Bounds on the growth rate and phase velocity of instabilities in non-divergent barotropic flow on a sphere: A semicircle theorem

By JOHN THUBURN\(^1\)* and PETER H. HAYNES\(^2\)

\(^1\)University of Reading, UK
\(^2\)University of Cambridge, UK

(Received 21 March 1995; revised 6 November 1995)

**SUMMARY**

Bounds are presented for the growth rate and angular phase velocity of unstable modes in non-divergent barotropic flow on a sphere. The bounds are given in terms of the minimum and maximum basic-flow angular velocity. The result is analogous to the earlier results of Howard and of Pedlosky for plane parallel shear flow and for quasi-geostrophic flow on a \(\beta\)-plane. An improvement to the earlier \(\beta\)-plane result, giving tighter bounds, is also presented.

**KEYWORDS:** Barotropic instability Rossby waves Semicircle theorem

1. **INTRODUCTION**

(a) **Bounds on the growth rates and phase velocities of unstable modes**

Only a few special cases can the growth rates and phase velocities of hydrodynamically unstable modes be calculated analytically; usually they must be found by numerical methods. However, there are some general results that give useful bounds on the growth rates and phase velocities of unstable modes in terms of simple basic-state quantities, such as the minimum and maximum basic-state velocity. For example, such bounds may be used to say almost immediately whether an observed disturbance in the atmosphere or ocean may be, or cannot possibly be, caused by a certain kind of normal mode instability, before embarking on a detailed investigation.

One famous example of such a set of bounds is given by Howard’s (1961) semicircle theorem. For incompressible flow on a plane, \((x, y)\), periodic or unbounded in the \(x\)-direction and with suitable boundary conditions in the \(y\)-direction, the complex phase velocity, \(c \equiv c_t + ic_i\), of small amplitude unstable disturbances in a basic flow, \(U(y)\), must satisfy the conditions,

\[
\left\{ c_t - \frac{1}{2}(U_{\text{min}} + U_{\text{max}}) \right\}^2 + c_i^2 \leq \left\{ \frac{1}{2}(U_{\text{max}} - U_{\text{min}}) \right\}^2
\]

\[
c_i > 0
\]

where \(U_{\text{min}}\) is the minimum value of \(U\) and \(U_{\text{max}}\) is the maximum value of \(U\). This means that \(c\) must lie within a semicircle in the complex plane, centred at \(c = \frac{1}{2}(U_{\text{min}} + U_{\text{max}})\) and of radius \(\frac{1}{2}(U_{\text{max}} - U_{\text{min}})\) (Fig. 1).

Pedlosky (1964) (see also Pedlosky 1987) extended this result to the case of three-dimensional quasi-geostrophic flow in a channel of width \(L\) on a \(\beta\)-plane for a basic state \(U(y, z)\). In this case the semicircle is somewhat larger, i.e.

\[
\left\{ c_t - \frac{1}{2}(U_{\text{min}} + U_{\text{max}}) \right\}^2 + c_i^2 \leq \left\{ \frac{1}{2}(U_{\text{max}} - U_{\text{min}}) \right\}^2 + \frac{\beta L^2}{2\pi^2}(U_{\text{max}} - U_{\text{min}})
\]

\[
c_i > 0
\]

(assuming that \(\beta > 0\)). He also showed that

\[
U_{\text{min}} - \frac{\beta L^2}{2\pi^2} \leq c_t \leq U_{\text{max}}
\]

so that the permitted region for \(c\) is a truncated semicircle (Fig. 2). In section 3 we shall improve these bounds by showing that a part of this apparently permitted region is actually forbidden.

* Corresponding author: Department of Meteorology, University of Reading, 2 Earley Gate, Whiteknights, Reading, Berkshire, RG6 2AU.
Figure 1. Schematic showing the permitted region for the complex phase velocity \( c = c_r + ic_i \) in Howard's semicircle theorem.

Figure 2. Schematic showing the permitted region for the complex phase velocity \( c = c_r + ic_i \) in Pedlosky's semicircle theorem. The shaded region corresponds to modes whose phase velocity is everywhere westward relative to the basic flow. The dashed line shows the reduction in the permitted region demonstrated by (36).

Note that these bounds apply only to unstable modes; neutral modes may lie outside the semicircles.

In section 2 of this paper a corresponding result will be derived for non-divergent barotropic flow in spherical geometry. This new result may be more relevant than the previous results to observed disturbances in the winter polar stratosphere, such as the '4-day wave' (Venne and Stanford 1979; Prata 1984; Lait and Stanford 1988) for which barotropic instability has been proposed as a possible mechanism (Hartmann 1983; Manney et al. 1988).

In spherical geometry it is the complex absolute angular phase velocity of the disturbance, \( \omega/m = (\omega_t + i\omega_i)/m \), where \( \omega_t \) is the angular frequency and \( m \) is the zonal wave-number, that is bounded in terms of the minimum and maximum absolute angular velocity of the basic flow, \( \Omega_{\text{min}} \) and \( \Omega_{\text{max}} \), viz.

\[
\left\{ \frac{\omega_t}{m} - \frac{1}{2} (\Omega_{\text{min}} + \Omega_{\text{max}}) \right\}^2 + \left\{ \frac{\omega_i}{m} \right\}^2 \leq \rho^2 \quad \omega_i > 0
\]

(4)

where \( \rho \) is the radius of the semicircle. In this case the size and shape of the permitted region for \( \omega/m \) depends on whether \( \Omega_{\text{min}} \) and \( \Omega_{\text{max}} \) have different signs or the same sign, and on the relative magnitudes of \( \Omega_{\text{min}} \) and \( \Omega_{\text{max}} \). If \( \Omega_{\text{min}} < 0 < \Omega_{\text{max}} \) then the permitted region for \( \omega/m \) is a semicircle similar to that in Howard's theorem:

\[
\rho^2 = \left\{ \frac{1}{2} (\Omega_{\text{max}} - \Omega_{\text{min}}) \right\}^2
\]

(5)

(Fig. 3).

If, on the other hand, \( \Omega_{\text{min}} \) and \( \Omega_{\text{max}} \) have the same sign, for example, as might occur on a rapidly rotating planet with small basic-state relative angular velocity, then the radius of the semicircle is larger.
NOTES AND CORRESPONDENCE

Figure 3. Schematic showing the permitted region for the complex angular phase velocity \( \omega/m = \omega_r/m + i\omega_i/m \) in the new semicircle theorem when \( \Omega_{\text{min}} < 0 < \Omega_{\text{max}} \).

Figure 4. Schematic showing the permitted region for the complex angular phase velocity \( \omega/m = \omega_r/m + i\omega_i/m \) in the new semicircle theorem when \( 0 < \Omega_{\text{min}} < \Omega_{\text{max}}/5 \). (When \( \Omega_{\text{max}}/5 < \Omega_{\text{min}} \) the semicircle is not truncated on the left-hand side.) The shaded region corresponds to modes whose angular phase velocity is everywhere westward relative to the basic flow.

Suppose, without loss of generality, that \( 0 < \Omega_{\text{min}} < \Omega_{\text{max}} \) and define \( r = \Omega_{\text{min}}/\Omega_{\text{max}} \). Then

\[
\rho^2 = \left\{ \frac{1}{2} (\Omega_{\text{max}} - \Omega_{\text{min}}) \right\}^2 + \frac{1}{2} \Omega_{\text{min}} \Omega_{\text{max}} \quad (0 < r < 3 - 2\sqrt{2})
\]

\[
= \left\{ \frac{1}{2} (\Omega_{\text{max}} - \Omega_{\text{min}}) \right\}^2 + \frac{1}{16} (\Omega_{\text{min}} + \Omega_{\text{max}})^2 \quad (3 - 2\sqrt{2} < r < 1/3)
\]

\[
= \left\{ \frac{1}{2} (\Omega_{\text{max}} - \Omega_{\text{min}}) \right\}^2 + \frac{1}{2} \Omega_{\text{min}} (\Omega_{\text{max}} - \Omega_{\text{min}}) \quad (1/3 < r < 1)
\]

(6)

In addition, \( \omega/m \) is confined to lie within a semicircle of radius \( \Omega_{\text{max}} - \Omega_{\text{min}} \) centred on \( \omega/m = \Omega_{\text{min}} \), i.e.

\[
\left\{ \frac{\omega_r}{m} - \Omega_{\text{min}} \right\}^2 = \left\{ \frac{\omega_i}{m} \right\}^2 \leq (\Omega_{\text{max}} - \Omega_{\text{min}})^2
\]

so that part of the region allowed by (4) and (6), including the region with \( \omega_r/m > \Omega_{\text{max}} \), is forbidden by (7). Furthermore, \( \omega_r/m \) cannot be smaller than \( \Omega_{\text{min}}/2 \), so that if \( r < 1/5 \) then the semicircle is truncated at the left-hand edge (Fig. 4).

The derivation of the new semicircle theorem is given in the next section. Some issues that arise are discussed in section 4.

2. DERIVATION OF THE SEMICIRCLE THEOREM

The derivation, essentially, follows the same steps as the derivation of Howard's and Pedlosky's results. The non-divergent barotropic vorticity equation on a non-rotating sphere (i.e. in an inertial frame) of unit radius, linearized about a steady zonally symmetric basic state, has the form

\[
\xi_t + \psi \xi_0 \mu - \xi_1 \psi_0 \mu = 0
\]

(8)
where $\psi$ is the stream function and $\zeta$ is the vorticity, $\lambda$ is longitude and $\mu$ is the sine of the latitude; subscript 0 indicates a basic-state quantity while a prime indicates a perturbation quantity. (For a rotating sphere it is simplest to transform immediately to a non-rotating frame by adding a 'solid body' rotation to the basic-state flow.) The vorticity and the stream function are related by

$$\zeta = \nabla^2 \psi \equiv \{(1 - \mu^2)\psi_{\mu}\}_{\mu} + \frac{\psi_{\lambda\lambda}}{(1 - \mu^2)}. \tag{9}$$

The perturbation can be written as a sum of disturbances of the form

$$\psi' = \text{Re}\{\hat{\psi}(\mu) e^{i(m \lambda - \omega t)}\}. \tag{10}$$

From now on restrict attention to a single growing mode, for which $\omega_i > 0$. Substituting (10) in (8) and writing $\Omega$ for the angular velocity of the basic flow, $-\psi_{\theta\mu}$, gives

$$(\Omega - \omega/m) \left[ \{(1 - \mu^2)\hat{\psi}_{\mu}\}_{\mu} - \frac{m^2 \hat{\psi}}{(1 - \mu^2)} \right] + \zeta_{\theta\mu} \hat{\psi} = 0. \tag{11}$$

Now define two new quantities, the (complex) relative angular phase velocity,

$$R(\mu) \equiv \{\omega/m - \Omega(\mu)\}, \tag{12}$$

and the parcel displacement parallel to the axis of symmetry of the basic flow, $n'$. A quantity $\hat{n}(\mu)$ can then be defined via an equation like (10), and this is related to $\hat{\psi}$ by

$$- R \hat{n} = \hat{\psi}. \tag{13}$$

($R$ is non-zero because $\omega_i > 0$). Substituting (12) and (13) in (11) gives

$$- R \left[ \{(1 - \mu^2)(\hat{R}n)_{\mu}\}_{\mu} - \frac{m^2 R \hat{n}}{(1 - \mu^2)} \right] + R \hat{n}\{(1 - \mu^2)(R - \omega/m)\}_{\mu\mu} = 0. \tag{14}$$

Next, with the aid of the identity

$$R \hat{n}\{(1 - \mu^2)R\}_{\mu\mu} - R\{(1 - \mu^2)(\hat{R}n)_{\mu}\}_{\mu} + (1 - \mu^2)^{-1/2} \{(1 - \mu^2)^{-1/2}\hat{n}\}_{\mu}(1 - \mu^2)^2 R^2 \}_{\mu} \equiv - \frac{\hat{n} R^2}{(1 - \mu^2)} \tag{15}$$

(14) can be put into self-adjoint form, viz.

$$-(1 - \mu^2)^{-1/2} \{(1 - \mu^2)^{-1/2}\hat{n}\}_{\mu}(1 - \mu^2)^2 R^2 \}_{\mu} + \frac{(m^2 - 1)}{(1 - \mu^2)} R^2 \hat{n} + 2 \left( \frac{\omega}{m} \right) R \hat{n} = 0. \tag{16}$$

Now multiply (16) by $\hat{n}^*$, the complex conjugate of $\hat{n}$, and integrate from pole to pole, integrating the first term by parts and using the fact that $\hat{n} \to 0$ and $\hat{n}_{\mu}$ is finite at the poles to eliminate the boundary terms, to obtain

$$\int_{-1}^{1} P R^2 \, d\mu + \frac{2\omega}{m} \int_{-1}^{1} Q R \, d\mu = 0. \tag{17}$$

Here $P$ is a real positive definite quantity, given by

$$P = \left| \frac{\hat{n}}{(1 - \mu^2)^{1/2}} \right|_{\mu}^2 \{(1 - \mu^2)^{-1/2}\hat{n}\}_{\mu}(1 - \mu^2)^2 R^2 \}_{\mu} + \frac{(m^2 - 1)}{(1 - \mu^2)} |\hat{n}|^2 \tag{18}$$

(note that $m$ must be an integer greater than or equal to 1), and $Q$ is another real positive definite quantity,

$$Q = |\hat{n}|^2. \tag{19}$$
Collecting the imaginary terms in (17), and using the assumption that \( \omega_t > 0 \), gives
\[
\frac{\omega_t}{m} = \frac{\int_{-1}^{1} (P + Q) \Omega \, d\mu}{\int_{-1}^{1} (P + 2Q) \, d\mu}.
\] (20)

Now define \( \sigma \) by the equation
\[
\sigma = \frac{\int_{-1}^{1} (P + Q) \, d\mu}{\int_{-1}^{1} (P + 2Q) \, d\mu},
\] (21)

and note that \( \frac{1}{2} \leq \sigma \leq 1 \). Then (20) implies that
\[
\sigma \Omega_{\text{min}} \leq \frac{\omega_t}{m} \leq \sigma \Omega_{\text{max}}.
\] (22)

In particular, if \( \Omega_{\text{min}} < 0 < \Omega_{\text{max}} \) then \( \Omega_{\text{min}} \leq \frac{\omega_t}{m} \leq \sigma \Omega_{\text{max}} \), but if \( 0 < \Omega_{\text{min}} < \Omega_{\text{max}} \) then \( \Omega_{\text{min}}/2 \leq \frac{\omega_t}{m} \leq \Omega_{\text{max}} \), and \( \omega_t/m \) may possibly lie between \( \Omega_{\text{min}}/2 \) and \( \Omega_{\text{min}} \).

The real terms in (17) give
\[
- \left\{ \left( \frac{\omega_t}{m} \right)^2 + \left( \frac{\omega_t}{m} \right)^2 \right\} \int_{-1}^{1} (P + 2Q) \, d\mu + \int_{-1}^{1} P \Omega^2 \, d\mu = 0.
\] (23)

From this point we can use two different approaches to derive semicircle inequalities. The ratio \( \Omega_{\text{min}}/\Omega_{\text{max}} \) determines which approach gives the tightest bound.

The first approach uses the fact that \( Q \Omega^2 \geq 0 \) to turn (23) into an inequality, viz.
\[
\left\{ \left( \frac{\omega_t}{m} \right)^2 + \left( \frac{\omega_t}{m} \right)^2 \right\} \int_{-1}^{1} (P + 2Q) \, d\mu \leq \int_{-1}^{1} (P + Q) \Omega^2 \, d\mu.
\] (24)

Then the use of the obvious inequality \( (\Omega - \Omega_{\text{min}})(\Omega - \Omega_{\text{max}}) \leq 0 \) allows this to be written as
\[
\left\{ \left( \frac{\omega_t}{m} \right)^2 + \left( \frac{\omega_t}{m} \right)^2 \right\} \int_{-1}^{1} (P + 2Q) \, d\mu \leq \int_{-1}^{1} (P + Q)(\Omega_{\text{min}} + \Omega_{\text{max}} - \Omega_{\text{min}} \Omega_{\text{max}}) \, d\mu,
\] (25)

which can be simplified using (20) and (21) to give
\[
\left( \frac{\omega_t}{m} \right)^2 + \left( \frac{\omega_t}{m} \right)^2 \leq \frac{\omega_t}{m} (\Omega_{\text{min}} + \Omega_{\text{max}}) - \sigma \Omega_{\text{min}} \Omega_{\text{max}}.
\] (26)

Rearranging this gives the final form:
\[
\left\{ \left( \frac{\omega_t}{m} - \frac{1}{2} (\Omega_{\text{min}} + \Omega_{\text{max}}) \right)^2 + \left( \frac{\omega_t}{m} \right)^2 \right\} \leq \left\{ \frac{1}{2} (\Omega_{\text{max}} - \Omega_{\text{min}}) \right\}^2 + (1 - \sigma) \Omega_{\text{min}} \Omega_{\text{max}}.
\] (27)

If \( \Omega_{\text{min}} \) and \( \Omega_{\text{max}} \) have different signs then the last term on the right-hand side is negative, and (27) implies (4) with \( \rho^2 \) given by (5), while if they have the same sign then the last term on the right-hand side is positive but less than or equal to \( \frac{1}{2} \Omega_{\text{min}} \Omega_{\text{max}} \), and (27) implies (4) with \( \rho^2 \) given by the first case of (6).

Returning to (23), the alternative approach is to rewrite it as
\[
\left\{ \left( \frac{\omega_t}{m} \right)^2 + \left( \frac{\omega_t}{m} \right)^2 \right\} = \frac{\int_{-1}^{1} P \Omega^2 \, d\mu}{\int_{-1}^{1} (P + 2Q) \, d\mu}.
\] (28)

Combining (28) with (20) gives
\[
\left( \frac{\omega_t}{m} - \Omega_0 \right)^2 + \left( \frac{\omega_t}{m} \right)^2 = \frac{\int_{-1}^{1} (P + 2Q)(\Omega - \Omega_0)^2 \, d\mu}{\int_{-1}^{1} (P + 2Q) \, d\mu} - \frac{\int_{-1}^{1} Q \Omega (\Omega - \Omega_0) \, d\mu}{\int_{-1}^{1} (P + 2Q) \, d\mu}
\] (29)
for any constant \( \omega_0 \). By choosing different values for \( \Omega_0 \) different sets of bounds can be found. Consider \( \Omega_0 = \frac{1}{2} (\Omega_{\min} + \Omega_{\max}) \), and note that

\[
(\Omega - \Omega_0)^2 \leq \left( \frac{1}{2} (\Omega_{\max} - \Omega_{\min}) \right)^2
\]

and that

\[
\Omega (\Omega - \Omega_0) \geq -\frac{1}{16} (\Omega_{\min} + \Omega_{\max})^2.
\]

In fact, if \( \Omega_{\min} \geq \frac{1}{3} \Omega_{\max} \) then a tighter bound holds, viz.

\[
\Omega (\Omega - \Omega_0) \geq -\frac{1}{2} \Omega_{\min} (\Omega_{\max} - \Omega_{\min}).
\]

Hence

\[
\left\{ \frac{\omega_r}{m} - \frac{1}{2} (\Omega_{\min} + \Omega_{\max}) \right\}^2 + \left\{ \frac{\omega_\theta}{m} \right\}^2 \leq \left( \frac{1}{2} (\Omega_{\max} - \Omega_{\min}) \right)^2 + \frac{1}{16} (\Omega_{\min} + \Omega_{\max})^2.
\]

and if \( \Omega_{\min} \geq \frac{1}{3} \Omega_{\max} \) then

\[
\left\{ \frac{\omega_r}{m} - \frac{1}{2} (\Omega_{\min} + \Omega_{\max}) \right\}^2 + \left\{ \frac{\omega_\theta}{m} \right\}^2 \leq \left( \frac{1}{2} (\Omega_{\max} - \Omega_{\min}) \right)^2 + \frac{1}{2} \Omega_{\min} (\Omega_{\max} - \Omega_{\min}).
\]

It may be verified that (27) gives the semicircle of smallest radius for \( \Omega_{\min} < (3 - 2\sqrt{2}) \Omega_{\max} \), that (33) gives the semicircle of smallest radius for \( (3 - 2\sqrt{2}) \Omega_{\max} \leq \Omega_{\min} < \frac{1}{2} \Omega_{\max} \), and that (34) gives the semicircle of smallest radius when it applies, namely when \( \frac{1}{2} \Omega_{\max} \leq \Omega_{\min} \).

Now consider (29) with \( \omega_0 = \Omega_{\min} \) for the case \( \Omega_{\min} > 0 \). Then \( (\Omega - \Omega_0)^2 \leq (\Omega_{\max} - \Omega_{\min})^2 \) and \( \Omega (\Omega - \Omega_0) \geq 0 \), implying (7). This concludes the derivation of the new semicircle theorem.

3. IMPROVED BOUNDS IN THE \( \beta \)-PLANE CASE

An equation analogous to (29) above can be obtained, essentially, by following the same steps for the case of quasi-geostrophic flow in a \( \beta \)-plane channel, viz.

\[
(c_r - U_0)^2 + c_\theta^2 = \frac{\int \text{d}y \text{d}z (U - U_0)^2 \text{d}y \text{d}z}{\int \text{d}y \text{d}z P} - \beta \frac{\int \text{d}y \text{d}z (U - U_0)Q \text{d}y \text{d}z}{\int \text{d}y \text{d}z P}.
\]

Here \( P \) and \( Q \) are positive definite quantities (quasi-geostrophic \( \beta \)-plane analogues of the quantities defined by (18) and (19)), and \( U_0 \) is any constant. Substituting \( U_0 = \frac{1}{2} (U_{\min} + U_{\max}) \) and using the fact that \( \int \text{d}y \text{d}z Q/\int \text{d}y \text{d}z P \) is bounded above by \( L^2/\pi^2 \) gives the semicircle found by Pedlosky, (2). Substituting \( U_0 = U_{\min} \) and noting that \( U - U_0 > 0 \) leads to another set of bounds, viz.

\[
(c_r - U_{\min})^2 + c_\theta^2 \leq (U_{\max} - U_{\min})^2.
\]

The inequality (36) forbid some of the region permitted by (2) and (3). The relevant part of the boundary of the permitted region is shown as a dashed curve in Fig. 2.

4. DISCUSSION

(a) Instability and critical layers

One corollary of Howard's theorem (actually proved earlier by Rayleigh) is that the real part of the phase velocity of an unstable mode, \( c_r \), must lie between \( U_{\min} \) and \( U_{\max} \); in particular, a marginally stable mode, one for which \( c_r \to 0^+ \), must have a critical layer. Lindzen and Tung (1978) and Lindzen, Farrell and Tung (1980) have argued strongly that the onset of barotropic and baroclinic instability can be understood in terms of overreflection of waves at critical layers. The issue has been debated, for example,
by Takahashi (1986, 1988), and Lindzen and Tung (1988). Equations (2) and (6) appear to permit the existence of unstable modes, including marginally stable modes, whose (angular) phase velocity lies outside the range of the basic flow (angular) velocity (the shaded regions in Figs. 2 and 4). Such modes could not be explained in terms of overreflection, since they would not have critical layers. This raises the question of whether the shaded regions in Figs. 2 and 4 have any physical significance, or whether they are merely artifacts of the derivations.

The shaded regions arise either when there is a \(\beta\)-effect or when there is a net background rotation of the basic state, which amounts to a \(\beta\)-effect. The \(\beta\)-effect provides a Rossby wave restoring mechanism, enabling waves, which could include unstable modes, to propagate westwards relative to the basic flow (assuming \(\beta > 0\), or \(\Omega_{\max} > \Omega_{\min} > 0\)).

As it happens, there are few known examples of flows with unstable modes whose complex (angular) phase velocity is everywhere westward relative to the basic flow. One recently studied and important example is the retrograde Bickley jet on a \(\beta\)-plane. An analytical and numerical investigation of the stability of this flow showed that it has some unstable modes and marginally stable modes with \(c_r < U_{\min}\) (Maslowe 1991). This shows that the shaded region in Fig. 2 does have physical significance and is not merely an artifact of the derivation. For these modes with phase velocity outside the range of the basic-flow velocity the onset of instability cannot be explained in terms of overreflection at critical layers. Maslowe explains it in terms of the coalescence of Rossby modes: as some parameter, for example \(\beta\), is varied, two Rossby modes with distinct structures and phase speeds in the stable régime become identical at the stability boundary and become a complex conjugate pair of growing and decaying modes in the unstable régime. In physical terms this coalescence might be interpreted as a resonant interaction between two different neutral wave modes. See, for example, Hoskins et al. (1985) and references therein.

A second example is given by Pedlosky and Polvani (1987) for a two-layer quasi-geostrophic \(\beta\)-plane channel model, for which the structure and phase velocity of the unstable modes can be calculated analytically. A third example is given by the Eady model on an \(f\)-plane with oppositely sloping upper and lower boundaries. From the governing equations, given, e.g. by Brindley and Moroz (1980), it is a straightforward task to find the dispersion relation for small amplitude disturbances and hence to show that there exist unstable modes whose phase velocity is everywhere westward relative to the basic-flow velocity when the depth decreases poleward. Here the sloping boundaries permit vortex stretching, which leads to a Rossby wave propagation mechanism analogous to the \(\beta\)-effect. Both of these examples, the quasi-geostrophic two-layer model and the Eady model with sloping boundaries, include cases of marginally stable modes that do not have critical layers. In both examples the onset of instability is again associated with the coalescence of Rossby modes.

It is just possible that both the second and third examples are pathological and are exceptions to the general case: the two-layer model does not have a continuous basic state, while for the Eady model the potential vorticity gradient in the interior of the flow is identically zero. However, the example of the retrograde Bickley jet is not open to either of these criticisms. These examples imply that overreflection at critical layers cannot be a universal explanation for the onset of barotropic and baroclinic instability. They also suggest that the shaded region in Fig. 4 admitted by the new semicircle theorem is almost certainly physically significant, and not merely an artifact of the derivation.

5. Conclusions

Howard's (1961) semicircle theorem, which gives bounds on the growth rate and phase velocity of unstable modes in two-dimensional incompressible shear flow on a plane, has been generalized to spherical geometry. The new result differs from Howard's in two ways. First, it is the angular velocity of the basic flow and the angular phase velocity of the unstable mode that are relevant in spherical geometry, compared with the linear velocities in the plane case. Second, under certain circumstances the new result permits unstable modes that propagate everywhere westward relative to the basic flow, as does Pedlosky's (1964) generalization of Howard's theorem to three-dimensional quasi-geostrophic flow on a \(\beta\)-plane.

The new result suggests that the zonal mean angular velocity, rather than linear velocity, may be a useful diagnostic for helping to interpret modelling results and observations. For example, barotropic instability has been proposed as a mechanism for the 4-day wave (Hartmann 1983; Manney et al. 1988): an eastward-moving disturbance observed in the winter polar stratosphere with a spectral peak corresponding to zonal wavenumber 1 and period close to four days. In synoptic temperature maps the 4-day wave appears as one or more warm pools moving around the pole (Lait and Stanford 1988). Figure 5 shows the zonal mean zonal wind, \(U\), for July in the southern hemisphere from the CIRA climatology (Fleming et al. 1990). It also shows the corresponding angular velocity, \(\Omega = U / (a \cos \phi)\), expressed as the time taken for an air parcel to go once around the pole, \(\tau = 2\pi / \Omega\). There is a range of periods close to four days in
the high-latitude upper stratosphere where the 4-day wave is observed. On the basis of this data and the new semicircle theorem, barotropic instability cannot be excluded as a mechanism for the 4-day wave. (But note that the possibility that the 4-day wave may be a passively advected feature cannot be excluded either.) Of course a more rigorous analysis would involve calculating similar diagnostics for the actual periods when the 4-day wave is observed. The possibility of unstable modes propagating everywhere westward relative to the basic flow also has implications for the interpretation of modelling results and observations. For example, Manney et al. (1988) assumed, following overreflection theories of instability in the literature, that the angular phase velocity of an unstable mode must match the angular velocity of the basic flow at some latitude near where the absolute vorticity gradient vanishes. The discussion above shows that this is not necessarily the case.

ACKNOWLEDGEMENTS

This work was begun while one author (JT) was a student at the Department of Atmospheric, Oceanic and Planetary Physics, University of Oxford, supported by a grant from the UK Natural Environment Research Council, and continued under the UK Universities Global Atmospheric Modelling Programme, also supported by the UK Natural Environment Research Council. The other author (PHH) is a member of the Centre for Atmospheric Science, a joint initiative of the Department of Applied Mathematics and Theoretical Physics and the Department of Chemistry, University of Cambridge.

REFERENCES

Hartmann, D. L. 1983 Barotropic instability of the polar night jet stream. J. Atmos. Sci., 40, 817–835
Pedlosky, J. and Polvani, L. M. 1987 Geophysical fluid dynamics (second ed.). Springer-Verlag