Mon. Not. R. Astron. Soc. **326,** 833–851 (2001)

Non-linear dynamics of the corotation torque

N. J. Balmforth[1] and D. G. Korycansky[2★]
[1]*Baskin School of Engineering, University of California, Santa Cruz, CA 95064, USA*
[2]*CODEP, Department of Earth Sciences, University of California, Santa Cruz, CA 95064, USA*

Accepted 2001 April 25. Received 2001 April 18; in original form 2000 December 22

ABSTRACT

The excitation of spiral waves by an external perturbation in a disc deposits angular momentum in the vicinity of the corotation resonance (the radius where the speed of a rotating pattern matches the local rotation rate). We use matched asymptotic expansions to derive a reduced model that captures non-linear dynamics of the resulting torque and fluid motions. The model is similar to that derived for forced Rossby wave critical layers in geophysical fluid dynamics. Using the model we explore the saturation of the corotation torque, which occurs when the background potential (specific) vorticity is redistributed by the disturbance. We also consider the effects of dissipation. If there is a radial transport of potential vorticity, the corotation torque does not saturate. The main application is to the creation, growth and migration of protoplanets within discs like the primordial solar nebula. The disturbance also nucleates vortices in the vicinity of corotation, which may spark further epochs of planet formation.

Key words: accretion, accretion discs – hydrodynamics – methods: analytical.

1 THE ASTROPHYSICAL PROBLEM

External perturbations generate disturbances in discs that take the form of spiral waves. These waves can transfer angular momentum and may therefore play an important role in shaping the disc itself. Moreover, the angular momentum transfer exerts a torque on the perturber that can cause the orbital migration of this object. As a result, it has been suggested that the dynamics of disturbed discs are relevant to how satellite galaxies might drive spiral structure (Goldreich & Tremaine 1979), to how planetary rings may be sculpted by moons (Goldreich & Tremaine 1980) and to accretion flow in circumstellar discs with binary companions (Lin & Papaloizou 1979). More recent applications extend beyond discs: the interaction of binary stars can be mediated by wave generation, leading to orbital circularization and synchronization (Goldreich & Nicholson 1989). However, the original theory has lately been brought back into prominence as a result of recent observations of extrasolar planets (Mayor & Queloz 1995), which has raised questions over the evolution of protoplanets inside discs, such as how Jupiter-sized objects could occur in relatively tight orbits.

In a fluid disc, angular momentum is deposited where the excited spiral waves are dissipated or at the 'corotation resonance', the radius for which the speed of a rigidly rotating spiral pattern matches the local fluid rotation rate. In the inviscid fluid equations, this radius appears as a singular point in linear theory (that is, for disturbances of infinitesimal amplitude); it is the analogue of the 'disturbing singularity' noted by Kelvin and now known as the 'critical level' in fluid mechanics. Though the singularity is an artefact of linear theory, the apparent divergence reflects how the region surrounding the critical level (the 'critical layer') is the site of enhanced fluid dynamical activity. By analogy, we should therefore expect that, in the slender annulus surrounding corotation, the fluid motions generated by external perturbations are especially strong and non-linear effects cannot be neglected there.

The theory of the torques exerted by evolving disturbances of infinitesimal amplitude and short radial wavelength was originally sketched out by Goldreich & Tremaine (1979) and Lin & Papaloizou (1979), and was subsequently elaborated upon by other authors (Papaloizou & Lin 1984; Ward 1989; Artymowicz 1993; Korycansky & Pollack 1993). More recently, numerical simulations have gone some way to extend the results into the non-linear regime (Lin & Papaloizou 1986; Korycansky & Papaloizou 1995; Papaloizou, Korycansky & Terquem 1995; Bryden et al. 1999; Miyoshi et al. 1999). Our aim in the current work is to explore the non-linear problem in more detail. Specifically, we provide some analytical and numerical results that complement and extend existing theory.

★E-mail: kory@es.ucsc.edu

Mathematically, there is a close relation of the current problem to that formulated in geophysical fluid mechanics to study the critical layers of forced Rossby waves, a problem of notable importance to mixing in the atmosphere (Stewartson 1978; Warn & Warn 1978; Killworth & McIntyre 1985). In that subject, modelling has advanced beyond linear theory and has explored the non-linear dynamics in some detail. Notably, inside the critical layer of the Rossby wave, vortices form that modify the momentum balance and ultimately halt the absorption of momentum. In the astrophysical context, a parallel result would suggest that the angular momentum deposition predicted by linear theory is purely a transient. In other words, the corotation torque saturates through non-linear effects. Here we derive such a result using a similar theoretical analysis to the developments of the geophysical problem.

2 A REDUCED MODEL

2.1 Governing equations

The disc has an equilibrium, axisymmetrical surface density and angular velocity profile, $\Sigma(r)$ and $\Omega(r)$, that are maintained by a central gravitational potential, $-GM_*/r$, where G is the gravitational constant and M_* is the mass of a central star (r is radius). To generate a disturbance in the disc, we further include a perturbing object that lies on a circular orbit of radius r_c, and has gravitational potential, $\Phi_g(r,\theta)$, stemming from a point mass, M_p. The perturber is relatively small, $M_p \ll M_*$, and is assumed not to influence the central star.

We begin the mathematical discussion with the non-linear equations for disturbances of a two-dimensional disc of inviscid, compressible fluid without self-gravity [we ignore three-dimensional effects, which might be important (Lubow & Ogilvie 1998)]. To prepare the way for our exploration of the problem, we quote these fluid equations in a dimensionless form in which lengths are measured by the unit r_c, speed by $V_c = r_c\Omega(r_c)$, time by $1/\Omega(r_c)$, gravitational potential by V_c^2 and surface density by $\Sigma(r_c)$. Then, in polar coordinates, the disturbance velocity $(u(r,\theta,t), v(r,\theta,t))$ and surface density $\sigma(r,\theta,t)$ satisfy the equations

$$\left(\frac{\partial}{\partial t}+\Omega\frac{\partial}{\partial\theta}\right)u + u\frac{\partial u}{\partial r}+\frac{v}{r}\frac{\partial u}{\partial\theta}-2\Omega v-\frac{v^2}{r}=-\frac{\partial w}{\partial r}-\epsilon^2\frac{\partial\Phi_g}{\partial r}, \tag{1}$$

$$\left(\frac{\partial}{\partial t}+\Omega\frac{\partial}{\partial\theta}\right)v + \frac{v}{r}\frac{\partial v}{\partial\theta}+\frac{u}{r}\frac{\partial}{\partial r}(r^2\Omega)+\frac{u}{r}\frac{\partial}{\partial r}(rv)=-\frac{1}{r}\frac{\partial w}{\partial\theta}-\frac{\epsilon^2}{r}\frac{\partial\Phi_g}{\partial\theta}, \tag{2}$$

$$\left(\frac{\partial}{\partial t}+\Omega\frac{\partial}{\partial\theta}\right)\sigma + \frac{1}{r}\frac{\partial}{\partial r}(r\sigma u)+\frac{1}{r}\frac{\partial}{\partial\theta}(\sigma v)+\frac{1}{r}\frac{\partial}{\partial r}(r\Sigma u)+\frac{1}{r}\frac{\partial}{\partial\theta}(\Sigma v)=0, \tag{3}$$

where $w = w(\sigma)$ is the enthalpy disturbance (expressed in units of V_c^2). By the non-dimensionalization, the perturbing potential can be written in the form

$$\Phi_g = \sum_{m=-\infty}^{\infty}\Phi_m\, e^{im\theta}, \tag{4}$$

where

$$\Phi_m = -(1-\delta_{m,0}/2)[b^m_{1/2}(r) - r\delta_{m,1}], \qquad b^m_{1/2}(r) = \frac{2}{\pi}\int_0^\pi \frac{\cos m\theta\, d\theta}{\sqrt{1+r^2-2r\cos\theta}}, \tag{5}$$

and becomes multiplied by a factor $\epsilon^2 \equiv M_p/M_*$. In this paper we exploit the mismatch in the masses of the central star and the perturber: $\epsilon \ll 1$. Hence ϵ is a small parameter which we use momentarily to organize an asymptotic expansion.

It is convenient to derive an equation for the potential vorticity (also called vortensity in astrophysics) of the disturbance:

$$\left(\frac{\partial}{\partial t}+\Omega\frac{\partial}{\partial\theta}\right)q + u\frac{\partial q}{\partial r}+\frac{v}{r}\frac{\partial q}{\partial\theta}+u\frac{dQ}{dr}=0, \tag{6}$$

where the total potential vorticity $Q + q$ is

$$Q+q = \frac{1}{\Sigma+\sigma}\left[\frac{1}{r}\frac{d}{dr}(r^2\Omega)+\frac{1}{r}\frac{\partial}{\partial r}(rv)-\frac{1}{r}\frac{\partial u}{\partial\theta}\right], \qquad Q = \frac{1}{r\Sigma}\frac{d}{dr}(r^2\Omega)\equiv\frac{2B}{\Sigma}. \tag{7}$$

We must impose radial boundary conditions on these governing equations. In disc theory, it is customary to imagine that the disturbance decays once it propagates into the central or peripheral regions of the disc either through non-linear shock formation, or because the disc's edges are non-reflective. Thus one applies an outgoing wave condition at some inner and outer radius (Korycansky & Pollack 1993; Korycansky & Papaloizou 1995). Alternatively, one can impose perfectly reflecting conditions at these radii, such as might arise at a free boundary where the surface density vanishes. We entertain both possibilities here.

2.2 The regular expansion

The perturbing potential rotates with angular velocity $\Omega_c = \Omega(1) = 1$. In addition to being weak, we also assume that the strength of the

perturber, if it develops, does so only slowly, at least on a time-scale of order ϵ^{-1} or longer: $\Phi_g \equiv \Phi_g(r, \theta - t, \epsilon t)$. This means that, in the frame of the perturber, the fluid will also develop slowly. Hence we transform into a corotating frame and introduce a slow time-scale, $T = \epsilon t$. Then, $\partial/\partial t \rightarrow -\partial/\partial\theta + \epsilon\partial/\partial T$.

To solve the equations, we use a matched asymptotic expansion. This begins with a regular perturbation expansion in powers of ϵ. We introduce the sequences

$$u = \epsilon^2(u_0 + \epsilon u_1 + \ldots), \qquad v = \epsilon^2(v_0 + \epsilon v_1 + \ldots), \qquad w = \epsilon^2(w_0 + \epsilon w_1 + \ldots), \tag{8}$$

and so on. To lowest order,

$$(\Omega - 1)u_{0\theta} - 2\Omega v_0 + (w_0 + \Phi_g)_r = 0, \qquad (\Omega - 1)v_{0\theta} + 2Bu_0 + \frac{1}{r}(w_0 + \Phi_g)_\theta = 0, \tag{9}$$

$$r(\Omega - 1)\sigma_{0\theta} + (r\Sigma u_0)_r + \Sigma v_{0\theta} = 0, \qquad (\Omega - 1)q_{0\theta} + u_0 Q_r = 0. \tag{10}$$

The potential vorticity is

$$q_0 = -\frac{Q\sigma_0}{\Sigma} + \frac{1}{r\Sigma}[(rv_0)_r - u_{0\theta}], \tag{11}$$

where $\sigma_0 = \Sigma w_0/c^2$ and c is the dimensionless sound speed (an inverse Mach number). These are the well-known linearized disc equations that have been studied many times in the past. By decomposing each variable into Fourier series in θ, the system may be solved for each Fourier mode separately. We quote the resulting ordinary differential equation for the variable ψ, defined by

$$r\Sigma u_0 = \psi_\theta \equiv \sum_m \mathrm{i}m\psi_m \, \mathrm{e}^{\mathrm{i}m\theta}. \tag{12}$$

It is

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{rc^2}{\Sigma\Upsilon}\frac{\mathrm{d}\psi_m}{\mathrm{d}r}\right) - \left\{\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{Qr^2(\Omega - 1)}{\Upsilon}\right] + \frac{\Sigma Q^2}{\Upsilon} + \frac{m^2}{r^2\Sigma} + \frac{Q_r}{r(\Omega - 1)}\right\}\psi_m = \frac{\Sigma Q\Phi_m}{\Upsilon} + \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{r^2(\Omega - 1)\Phi_m}{\Upsilon}\right], \tag{13}$$

where $\Upsilon = c^2 - r^2(\Omega - 1)^2$ and $\kappa^2 = 4\Omega B$.

Despite the reduction to a linear, ordinary differential equation, there is a problem lurking in the leading-order solution: The equation has a singular point at corotation where Frobenius analysis indicates that there is a regular and a singular solution. The former is of the form $\psi_m \sim (r - 1)$ as $r \rightarrow 1$. However, this regular solution does not, in general, satisfy the boundary conditions. Thus our leading-order solution must contain the second Frobenius solution and therefore be irregular at corotation; this problematic feature is the analogue of Kelvin's disturbing singularity.

The singular Frobenius solution has the form

$$\psi_m \sim 1 + \frac{Q_c'}{\Omega_c'}(r - 1)\log|r - 1|, \tag{14}$$

which signifies that $q_0 \sim (r - 1)^{-1}$. Worse still, the situation is aggravated at higher order where a continuation of the expansion shows that $q_1 \sim (r - 1)^{-2}$. Thus, near corotation the asymptotic expansion seriously breaks down. In particular, for $r - 1 \sim O(\epsilon)$, the asymptotic sequence of terms becomes disordered. This signifies an inner region, a slender annulus surrounding corotation, in which we must search for another solution. This inner solution accounts explicitly for the detailed dynamics of corotation.

Because the gravitational potential of a point mass is logarithmically singular, there can be a further source of irregularity in the outer equations. It is, however, straightforward to show that this singularity is weaker. More specifically, we may take the particular solution stemming from this singular inhomogeneous term to have the form $\psi \sim (r - 1)^2\log|r - 1|$. Though this is important for the detailed outer solution, given next, it does not affect the structure of the asymptotic expansion.

2.3 Construction of the outer solution

Although there is a critical, annular region around corotation, we must still solve the leading-order equations in the remainder of the disc. Motivated by the form of the Frobenius solution, we take ψ_m to be continuous across $r = 1$. However, because the radial derivative is singular, it is not possible to enforce continuity on $\mathrm{d}\psi_m/\mathrm{d}r$. Thus, we must allow a jump in $\mathrm{d}\,\psi_m/\mathrm{d}\,r$, or equivalently v_0 across corotation. This jump is actually determined by what happens inside the critical region; in effect, the perturber generates what appears to be a ring-like source of potential vorticity at $r = 1$, which sheds spiral waves into the outer regions.

Practically, we solve the outer equations numerically using the shooting method described by Korycansky & Pollack (1993). In this method, the integration of the equations for each value of m begins from a position close to corotation; we integrate three distinct (dimensionless) solutions for ψ_m, two homogeneous solutions, u_m^α and u_m^β, and a particular solution, u_m^p, using the initial conditions

$$u_m^\alpha \rightarrow \psi_\mathrm{s}(r), \quad u_m^\alpha(1) = 1; \qquad u_m^\beta = 0, \quad \frac{\mathrm{d}u_m^\beta}{\mathrm{d}r} = 1; \qquad \text{and} \qquad u_m^\mathrm{p} = \frac{\mathrm{d}u_m^\mathrm{p}}{\mathrm{d}r} = 0. \tag{15}$$

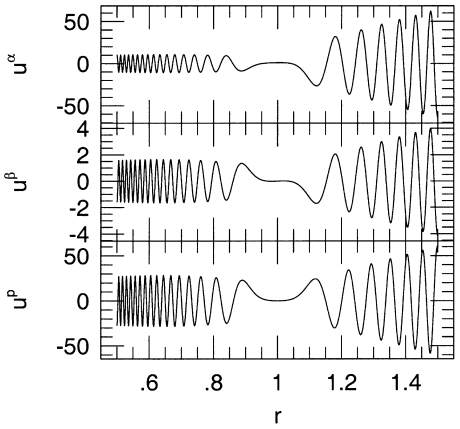

Figure 1. The dimensionless outer solutions u_m^α, u_m^β and $u_m^{\rm p}$ for $m = 10$ and sound speed $c = 10^{-1.5}$. The potential is softened with a length $r_{\rm p} = 10^{-4}$ to ease calculation; the softening is not essential. The inner boundary is $r_{\rm i} = 0.5$ and the outer boundary is $r_{\rm o} = 1.5$.

Thus u_m^α matches on to the singular Frobenius solution at corotation and is scaled to unit amplitude there, u_m^β limits to the regular solution, and the particular solution has the leading-order behaviour $u_m^{\rm p} \sim (r-1)^2 \log|r-1|$. A typical set of solutions for u is shown in Fig. 1.

Given u_m^α, u_m^β and $u_m^{\rm p}$, we construct a general solution

$$\psi_m(r) = \Psi_m u_m^\alpha(r) + J_m^\pm u_m^\beta(r) + u_m^{\rm p}(r), \tag{16}$$

where the $\pm$ refer to inside and outside corotation. With this construction, $\psi_m(r)$ contains a jump in the regular part of its derivative of $(J_m^+ - J_m^-)$, and $\psi_m(1) = \Psi_m$.

The boundary conditions at the inner and outer radii, $r_{\rm i}$ and $r_{\rm o}$ respectively, can be written in the form

$$\frac{\rm d}{{\rm d}r}\psi_m(r_{\rm i}) + C_m^{\rm i}\psi_m(r_{\rm i}) = 0 \qquad \text{and} \qquad \frac{\rm d}{{\rm d}r}\psi_m(r_{\rm o}) + C_m^{\rm o}\psi_m(r_{\rm o}) = 0, \tag{17}$$

where $C_m^{\rm i,o}$ are constants depending on the equilibrium disc structure. The coefficients are derived from a match to a second-order WKB solution as used by Korycansky & Papaloizou (1995).

Hence,

$$[\Psi_m u_m^\alpha(r_{\rm i}) + J_m^+ u_m^\beta(r_{\rm i}) + u_m^{\rm p}(r_{\rm i})]_r - C_m^{\rm i}[\Psi_m u_m^\alpha(r_{\rm i}) + J_m^+ u_m^\beta(r_{\rm i}) + u_m^{\rm p}(r_{\rm i})] = 0 \tag{18}$$

and

$$[\Psi_m u_m^\alpha(r_{\rm o}) + J_m^- u_m^\beta(r_{\rm o}) + u_m^{\rm p}(r_{\rm o})]_r - C_m^{\rm o}[\Psi_m u_m^\alpha(r_{\rm o}) + J_m^- u_m^\beta(r_{\rm o}) + u_m^{\rm p}(r_{\rm o})] = 0. \tag{19}$$

We rearrange these expressions into the formula

$$\Psi_m = R_m(J_m^+ - J_m^-) + F_m, \tag{20}$$

where

$$R_m = \left(\frac{\mathcal{U}_1^{\rm i}}{\mathcal{U}_2^{\rm i}} - \frac{\mathcal{U}_1^{\rm o}}{\mathcal{U}_2^{\rm o}}\right)^{-1} \qquad \text{and} \qquad F_m = -\left(\frac{\mathcal{U}_{\rm p}^{\rm i}}{\mathcal{U}_2^{\rm i}} - \frac{\mathcal{U}_{\rm p}^{\rm o}}{\mathcal{U}_2^{\rm o}}\right)\left(\frac{\mathcal{U}_1^{\rm i}}{\mathcal{U}_2^{\rm i}} - \frac{\mathcal{U}_1^{\rm o}}{\mathcal{U}_2^{\rm o}}\right)^{-1}, \tag{21}$$

with

$$\mathcal{U}_1^{\rm i,o} = \frac{\rm d}{{\rm d}r}u_m^\alpha(r_{\rm i,o}) - C_m^{\rm i,o}u_m^\alpha(r_{\rm i,o}), \qquad \mathcal{U}_2^{\rm i,o} = \frac{\rm d}{{\rm d}r}u_m^\beta(r_{\rm i,o}) - C_m^{\rm i,o}u_m^\beta(r_{\rm i,o}) \tag{22}$$

and

$$\mathcal{U}_{\rm p}^{\rm i,o} = \frac{\rm d}{{\rm d}r}u_m^{\rm p}(r_{\rm i,o}) - C_m^{\rm i,o}u_m^{\rm p}(r_{\rm i,o}). \tag{23}$$

Thus, the outer solutions yield the R_m and F_m coefficients for use in the inner solution as described below.

2.4 The critical region

To resolve the inner region we introduce a stretched inner coordinate $Y = (r-1)/\epsilon$ and time variable $T : \partial/\partial t \longrightarrow -\partial/\partial\theta + \epsilon\partial/\partial T$. We further set

$$u = \epsilon^2 U(Y,\theta,T), \qquad v = \epsilon^2 V(Y,\theta,T), \qquad \sigma = \epsilon^2 S(Y,\theta,T), \qquad w + \Phi_{\rm g} = \epsilon^2 W(Y,\theta,T), \qquad q = \epsilon\zeta(Y,\theta,T). \tag{24}$$

Equations (1–3) and (6) are now

$$\epsilon^3\left[\frac{\partial}{\partial T}+\frac{(\Omega-1)}{\epsilon}\frac{\partial}{\partial\theta}\right]U+\epsilon^3U\frac{\partial U}{\partial Y}+\frac{\epsilon^4V}{1+\epsilon Y}\frac{\partial U}{\partial\theta}-2\epsilon^2\Omega U-\frac{\epsilon^4V^2}{1+\epsilon Y}=-\epsilon\frac{\partial W}{\partial Y}, \tag{25}$$

$$\epsilon^3\left[\frac{\partial}{\partial T}+\frac{(\Omega-1)}{\epsilon}\frac{\partial}{\partial\theta}\right]V+2\epsilon^3U\frac{\partial V}{\partial Y}+\frac{\epsilon^4V}{1+\epsilon Y}\frac{\partial V}{\partial\theta}+\epsilon^2QU=-\frac{\epsilon^2}{1+\epsilon Y}\frac{\partial W}{\partial\theta}, \tag{26}$$

$$\epsilon^3\left[\frac{\partial}{\partial T}+\frac{(\Omega-1)}{\epsilon}\frac{\partial}{\partial\theta}\right]S+\frac{\epsilon}{1+\epsilon Y}\frac{\partial}{\partial Y}[(1+\epsilon Y)(\Sigma+\epsilon^2S)U]+\frac{\epsilon^2}{1+\epsilon Y}\frac{\partial}{\partial\theta}[(\Sigma+\epsilon^2S)V]=0 \tag{27}$$

and

$$\epsilon^2\left[\frac{\partial}{\partial T}+\frac{(\Omega-1)}{\epsilon}\frac{\partial}{\partial\theta}\right]\zeta+\epsilon^2U\frac{\partial\zeta}{\partial Y}+\frac{\epsilon^3V}{1+\epsilon Y}\frac{\partial\zeta}{\partial\theta}+\epsilon^2UQ'=0, \tag{28}$$

in which $\Omega-1=\Omega(1+\epsilon Y)-1\approx\epsilon Y\Omega'(1)$, $Q=Q(1+\epsilon Y)$ and $\Sigma=\Sigma(1+\epsilon Y)$. Therefore, to leading order,

$$\frac{\partial W}{\partial Y}=0,\qquad Q_cU=-\frac{\partial W}{\partial\theta},\qquad \frac{\partial U}{\partial Y}=0, \tag{29}$$

where the subscript 'c' indicates the value of the equilibrium variable at corotation ($r=1$). These relations imply that U and W are independent of Y (and so they can be taken to be the leading terms of the Taylor expansion of the outer solutions, u_0 and φ_0, at $r=1$).

Also to leading order, the potential vorticity equation becomes

$$\zeta_T+\Omega_c'Y\zeta_\theta+U\zeta_Y+UQ_c'=0. \tag{30}$$

Finally, from the definition of q in equation (7), we have

$$\zeta=\frac{\partial V}{\partial Y} \tag{31}$$

to ϵ^2. The critical region potential vorticity equation (30) is fully non-linear; the only simplification in the equation is the replacement of the non-linear advection term by one stemming from a streamfunction variable that does not depend on the coordinate Y. As in the Rossby wave critical layer problem this is one of the few simplifications that one can make in order to incorporate fully the physics of the critical region.

2.5 Matching

The outer and inner expansions provide the asymptotic solutions in the bulk of the disc and inside the critical region. These solutions must be matched to determine the complete solution, which amounts to equating the two asymptotic sequences over an intermediate matching region. This matching region is defined by $1\gg r-1=\delta\gg\epsilon$ and $\epsilon^{-1}\gg\Delta=(r-1)/\epsilon\gg1$ (so that $\delta=\epsilon\Delta$).

In the matching region, the outer solution is given by the Frobenius solutions:

$$\psi\sim\sum_{m=0}^{\infty}\left[\Psi_m\left(1+\frac{Q_c'}{\Omega_c'}\delta\log|\delta|\right)+J_m^{\pm}\delta\right]e^{im\theta}. \tag{32}$$

Hence,

$$u\sim\epsilon^2\sum_{m=1}^{\infty}im\Psi_m e^{im\theta},\qquad w+\Phi_g\sim-\epsilon^2Q_c\sum_{m=1}^{\infty}\Psi_m e^{im\theta}, \tag{33}$$

$$v\sim\epsilon^2\sum_{m=1}^{\infty}\left[J_m^{\pm}-\frac{Q_c'}{\Omega_c'}\Psi_m(1+\log|\delta|)\right]e^{im\theta} \tag{34}$$

and

$$q\sim-\frac{\epsilon^2}{\delta}\sum_{m=1}^{\infty}\frac{Q_c'}{\Omega_c'}\Psi_m e^{im\theta}. \tag{35}$$

The inner solution, on the other hand, becomes

$$u\sim\epsilon^2U,\qquad w+\Phi_g\sim\epsilon^2W,\qquad v\sim\frac{1}{Q_c}\int^{\Delta}\zeta\,dY, \tag{36}$$

and (by taking the leading-order balance in the critical region potential vorticity equation for large Y)

$$\zeta_\theta\sim-\frac{\epsilon Q_c'U}{\Omega_c'\Delta}\equiv-\frac{\epsilon^2Q_c'U}{\Omega_c'\delta}. \tag{37}$$

The matching of u, W and q is straightforward and implies that

$$U = \sum_{m=1}^{\infty} \mathrm{i}m\Psi_m \, \mathrm{e}^{\mathrm{i}m\theta} \equiv \Psi_\theta. \tag{38}$$

Next we must match v. However, the logarithmic terms in the outer solution uncover a technical point that we have until now ignored: In principle, to match the inner and outer forms for v, we should include $\log \epsilon$ terms in the inner asymptotic sequences. In fact, this technicality merely complicates the expansion without adding any important details, and may be avoided by matching the jump in v symmetrically across the critical region – that is, by equating $[v]_{1-\delta}^{1+\delta}$ with $[V]_{-\Delta}^{\Delta}$. The logarithmic terms of the outer solution then cancel, leaving

$$\sum_{m=1}^{\infty} (J_m^+ - J_m^-) \, \mathrm{e}^{\mathrm{i}m\theta} = \int_{-\Delta}^{\Delta} \zeta \, \mathrm{d}Y. \tag{39}$$

Equivalently,

$$J_m^+ - J_m^- = \int_0^{2\pi} \int_{-\Delta}^{\Delta} \mathrm{e}^{-\mathrm{i}m\theta} \zeta \, \mathrm{d}Y \frac{\mathrm{d}\theta}{2\pi}. \tag{40}$$

Lastly, in the asymptotic scheme, $\epsilon \to 0$ and so we may take $\Delta \to \infty$. Hence,

$$J_m^+ - J_m^- = \langle \mathrm{e}^{-\mathrm{i}m\theta} \zeta \rangle \qquad \text{where} \qquad \langle \chi \rangle = \frac{1}{2\pi} \int_0^{2\pi} \int_{-\infty}^{\infty} \chi(Y, \theta) \, \mathrm{d}Y \, \mathrm{d}\theta, \tag{41}$$

in which we must be careful to interpret the logarithmic divergence of the integral at the upper and lower limits (where $\zeta \sim Y^{-1}$) in terms of a principal value. Together with the critical region potential vorticity equation, this relation is all we need to close the system of equations because the outer solution provides the relation between $J_m^+ - J_m^-$ and Ψ_m, giving:

$$\Psi_m = F_m + R_m \langle \mathrm{e}^{-\mathrm{i}m\theta} \zeta \rangle. \tag{42}$$

2.6 Canonical system

We now scale the equations to place them in a convenient form. Let

$$Y = \frac{1}{\sqrt{|\Omega_{\mathrm{c}}'|}} \tilde{Y}, \qquad \zeta = \tilde{\zeta} \sqrt{|\Omega_{\mathrm{c}}'|} \qquad \text{and} \qquad T = \frac{1}{\sqrt{|\Omega_{\mathrm{c}}'|}} \tilde{T}. \tag{43}$$

Then, after various algebraic manipulations (and dropping the tilde decoration), we arrive at the reduced model:

$$\zeta_T + Y\zeta_\theta - \Psi_\theta \zeta_Y + \beta \Psi_\theta = 0 \qquad \text{and} \qquad \Psi_m = F_m + R_m \int_{-\infty}^{\infty} \zeta_m(Y, T) \, \mathrm{d}Y, \tag{44}$$

where

$$\beta = \frac{Q_{\mathrm{c}}'}{|\Omega_{\mathrm{c}}'|} \tag{45}$$

is a measure of the potential vorticity gradient; for an undisturbed disc with constant surface density in a state of Keplerian rotation, $\beta = 1/3$. This model system describes how a global wave pattern is excited by the perturber, how the waves then advect and twist up the potential vorticity near corotation, and how this in turn affects the wave generation. The model system is similar to those derived for the forced Rossby wave critical layer problem (Stewartson 1978; Warn & Warn 1978).

Note that the magnitude of the perturbing potential does not appear in the final model equations; the only parameters are β, F_m and R_m, which depend on the disc structure. The mass of the perturber enters the problem through the small parameter ϵ, and once we are done with the formal expansion, this quantity is needed only to reconstruct the original field variables. Moreover, the reduced model has the property that the transformation, $T \to T/a$, $Y \to aY$ and $\Psi_m \to a^2 \Psi_m$, leaves the equations invariant, but for the replacement $F_m \to F_m/a^2$. Thus, we can always rescale the coefficients F_m for any disc model or, equivalently, fix the size of the domain in Y we require for computations with the model.

For practical purposes, we select disc models with constant surface density, Keplerian rotation and absorbing edges. The only remaining disc parameters are then the disc's inner and outer radii, and the sound speed c. Table 1 and Fig. 2 review some specific values and properties of the coefficients F_m (assumed constant in time) and R_m, for disc models of this kind. The inner disc radius is at 0.5, and the outer radius is 1.5, but changes in the locations of these edges lead to insignificant differences in R_m and F_m, indicating that these variables are not important. The effect of the sound speed, however, is more significant. From the table and figure we observe that the real parts of the coefficients $R_{m\mathrm{r}} \sim -1/(2m)$ and $F_{m\mathrm{r}} \sim \log(m)/(m^2 c^2)$ for $m \gg 1$, and that (approximately) $R_{m\mathrm{r}} \to -2c$ and $F_{m\mathrm{r}} \to$ constant for $m \to 1$. These properties are expected from the short-wavelength arguments given in Appendix A. The imaginary parts of the coefficients have a more complicated dependence on m, but they are also much smaller. As indicated below, the quantities $R_{m\mathrm{i}}$ and $F_{m\mathrm{i}}$ measure the amount of wave leakage through the boundaries; for discs with reflective boundaries, $R_{m\mathrm{i}} = F_{m\mathrm{i}} = 0$.

Table 1. The coefficients R_m and F_m for two discs with constant surface density and sound speed $c = 10^{-1}$. The outer solutions are also computed using a standard, softened potential for the perturber; the potential softening length is $r_p = 10^{-4}$. The softening of the potential makes the computation simpler but is not essential; the final results are insensitive to the softening length. The inner boundary is $r_i = 0.5$ and the outer boundary is $r_o = 1.5$.

m	R_{mr}	R_{mi}	F_{mr}	F_{mi}
1	-2.2255×10^{-1}	0	4.1736	0
2	-1.9541×10^{-1}	2.7213×10^{-4}	2.7416	-1.5176×10^{-2}
3	-1.6211×10^{-1}	3.0553×10^{-4}	1.7622	-2.4319×10^{-2}
4	-1.3247×10^{-1}	-1.4563×10^{-3}	1.1400	4.1117×10^{-2}
5	-1.0952×10^{-1}	-1.8579×10^{-3}	7.6471×10^{-1}	4.2458×10^{-2}
10	-5.2750×10^{-2}	-1.5454×10^{-3}	1.3523×10^{-1}	2.5187×10^{-2}
15	-3.3277×10^{-2}	-6.7317×10^{-4}	2.5768×10^{-2}	9.5377×10^{-3}
20	-2.4342×10^{-2}	-2.8121×10^{-4}	1.3725×10^{-3}	3.6876×10^{-3}
30	-1.6134×10^{-2}	-5.0620×10^{-5}	-4.0720×10^{-3}	6.0625×10^{-4}
40	-1.2206×10^{-2}	-9.0206×10^{-6}	-2.4548×10^{-3}	9.9890×10^{-5}
50	-9.8418×10^{-3}	-1.3146×10^{-6}	-1.2739×10^{-3}	1.2222×10^{-5}
100	-4.9834×10^{-3}	1.2609×10^{-8}	-1.1265×10^{-4}	-1.6665×10^{-7}
150	-3.3287×10^{-3}	8.4701×10^{-11}	-4.2283×10^{-5}	-1.0624×10^{-9}
200	-2.4981×10^{-3}	4.9936×10^{-13}	-2.6022×10^{-5}	-6.1473×10^{-12}
300	-1.6661×10^{-3}	1.4678×10^{-17}	-1.5044×10^{-5}	-1.8919×10^{-16}

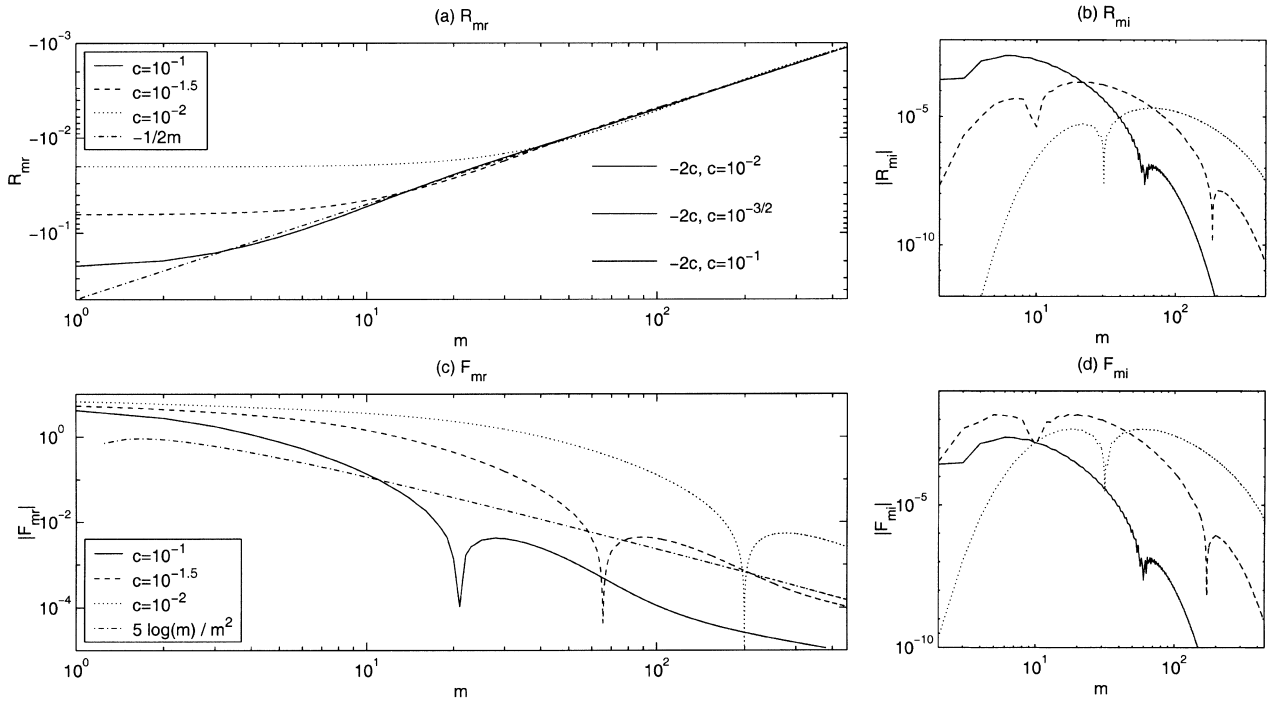

Figure 2. Plots of the coefficients R_m and F_m for discs with constant surface density and three values of the sound speed. The potential is softened with a length $r_p = 10^{-4}$. The inner boundary is $r_i = 0.5$ and the outer boundary is $r_o = 1.5$.

It is also informative to consider the shape of the function $F(\theta) \equiv \sum_m F_m \, e^{im\theta}$, which is illustrated in Fig. 3 for the three disc models. This function describes how the perturber forces the global streamfunction in the absence of any feedback from the critical region. As shown by Fig. 3, this 'perturbing potential' has a flat well and peaks at the position of the perturber. The sharpness of the peaks is controlled by the sound speed, which is the key effect of this variable.

3 INVISCID LINEAR THEORY

We begin our discussion of the reduced model by looking into the linear dynamics. After dropping the non-linear term and decomposing into Fourier modes, we find

$$\zeta_{mT} + im\zeta_m = -im\beta\Psi_m. \tag{46}$$

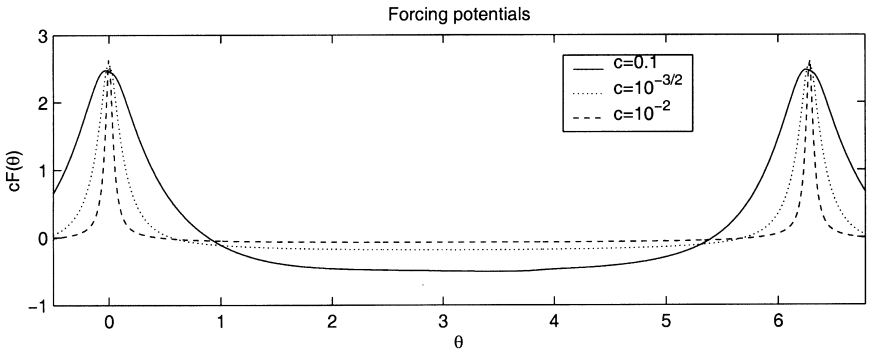

Figure 3. The 'forcing potential' for the three discs with constant surface density.

With the initial condition, $\zeta_m(Y, T = 0) = 0$, this equation has the integral

$$\zeta_m = -\,\mathrm{i}m\beta \int_0^T \mathrm{e}^{-\,\mathrm{i}mY(T-s)}\Psi_m(s)\,\mathrm{d}s. \tag{47}$$

Thence,

$$\int_{-\infty}^{\infty} \zeta_m(Y, T)\,\mathrm{d}Y = -\,\mathrm{i}\pi\beta\Psi_m\,\mathrm{sgn}(m), \tag{48}$$

and finally

$$\Psi_m = \frac{F_m}{1 + \mathrm{i}\pi\beta R_m\,\mathrm{sgn}(m)}. \tag{49}$$

Given this solution, from the following equation we may estimate directly the angular momentum deposited over the critical region according to linear theory:

$$\left[r^2\Sigma \int_0^{2\pi} uv\,\mathrm{d}\theta\right]_{1^-}^{1^+} \equiv \int_0^{2\pi} [UV]_{Y\to-\infty}^{Y\to\infty}\,\mathrm{d}\theta = -2\pi\langle\Psi\zeta_\theta\rangle = \sum_{m>0}\frac{4\pi^2 m Q_{\mathrm{c}}'}{|\Omega_{\mathrm{c}}'|}|\Psi_m|^2, \tag{50}$$

which is equivalent to a formula given by Korycansky & Pollack (1993) and, on using the tight-winding approximation, can be reduced to Goldreich & Tremaine's celebrated formula.

4 CONSERVATION LAWS

The reduced model has a number of conservation laws, obtained by taking 'moments' of the vorticity equation with respect to Y, Ψ and ζ:

$$\langle\zeta\rangle_T = 0, \qquad \langle Y\zeta\rangle_T = \langle\Psi\zeta_\theta\rangle, \tag{51}$$

$$\left\langle \left(\tfrac{1}{2}Y^2 + \Psi\right)\zeta\right\rangle_T - \langle\Psi_T\zeta\rangle = 0 \qquad \text{and} \qquad \tfrac{1}{2}\langle\zeta^2\rangle_T = \beta\langle\Psi\zeta_\theta\rangle. \tag{52}$$

These correspond to conservation of potential vorticity, angular momentum, energy and enstrophy within the critical layer. There is also an infinite number of Casimir invariants, $\langle\mathcal{F}(\zeta + \beta Y)\rangle$, given by any function $\mathcal{F}$ of the total vorticity $\zeta + \beta Y$.

The angular momentum (and enstrophy) balance contains the integral $\langle\Psi\zeta_\theta\rangle$, which is related to the change in the angular momentum flux across corotation, as in equation (50). We may also use the relation between the vorticity and streamfunction to write:

$$\langle\Psi\zeta_\theta\rangle = \sum_m \mathrm{i}mF_m\langle\mathrm{e}^{\mathrm{i}m\theta}\zeta\rangle - 2\sum_{m>0} mR_{mi}|\langle\mathrm{e}^{\mathrm{i}m\theta}\zeta\rangle|^2 = \sum_m \mathrm{i}m\frac{\Psi_m^* F_m}{R_m} - 2\sum_{m>0} mR_{mi}\frac{|\Psi_m|^2}{|R_m|^2}. \tag{53}$$

Because there are no time derivatives appearing in the outer equations at leading order, over the time-scale on which the vorticity is rearranged inside the critical region, the outer region is in a quasi-steady balance. Thus, if spiral waves are generated in the main body of the disc, they propagate to the edge and are absorbed immediately if the edges are non-reflective. This leads to an instantaneous transfer of angular momentum, which is given by the second term on the right-hand side of equation (53). (Hence, the angular momentum transfer by free waves is determined by the imaginary part of R_m.)

In linear theory, the angular momentum transfer becomes

$$\langle\Psi\zeta_\theta\rangle = \sum_{m>0} 2\pi m\beta|\Psi_m|^2, \tag{54}$$

which is of definite sign (the sign of β). Thus, linear theory predicts that angular momentum is continually absorbed at corotation, and the total absorption,

$$M = \int_0^T \langle \Psi_\theta \zeta \rangle \, \mathrm{d}T, \tag{55}$$

increases without bound. By contrast, from equation (51),

$$M \equiv [\langle Y\zeta \rangle]_0^T. \tag{56}$$

Hence, provided $\langle Y\zeta \rangle$ does not diverge, and the potential vorticity remains localized to the critical region, $\langle \Psi_\theta \zeta \rangle \to 0$ over long times and angular momentum *cannot* continue to be absorbed. In this circumstance, the total absorption is finite as $T \to 0$, and so inviscid linear theory cannot remain valid indefinitely. In other words, the corotation torque must saturate. [This argument is adapted from a general theorem by Killworth & McIntyre (1985).]

5 NON-LINEAR DYNAMICS

5.1 The numerical scheme

To explore the non-linear dynamics of excited disturbances, we solve the equations of the reduced model numerically. The scheme is based on an operator-splitting algorithm proposed by Cheng & Knorr (1976), and is similar to that described by Balmforth, Llewellyn Smith & Young (2001) and Balmforth & Piccolo (2001). Briefly, the system is evolved using two steps, advection in θ followed by advection in Y. The equations are solved on a finite domain with boundary conditions given by the far-field form of the vorticity [which is also used to provide the far-field contribution to the integral $\langle \mathrm{e}^{-\mathrm{i}m\theta}\zeta \rangle$ (Haynes 1985)]. For convenience, we rescale the coefficients F_m by a factor of 10 (as described in Section 2.6), so that the associated potential has order unity values, and use domain sizes in Y of 10 or 20. The code has a resolution of 512×1025, and a time-step of 10^{-3} (we checked that the spatial and temporal resolution was adequate).

To introduce the perturbing potential smoothly, we gradually turn on the forcing potential:

$$F_m = \tilde{F}_m(1 - \mathrm{e}^{-5T^2}), \tag{57}$$

where $\tilde{F}_m$ denotes the time-independent disc constants computed from the outer solution of the matched asymptotics. This functional form for F_m smoothly turns on the perturber over the interval [0,1], and is selected arbitrarily because we do not account for the origin of the perturber.

A further crucial ingredient in the numerical scheme is the addition of explicit dissipation via the terms $\lambda\zeta_{YY}$ and $\nu\partial_x^8\zeta$. This addition is not strictly necessary and the inviscid code runs without problems. However, due to the creation of increasingly sharp spatial scales, the code quickly runs out of resolution whereupon an intrinsic dissipation operates that is difficult to quantify and control. Very roughly, the intrinsic dissipation is equivalent to explicit viscosities of $\lambda \sim 10^{-7}$ and $\nu \sim 10^{-15}$. The effect of large dissipative terms is explored in the next section.

5.2 An example

By way of illustration of the non-linear dynamics, we select the constant surface density disc with $c = 0.1$. Of the three displayed in Fig. 2, this disc has the least sharp forcing potential (Fig. 3) and is therefore the simplest to deal with from a numerical perspective. We also fix $\lambda = 10^{-6}$ and $\nu = 3 \times 10^{-14}$. The results are shown pictorially in the sequence of Figs 4–6. We comment later on how the dynamics changes with sound speed.

Fig. 4 shows the initial development of structure within the critical region as the forcing is turned on. The relatively sharp form of the potential near the perturber ($\theta = Y = 0$) quickly generates a response in the potential vorticity distribution much like the shock structures seen in other computations (Korycansky & Papaloizou 1996; Bryden et al. 1999). From this response, an extended, elongated vortex rolls up and begins to circulate around the critical region. This vortex generates a significant feedback on the streamfunction $\Psi(\theta)$, and the dynamics then develops unsteadily.

Shortly after the nucleation of the elongated vortex, narrow filaments form as fluid is drawn past the perturber, which strains and intensifies them. The filaments subsequently lose stability and begin to roll up into secondary vortices in a process much like the classical Kelvin–Helmholtz instability (Fig. 5). The secondary vortices then interact with one another and merge. Occasionally, the secondary vortices undergo a close encounter with the main vortex and are sheared out and cannibalized. This epoch of vortex formation and interaction is very reminiscent of numerical experiments with two-dimensional turbulence (McWilliams 1990). Overall, the filamentation and twisting up of the vorticity distribution vigorously stirs the fluid over the critical region.

The roll-up of the filaments is the correspondence in the current problem of the secondary instabilities studied in the forced Rossby wave critical layer problem (Killworth & McIntyre 1985; Haynes 1985, 1989). In either case, the forced disturbance winds up the vorticity into a distribution that has the reversals in vorticity gradient required to produce Rayleigh's classical shear instability. The instabilities typically take the form of shorter-wavelength perturbations that mix up the vorticity distribution yet further. The complicated fluid dynamics that results was dubbed 'critical layer turbulence' by Haynes (1989).

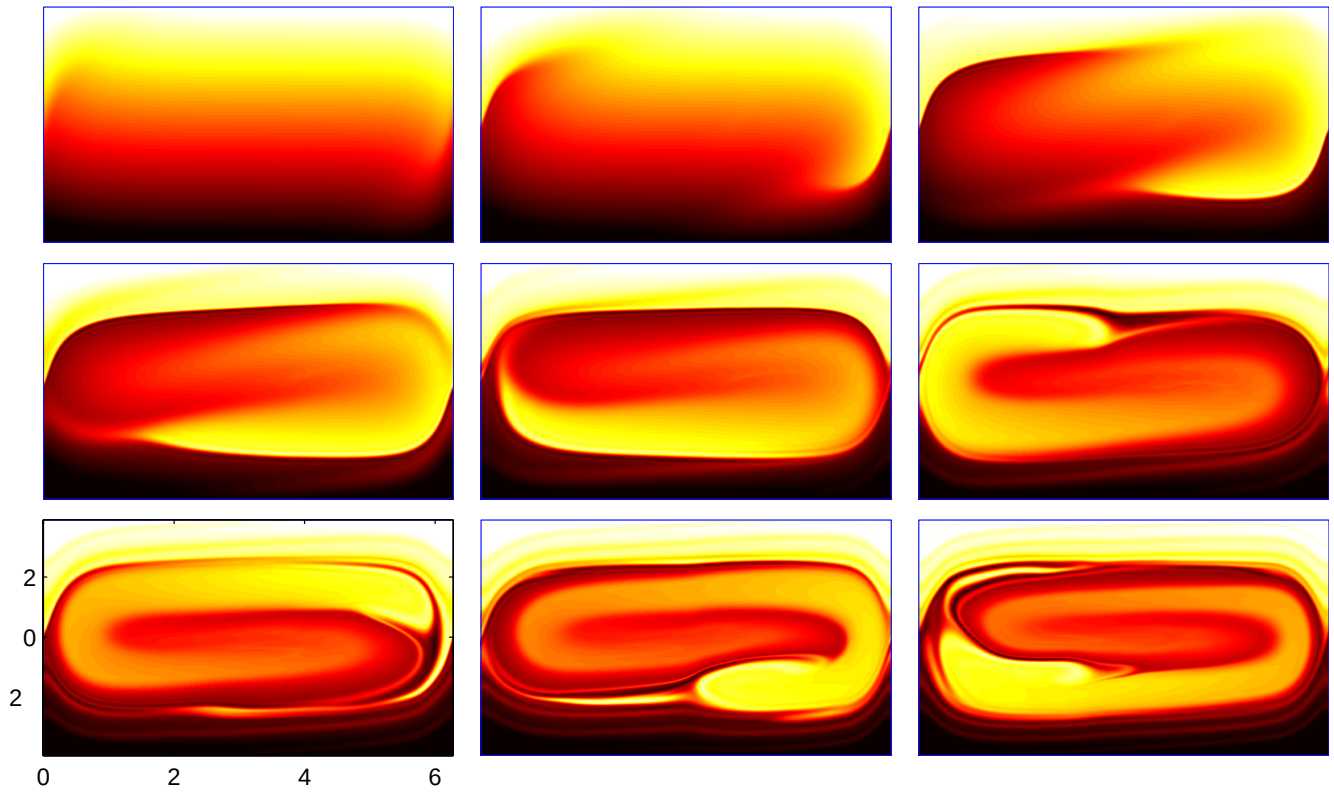

Figure 4. Snapshots of the total vorticity, $q = \zeta + \beta y$, drawn as densities on the (θ, Y) plane. The snapshots are taken at $T = 0.5$, 1, 2, 3, 4, 6, 8, 10 and 12. The shading is given by the key in Fig. 6.

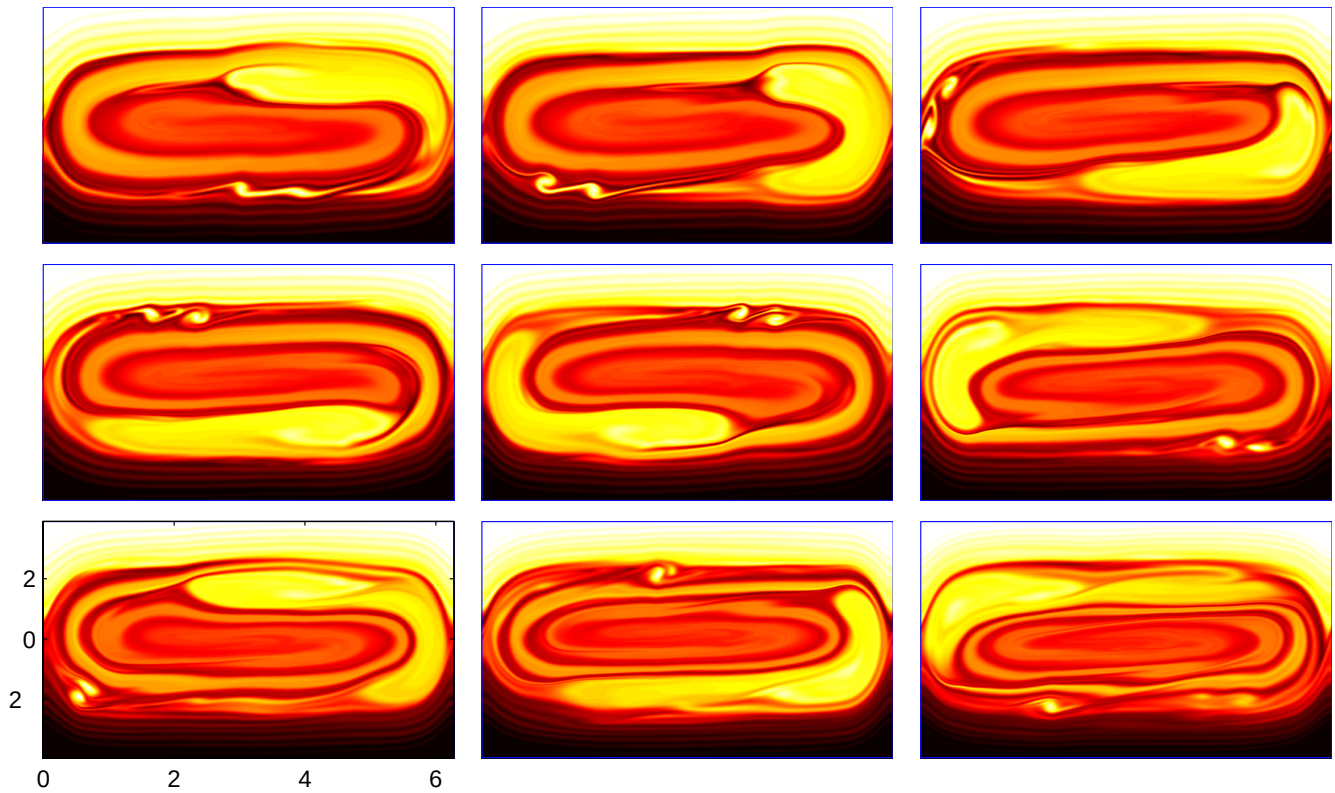

Figure 5. A continuation of Fig. 4, showing an epoch of secondary instability and vortex formation. The snapshots are taken at $T = 16$, 17, 18, 19, 20, 22, 24, 26 and 30. The shading is given by the key in Fig. 6.

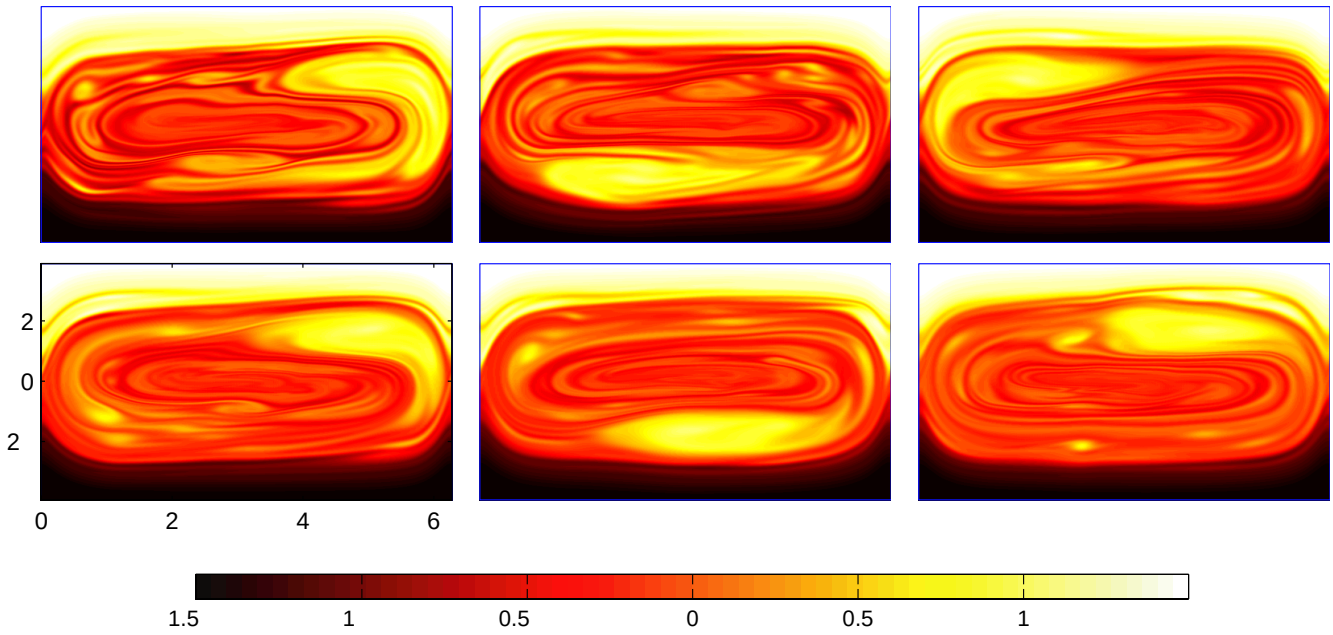

Figure 6. A longer-time continuation of Figs 4 and 5. The snapshots are taken at $T = 40$, 50, 60, 70, 80 and 100. The shading is given by the key.

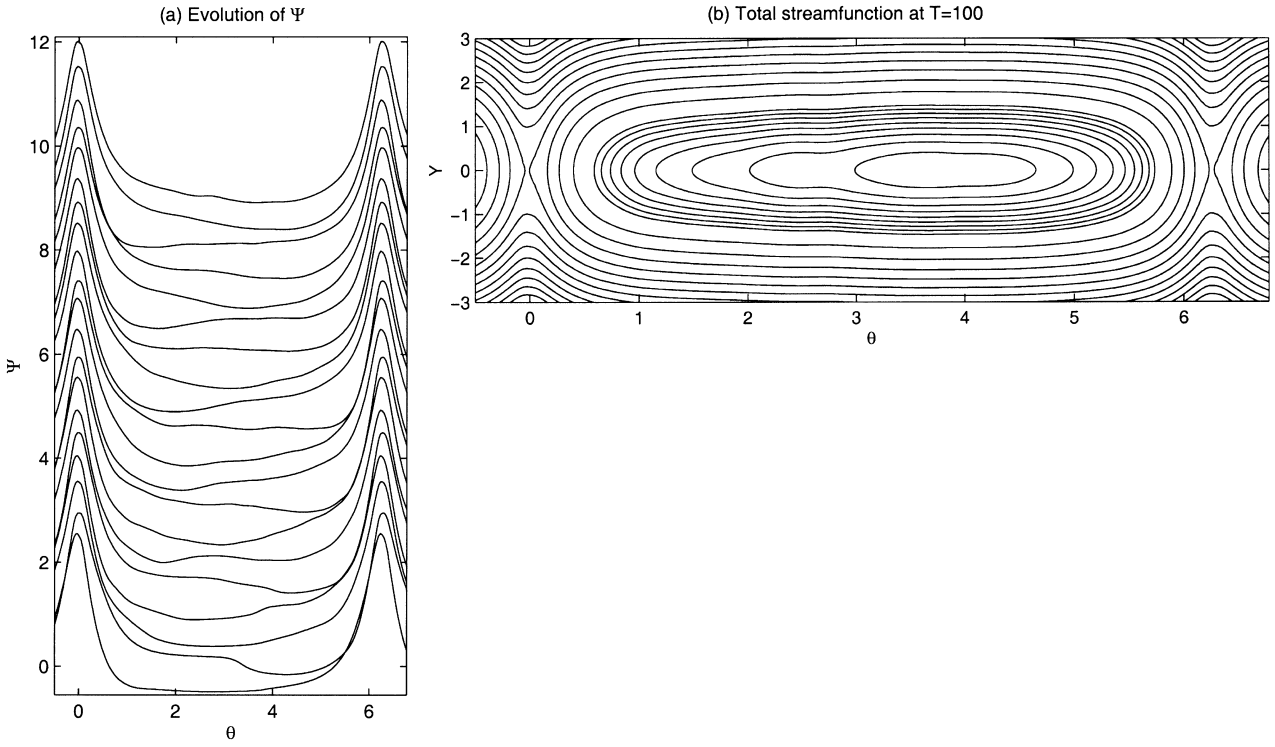

Figure 7. Further details of the solution. (a) Snapshots of the streamfunction $\Psi(\theta)$ at times $T = 10j$ for $j = 1, 2, \ldots, 10$. Each curve is offset by 0.5 for clarity. (b) The total streamfunction, $\Psi(\theta) + Y^2/2$, at $T = 100$.

Evolution over longer time-scales is displayed in Fig. 6. The filamentation of the vorticity continues, and vortices are again generated and interact. However, the sharpness of the filaments gradually degrades as the scales in the critical region decrease to the viscous scales. In other words, after a period, the simulation enters a regime of significant dissipation wherein further filamentation and vortex formation is impeded. The distribution then evolves to a smoother state, although the primary vortex maintains its identity over the entire simulation. Given higher resolution (and reduced values for the viscous coefficients), the unsteady complicated dynamics would presumably continue for longer.

Further details of the solution are shown in Fig. 7. Panel (a) displays regularly spaced snapshots of the streamfunction, $\Psi(\theta)$. As the primary vortex circulates around the critical region, the wide, flat 'floor' of the streamfunction tilts back and forth. The titling is at least

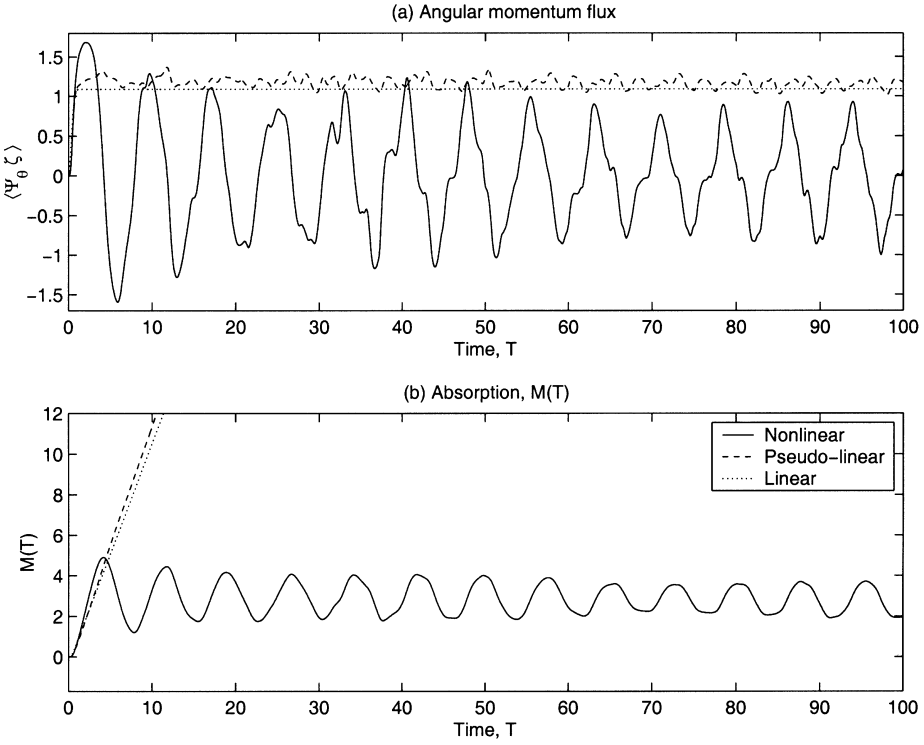

Figure 8. (a) The history of the angular momentum flux into the critical region, as measured by $\langle \Psi_\theta \zeta \rangle$. (b) The absorption, $M(T)$. Also plotted are the results predicted by the exact linear theory (dotted curves), and by a 'pseudo-linear theory' in which we evaluate the integral $\langle \zeta e^{im\theta} \rangle$ according to linear theory, giving $\sum_m 4\pi^2 m Q_c' |\Psi_m|^2 / |\Omega_c'|$, but then use the non-linearly evolved streamfunction components, $\Psi_m(T)$ (dashed curves).

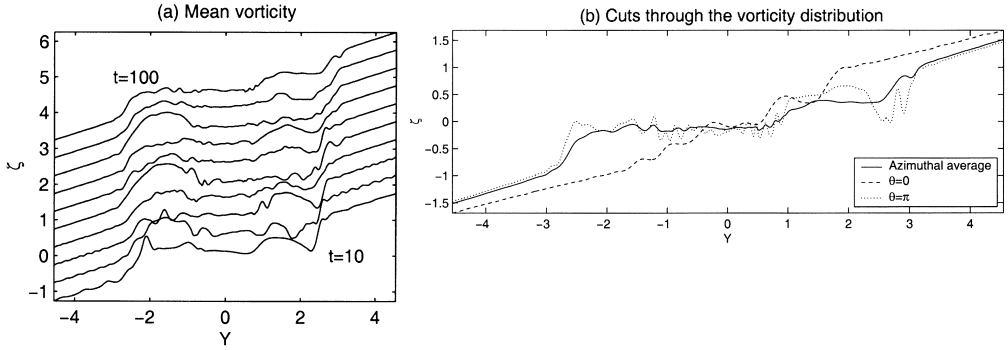

Figure 9. The azimuthally averaged potential vorticity distribution. (a) Snapshots at intervals of 10 time units. (b) The final mean distribution along with cuts through $q(Y,\theta,100)$ at $\theta = 0$ and π.

partially responsible for drawing the filaments of fluid past the perturber. Also shown in Fig. 7(b) is the final, total streamfunction $\Psi(\theta) + Y^2/2$. In a steady situation, the shape of this function shows the phase portrait of the motion of fluid elements. Although the fluid here is not steady, the contours of constant $\Psi(\theta) + Y^2/2$ provide some rationalization for the structure that develops inside the critical region.

The history of the angular momentum flux into the critical region, $\langle \Psi_\theta \zeta \rangle$, and the absorption, $M(T)$, during the simulation are shown in Fig. 8. As anticipated by the arguments of Section 4, there is initially a flux into the region approximately given by the linear theory, but this flux does not continue indefinitely. Instead, the flux declines and oscillates erratically about zero, signifying an alternating sequence of angular momentum absorption and emission. As a result, the absorption, $M(T)$, remains bounded as $T \to \infty$.

The oscillations in the angular momentum flux are closely connected to the circulation of the primary vortex within the critical region, and, like that vortical structure, show little sign of decay. It seems entirely possible that the vortex and oscillations persist indefinitely, and so the critical region never settles down to a steady state. Dissipation, however, may eventually take its toll.

The net effect of the rearrangement of the potential vorticity on the mean profile within the critical region is shown in Fig. 9. This illustrates how the disturbance mixes and flattens the total potential vorticity over this region. It is partly as a result of this effect that the angular momentum flux into the critical region declines: the fluid can only absorb angular momentum at corotation provided there is an effective potential vorticity gradient there. Some further properties of the asymptotic, quasi-steady state are described in Appendix B.

5.3 Variations

We have also experimented with other choices for the parameters. Several of the features described above appear generic for the disc response: a large-scale vortex appears and circulates around the critical region; the potential vorticity distribution is stirred and mixed; the momentum flux declines and then oscillates in tandem with the circulation of the vortex; and the absorption of angular momentum saturates. Secondary instabilities are also common, although they do not appear in every simulation (see below).

For example, if we artificially set $R_{mi} = F_{mi} = 0$, the results are essentially unchanged. This simulation mimics a disc with reflective inner and outer edges, and the lack of any change shows how the angular momentum flux through the boundaries has a negligible effect on the critical layer dynamics [actual solutions of equation (13) with reflective boundary conditions yield real R_m and F_m that are basically the same, but for sharp peaks in the coefficients related to quantization conditions for resonance with normal modes]. This is not to say that effects of wave reflection are completely unimportant. In fact, without non-reflective edges, the disc can support unstable normal modes due to the mechanism of over-reflection (Papaloizou & Pringle 1987; Narayan, Goodman & Goldreich 1987). Such self-amplified waves could, in principle, grow to sufficient amplitude to dominate the disc's dynamics and control the redistribution of material and angular momentum. The main effect of the boundary conditions at the disc's edges is therefore to eliminate these modes.

Another potentially important variable is the time-scale over which the perturber is turned on. To gauge its effect, we ran several computations with turn-on times as low as 1/2 and as large as 20. In none of the cases was there a qualitative change in the dynamics, although there appeared to be a greater degree of secondary instability when the perturber was turned on more slowly. The insensitivity to the turn-on time can be understood on referring back to the scaling transformation mentioned earlier ($T \to T/a$, $Y \to aY$, $\Psi_m \to a^2\Psi_m$, $F_m \to F_m/a^2$). This scaling indicates that, provided the disc response remains unimportant, a given instantaneous perturber amplitude corresponds to a certain disc time-scale (and critical region thickness). This intrinsic time-scale is infinite when the perturber's mass is zero, but decreases as the perturber is turned on (for the final amplitude of the perturber used in the simulations, the corresponding disc time-scale is about 6, so the range of simulations span cases with relatively short and long turn-on times). Thus there is an initial period over which the strength of the perturber grows more quickly than the disc can respond, however the perturber is introduced. Thereafter, if the perturber grows relatively rapidly, the disc response is always relatively slow, and so the turn-on is effectively instantaneous. If the turn-on time is slower, however, there is a later stage in which the perturber grows more slowly than the disc responds. In this instance, the disc can evolve into a quasi-steady state whilst the perturber continues to grow, but the flow structure formed in the early stages will not change dramatically from then on. (The latter situation is perhaps more analogous to planet formation because the accretion time-scales of planetary cores are extremely long compared to Keplerian circulation periods.) Nevertheless, in both cases, the evolution begins with a growing perturber and ineffectual disc response, a phase that continues until either the perturber is fully turned on, or the disc begins to respond more quickly. At that stage, the disc responds as though the perturber was suddenly turned on with a certain size, and the history of the perturber is unimportant. Thus, we expect (in the absence of other factors such as dissipation) that the flow characteristics should be relatively insensitive to the perturber's growth time once we factor out any change in scale.

The most significant variable appears to be the sound speed of the undisturbed disc, which controls the sharpness of the forcing potential (see Fig. 3). As a result, in discs with lower sound speed, the primary vortex has a much more elongated appearance, and spiral filaments of vorticity form that become tightly wound as time proceeds – see Fig. 10. These filaments are strongly strained by passage near the perturber, and presumably as a result do not suffer any secondary instability as in the disc with higher sound speed. It is not clear that the filaments would remain stable if the dissipation were less; further experimentation suggested that secondary instability might eventually occur. However, secondary instability does not always occur in the forced Rossby wave critical layer problem either (Haynes 1985).

6 DISSIPATIVE EFFECTS

To regularize the computations described above, we have added dissipation into the reduced model. In principle, such terms ought to result from the introduction of suitable dissipative terms into the governing equations. For example, if we include viscous stresses in the momentum equations with a viscosity $\nu = \epsilon^3\nu_3$, the regularizing radial diffusion, $\lambda\zeta_{YY}$, is the leading-order descendant in the reduced model. In this section, we explore the effect of varying the coefficient, λ. The hyperdiffusion in θ is harder to justify from first principles, and so we keep it chiefly as a numerical device to ensure that the vorticity does not develop too sharp an azimuthal gradient and remains resolved.

The angular momentum flux and absorption for computational runs with different values for λ are shown in Fig. 11. The increased dissipation acts to remove the persistent oscillations observed in the flux for smaller λ. This is connected to the smoothing of the vortical structures that generate those oscillations. As a result, the vorticity distribution inside the critical region becomes featureless, as shown in Fig. 12 for $\lambda = 10^{-3}$. At this stage, asymptotic arguments (Appendix B.1) indicate that the total vorticity q should become a function of the total streamfunction, $\Phi = \Psi + Y^2/2$, which is, indeed, observable in Fig. 12.

The viscous runs also do not show a saturation of the absorbed flux. This is somewhat surprising because, when we add radial viscosity to the reduced model, the momentum conservation laws are not altered (the angular momentum balance remains $\langle Y\zeta\rangle_T = \langle\Psi_\theta\zeta\rangle$). Therefore, the arguments of Section 4 remain valid even in the dissipative problem, and so the corotation torque should again saturate when $\lambda \neq 0$, in contrast to the numerical results. In fact, the larger the viscosity, the greater the absorption, $M(T)$. The apparent contradiction arises because the viscosity induces a radial diffusion of vorticity out of the critical region (Brown & Stewartson 1978). As a result, $\langle Y\zeta\rangle$ no longer remains bounded, even though the total potential vorticity is fixed. We see this effect especially transparently in the limit of large λ, which we explore

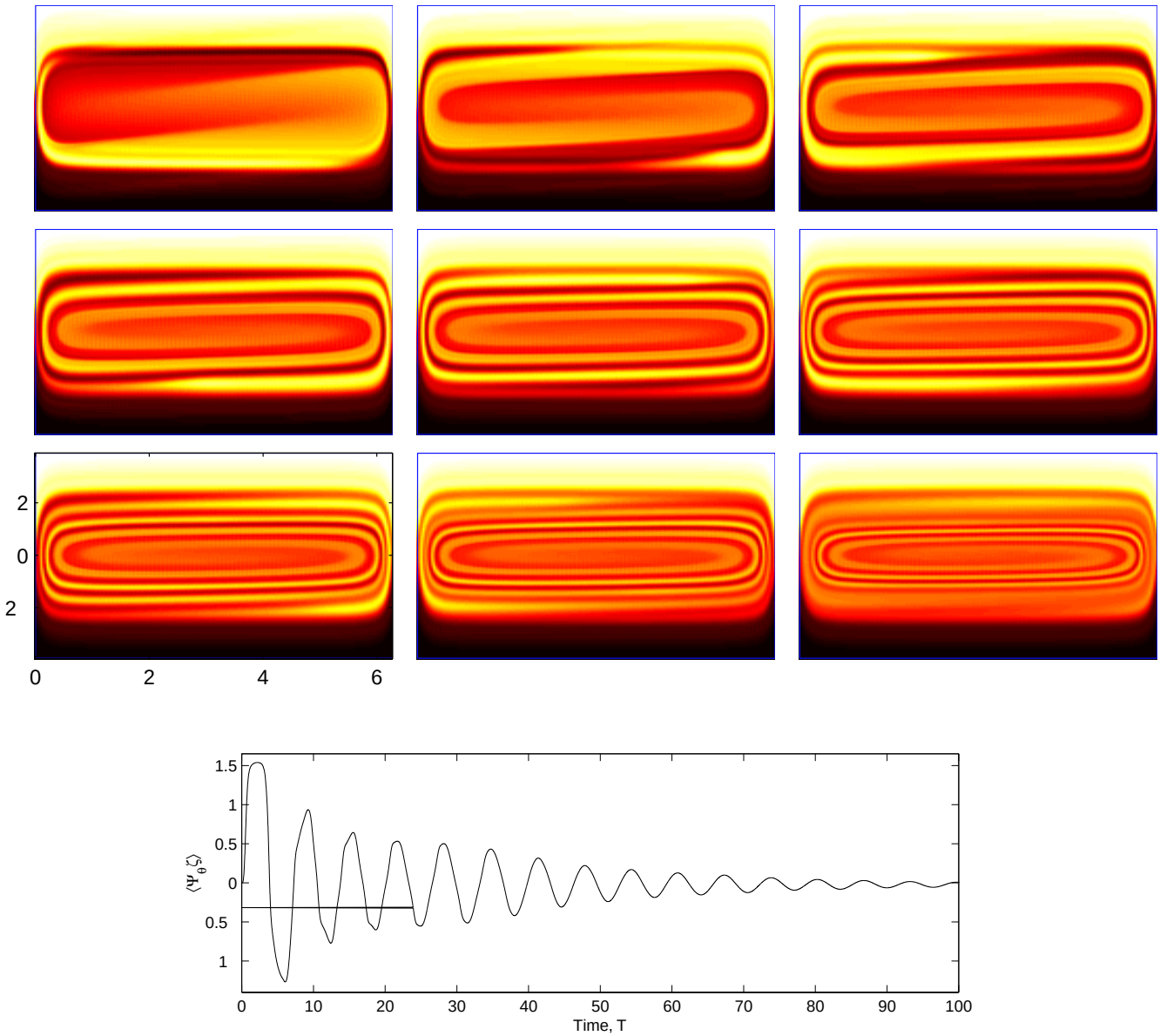

Figure 10. A time sequence showing the twist-up of the potential vorticity distribution in a disc with $c = 10^{-3/2}$. The times of the snapshots are 10 to 80 in steps of 10, and then finally 100. The colour scheme is given by the key of Fig. 6. The lowest panel shows the history of the angular momentum flux, $\langle \Psi_\theta \zeta \rangle$.

in Appendix B.2. In fact, the arguments indicate that the critical layer vorticity always spreads whenever λ is non-zero, and so the angular momentum flux into the critical region is always finite in the dissipative problem (though small when $\lambda \ll 1$, see Appendix B.1).

7 FINAL COMMENTS

A major question we have addressed in this paper is how non-linear effects influence the torque exerted at corotation by a perturbing potential. In this regard, we have been guided by a related problem in geophysical fluid dynamics pertaining to the critical layers of forced Rossby waves. In the related geophysical problem, it is found that wave action is initially absorbed in the critical region, but, as the vorticity distribution locally turns over there in response, the absorption declines and saturates, in contrast to linear theory. Here, we have verified that one can paint a similar picture for the astrophysical problem.

The possibility that the corotation torque can saturate has been discussed previously in astrophysics (Goldreich & Tremaine 1981; Ward 1989), in terms of the idea that the absorption of angular momentum creates librating particles near corotation. No torque can be exerted on such librating particles, which are analogous to fluid elements circulating on closed streamlines within the critical region in the image proposed here. In fact, using the conservation laws of the governing equations, it is possible to establish rigorously that the corotation torque must saturate in inviscid discs.

We have further explored dissipative effects on the non-linear dynamics. Provided such effects produce a radial transport of vorticity out

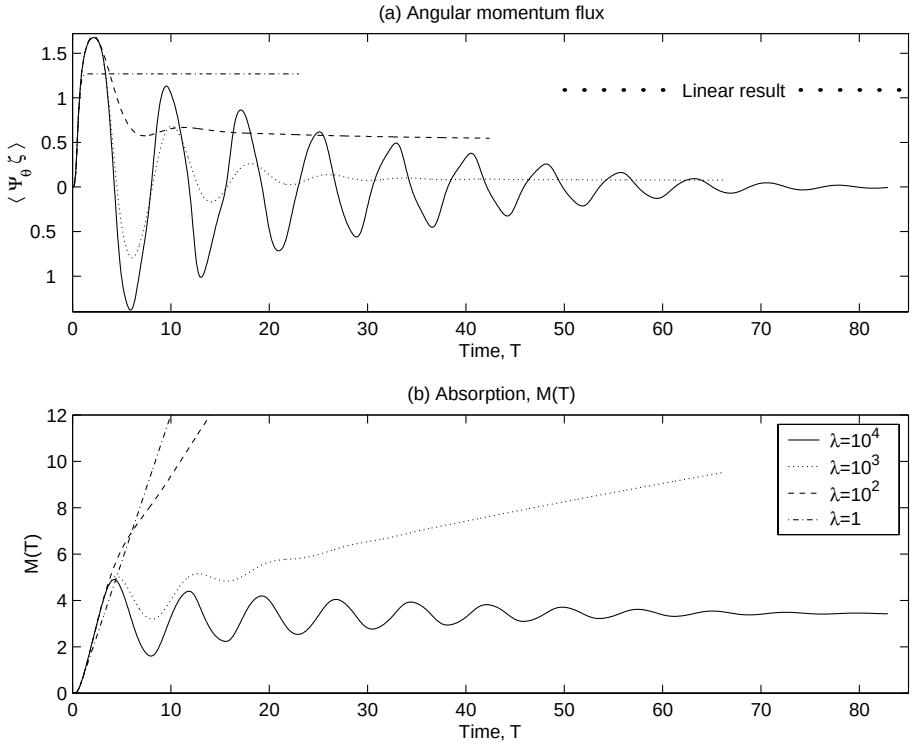

Figure 11. Angular momentum flux into the critical region for runs with different radial viscosities, λ.

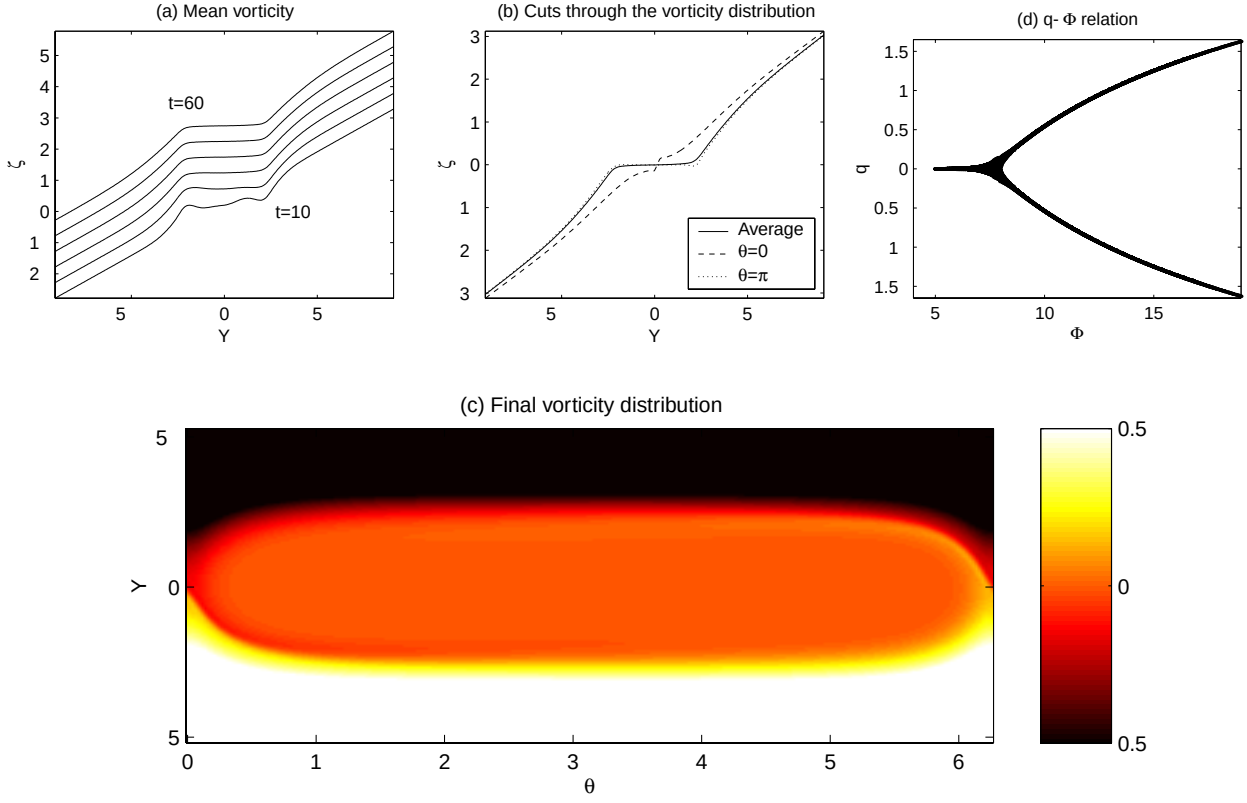

Figure 12. (a) The azimuthally averaged potential vorticity distribution for the run with $\lambda = 10^{-3}$. (b) The final mean distribution along with cuts through $q(Y, \theta, 100)$ at $\theta = 0$ and π. (c) The vorticity distribution as a whole. (d) The values of (Φ, q) from points taken over the domain, which collapse closely on to a multi-branched curve.

of the critical region, the dissipation halts the saturation of the corotation torque and permits a continued absorption of angular momentum. Ward (1997) has also suggested that saturation can be avoided if there is significant dissipation in the disc.

A significant limitation of the current study is that we have imposed a perturbation without allowing any feedback of the disc response on the perturber. In the astrophysical application, the transfer of angular momentum to the disc exacts an equal and opposite reaction on the perturber, which drifts radially as a result. This effect is missing from our current computations; the perturber's orbit is fixed. Hence, whilst our results may still be indicative, we cannot truly address the important astrophysical issues. A future, and more self-consistent, calculation will include the interaction between the disc and the perturber, and will also model the growth of the planet on a realistic time-scale.

Finally, although the main application of the theory presented here is to planet–disc interactions, it has also been proposed that disc turbulence generates persistent, large-scale vortices. Moreover, as a result of their ability to concentrate heavy particles suspended in the flow, these structures may play an important role in planet formation (Barge & Sommeria 1995; Bracco et al. 1999a,b). This idea has been criticized because it is not clear whether discs can spontaneously generate vortices (Hawley, Balbus & Winters 1999) or whether they can survive for long periods if formed. However, the current problem highlights the emergence of persistent vortices as a natural fluid mechanical phenomenon in the vicinity of corotation (at least when dissipation is small). Should these vortices be the setting of the birth of further protoplanets, the stage is set for interesting orbital dynamics once these objects grow large enough to interact with the parent body.

In conclusion, we stress that the analysis presented here has several advantages over simulations of the full equations for the disc. For example, the matched asymptotics filters the relatively fast propagation of spiral waves in the bulk of the disc and naturally focuses attention on the location where most action occurs – the critical region. As a result, we are able to evolve the fluid motions over relatively long time-scales. Moreover, we place a substantial number of grid points over the special region and resolve the fluid motions in great detail, which further allows us to access the inviscid limit. Full numerical simulations, on the other hand, are far more dissipative and technically complicated because of the need to resolve faster time-scales and the entire disc (including the inner and outer edges). Of course, any asymptotic analysis has limited validity; here, our theory is only applicable to fairly weak external perturbations. Nevertheless, we view the type of approach taken here to be useful and complementary to the more common approach of full numerical simulation.

ACKNOWLEDGMENTS

We thank D. N. C. Lin and J. C. B. Papaloizou for useful discussions, as well as the referee, P. Artymowicz, for helpful comments.

REFERENCES

Artymowicz P., 1993, ApJ, 419, 155
Balmforth N. J., Piccolo C., 2001, J. Fluid Mech., 426, 95
Balmforth N. J., Llewellyn Smith S. G., Young W. R., 2001, J. Fluid Mech., in press
Barge P., Sommeria J., 1995, A&A, 295, L1
Bracco A., Chavanis P. H., Provenzale A., Spiegel E. A., 1999a, Phys. Fluids A, 11, 2280
Bracco A., Provenzale A., Spiegel E. A., Yecko P., 1999b, in Abramowicz M., Björnsson G., Pringle J., eds, Theory of Black Hole Accretion Disks. Cambridge Univ. Press, Cambridge
Brown S. N., Stewartson K., 1978, Geophys. Astrophys. Fluid Dyn., 10, 1
Bryden G., Chen X., Lin D. N. C., Nelson R. P., Papaloizou J. C. B., 1999, ApJ, 514, 344
Cheng C. Z., Knorr G., 1976, J. Comput. Phys., 22, 330
Goldreich P., Nicholson P. D., 1989, ApJ, 342, 1079
Goldreich P., Tremaine S., 1979, ApJ, 233, 857
Goldreich P., Tremaine S., 1980, ApJ, 241, 425
Goldreich P., Tremaine S., 1981, ApJ, 243, 1062
Hawley J. F., Balbus S. A., Winters W. F., 1999, ApJ, 518, 394
Haynes P. H., 1985, J. Fluid Mech., 161, 493
Haynes P. H., 1989, J. Fluid Mech., 207, 231
Killworth P. D., McIntyre M. E., 1985, J. Fluid Mech., 161, 449
Korycansky D. G., Papaloizou J. C. B., 1995, MNRAS, 274, 85
Korycansky D. G., Papaloizou J. C. B., 1996, ApJS, 105, 181
Korycansky D. G., Pollack J. B., 1993, Icarus, 102, 150
Lin D. N. C., Papaloizou J. C. B., 1979, MNRAS, 186, 799
Lin D. N. C., Papaloizou J. C. B., 1986, ApJ, 307, 395
Lubow S. H., Ogilvie G. I., 1998, ApJ, 504, 983
McWilliams J. C., 1990, J. Fluid Mech., 219, 361
Mayor M., Queloz D., 1995, Nat, 378, 355
Miyoshi K., Takeuchi T., Tanaka H., Ida S., 1999, ApJ, 516, 451
Narayan R., Goodman J., Goldreich P., 1987, MNRAS, 228, 1
Papaloizou J. C. B., Lin D. N. C., 1984, ApJ, 285, 818
Papaloizou J. C. B., Pringle J. E., 1987, MNRAS, 225, 267
Papaloizou J. C. B., Korycansky D. G., Terquem C., 1995, Ann. N.Y. Acad. Sci., 773, 261
Stewartson K., 1978, Geophys. Astrophys. Fluid Dyn., 9, 185
Ward W. R., 1989, ApJ, 336, 526

Ward W. R., 1997, Icarus, 126, 261
Warn T., Warn H., 1978, Stud. Appl. Math., 59, 37

APPENDIX A: ASYMPTOTIC VALUES OF THE TRANSFER FUNCTION

In certain limits of the problem we can analytically construct the transfer function, R_m, and the forcing, F_m. The essential detail that facilitates the construction is that in these limits the particular solution becomes a slowly decaying function with radius, but the homogeneous solutions vary exponentially quickly. As a result, inside the evanescent region surrounding corotation, the homogeneous solutions take the form $\psi_m = \exp[\pm K_m(r-1)]$, and the corotation singularity only appears at higher order. Consequently, the relevant homogeneous solution that satisfies the boundary conditions is

$$\psi_m \sim \exp(-K_m|r-1|); \tag{A1}$$

the exponential decay away from corotation is sufficient to match the boundary conditions to leading order, independently of the detailed propagation characteristics in the regions beyond the evanescent zone. Hence, for $\delta \to 0$,

$$[\psi_{mr}]_{1-\delta}^{1+\delta} \sim -2K_m\psi_m(1). \tag{A2}$$

For simplicity, we sketch out the details of two special limits for a disc with constant surface density.

A1 Large *m*

When $m \gg 1$, the equation for ψ_m becomes

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{rc^2}{\Upsilon}\frac{\mathrm{d}\psi_m}{\mathrm{d}r}\right) - \frac{m^2}{r^2}\psi_m \sim \frac{\Sigma^2 Q\Phi_m}{\Upsilon} + \frac{\Sigma}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{r^2(\Omega-1)\Phi_m}{\Upsilon}\right]. \tag{A3}$$

Now, $\Phi_m \approx p_m \log|r-1| + q_m$ near $r=1$, where $p_m = -2/\pi$ and $q_m = (2/\pi)[\gamma - \ln(m/2)]$, and γ is Euler's constant (Goldreich & Tremaine 1980). For a disc with constant surface density in Keplerian rotation the (scaled or dimensionless) asymptotic solution near $r=1$ is

$$\psi_m = \Psi_m \mathrm{e}^{-m|r-1|} + \frac{3q_m}{2m^2c^2}(1 - \mathrm{e}^{-m|r-1|}). \tag{A4}$$

Hence

$$J_m^+ - J_m^- = -2m\left(\Psi_m - \frac{3q_m}{2m^2c^2}\right), \tag{A5}$$

or

$$\Psi_m = \frac{3q_m}{2m^2c^2} - \frac{1}{2m}(J_m^+ - J_m^-). \tag{A6}$$

A2 Tight-winding approximation

When $c \to 0$, we use an alternative equation for the variable, $\varphi_m = w_m + \Phi_m$:

$$\frac{\mathrm{d}^2\varphi_m}{\mathrm{d}r^2} + \frac{\mathrm{d}}{\mathrm{d}r}\left[\ln\left(\frac{r\Sigma}{D}\right)\right]\frac{\mathrm{d}\varphi_m}{\mathrm{d}r} + \left[\frac{2\Omega}{r(\Omega-1)}\frac{\mathrm{d}}{\mathrm{d}r}\ln\left(\frac{\Sigma\Omega}{D}\right) - \frac{m^2}{r^2} + \frac{D}{c^2}\right]\varphi_m = \frac{D}{c^2}\Phi_m, \tag{A7}$$

where $D = m^2(\Omega - \Omega_c)^2 - \kappa^2$. Near corotation,

$$\frac{\mathrm{d}^2\varphi_m}{\mathrm{d}r^2} - \frac{\kappa^2}{c^2}\varphi_m = -\frac{\kappa^2}{c^2}\Phi_m. \tag{A8}$$

Hence,

$$\varphi_m \sim \Pi_m \mathrm{e}^{-|r-1|/c} + (q_m - \tfrac{3}{2}p_m)(1 - \mathrm{e}^{-|r-1)|/c}), \tag{A9}$$

and so, for $r \to 1$,

$$\Pi_m = q_m - \tfrac{3}{2}p_m - \tfrac{1}{2}c[\varphi_{mr}]_{1-\delta}^{1+\delta}. \tag{A10}$$

Finally, since $\psi_m = -2\varphi_m$ and $\psi_{mr} = -\varphi_{mr}/2$ near corotation [from $\Sigma(Q\psi_m + \varphi_m) \approx (\Omega-1)\psi_{mr}$],

$$\Psi_m = 3p_m - 2q_m - 2c(J_m^+ - J_m^-). \tag{A11}$$

APPENDIX B: LIMITS OF THE DISSIPATIVE PROBLEM

In this Appendix we consider the reduced model with a radial viscous stress:

$$\zeta_T + Y\zeta_\theta - \Phi_\theta\zeta_Y + \beta\Psi_\theta = \lambda\zeta_{YY}. \tag{B1}$$

B1 The quasi-steady critical region

The numerical solutions presented in the main text illustrate how the vorticity distribution approaches a steady state over long times. If the viscosity is relatively weak, we can construct this state analytically by perturbation theory. After discarding the time derivative, switching to the total vorticity, $q = \zeta - \beta Y$, and introducing the coordinate transformation, $(Y, \theta) \rightarrow (\Phi, \theta)$, where $\Phi = \Psi + Y^2/2$ denotes the total streamfunction, we arrive at another version of the vorticity equation:

$$q_\theta = \lambda(Yq_\Phi)_\Phi. \tag{B2}$$

The left-hand term of this equation is small to leading order, and so $q \approx Q(\Phi)$, where $Q(\Phi)$ is to be determined. In other words, the action of the advection of the flow is to mix the vorticity distribution along streamlines; the viscous diffusion across these streamlines is slower and captured by the next-order terms. The streamline pattern shown in Fig. 7 is typical.

To find the function $Q(\Phi)$, we average (B2) over each streamline. Equivalently, we integrate equation (B2) with respect to θ; if the streamline is open, we integrate over $[0,2\pi]$, and if it is one of the closed streamlines, we integrate around the contour (see Fig. 7). The right-hand side then vanishes, leaving

$$\frac{\mathrm{d}}{\mathrm{d}\Phi}\left[I(\Phi)\frac{\mathrm{d}q}{\mathrm{d}\Phi}\right] = 0, \qquad I(\Phi) = \int_C \sqrt{2(\Phi - \Psi)}\,\mathrm{d}\theta, \tag{B3}$$

where C denotes the relevant integration range.

To leading order, we may therefore write

$$Q = K_1^\alpha \int_{\Phi_S}^{\Phi} \frac{\mathrm{d}\Phi'}{I(\Phi')} + K_2^\alpha, \tag{B4}$$

where we fix the lower limit of integration, Φ_S, to be the value of the total streamfunction on the separatrices that divide the regions of open streamlines from the closed streamline region ($\Phi_S \equiv \Psi_{max}$). The integration constants, K_1^α and K_2^α, need not have the same values in these two regions (hence we add the superscript α to distinguish them). However, the vorticity is continuous across the separatrices (Brown & Stewartson 1978) and must be antisymmetrical in Y (Φ is even in Y). Hence $K_1^\alpha = K_2^\alpha = 0$ inside the closed streamline region, and $K_2^\alpha = 0$ outside, giving

$$Q = \begin{cases} K_1^\alpha \int_{\Phi_S}^{\Phi} \mathrm{d}\Phi'/I(\Phi') & \text{for } \Phi > \Phi_S, \\ 0 & \text{for } \Phi \le \Phi_S. \end{cases} \tag{B5}$$

Finally, the boundary conditions require that $q \rightarrow -\beta Y$ for $Y \rightarrow \pm\infty$. This determines the remaining constants: $K_1^\alpha = -2\pi\beta\,\mathrm{sgn}(Y)$.

Although we apply the boundary condition to determine K_1^α in the open streamline region, the resulting solution does not actually satisfy the true boundary condition, $\zeta \rightarrow 0$ as $Y \rightarrow \pm\infty$. In fact,

$$\zeta \rightarrow \pm\beta\left\{\sqrt{2\Phi_S} - \int_{\Phi_S}^{\infty}\left[\frac{2\pi}{I(\Phi)} - \frac{1}{\sqrt{2\Phi}}\right]\mathrm{d}\Phi\right\}, \tag{B6}$$

which is certainly not zero. This feature of the solution arises because the radial viscosity causes the vorticity inside the critical region to spread outwards (Brown & Stewartson 1978). This creates an expanding diffusion layer of the kind considered in the second half of this Appendix. In the outer diffusion layer, the vorticity is brought back to zero, but the solution develops a vorticity jump across the inner parts of the critical region.

There is a notable consequence of the development of the vorticity jump: In a quasi-steady state, the angular momentum balance within the critical region is now replaced by

$$\langle\Psi_\theta\zeta\rangle = -\lambda[\zeta]_{-\infty}^{\infty} \equiv 2\lambda\beta\left\{\int_{\Phi_S}^{\infty}\left[\frac{2\pi}{I(\Phi)} - \frac{1}{\sqrt{2\Phi}}\right]\mathrm{d}\Phi - \sqrt{2\Phi_S}\right\}. \tag{B7}$$

Thus, the angular momentum flux into the critical region is finite in the quasi-steady situation, a circumstance that arises because of the viscous spread of the vorticity. Importantly, this means that the corotation torque will not saturate whenever there is radial diffusion. However, the torque is $O(\lambda)$ for $\lambda \ll 1$.

To complete the leading-order solution, we finally need to compute the integral $\langle \zeta \mathrm{e}^{-\mathrm{i}m\theta} \rangle$ using the vorticity–streamfunction relation, $\zeta = \beta Y + Q(\Phi)$. This leads to a functional equation for $\Psi(\theta)$ that must be solved to specify the solution. Here we will not explicitly complete this step.

B2 The limit of large dissipation

When $\lambda \gg 1$, in order to balance the large dissipative term on the right-hand side of the vorticity equation, we rescale the coordinates:

$$T \rightarrow \lambda^{-1/3}\tau, \qquad Y \rightarrow \lambda^{1/3} z \qquad \text{and} \qquad \zeta(Y, \theta, T) \rightarrow \lambda^{-1/3}\tilde{\zeta}(z, \theta, \tau). \tag{B8}$$

Then, on dropping a tilde,

$$\zeta_\tau + z\zeta_\theta - \zeta_{zz} = -\beta\Psi_\theta + \lambda^{-2/3}\Phi_\theta\zeta_Y. \tag{B9}$$

Following Brown & Stewartson (1978), we introduce the asymptotic sequences

$$\zeta = \zeta_0 + \lambda^{-2/3}\zeta_1 + \ldots \qquad \text{and} \qquad \Psi = \Psi_0 + \lambda^{-2/3}\Psi_1 + \ldots. \tag{B10}$$

At leading order, we recover the viscous version of the linear vorticity equation:

$$\zeta_{0\tau} + z\zeta_{0\theta} - \zeta_{0zz} = -\beta\Psi_{0\theta}. \tag{B11}$$

We solve this equation by decomposing into the azimuthal Fourier modes and then Fourier transforming in z. The first result we obtain is that

$$\Psi_{0m} = \frac{F_m}{1 + \mathrm{i}\pi\beta R_m}, \tag{B12}$$

which repeats the inviscid linear theory, and illustrates how the linear dynamics are insensitive to the details of the dissipative mechanism. In the problem we consider, the perturber is also turned on over a short interval $[0,T_\mathrm{o}]$ (with $T_\mathrm{o} = 1$ for our example of Section 5). To simplify matters, we replace the smooth turn-on by a sudden one:

$$F_m(T) \rightarrow \tilde{F}_m H(\tau) \qquad \text{and} \qquad \Psi_{0m} \rightarrow \tilde{\Psi}_{0m} H(\tau). \tag{B13}$$

Our second result is then

$$\zeta_0 = -\mathrm{i}\beta\Psi_{0m}\int_0^{m\tau} \mathrm{e}^{-q^3/3m - \mathrm{i}qz}\,\mathrm{d}q. \tag{B14}$$

The important term in the asymptotic solution that reveals the viscous spread of the critical region vorticity is the azimuthal mean of the next-order correction, $\bar{\zeta}_1(z,\tau)$ (where the bar refers to the azimuthal average). This quantity satisfies the relation

$$\bar{\zeta}_{1\tau} - \bar{\zeta}_{1zz} = (\overline{\Psi_{0\theta}\zeta_0})_z. \tag{B15}$$

More algebraic manipulations then furnish the result,

$$\bar{\zeta}_1 = 2\beta\sum_{m>0} m|\Psi_{0m}|^2 \int_0^{m\tau} \frac{\sin qz}{q}\left[\mathrm{e}^{-q^3/3m} - \mathrm{e}^{2q^3/3m - q^2\tau}\right]\mathrm{d}q \approx \pi\beta\left[1 - \mathrm{Erf}\left(\frac{|z|}{2\sqrt{\tau}}\right)\right]\mathrm{sgn}(z)\sum_{m>0} m|\Psi_{0m}|^2, \tag{B16}$$

for large $|z|$ and τ with $z^2/T = O(1)$ (Brown & Stewartson 1978). Therefore, for any finite time, $\bar{\zeta}_1 \rightarrow 0$ as $z \rightarrow 0$, in line with the statement of the original boundary conditions. However, for $\tau \rightarrow \infty$, $\bar{\zeta}_1 \rightarrow \pi\beta\,\mathrm{sgn}(z)\sum_{m>0} m|\Psi_{0m}|^2$ as we move out of the critical region. In other words, there is an expanding diffusion layer that carries vorticity out into the bulk of the disc, and eventually creates a jump in vorticity across the critical region.

This paper has been typeset from a TEX/LATEX file prepared by the author.