axis.

Using symmetry to get rid of the lower half of the range of Y-integration in (A 3), and making the substitution (A 4), we now have

$$\begin{aligned}\hat{\alpha}(\infty) &= -\frac{1}{\pi}\int_0^\infty \mathrm{d}Y \int_{-\pi}^{\pi} \mathrm{d}\hat{x}\; Y(\hat{Q}_R - Y) \\ &= -\frac{1}{\pi}\int_0^\infty \mathrm{d}Y \int_{-\pi}^{\pi} \mathrm{d}\hat{x}\; Y\left(\frac{\pi H(1-m)}{m^{\frac{1}{2}}K(m)} - Y\right). \end{aligned} \tag{A 11}$$

Here $H(1-m)$ defined to be 1 when $m < 1$, i.e. outside the cat's eyes, and 0 when $m > 1$, i.e. inside the cat's eyes. The $\hat{x}$-integration is to be carried out at fixed Y (and not, it should be noted, at fixed Ψ); and

$$\begin{aligned} m = m(\hat{x}, Y) &= \frac{2}{1-\Psi(\hat{x}, Y)} \\ &\simeq 4Y^{-2} + O(Y^{-4}) \quad \text{as } Y \to \infty. \end{aligned} \tag{A 12}$$

Also
$$m^{-\frac{1}{2}} = \tfrac{1}{2}Y\left(1 + \frac{1-\cos\hat{x}}{Y^2} + O(Y^{-4})\right) \quad \text{as } Y \to \infty. \tag{A 13}$$

We note from (A 9) that

$$\begin{aligned} K(m) &= \tfrac{1}{2}\pi\{1 + \tfrac{1}{4}m + O(m^2)\} \quad \text{as } m \to 0 \\ &= \tfrac{1}{2}\pi\{1 + Y^{-2} + O(Y^{-4})\} \quad \text{as } Y \to \infty. \end{aligned} \tag{A 14}$$

Thus (for $Y > 0$),

$$\begin{aligned} Y\hat{Q}_{\mathrm{R}}(\Psi) &= \pi Y m^{-\frac{1}{2}}\{K(m)\}^{-1} \\ &= Y^2\left(1 + \frac{1-\cos\hat{x}}{Y^2} + O(Y^{-4})\right)\left\{1 - \frac{1}{Y^2} + O(Y^{-4})\right\}. \end{aligned} \tag{A 15}$$

We see from this that the integrand of (A 11) is $O(Y^{-2})$ *after* averaging with respect to $\hat{x}$, which verifies that the Y-integration is then convergent.

The rest is routine calculation. We find from (A 11) that

$$\hat{\alpha}(\infty) = -8\int_0^1 m^{-\frac{5}{2}}\left(\tfrac{1}{2}\pi\{K(m)\}^{-1} - \frac{2}{\pi}E(m)\right)\mathrm{d}m + \frac{32}{3\pi}\int_0^{\frac{1}{2}\pi} \cos^3\theta\, \mathrm{d}\theta, \tag{A 16}$$

where $E(m)$ is the second complete elliptic integral

$$E(m) = \int_0^{\frac{1}{2}\pi} (1 - m\sin^2\theta)^{\frac{1}{2}}\,\mathrm{d}\theta \tag{A 17}$$

in the notation of Abramowitz & Stegun (1965, 17.3.3). The second integral in (A 16) has the value $\frac{2}{3}$; the first was evaluated numerically and found to be equal to -0.10278. Thus

$$\hat{\alpha}(\infty) = 8 \times 0.10278 + \frac{32}{3\pi} \times 0.66667 = 3.0858, \tag{A 18}$$

of which the dimensional equivalent is (1.2). The first term is the contribution to (A 11) from outside the cat's eyes, and the second term the contribution from within them.

P. H. Haynes (personal communication) has derived the following alternative expression for $\hat{\alpha}(\infty)$:

$$\hat{\alpha}(\infty) = -8\int_0^1 m^{-\frac{5}{2}}(\tfrac{1}{2}\pi\{K(m)\}^{-1} - (1-\tfrac{1}{2}m)^{\frac{1}{2}})\,\mathrm{d}m + \tfrac{4}{3}\sqrt{2}. \qquad \text{(A 19)}$$

Numerical evaluation of this expression confirms the value 3.0858, which provides an excellent check on the original computation from (A 16).

REFERENCES

Abramowitz, M. & Stegun, I. A. 1965 *Handbook of Mathematical Functions*. Dover. 1046 pp.

Al-Ajmi, D. N., Harwood, R. S. & Miles, T. 1985 A sudden warming in the middle atmosphere of the southern hemisphere. *Q. J. R. Met. Soc.* **111**, 359–389.

Arnol'd, V. I. 1966 On an a priori estimate in the theory of hydrodynamical stability. *Izv. Vyssh. Uchebn. Zaved. Matematika* **54**, no. 5, 3–5. (English transl.: *Amer. Math. Soc. Transl., Series* 2, **79**, 267–269 (1969).)

Béland, M. 1976 Numerical study of the nonlinear Rossby wave critical level development in a barotropic flow. *J. Atmos. Sci.* **33**, 2066–2078.

Béland, M. 1978 The evolution of a nonlinear Rossby wave critical level: effects of viscosity. *J. Atmos. Sci.* **35**, 1802–1815.

Benney, D. J. & Bergeron, R. F. 1969 A new class of nonlinear waves in parallel flows. *Stud. Appl. Maths* **48**, 181–204.

Brown, S. N. & Stewartson, K. 1978 The evolution of the critical layer of a Rossby wave, part II. *Geophys. Astrophys. Fluid Dyn.* **10**, 1–24.

Brown, S. N. & Stewartson, K. 1980 On the algebraic decay of disturbances in a stratified linear shear flow. *J. Fluid Mech.* **100**, 811–816.

Charney, J. G. & Stern, M. E. 1962 On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.* **19**, 159–172.

Clough, S. A., Grahame, N. S. & O'Neill, A. 1985 Potential vorticity in the stratosphere derived using data from satellites. *Q. J. R. Met. Soc.* **111**, 335–358.

Davis, R. E. 1969 On the high Reynolds number flow over a wavy boundary. *J. Fluid Mech.* **36**, 337–346.

Dickinson, R. E. 1970 Development of a Rossby wave critical level. *J. Atmos. Sci.* **27**, 627–633.

Edmon, H. J., Hoskins, B. J. & McIntyre, M. E. 1980 Eliassen–Palm cross-sections for the troposphere. *J. Atmos. Sci.* **37**, 2600–2616. (See also Corrigendum. *J. Atmos. Sci.* **38**, 1115, especially second last item).

Eliassen, A. & Palm, E. 1961 On the transfer of energy in stationary mountain waves. *Geofys. Publ.* **22**, no. 3, 1–23.

Fletcher, N. H. 1979 Air flow and sound generation in musical wind instruments. *Ann. Rev. Fluid Mech.* **11**, 123–146.

Foote, J. R. & Lin, C. C. 1950 Some recent investigations in the theory of hydrodynamic stability. *Q. Appl. Maths* **8**, 265–280.

Green, J. S. A. 1970 Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Q. J. Met. Soc.* **96**, 157–185.

Haberman, R. 1972 Critical layers in parallel flows. *Stud. Appl. Math.* **51**, 139–161.

Haidvogel, D. B. & Rhines, P. B. 1983 Waves and circulation driven by oscillatory winds in an idealized ocean basin. *Geophys. Astrophys. Fluid Dyn.* **25**, 1–63.

Hamilton, K. 1982 Some features of the climatology of the northern hemisphere stratosphere revealed by NMC upper atmosphere analyses. *J. Atmos. Sci.* **39**, 2737–2749.

Haynes, P. H. 1985 Nonlinear instability of a Rossby-wave critical layer. *J. Fluid Mech.* **161**, 493–511.

Held, I. M. 1983 Stationary and quasi-stationary eddies in the extratropical troposphere: theory. In *Large-scale dynamical processes in the atmosphere* (ed. R. P. Pearce & B. J. Hoskins), pp. 127–168. Academic, 397 pp.

HOLLIDAY, D. & MCINTYRE, M. E. 1981 On potential energy density in an incompressible, stratified fluid. *J. Fluid Mech.* **107**, 221–225.

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.

KAROLY, D. J. 1982 Eliassen–Palm cross sections for the northern and southern hemispheres. *J. Atmos. Sci.* **39**, 178–182.

KELVIN, LORD (W. THOMSON) 1887 Stability of fluid motion – rectilineal motion of viscous fluid between two parallel planes. *Phil. Mag.* **24**, 188–196.

LEOVY, C. B., SUN, C.-R., HITCHMAN, M. H., REMSBERG, E. E., RUSSELL, H. M., GORDLEY, L. L., GILLE, J. C. & LYJAK, L. V. 1985 Transport of ozone in the middle stratosphere: evidence for planetary wave breaking. *J. Atmos. Sci.* **42**, 230–244.

LESSER, M. B. & CRIGHTON, D. J. 1975 Physical acoustics and the method of matched asymptotic expansions. In *Physical Acoustics*, vol. 11 (ed. W. P. Mason & R. N. Thurston), pp. 69–149. Academic.

LIN, C. C. 1955 *The Theory of Hydrodynamic Stability.* Cambridge University Press. 155 pp.

MASLOWE, S. A. 1981 Shear flow instabilities and transition. In *Hydrodynamic Instabilities and the Transition to Turbulence* (ed. H. L. Swinney & J. P. Gollub), pp. 181–228. Springer, 292 pp.

MCINTYRE, M. E. 1982 How well do we understand the dynamics of stratospheric warmings? *J. Met. Soc. Japan* **60**, 37–65.

MCINTYRE, M. E. & PALMER, T. N. 1983 Breaking planetary waves in the stratosphere. *Nature* **305**, 593–600.

MCINTYRE, M. E. & PALMER, T. N. 1984 The 'surf zone' in the stratosphere. *J. Atmos. Terr. Phys.* **46**, 825–849.

MCINTYRE, M. E., SCHUMACHER, R. T. & WOODHOUSE, J. 1983 On the oscillations of musical instruments. *J. Acoust Soc. Am.* **74**, 1325–1345.

MCINTYRE, M. E. & SHEPHERD, T. G. 1986 An exact conservation theorem for finite-amplitude disturbances to nonparallel shear flows, with remarks on Hamiltonian structure and on Arnol'd's stability theorems. *J. Fluid Mech.*, submitted.

NIGAM, S. & HELD, I. M. 1983 The influence of a critical latitude on topographically forced stationary waves in a barotropic model. *J. Atmos. Sci.* **40**, 2610–2622.

RHINES, P. B. 1977 The dynamics of unsteady currents. In *The Sea: Ideas and observations on progress in the study of the seas* (ed. E. D. Goldberg *et al.*), vol. 6, pp. 189–318. Wiley.

RHINES, P. B. 1979 Geostrophic turbulence. *Ann. Rev. Fluid Mech.* **11**, 401–441.

RHINES, P. B. & YOUNG, W. R. 1983 How rapidly is a passive scalar mixed within closed streamlines? *J. Fluid Mech.* **133**, 133–145.

RITCHIE, H. 1985 Rossby wave resonance in the presence of a nonlinear critical layer. *Geophys. Astrophys. Fluid Dyn.* **31**, 49–92.

RUTTENBERG, S. 1980 *Collection of Extended Abstracts Presented at ICMUA Sessions and IUGG Symposium 18*, pp. 199–204. ICMUA (Ruttenberg), NCAR, Box 3000, Boulder, CO 80307.

SHEPHERD, T. G. 1985 On the time development of small disturbances to plane Couette flow. *J. Atmos. Sci.* **42**, 1868–1871.

SMITH, F. T. & BODONYI, R. J. 1982 Nonlinear critical layers and their development in streaming-flow stability. *J. Fluid Mech.* **118**, 165–185.

STARR, V. P. 1968 *Physics of Negative Viscosity Phenomena.* McGraw-Hill. 256 pp.

STEWARTSON, K. 1978 The evolution of the critical layer of a Rossby wave. *Geophys. Astrophys. Fluid Dyn.* **9**, 185–200.

STEWARTSON, K. 1981 Marginally stable inviscid flows with critical layers. *IMA J. Appl. Math.* **27**, 133–175.

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

TUNG, K. K. 1979 A theory of stationary long waves. Part III: Quasi-normal modes in a singular waveguide. *Mon. Weather Rev.* **107**, 751–774.

TUNG, K. K. & LINDZEN, R. S. 1979*a* A theory of stationary long waves. Part I: A simple theory of blocking. *Mon. Weather Rev.* **107**, 714–734.

TUNG, K. K. & LINDZEN, R. S. 1979*b* A theory of stationary long waves. Part II: Resonant Rossby waves in the presence of realistic vertical shears. *Mon. Weather Rev.* **107**, 735–750.

VAN DYKE, M. 1975 *Perturbation Methods in Fluid Mechanics* (annotated edition). Parabolic, 271 pp.

WARN, T. & WARN, H. 1976 On the development of a Rossby wave critical level. *J. Atmos. Sci.* **33**, 2021–2024.

WARN, T. & WARN, H. 1978 The evolution of a nonlinear critical level. *Stud. Appl. Math.* **59**, 37–71.

WHITTAKER, E. T. 1937 *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, 4th edn. Cambridge University Press, 456 pp.

YAMAGATA, T. 1976 On trajectories of Rossby wave-packets released in a lateral shear flow. *J. Ocean. Soc. Japan.* **32**, 162–168.