

Appendix I: Rossby-wave propagation and shear instability

M. E. McIntyre

1. Introduction

This Appendix explores Rayleigh’s inviscid shear instability problem and its relation to the Rossby-wave propagation mechanism, or ‘Rossby quasi-elasticity’. The instability problem is the simplest of those solved by Rayleigh in his pioneering work last century on the undular instability of jets and shear layers. It provides us with a robust paradigm for a very basic fluid-dynamical process. It is robust in the sense that it gives essentially the same result as the $U \propto \tanh(y/b)$ case and practically all the other inviscid shear-layer profiles that you can explore for yourself in computer demonstration 10. These include almost *any* moderately smooth shear-layer profile drawn with the mouse, with y -scale b somewhat less than the computational domain size L .

The essential qualitative result, well documented in many places in a vast research literature, is that almost any shear layer sandwiched between constant-velocity regions is unstable to small sideways undular displacements, with a fastest exponential growth rate equal to a modest fraction, often a fifth or so, of the typical shear. The fastest-growing instability has a radian wavelength of the same order as the shear layer thickness $2b$, where radian wavelength means full wavelength $/2\pi$.

As already remarked in the lectures, the same qualitative result applies also to the ‘KH instability’ (Kelvin–Helmholtz, sometimes called Taylor–Goldstein, instability) of a stratified shear layer $U(z), N^2(z)$ at sufficiently small Richardson number $Ri = (N/U_z)^2$. This instability is sometimes visible in the sky as groups of long-crested ‘billow clouds’ having lifetimes of order ten minutes. The instability commonly occurs when larger-scale disturbances tilt a strongly stratified layer and produce sufficient vertical shear U_z in the layer (via the horizontal gradient of the buoyancy acceleration, in the vorticity equation, equivalently the term $\propto \nabla\rho \times \nabla p$, cf. Dr Linden’s lectures) to bring the local value of Ri well below 0.25. You can use the computer’s Movie Viewer (load KH.IMG). to see the evolution to finite amplitude of a typical KH instability, produced in the laboratory by tilting a thin stratified layer in a long tank (S.A. Thorpe 1973, J. Fluid Mech., **61**, 731)†

It can be shown that, for laboratory-scale shear layers in which viscosity might be directly significant, the fastest instabilities are not much affected until Reynolds numbers Ub/ν are down

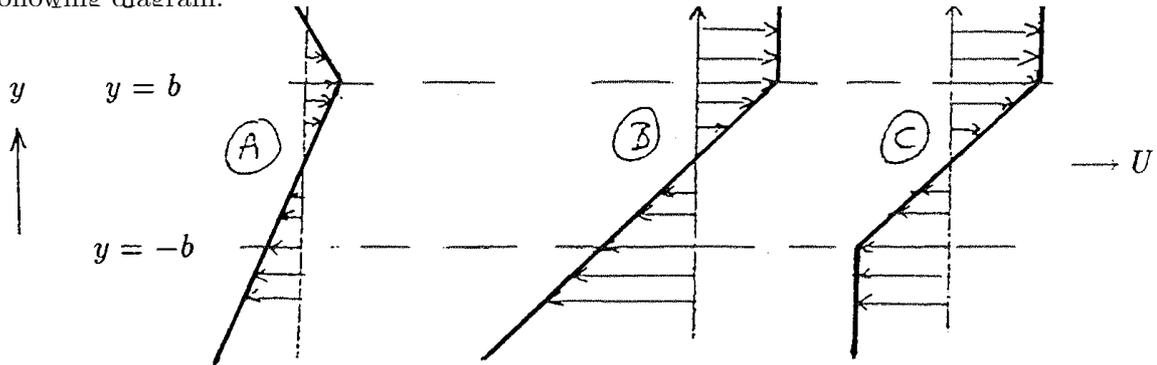
† You might like to think about how it is that minimum $Ri = (N/U_z)^2$ tends to occur in the most strongly stratified layers, i.e., where N^2 is *largest*. This is true both in the tilted-tank experiment and in most naturally-occurring situations. The tilting envisaged is one in which some larger-scale disturbance tilts a relatively thin but horizontally extensive layer of relatively strong stratification. Such a layer, by definition, has a strong maximum in N considered as a function of z . The key point is that if the layer tilts approximately as a plane, making a small time-dependent angle $\alpha(t)$ with the horizontal, then the horizontal gradient $\nabla_H\sigma$ of the buoyancy acceleration σ has approximate magnitude $\alpha N^2(z)$, a strong function of z . This gives rise to horizontal vorticity, appearing mainly as vertical shear U_z (because of the large ratio of horizontal to vertical scales in this situation), and having the z -dependence of N^2 , not N . Specifically, $U_z \simeq \gamma(t)N^2(z)$ where, if Coriolis forces are negligible, as in the tilted-tank experiment, $\gamma(t)$ is simply the time integral of $\alpha(t)$. Then $Ri = (N/U_z)^2 = (\gamma N)^{-2}$. So, when $\gamma(t)$ increases, Ri becomes smallest soonest at a maximum, not a minimum, of $N(z)$. The same formula $Ri = (\gamma N)^{-2}$, hence the same conclusion, can be shown to hold far more generally with suitably modified $\gamma(t)$. For instance, in the opposite-extreme case of geostrophic balance (see Haynes’ lectures), the formula $Ri = (\gamma N)^{-2}$ still holds but with $\gamma(t) = \alpha(t)/f$ where f is the Coriolis parameter.

to very modest values, of the order of 10 or less — another striking indication of the robustness of the shear instability mechanism in these simplest cases.

Rayleigh’s problem is the case of small-amplitude disturbances to an exactly inviscid layer with exactly constant shear, the ‘vorticity strip’ first mentioned in the lectures at transparency MEM 35; see also computer demonstrations 6, 7. In this particular case the problem can be explicitly solved in detail, with no more than exponential functions and a modicum of patience — apart from just one tricky technicality, the derivation of equation (6) below. Section 4 gives the full analysis, and shows how it illustrates the fundamental relation between shear instability and the Rossby-wave propagation mechanism. An equivalent visual-verbal description is given in section 5, following the review by Hoskins et al (1985, Q. J. Roy. Meteorol. Soc., **111**, 877–946 and **113**, 402–404). To prepare the way for sections 4 and 5, the appropriate case of simple Rossby waves is analyzed first (sections 2 and 3). The qualitative understanding thus gained shows *why* the instability mechanism is robust, particularly as regards its finite-amplitude consequences illustrated in the lectures at transparency MEM 35.

Moreover, that understanding can be extended immediately to the fastest — and likewise robust — three-dimensional ‘baroclinic instabilities’ on horizontal temperature gradients (Dr Haynes’ lectures), which are usually thought of as accounting for the existence of the mid-latitude cyclones and anticyclones that are conspicuous features of atmospheric weather patterns. Mid-latitude cyclones and anticyclones have horizontal length scales $L \lesssim 10^3$ km; their oceanic counterparts (scales $L \lesssim 10^2$ km) place severe requirements on numerical resolution for eddy-resolving ocean circulation models. The extension to three-dimensional baroclinic problems is obtained simply by *replacing vorticity with potential vorticity* and replacing ‘vorticity inversion’ (the inverse Laplacian operator) with ‘potential vorticity inversion’; see also Appendix II. Rayleigh’s instability itself has direct relevance to some atmospheric weather developments, and ocean-current instabilities, associated with horizontal shear, and in this context is often referred to as a ‘barotropic shear instability’ or a ‘Rayleigh-Kuo instability’. The description of the instability mechanism in section 5 is written so as to apply, suitably interpreted, both to the barotropic and to the baroclinic cases. On first reading, however, it can be viewed simply as a summary of what happens in Rayleigh’s problem, underpinned by the detailed justification available, for those interested, in sections 3 and 4.†

Rayleigh’s problem, then, is to find the inviscid, exponentially-growing small disturbances, if any, to the unidirectional basic or background velocity profile $(u, v) = \{U(y), 0\}$ shown as © in the following diagram:



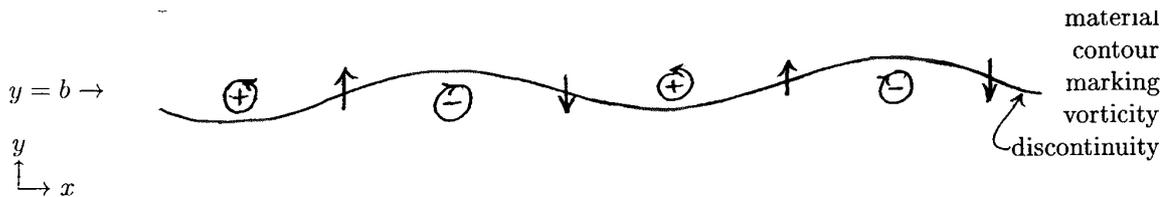
† There is a vast and highly technical literature on the linearized theory of shear instabilities that are more complicated, slower-growing, and less robust as regards their finite-amplitude consequences — hence less likely to be practically important, albeit sometimes mathematically intriguing. Some of these more complicated instabilities can be understood in terms of a phenomenon called ‘over-reflection’, as first, I believe, clearly illustrated by A. E. Gill (1965, Phys. Fluids., **8**, 1428–1430), who analyzed an instability arising from the over-reflection of sound waves between two vortex sheets.

The shear $U_y = dU/dy$ is piecewise constant in each profile shown. The domain is unbounded. The planetary vorticity gradient $\beta = df/dy$ is taken to be zero since we are interested, at first, in a paradigm that is equally relevant to large-scale and small-scale flow.

2. Rossby-wave propagation on a concentrated vorticity gradient

First consider profiles ① and ②, which are *stable*. They provide another illustration of the Rossby wave propagation mechanism or ‘Rossby-wave quasi-elasticity’, which is basic to most problems in atmosphere–ocean dynamics and, for instance, has an important role in the approximate chemical isolation of ‘Meddies’ (Atlantic Mediterranean Eddies) from their surroundings, and similarly the chemical isolation of the stratospheric polar vortex and ozone hole (lecture transparency MEM 7). (Epigrammatically, “Strong vortices have strongly Rossby-elastic edges.”)

Here the vorticity gradient to which the wave propagation owes its existence is concentrated on a single material contour, namely that material contour whose undisturbed position is $y = b$. (The distinction between these Rossby waves and those on a constant, or smoothly varying, basic vorticity or potential-vorticity gradient may be compared to the distinction between surface gravity waves and internal gravity waves. The propagation of surface gravity waves, or ordinary ‘water waves’, can be described as owing its existence to a density or buoyancy gradient concentrated at the water surface, as compared with internal gravity waves on a continuous buoyancy gradient.) The following sketch reminds us of the basic Rossby-wave mechanism:



The encircled signs indicate the sense of the vorticity anomalies q' caused by displacing the contour; note that the basic-state vorticity is more positive, or less negative, on the positive- y or ‘northern’ side of the contour). The straight arrows indicate the sense of the induced disturbance *velocity* field, i.e. the velocity field resulting from inversion of q' . The phase of the velocity pattern is important. The velocity pattern is a quarter wavelength out of phase with the material *displacements*, marked by the undular shape of the material contour itself. What follows from this is a matter of simple kinematics. If one makes a movie of the situation in one’s mind’s eye, as viewed from a frame of reference moving with the basic flow U at $y = b$, one can see that the undulations must be propagating relative to the basic flow. (The notion of vorticity inversion allows one to say, epigrammatically, that the undulations are *caused* to propagate *by* the disturbance vorticity anomalies $\oplus \ominus \dots$.) The propagation is toward the left in this case; generally it is in whichever direction has the most positive, or least negative, basic-flow vorticity or potential vorticity on the right. This is sometimes called ‘pseudo-westward’ or, more aptly, ‘quasi-westward’.

The all-important quarter-wavelength phase shift is easily understandable from the properties of vorticity inversion, for instance as visualized by the electrostatic and soap-film analogies described in the lectures. For instance a soap film being pushed and pulled in an x -periodic pattern will show a corresponding pattern of hills and valleys; this tells us that the streamfunction anomalies ψ' are 180° out of phase with the vorticity anomalies, and hence *in* phase with the material contour displacements. You can also verify this picture from the computer demonstrations.*

* Demonstration 7 (look at one colour only) is probably more convenient for this purpose than demonstration 6.

The next section verifies the correctness of the foregoing picture in an independent way, by using the traditional linearized mathematical theory for small displacements. This prepares the way for a similar mathematical treatment of Rayleigh's problem. If you are happy to take the theory on trust on first reading, you can skip to section 5 at this point.

3. Mathematical verification of the Rossby propagation mechanism

We regard all the basic velocity profiles as limiting cases of smooth profiles with continuous derivatives. This is one way of being sure to get the correct jump conditions across the discontinuities in U_y — the only tricky point, equation (6) below, in an otherwise straightforward analysis. For smooth profiles $U(y)$, with viscosity neglected, the linearized disturbance equation can be written

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x}\right) q' - U_{yy} \frac{\partial \psi'}{\partial x} = 0, \quad (1)$$

where $-U_{yy} = -d^2U/dy^2$, the basic or background vorticity gradient giving rise to the Rossby-wave mechanism, and q' and ψ' are respectively the disturbance contributions to the vorticity and streamfunction, with the convention

$$(u', v') = (-\partial\psi'/\partial y, \partial\psi'/\partial x) \quad . \quad (2)$$

for the disturbance velocity. The relation between ψ' and q' appears in this notation as

$$\nabla^2 \psi' = \frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2} = q' \quad (3a)$$

$$\left(\psi' \text{ periodic in } x; \quad \psi' \text{ and } \psi'_y \rightarrow 0 \text{ as } |y| \rightarrow \infty\right) \quad , \quad (3b)$$

ψ'_y being shorthand for $\partial\psi'/\partial y$ as usual. We may summarize the content of (3) more succinctly as

$$\psi' = \nabla^{-2} q' \quad , \quad (4)$$

making explicit the idea of vorticity inversion as used just now in section 2, in the lectures, and in Appendix III equation (14b), with boundary conditions of evanescence in y and periodicity in x understood here.

In the limit of piecewise constant shear dU/dy , we have $U_{yy} = 0$ for $y \neq \pm b$. Hence (1) implies that

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x}\right) q' = 0 \quad (y \neq \pm b) \quad . \quad (5)$$

By a careful consideration of the limit near $y = \pm b$ it can also be shown* that (1) implies

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x}\right) \left[\frac{\partial \psi'}{\partial y}\right]_+^+ - \left[\frac{dU}{dy}\right]_+^+ \frac{\partial \psi'}{\partial x} = 0 \quad \text{at } y = \pm b \quad , \quad (6)$$

* One way of deriving (6) from (1) is to use the concept of Dirac delta or 'point-charge' functions. When we take the limit in (1), U_{yy} will tend to a delta function of strength $[U_y]_+^+$ centred on $y = b$. In the case of profile © one must add a similar contribution centred on $y = -b$. Equation (1) can be satisfied in the limit only if the ψ'_{yy} contribution to $\nabla^2 \psi'$ [see (3a)] likewise tends to a delta function, with ψ' continuous and ψ'_y piecewise continuous.

where the square brackets denote jumps or differences across $y = \pm b$. That is, $[F(y)]_{\pm}^{\pm}$ at $y = b$ means $F(y+) - F(y-)$ where $F(y+) = \lim_{y \downarrow b} F(y)$ and $F(y-) = \lim_{y \uparrow b} F(y)$, for any function $F(y)$ such that the limits exist. The condition (6) says, in a rather inscrutable way, that the initial vorticity distribution moves with the undulating material contour. The inscrutability arises from using the *linearized Eulerian* description — for reasons of mathematical convenience — to describe something that appears simple only in the *exact Lagrangian* description.†

Both for profile ① and for profile ② we have solutions of the form

$$\psi' = \hat{\psi}(y)e^{ik(x-ct)} \quad (7)$$

where

$$\hat{\psi}(y) \propto e^{-|k(y-b)|} \quad , \quad (8)$$

making q' vanish for $y \neq b$ and hence satisfying (5). Then (6) gives

$$ik(U - c)(-2|k|) = [U_y]_{\pm}^{\pm} ik \quad \text{at } y = b \quad ,$$

so that the intrinsic phase speed is

$$c - U(b) = -\frac{1}{2}G|k|^{-1} \quad (G > 0) \quad (9)$$

where $G = -[U_y]_{\pm}^{\pm}$ at $y = b$ ($G > 0$ for profiles ① and ②). Thus a disturbance of x -wavelength $2\pi/k$ centred on $y = b$ propagates to the *left* with phase speed (9), relative to $U(b)$, as anticipated in the lectures. (This agreement between equations and pictures is a good check that we have the sign right in (9). Note also that the intrinsic *frequency* $-\frac{1}{2}G \operatorname{sgn} k$ is independent of $|k|$ — inevitable on dimensional grounds, as with internal gravity waves, since profiles ① and ② have no length scale, and the only relevant property of the basic flow, $G = -[U_y]_{\pm}^{\pm}$, is a constant having the dimensions of frequency. It follows incidentally that the intrinsic group velocity is zero.)

[Exercise: Verify that, in this case, Rossby waves that have less ‘room’ to propagate will propagate more slowly, in the sense of having smaller intrinsic phase speeds. Take for instance the case in which rigid boundaries are introduced at $y = b \pm a$ for some positive constant a ; it is easy to show that this always replaces the $\frac{1}{2}|k|^{-1}$ in (9) by a smaller quantity.]

† More precisely, the inscrutability of (6) is connected with the noninterchangeability of the two limits involved in deriving it, the first being the limit of small disturbance amplitude, already taken in (1) through the omission of terms like $q'_y \psi'_x$, and the second being the limit of infinitely steep vorticity gradients at the material contour! One way to make sense of (6), independently of its derivation from (1), is to recognize that although (for reasons of mathematical convenience) (6) refers to values exactly at $y = +b$, for instance, it actually represents physical conditions *at the displaced position*, $y = b + \eta$, say, of the material contour. The total velocity field on each side of the contour has been, in effect, extrapolated back to $y = b$ using one-term Taylor expansions, again neglecting products of small quantities like $\eta u'_y$ and *ignoring the fact that $y = b$ may be on the wrong side of the contour*. Now if the simple vorticity discontinuity of profile ① or ② moves with the undulating material contour $y = b + \eta$, then the velocity jump $\Delta_{\eta} u$ across the contour $y = b + \eta$ must vanish, $\Delta_{\eta} u = 0$, for otherwise a sheet of infinite vorticity would have appeared from nowhere (see also section 7 below). The small-amplitude approximation to $\Delta_{\eta} u$, expressed in terms of the fields extrapolated back to $y = b$, is $[u' + \eta U_y]_{\pm}^{\pm}$; so we must have $[u' + \eta U_y]_{\pm}^{\pm} = 0$. Therefore we must also have $(\partial/\partial t + U\partial/\partial x)[u' + \eta U_y]_{\pm}^{\pm} = 0$. This is (6), because η is a continuous function of y (and a differentiable function of x) such that $(\partial/\partial t + U\partial/\partial x)\eta = v' = \partial\psi'/\partial x$.

4. Mathematical analysis of the instability mechanism

What happens if we add another region of concentrated vorticity gradient, with the opposite sign, as in profile ©? The Rayleigh–Kuo and Fjørtoft theorems (see computer demonstration notes, page 10–3) now suggest that instability is possible: G changes sign between regions, evading the Rayleigh stability condition, and U also changes sign, in the sense required to evade the Fjørtoft stability condition. This still does not guarantee instability; see section 6 below. But we now show directly, following Rayleigh, that profile © does, in fact, have unstable modes provided $|kb|$ is not too large. (One can see from the decaying exponential structure in (8) that instability will certainly not be found for $|kb| \gg 1$. For if $|kb| \gg 1$, then (8) and a similar solution $\hat{\psi} \propto e^{-|k(y+b)|}$ will apply with exponentially small error; the neighbourhoods of $y = \pm b$ are too far away from each other ($b \gg |k|^{-1}$) to interact significantly, and will behave independently.)

For general $|kb|$, we have (taking $k > 0$ to save having to write $|k|$ all the time):

$$\hat{\psi} = \begin{cases} A \sinh(2kb).e^{-k(y-b)} & (y > b) \\ A \sinh k(y+b) + B \sinh k(b-y) & (-b < y < b) \\ B \sinh(2kb).e^{k(y+b)} & (y < -b) \end{cases} \quad (10)$$

for some pair of constant coefficients A and B . The form of (10) has again been chosen to make q' vanish for $y \neq b$ and hence to satisfy (5), and also to make $\hat{\psi}$ continuous at $y = \pm b$. The ratio A/B , and the constant c , are still available to satisfy (6) at each interface. Writing $U_y = \Lambda$ (positive constant) for $|y| < b$, dividing (6) by ik , and writing sh for $\sinh 2kb$ and ch for $\cosh 2kb$, we have

$$\begin{cases} \text{At } y = b : & (\Lambda b - c)[\hat{\psi}_y]_{-}^{+} + \Lambda \hat{\psi} = 0; & \text{also } \hat{\psi} = A \text{ sh}, & [\hat{\psi}_y]_{-}^{+} = -Ak \text{ sh} - (Ak \text{ ch} - Bk) \\ \text{At } y = -b : & (-\Lambda b - c)[\hat{\psi}_y]_{-}^{+} - \Lambda \hat{\psi} = 0; & \text{also } \hat{\psi} = B \text{ sh}, & [\hat{\psi}_y]_{-}^{+} = (Ak - Bk \text{ ch}) - Bk \text{ sh} . \end{cases}$$

The first line gives

$$-Ak \text{ sh} - Ak \text{ ch} + Bk + \frac{\Lambda}{\Lambda b - c} A \text{ sh} = 0 \quad . \quad (11a)$$

The second gives

$$Ak - Bk \text{ ch} - Bk \text{ sh} + \frac{\Lambda}{\Lambda b + c} B \text{ sh} = 0 \quad (11b)$$

A nontrivial solution for $A : B$ exists if and only if the determinant vanishes; write $\text{sh} + \text{ch} = \exp = \exp 2kb$ (since $\text{ch} = \frac{1}{2}\{\exp + (1/\exp)\}$ and $\text{sh} = \frac{1}{2}\{\exp - (1/\exp)\}$):

$$\left(-k \exp + \frac{\Lambda \text{ sh}}{\Lambda b - c}\right) \left(-k \exp + \frac{\Lambda \text{ sh}}{\Lambda b + c}\right) - k^2 = 0 \quad . \quad (12)$$

This will give c (for real, prescribed k); it also gives a quick check that we have done our sums correctly so far, since in the large- k limit both $k \exp$ and sh are overwhelmingly greater than k^2 , so that in (12) we have $() () = 0$ to an excellent approximation, so that one or other factor must vanish, again to an excellent approximation. The vanishing of the first factor gives $\Lambda b - c \simeq \frac{1}{2}\Lambda k^{-1}$, equivalent to (9) (isolated Rossby wave on interface $y = +b$). The second factor similarly gives the wave on $y = -b$.

Multiplying-out the product $() ()$ in (12) and noting that $(\exp)^2 - 1 = 2 \text{ sh} \cdot \exp$, we have

$$2k^2 \text{ sh} \cdot \exp - k\Lambda \text{ sh} \cdot \exp \cdot \left(\frac{1}{\Lambda b - c} + \frac{1}{\Lambda b + c}\right) + \frac{\Lambda^2 \text{ sh}^2}{\Lambda^2 b^2 - c^2} = 0 \quad .$$

The quantity in parentheses is equal to $2\Lambda b/(\Lambda^2 b^2 - c^2)$. Therefore

$$\Lambda^2 b^2 - c^2 = \frac{2kb\Lambda^2 \text{ sh. exp} - \Lambda^2 \text{ sh}^2}{2k^2 \text{ sh. exp}} \quad \left(= \frac{\Lambda^2 b^2}{kb} - \frac{\Lambda^2 b^2 \text{ sh}}{2k^2 b^2 \text{ exp}} \right)$$

or

$$c^2 = \Lambda^2 b^2 \left(1 - \frac{1}{kb} + \frac{\text{sh}}{2k^2 b^2 \text{ exp}} \right) . \quad (13)$$

If we make c dimensionless with respect to the total velocity difference $\Delta U = 2b\Lambda$, and k with respect to the shear layer width $2b$, say $C = c/2b\Lambda$, $K = 2kb$, then (13) becomes

$$C^2 = \left(\frac{c}{2b\Lambda} \right)^2 = \frac{1}{4} - \frac{1}{2}K^{-1} + \frac{\sinh(K)}{2K^2 \exp(K)} = \frac{1}{4K^2} [(K-1)^2 - \exp(-2K)] . \quad (14)$$

Note that $C^2 = -\frac{1}{4} + \frac{1}{3}K + O(K^2)$ as $K \rightarrow 0$ (by Taylor-expanding $\exp(-2K) = 1 - 2K + 2K^2 - \frac{4}{3}K^3 + O(K^4)$); note that the first two orders cancel, so that

$$C = \pm \frac{1}{2}i(1 - \frac{2}{3}K + O(K^2)) \quad \text{as } K \rightarrow 0 . \quad (15)$$

This demonstrates the existence of instability for some range of K : there exists a mode with $\text{Im } c > 0$, at least for sufficiently small K . At this point we get another check that the algebra is correct; the limiting values $C \simeq \pm \frac{1}{2}i$, or $c \simeq \pm \frac{1}{2}\Delta U i$, agree with those implied by the theory of waves on a single vortex sheet (e.g. Batchelor's textbook, eq. (7.1.20)). The fact that small K means radian wavelength $k^{-1} \gg 2b$ suggests that the disturbance should see the whole shear layer as being thin; i.e. as a vortex sheet. Similarly, any other $U(y)$ profile that goes monotonically between two constant values should give the same long-wave behaviour in an infinite domain. For instance the tanh profile also checks out in this respect ($c \sim \frac{1}{2}i \times$ total change in U). You could try some mouse-drawn profiles as well, with small-ish but finite k , but it will be necessary to make the domain size L somewhat larger than k^{-1} .

Next we note the *phase relations* implied by (10), (11) and (14) — crucial to a full understanding of what is going on! From (11a, b) respectively we get

$$\frac{B}{A} = \left(\exp - \frac{\Lambda \text{ sh}}{k(\Lambda b - c)} \right) = \left(\exp - \frac{\Lambda \text{ sh}}{k(\Lambda b + c)} \right)^{-1} . \quad (16)$$

We are interested only in cases where k and K are real and c is pure imaginary, i.e. (14) is negative-valued. The two expressions in large parentheses are then complex conjugates of each other, since they differ only in the sign of $\pm c$. It follows that

$$\left| \frac{B}{A} \right| = 1 , \quad (17)$$

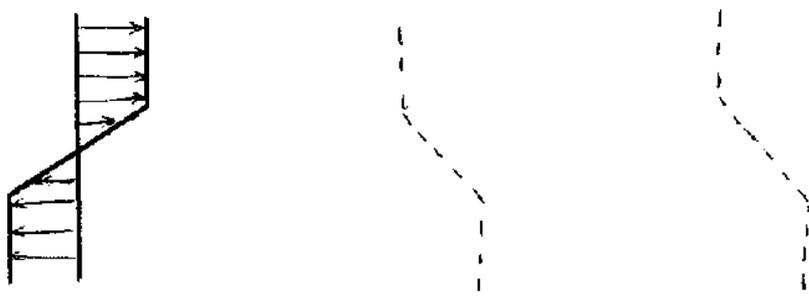
the simplest result consistent with the symmetry of the problem. The relative phases of ψ' and therefore of $v' = \psi'_x$ are (when $c = ic_i$, pure imaginary):

$$\arg \left(\frac{B}{A} \right) = \arcsin \text{Im} \left(\frac{B}{A} \right) = \arcsin \text{Im} \left(- \frac{\Lambda \text{ sh}}{k(\Lambda b - ic_i)} \right) = \arcsin \left(\frac{-\Lambda c_i \text{ sh}}{k(\Lambda^2 b^2 + c_i^2)} \right) , \quad (18a)$$

or in dimensionless form, dividing numerator and denominator by $2b\Lambda^2$,

$$\arg \left(\frac{B}{A} \right) = \arcsin \left(\frac{-C_i \sinh(K)}{K(\frac{1}{4} + C_i^2)} \right) , \quad (18b)$$

This gives the phase angle, or fraction-of-a-wavelength times 2π , by which the pattern in v' at $y = b$ leads that at $y = -b$. It is *negative* for the growing mode, $C_i > 0$, so in our picture, with x pointing to the right, the v' pattern at $y = b$ is shifted to the *left* of that at $y = -b$. The constant-phase lines ‘tilt oppositely to the shear’:



The phase shift (18b) across the shear tends* to $\arcsin(-1) = -\pi/2$, corresponding to a quarter of a wavelength, as $K \downarrow 0$ and $C_i \uparrow \frac{1}{2}$. Since the complex displacement amplitude $\hat{\eta} = \hat{\psi}/(U - c)$, from $(\partial/\partial t + U\partial/\partial x)\eta = v'$ and $v' = \hat{\psi}'_x$, and since the q' pattern at each interface $\propto \mp \eta$ at $y = \pm b$, the phase shift in the q' pattern is given by

$$\begin{aligned} \arg\left(\frac{B/(-\Lambda b - c)}{-A/(\Lambda b - c)}\right) &= \arg\left(\frac{B(\Lambda b - c)}{A(\Lambda b + c)}\right) = \arg\left(\frac{B}{A}\right) + \arg\left(\frac{1 - 2C}{1 + 2C}\right) \\ &= \arg\left(\frac{B}{A}\right) + 2 \arctan(-2C_i) \quad \text{if } C = iC_i, \text{ pure imaginary} \end{aligned} \quad (19)$$

Thus the q' pattern has a phase shift in the same sense, but *bigger*. On the next page are some numerical values showing how the quantities of interest vary as function of dimensionless wavenumber K . From left to right: dimensionless wavenumber K , imaginary part C_i of dimensionless phase speed (real part being zero), dimensionless growth rate KC_i , phase shift for v' or ψ' , phase shift for q' , the last two being expressed as fractions of a wavelength.

Note from (15) (18b), and (19) that the phase shift for the q' pattern $\rightarrow -\pi$, corresponding to half a wavelength, as $K \downarrow 0$. This again is consistent with expectation (and with well known results) for effectively thin shear layers, or vortex sheets. It says that, in the long-wave limit, the displacements η ($\propto \mp q'$) are almost exactly in phase across the shear layer; that is, the layer does undulate almost as a single entity.

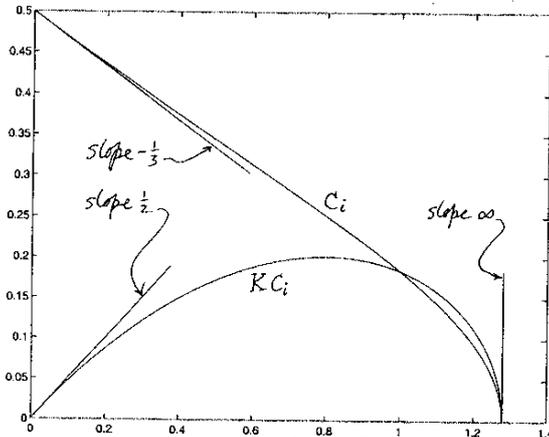
* Note that C_i is real, and $\rightarrow \frac{1}{2}$ as $K \rightarrow 0$, and that $\sinh(K)/K \rightarrow 1$ as $K \rightarrow 0$.

Some numerical results for Rayleigh's shear instability problem:
 (Dimensionless growth rate $KC_i = kc_i/\Lambda$ peaks at $K = 0.797$,
 with max. value $KC_i = 0.20119$:)

K	C_i	KC_i	shifts	
			u'	v'
0.100	0.467	0.047	-0.244	-0.484
0.200	0.435	0.087	-0.238	-0.466
0.300	0.404	0.121	-0.231	-0.447
0.400	0.374	0.149	-0.222	-0.426
0.500	0.343	0.172	-0.213	-0.404
0.600	0.313	0.188	-0.202	-0.380
0.700	0.283	0.198	-0.189	-0.353
0.800	0.251	0.201	-0.175	-0.323
0.900	0.219	0.197	-0.158	-0.290
1.000	0.184	0.184	-0.138	-0.250
1.100	0.144	0.159	-0.112	-0.201
1.200	0.094	0.113	-0.075	-0.134

```

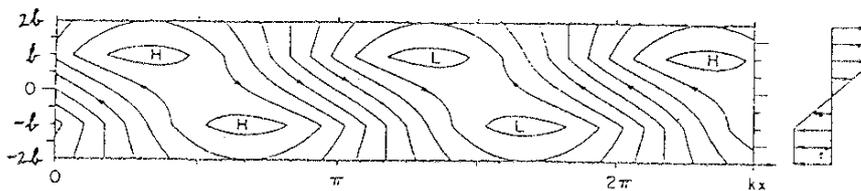
2INPUT"FROM "K0
3INPUT"TO "K1
4INPUT"STEP "S
8@%=&01020307:REM print format
10FOR K=K0 TO K1 STEP S
20GOSUB102
50PRINTK,C,K*C,FRACV,FRACQ
60NEXT
99END
102 EX=EXP(K)
103 EM=1/EX
105 SH=.5*(EX-EM)
115 K2=K*K
140 DEN=2*K2*EX
150 C2=.25-(.5/K)+(SH/DEN) (14)
155 C=SQR((ABS(C2))) (18)
160 ARGV=ASN(-C*SH/(K*(.25-C2)))
170 FRACV=ARGV/(2*PI)
180 ARGQ=ARGV+2*ATN(-2*C)
190 FRACQ=ARGQ/(2*PI)
200RETURN
    
```



C^2 changes sign just before

K reaches 1.28:	50PRINTK,C
1.2783	4.23329659E-3
1.2784	2.65127772E-3
1.2785	1.96502458E-3
1.2786	3.84073147E-3
1.2787	5.06361944E-3

The following picture, taken from Gill's book, gives $\psi' = \hat{\psi}(y)e^{ik(x-ct)}$ for the fastest growing mode, as a function of x and y when the arbitrary constant A is taken such that AB is pure imaginary:
 The following picture, taken from Gill's book, gives $\psi' = \hat{\psi}(y)e^{ik(x-ct)}$ for the fastest growing mode, as a function of x and y when the arbitrary constant A is taken such that AB is pure imaginary:



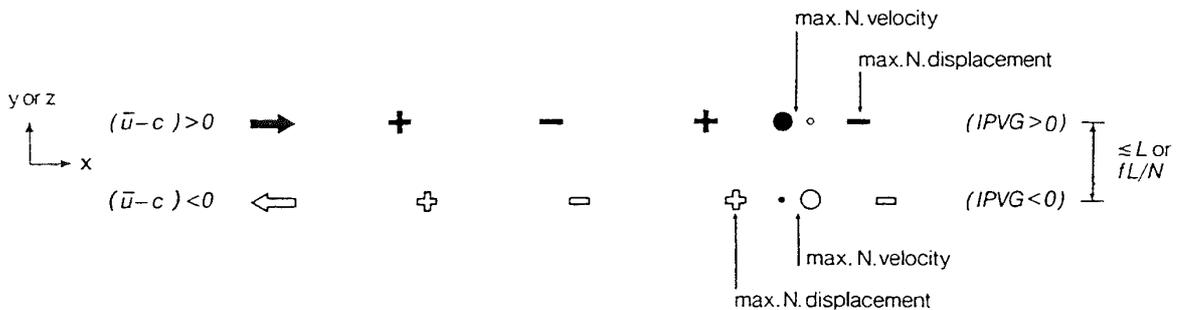
The next section gives a verbal description that serves to summarize the key points about the picture that has emerged. It also tries to make the qualitative robustness of the whole picture more evident, including the pattern of phase shifts and how they are related to the exponential growth with time. As mentioned earlier it can be understood as applying to more than one type of large and small-scale shear instability occurring in the atmosphere and oceans; but on first reading it can be understood simply as a summary of the picture just derived for Rayleigh’s problem.

5. The essentials of the instability mechanism

This follows section 6*d* of the review article by Hoskins *et al.* cited in section 1. Those of you with a particular interest in large-scale atmospheric and oceanic eddies might be interested not only in the wider meaning of what is to follow, but also in other parts of the review such as the description of what happens, in certain cases, when large amplitudes are attained, and the relationship to cyclogenesis in the real atmosphere (as hinted at in my last lecture). In this connection you should note one point about terminology. Phrases like ‘IPV maps’, ‘IPV distributions’, ‘IPV anomalies’, etc., are used in the review article as a handy abbreviation to signify isentropic or isopycnic maps, distributions, anomalies, etc, of PV, where ‘PV’ means the quantity $Q = \rho^{-1} \zeta_{\text{abs}} \cdot \nabla \theta$ defined in the lectures and in equation (12) of Appendix III, i.e., the Rossby–Ertel potential vorticity. (The full name follows the historical precedents dating from a paper by Rossby published in 1936, ref. [99] in Appendix III.) As will be explained in the lectures, it is isentropic or isopycnic *distributions* of PV — and, for Rossby waves and shear instabilities, isentropic or isopycnic *gradients*, and *anomalies*, of PV — that are dynamically significant. They play the role of vorticity gradients and anomalies in two-dimensional vortex dynamics. In the diagram below, taken from the review article, ‘IPVG’ means (northward) isentropic gradient of PV (and ‘N’ or ‘northward’ corresponds to $+y$ above). Since the review was published it has become apparent, however, that phrases like ‘IPV gradient’ can be too easily misread as signifying a gradient of something called ‘IPV’. Therefore in these notes I shall use phrases like ‘PV gradients’, leaving tacit the important fact that, in the case of layerwise-two-dimensional stratified flow, this must be understood to mean *isentropic* or *isopycnic* gradients.

On first reading, as suggested, references to baroclinic phenomena can be ignored, and, as appropriate for the case of the strictly two-dimensional flow that is our immediate concern, ‘PV gradient’ can be read as meaning vorticity gradient (absolute vorticity gradient if in a rotating frame), ‘PV anomaly’ as meaning q' , and so on.

As already suggested, the simplest instabilities — by which we mean those with the simplest spatial structures — are also, in many cases, those with the fastest growth rates. These simplest instabilities, including that arising in Rayleigh’s problem, are all characterized by a pattern of PV anomalies (q' anomalies) of the general sort shown schematically by the plus and minus signs in the following diagram:



The pattern can be thought of as a pair of Rossby waves propagating side by side, or one above the other, depending on whether a barotropic (Rayleigh-like) or a baroclinic instability is in question.

Viewed in a reference frame moving with the zonal phase speed c of the disturbance, each Rossby wave propagates against, and is held stationary by, the local basic flow. From the nature of the Rossby propagation mechanism (recall diagram on page 3), this is dynamically possible if the sign of the basic PV gradient is positively correlated with that of the relative zonal flow ($U - c$), i.e. both signs positive, as in the top half of the diagram, or both signs negative, as in the bottom half. This is evidently the simplest configuration consistent with the Rayleigh–Kuo and Fjørtoft necessary conditions for instability. It will be noticed that if the basic zonal flow U has a continuous profile then a ‘steering level’ or ‘critical line’ will be present, where by definition $U - c = 0$. We shall assume that the basic PV gradient is small or negligible in some region containing the critical line (as in Rayleigh’s problem); the more general case is discussed in the review article. Moreover, for expository purposes we shall restrict attention at first to patterns whose spatial scale is such that, if the induced velocity field associated with each Rossby wave in the diagram did not affect the other, then their phase propagation would be somewhat too slow to hold them stationary against the basic zonal flow.

The essence of the instability mechanism is that the induced velocity fields do, however, overlap significantly. That is why the width $2b$ in Rayleigh’s problem, if instability is to occur at a given wavelength $2\pi/k$, has to be of order k^{-1} or less. Similarly, in order to get a baroclinic instability of horizontal scale L , say, and simple spatial structure, the vertical separation between the two rows of PV anomalies has to be of the order of one Rossby height fL/N or less, as illustrated in Dr Haynes’ lectures by the Eady baroclinic instability problem. The overlapping of the induced velocity fields has the following consequences, under the assumed conditions:

- (i) Inasmuch as the PV anomaly patterns are less than a quarter wavelength out of phase with each other, the case shown in the diagram, each half helps the other to propagate against the basic zonal flow. That is, the contributions to the northward velocity induced by each PV pattern partially reinforce each other, making the phase of each pattern propagate upstream faster than it would in isolation. This is how the patterns hold themselves stationary against the basic flow, under the assumed conditions.
- (ii) Because of this interdependence between the two counterpropagating Rossby waves, their relative phase tends to lock on to a configuration like that shown. For if the PV patterns were each to shift slightly downstream, i.e. the upper pattern towards the right and the lower towards the left, so as to be more nearly in phase, then each half would help the other to propagate still more strongly, moving the patterns back upstream towards their original relative positions. Conversely, if the patterns were shifted upstream, so as to be more out of phase, then propagation would be weakened, and advection by the basic zonal flow would tend to restore the original phase relation.
- (iii) Just as in the diagram on page 3, the northward velocity induced by the upper PV pattern alone is a quarter wavelength out of phase with that pattern. The large black dot in the diagram marks the position of the northward velocity maximum induced by the upper PV pattern alone, for the right-hand-most wave period. This is *less than a quarter wavelength* out of phase with the bottom PV pattern, and therefore with the bottom displacement pattern, as indicated by the position of the small black dot directly below. If we add the velocities induced by the bottom PV pattern (open dots) to get the total velocity field, we see at once that the total velocity is also less than a quarter wavelength out of phase with the displacement pattern. This is true on the top level as well as on the bottom level.
- (iv) It follows that the total northward velocity field in each half of the disturbance can be regarded as a sum of sinusoidal contributions in phase with, and a quarter-wavelength

out of phase with, the northward displacement field. Moreover, the in-phase contribution has the same sign as the displacement. A velocity in phase with the corresponding displacement implies, by simple kinematics, that both must be growing.

The instability mechanism just described can be summarized in one sentence, by saying that

‘The induced velocity field of each Rossby wave
keeps the other in step, and makes the other grow.’

These two effects of the induced velocity field are associated respectively with its in-quadrature and in-phase contributions. The pure, exponentially-growing normal mode of linear instability theory describes a situation in which the two PV anomaly patterns have locked on to each other and settled down to a common phase speed c , such that the rates of growth which each induces in the other are precisely equal, allowing the shape of the pattern as a whole to become precisely fixed, and the growth of all disturbance quantities precisely exponential.

Cases in which the spatial scale is sufficiently large that each wave in isolation would propagate *faster* than the basic zonal flow can be understood in the essentially same way. The main changes needed are in statement (i) of the foregoing, where ‘help’ is replaced by ‘hinder’, ‘faster’ by ‘slower’, and so on. Whereas in the ‘helping’ case the phase shift between the two PV patterns is less than 0.25 of a wavelength, as shown in the diagram, in the ‘hindering’ case the phase shift lies between 0.25 and 0.5 of a wavelength. The relative phase tends to lock on just as before, and the summarizing statement (69) remains true.

In fact this latter case is usually the one which exhibits the largest growth rates, as would generally be expected from the fact that a larger phase shift between the two PV anomaly patterns enables the total induced velocity to be more nearly in phase with the displacement, tending to give a larger growth rate. This is exemplified both by the Rayleigh and by the Eady problem. It can also be checked, as already done for the Rayleigh problem above, that the phase shifts in the northward velocity and geopotential height anomaly patterns are indeed substantially less than those in the corresponding PV anomaly patterns (respectively 0.18 and 0.25 of a wavelength at maximum growth rate, in the two examples), as suggested by the diagram. This can be looked upon as another consequence of the smoothing property of the inversion operator.

6. Suppression of shear instabilities by boundary constraints

Arnol’d’s second stability theorem (discovered in the 1960s, but not widely known until the 1980s) proves that there are cases where neither the Rayleigh–Kuo theorem nor the Fjørtoft theorem rules out instability, yet where the flow is stable (indeed, stable in a certain finite-amplitude sense). These are cases with side boundaries so close to the shear layer that the Rossby-wave propagation mechanism does not have room to operate sufficiently strongly to hold a phase-locked configuration. (This is again a manifestation of the *scale effect* in the vorticity inversion operator. It shows up also in the simple plane-wave dispersion relation $c - U = -\beta/(k^2 + \ell^2)$; when y -wavenumbers ℓ become large, as would be necessary to fit the waves into a narrow channel, intrinsic phase speeds $c - U$ become small. See also the Exercise at the end of section 3.)

A relevant case is where $\beta = 0$ and $U(y)$ is of the form $\sin(ay)$, in which case the critical channel width $2L = \pi/a$. So for instance if $|a| = 0.25\pi$ — note that you can type `sin(.25*pi*y)` when inputting the U profile — then Arnol’d’s second theorem implies that making the half channel width L less than its default value 2 stabilizes the flow (e.g. McIntyre and Shepherd 1987, *J. Fluid Mech.* **181**, pp. 542, 543). (The result is well known to specialists in instability theory, albeit missed by at least one of the standard monographs on hydrodynamic instability theory!) When L just exceeds 2, only the longest wavelengths (smallest k values) are unstable, and only weakly. Try for instance k values between 0.01 and 0.1, and $L = 2.01, 2.02, 2.05$. You may need a relatively fine grid value, say 38. This is quite a delicate check on the correctness of the computer program!

7. The continuous spectrum of singular neutral modes

You may be wondering about the origin of the large number of neutral modes that are always found in the computer demonstration (modes with real c and therefore neither growing nor decaying). A few of these may be ordinary Rossby waves, especially for the larger values of β ; but the majority are likely to correspond, within numerical discretization error, to what is referred to in the literature as the ‘continuous spectrum’ of singular neutral modes. Being singular, these cannot be properly represented by a general-purpose numerical method; but their presence in the continuous problem is likely to be the main reason for the appearance of many neutral modes in the discretized problem.

Quite unlike the instability in which we are interested (which begins as an undulation of the pre-existing vorticity distribution), the continuous spectrum modes, and their superpositions including what are called ‘sheared disturbances’, correspond to *artificially changing the initial vorticity distribution* — more precisely, artificially changing the vorticities of fluid elements by small amounts that have an oscillatory x -dependence — and then letting the system evolve freely. Such (weak) vorticity distributions tend to be sheared over, and thus tend to develop increasingly fine scales in the y direction, as one might expect of a quantity advected by a total velocity field that is close to pure shear. A consideration of such disturbances is necessary for a full mathematical understanding of the instability problem with arbitrary initial conditions, but is not of primary interest here.

However, it is easy to say simply but precisely what the continuous-spectrum neutral modes are, which may be useful since, despite the clear explanation in Rayleigh’s *Theory of Sound*, p. 391, the subsequent literature contains a certain amount of confusion over what is fundamentally a simple technical point. A singular neutral mode of the continuous spectrum, for given wavenumber k , is a disturbance with a non-zero velocity jump $\Delta_\eta u$ on a single material contour $y = y_0 + \eta$, with $\Delta_\eta u$ varying like $\sin\{kx - kU(y_0)t + \text{constant}\}$ along the contour. In other words, it corresponds to a frozen, sinusoidally varying *sheet of vorticity* inserted as an initial disturbance on exactly the one material contour. Here η is the displacement in the y direction as before. (In order to be a normal mode, i.e. to have constant spatial shape as time goes on, the whole disturbance generally has to involve undulations of the remaining material contours and hence, in general, a smoothly-varying distribution of disturbance vorticity at any other $y \neq y_0$.) Such modes are said to belong to a ‘continuous spectrum’ because they have frequencies $kU(y_0)$ that vary continuously as y_0 varies. Of course it takes at least two such modes, with vorticity sheets located at two values $y_1 \neq y_0$ of y and advected at different speeds $U(y_1) \neq U(y_0)$, to begin to describe the ‘shearing-over’ effect; more usually, one has a continuous superposition expressed by an integral.

8. Other basic instabilities, especially a 3D one recently discovered

It is arguable that the inviscid shear instability described and analyzed in §§1–6 above is representative of one of the most basic, quintessentially fluid-dynamical classes of instabilities, underlying much high-Reynolds-number fluid-dynamical behaviour both large and small scale, stratified and unstratified, layerwise-two-dimensional (moderate to large Richardson number) and fully three-dimensional (small to zero Richardson number). That is why I have concentrated on it. There are of course many other kinds of fluid instabilities, some of an obvious kind, such as the convective or Rayleigh-Taylor instability associated with negative stratification ($N^2 < 0$), and others less so, such as the ‘elliptic instability’ recently discovered by Pierrehumbert and Bayly* which, although

* A Soviet colleague, Dr V.A. Vladimirov, has pointed out that the instability is essentially the same as that discovered by Tsai and Widnall (JFM **73**, no 4, 1976) and by Gledzer et al (Izv. Akad. Nauk SSSR, FAO, **11**, no 10, 1975). See also Gledzer, E. B., Ponomarev, V. M., 1992, J. Fluid Mech. **240**, 1–30.

its significance is still being assessed, seems likely to be another robust paradigm and very basic to an understanding of many fully three-dimensional flows, such as small-scale turbulent mixing (but not layerwise-two-dimensional flows).

A discussion of the elliptic and related instabilities is beyond the scope of the present lectures, although it might be brought in on future occasions, especially if we can develop some suitable computer demonstrations. The physical mechanism is entirely different from the above — it appears to be more closely akin to the Mathieu parametric instability of a pendulum whose point of support is oscillated, and to the ‘resonant triad’ wave–wave interactions that have been extensively studied in connection with oceanic surface and internal gravity waves. In the meantime the interested reader may consult the review by Bayly et al, 1988, *Ann. Rev. Fluid Mech.*, **20**, especially pages 381–384.

9. Suppression of shear instabilities by a large-scale strain field

An important paper analyzing this effect is

Dritschel, D. G., Haynes, P. H., Jukes, M. N., Shepherd, T. G., 1991, *J. Fluid Mech.*, **230**, 647–665.

Such a suppression of shear instabilities is essential to understand the existence of thin filaments of vorticity that appear in simulations such as that shown in my lecture notes, transparencies MEM 35–6, 57, etc. (Meticulous checks were done in these cases to make sure that any small-scale shear instability would be resolved numerically if it occurred.)

This implies an important qualification to the earlier remarks about robustness. Shear instability is robust to finiteness of disturbance amplitude, but not to large-scale strain fields that are stretching the filaments. Stabilization by such stretching can occur for strain rates only a modest fraction, often a sixth or so, of the vorticity contrast in the shear flow. This fact, and its generalization to baroclinic cases, is often critical to the ‘mesoscale developments’, or lack thereof, that in turn can be critical to weather forecasting.