FORMATION, SURVIVAL, AND DESTRUCTION OF VORTICES IN ACCRETION DISKS

Yoram Lithwick
CITA, Toronto, ON, Canada; yoram@cita.utoronto.ca

ABSTRACT

Two-dimensional hydrodynamical disks are nonlinearly unstable to the formation of vortices. Once formed, these vortices essentially survive forever. What happens in three dimensions? We show with incompressible shearing box simulations that in three dimensions, a vortex in a short box forms and survives just as in two dimensions. But a vortex in a tall box is unstable and is destroyed. In our simulation, the unstable vortex decays into a transient turbulent-like state that transports angular momentum outward at a nearly constant rate for hundreds of orbital times. The three-dimensional instability that destroys vortices is a generalization of the two-dimensional instability that forms them. We derive the conditions for these nonlinear instabilities to act by calculating the coupling between linear modes, and thereby derive the criterion for a vortex to survive in three dimensions as it does in two dimensions: the azimuthal extent of the vortex must be larger than the scale height of the accretion disk. When this criterion is violated, the vortex is unstable and decays. Because vortices are longer in azimuthal than in radial extent by a factor that is inversely proportional to their excess vorticity, a vortex with given radial extent will only survive in a three-dimensional disk if it is sufficiently weak. This counterintuitive result explains why previous three-dimensional simulations always yielded decaying vortices: their vortices were too strong. Weak vortices behave two-dimensionally even if their width is much less than their height because they are stabilized by rotation, and behave as Taylor–Proudman columns. We conclude that in protoplanetary disks, weak vortices can trap dust and serve as the nurseries of planet formation. Decaying strong vortices might be responsible for the outward transport of angular momentum that is required to make accretion disks accrete.

Key words: accretion, accretion disks – instabilities – solar system: formation – turbulence

Online-only material: color figures

1. INTRODUCTION

Matter accretes onto a wide variety of objects, such as young stars, black holes, and white dwarfs, through accretion disks. In highly ionized disks, magnetic fields are important, and they trigger turbulence via the magnetorotational instability (Balbus & Hawley 1998). However, many disks, such as those around young stars or dwarf novae, are nearly neutral (e.g., Sano et al. 2000; Gammie & Menou 1998). In these disks, the fluid motions are well described by hydrodynamics.

Numerical simulations of hydrodynamical disks in two dimensions—in the plane of the disk—often produce long-lived vortices (Godon & Livio 1999; Umurhan & Regev 2004; Johnson & Gammie 2005). If vortices really exist in accretion disks, they can have important consequences. First and foremost, they might generate turbulence. Since turbulence naturally transports angular momentum outward,[1] as is required for mass to fall inward, it might be vortices that cause accretion disks to accrete. Second, in disks around young stars, long-lived vortices can trap solid particles and initiate the formation of planets (Barge & Sommeria 1995).

Why do vortices naturally form in two-dimensional simulations? Hydrodynamical disks are stable to linear perturbations. However, they are nonlinearly unstable, despite some claims to the contrary in the astrophysical literature. In two dimensions, the incompressible hydrodynamical equations of a disk are equivalent to those of a nonrotating linear shear flow (e.g., Lithwick 2007, hereafter L07). And it has long been known

that such flows are nonlinearly unstable (Gill 1965; Lerner & Knobloch 1988; L07). This nonlinear instability is just a special case of the Kelvin–Helmholtz instability. Consider a linear shear flow extending throughout the x–y plane with a velocity profile $\boldsymbol{v} = -qx\hat{\boldsymbol{y}}$, where $q > 0$ is the constant shear rate, so that $-q$ is the flow's vorticity. (In the equivalent accretion disk, the local angular speed is $\Omega = 2q/3$.) This shear flow is linearly stable to infinitesimal perturbations. But if the shear profile is altered by a small amount, the alteration can itself be unstable to infinitesimal perturbations. To be specific, let the alteration be confined within a band of width Δx, and let it have vorticity $\omega = \omega(x)$ (with $|\omega| \lesssim q$), so that it induces a velocity field in excess of the linear shear with components $u_y \sim \omega \Delta x$ and $u_x = 0$. Then this band is unstable to infinitesimal nonaxisymmetric (i.e., nonstream-aligned) perturbations provided roughly that

$$|k_y| \lesssim \frac{1}{q}\frac{|\omega|}{\Delta x} \quad \Rightarrow \text{two-dimensional instability,} \qquad (1)$$

where k_y is the wavenumber of the nonaxisymmetric perturbation.[2] For any value of $|\omega|$ and Δx, the band is always unstable to perturbations with a long enough wavelength. Remarkably, instability even occurs when $|\omega|$ is infinitesimal. Hence,

---

[1] Energy conservation implies that turbulence transports angular momentum outward; see Section 3. Nonetheless, if an external energy source (e.g., the radiative energy from the central star) drives the turbulence, then angular momentum could, in principle, be transported inward.

[2] More precisely, the necessary and sufficient condition for instability in the limit $|\omega| \ll q$ is that $|k_y| < \frac{1}{2q}\int_{-\infty}^{\infty}\frac{d\omega/dx}{x-x_0}dx$, where x_0 is any value of x at which $d\omega/dx = 0$ (Gill 1965; Lerner & Knobloch 1988; L07). For arbitrarily large ω, Rayleigh's inflection point theorem and Fjørtoft's theorem give necessary (though insufficient) criteria for instability (Drazin & Reid 2004). The former states that for instability, it is required that $d\omega/dx = 0$ somewhere in the flow, that is, the velocity field must have an inflection point. Lovelace et al. (1999) generalized Rayleigh's inflection point theorem to compressible and nonhomentropic disks.

we may regard this as a true nonlinear instability. Balbus & Hawley (2006) asserted that detailed numerical simulations have not shown evidence of nonlinear instability. The reason many simulations fail to see it is that their boxes are not long enough in the y-direction to encompass a small enough nonzero $|k_y|$.

In two dimensions, the outcome of this instability is a long-lived vortex (e.g., L07). A vortex that has been studied in detail is the Moore–Saffman vortex, which is a localized patch of spatially constant vorticity superimposed on a linear shear flow (Saffman 1995). When $|\omega| \lesssim q$, where ω here refers to the spatially constant excess vorticity within the patch, and when the vorticity within the patch ($\omega - q$) is stronger than that of the background shear, then the patch forms a stable vortex that is elongated in y relative to x by the factor

$$\frac{\Delta y}{\Delta x} \sim \frac{q}{|\omega|}. \tag{2}$$

This relation applies not only to Moore–Saffman vortices, but also to vortices whose ω is not spatially constant. It may be understood as follows. A patch with characteristic excess vorticity $\sim \omega$ and with $\Delta y \gg \Delta x$ induces a velocity field in the x-direction with amplitude $u_x \sim |\omega| \Delta x$, independent of the value of Δy (e.g., Section 6 in L07). As long as $|\omega| \lesssim q$, the y-velocity within the vortex is predominantly due to the background shear, and is $\sim q \Delta x$. Therefore, the time to cross the width of the vortex is $t_x \sim \Delta x / u_x \sim 1/|\omega|$ and the time to cross its length is $t_y \sim \Delta y / (q \Delta x)$. Since these times must be comparable in a vortex, Equation (2) follows. Equation (2) is very similar to Equation (1). The two-dimensional instability naturally forms into a two-dimensional vortex. Furthermore, the exponential growth rate of the instability is $\sim |\omega|$, which is comparable to the rate at which fluid circulates around the vortex.

More generally, an arbitrary axisymmetric profile of $\omega(x)$ tends to evolve into a distinctive banded structure. Roughly speaking, bands where $\omega < 0$ contain vortices, and these are interspersed with bands where $\omega > 0$, which contain no vortices. (Recall that we take the background vorticity to be negative; otherwise, the converse would hold.) The reason for this is that only regions that have $\omega < 0$ can be unstable, as may be inferred either from the integral criterion for instability given in footnote 2 or from Fjørtoft's theorem. For more detail on vortex dynamics in shear flows, see the review by Marcus (1993).

What happens in three dimensions? To date, numerical simulations of vortices in three-dimensional disks have been reported in two papers. Barranco & Marcus (2005) initialized their simulation with a Moore–Saffman vortex and solved the anelastic equations in a stratified disk. They found that this vortex decayed. As it decayed, new vortices were formed in the disk's atmosphere, two scale heights above the midplane. The new vortices survived for the duration of the simulation. Shen et al. (2006) performed both two-dimensional and three-dimensional simulations of the compressible hydrodynamical equations in an unstratified disk, initialized with large random fluctuations. They found that whereas the two-dimensional simulations produced long-lived vortices, in three dimensions vortices rapidly decayed.

Intuitively, it seems clear that a vortex in a very thin disk will behave as it does in two dimensions. And from the three-dimensional simulations described above, it may be inferred that placing this vortex in a very thick disk will induce its decay. Our main goal in this paper is to understand these two behaviors and the transition between them. A crude explanation of our

final result is that vortices decay when the two-dimensional vortex motion resonantly couples to three-dimensional modes, that is, to modes that have a vertical wavenumber $k_z \neq 0$. As described above, a vortex with excess vorticity $|\omega|$ has circulation frequency $\sim |\omega|$, and $k_y / k_x \sim |\omega| / q$, where k_x and k_y are its "typical" wavenumbers. Furthermore, it is well known that the frequency of axisymmetric ($k_y = 0$) inertial waves is $\Omega k_z / \sqrt{k_x^2 + k_z^2}$ (see Equation (41)). Equating the two frequencies, and taking the k_x of the three-dimensional mode to be comparable with the k_x of the vortex, as well as setting $q = 3\Omega/2$ for a Keplerian disk, we find

$$k_z \sim k_y \tag{3}$$

as the condition for resonance. Therefore, a vortex with length Δy will survive in a box with height $\Delta z \lesssim \Delta y$, because in such a box, all three-dimensional modes have too high a frequency to couple with the vortex, that is, all nonzero k_z exceed the characteristic $k_y \sim 1/\Delta y$. But when $\Delta z \gtrsim \Delta y$, there exist k_z in the box that satisfy the resonance condition (3), leading to the vortex's destruction. This conclusion suggests that vortices live indefinitely in disks with scale height less than their length ($h \lesssim \Delta y$) because in such disks, all three-dimensional modes have too high a frequency for resonant coupling. This conclusion is also consistent with the simulations of Barranco & Marcus (2005) and Shen et al. (2006). Both of these works initialized their simulations with strong excess vorticity $|\omega| \sim q$, corresponding to nearly circular vortices. Both had vertical domains that were comparable with the vortices' width. Therefore, both saw that their vortices decayed. Had they initialized their simulations with smaller $|\omega|$, and increased the box length L_y to encompass the resulting elongated vortices, both would have found long-lived three-dimensional vortices. Barranco & Marcus's discovery of long-lived vortices in the disk's atmosphere is simple to understand because the local scale height is reduced in inverse proportion to the height above the midplane. Therefore, higher up in the atmosphere, the dynamics becomes more two dimensional, and a given vortex is better able to survive the higher it is.[3]

1.1. Organization of the Paper

In Section 2, we introduce the equations of motion, and in Section 3, we present two pseudospectral simulations. One illustrates the formation and survival of a vortex in a short box and the other illustrates the destruction of a vortex in a tall box. In Sections 4 and 5, we develop a theory explaining this behavior. The reader who is satisfied by the qualitative description leading to Equation (3) may skip those two sections. The theory that we develop is indirectly related to the transient amplification scenario for the generation of turbulence. Even though hydrodynamical disks are linearly stable, linear perturbations can be transiently amplified before they decay, often by a large factor. It has been proposed that sufficiently amplified modes might couple nonlinearly, leading to turbulence (e.g., Chagelishvili et al. 2003; Yecko 2004; Afshordi et al. 2005). However, to make this proposal more concrete, one must work out how modes couple nonlinearly. In L07, we did that in two dimensions. We showed that the two-dimensional nonlinear instability of Equation (1) is a consequence of the coupling of an axisymmetric mode with

---

[3] However, Barranco & Marcus (2005) also included buoyancy forces in their simulations, which we ignore here. How buoyancy affects the stability of vortices is a topic for future work. But see Section 6 for some speculations.

a transiently amplified mode, which may be called a "swinging mode" because its phasefronts are swung around by the background shear. In Section 5, we show that the three-dimensional instability responsible for the destruction of vortices is a generalization of this two-dimensional instability. It may be understood by examining the coupling of a three-dimensional swinging mode with an axisymmetric one. Three-dimensional modes become increasingly unstable as $|k_z|$ decreases, and in the limit that $k_z \to 0$, the three-dimensional instability smoothly matches onto the two-dimensional one. Thicker disks are more prone to three-dimensional instability because they encompass smaller $|k_z|$.

2. EQUATIONS OF MOTION

We solve the "shearing box" equations, which approximate the dynamics in an accretion disk on lengthscales much smaller than the distance to the disk's center. We assume incompressibility, which is a good approximation when relative motions are subsonic. We also neglect vertical gravity, and hence stratification and buoyancy, which is an oversimplification. To fully understand vortices in astrophysical disks, one must consider the effects of vertical gravity in addition to shear and rotation. In this paper, we consider only two pieces of this puzzle—shear and rotation. Adding the third piece—vertical gravity—is a topic that we leave for future investigations. See also Section 6 for some speculations.

An unperturbed Keplerian disk has an angular velocity profile $\Omega(r) \propto r^{-3/2}$. In a reference frame rotating at constant angular speed $\Omega_0 \equiv \Omega(r_0)$, where r_0 is a fiducial radius, the incompressible shearing box equations of motion read as

$$\partial_t \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{\nabla} \boldsymbol{v} = -2\Omega_0 \hat{\boldsymbol{z}} \times \boldsymbol{v} + 2q\Omega_0 x \hat{\boldsymbol{x}} - \boldsymbol{\nabla} P/\rho, \quad (4)$$
$$\boldsymbol{\nabla} \cdot \boldsymbol{v} = 0, \quad (5)$$

adopting Cartesian coordinates x, y, z, which are related to the disk's cylindrical r, θ via $x \equiv r - r_0$ and $y \equiv r_0(\theta - \Omega_0 t)$; $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{z}}$ are unit vectors, and

$$q \equiv -\left.\frac{d\Omega}{d \ln r}\right|_{r_0} = \frac{3}{2}\Omega_0. \quad (6)$$

We retain q and Ω_0 as independent parameters because they parameterize different effects: shear and rotation, respectively. The first term on the right-hand side of Equation (4) is the Coriolis force and the second is what remains after adding centrifugal and gravitational forces. Decomposing the velocity into

$$\boldsymbol{v} = -qx\hat{\boldsymbol{y}} + \boldsymbol{u}, \quad (7)$$

where the first term is the shear flow of the unperturbed disk, yields

$$\left(\partial_t - qx\partial_y\right)\boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{\nabla}\boldsymbol{u} = qu_x\hat{\boldsymbol{y}} - 2\Omega\hat{\boldsymbol{z}} \times \boldsymbol{u} - \boldsymbol{\nabla} P/\rho \quad (8)$$
$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0, \quad (9)$$

dropping the subscript from Ω_0, as we shall do in the remainder of this paper. An unperturbed disk has $\boldsymbol{u} = 0$.

In addition to the above "velocity–pressure" formulation, an alternative "velocity–vorticity" formulation will prove convenient. It is given by the curl of Equation (8):

$$(\partial_t - qx\partial_y)\boldsymbol{\omega} = -q\hat{\boldsymbol{y}}\omega_x + (2\Omega - q)\partial_z\boldsymbol{u} + \boldsymbol{\nabla} \times (\boldsymbol{u} \times \boldsymbol{\omega}), \quad (10)$$

where

$$\boldsymbol{\omega} \equiv \boldsymbol{\nabla} \times \boldsymbol{u} \quad (11)$$

is the vorticity of $\boldsymbol{u}$. Equation (10), together with the inverse of Equation (11)

$$\boldsymbol{u} = -\nabla^{-2}\boldsymbol{\nabla} \times \boldsymbol{\omega}, \quad (12)$$

forms a complete set.

Equation (10) implies that the total vorticity field is frozen into the fluid, because it is equivalent to

$$\partial_t \boldsymbol{\omega}_{\text{tot}} = \boldsymbol{\nabla} \times (\boldsymbol{v} \times \boldsymbol{\omega}_{\text{tot}}), \quad (13)$$

where

$$\boldsymbol{\omega}_{\text{tot}} \equiv (2\Omega - q)\hat{\boldsymbol{z}} + \boldsymbol{\omega} \quad (14)$$

is the total vorticity; note that $-q\hat{\boldsymbol{z}}$ is the vorticity of the unperturbed shear flow in the rotating frame and hence $(2\Omega - q)\hat{\boldsymbol{z}}$ is the unperturbed vorticity in the nonrotating frame. The vorticity–velocity picture is similar to magnetohydrodynamics (MHD), where it is the magnetic field that is frozen-in because it satisfies Equation (13) in place of $\boldsymbol{\omega}_{\text{tot}}$. However, in MHD the velocity field has its own dynamical equation, whereas in incompressible hydrodynamics it is directly determined from the vorticity field via Equation (12).

3. TWO PSEUDOSPECTRAL SIMULATIONS

The pseudospectral code is described in detail in the appendix of L07. It solves the velocity–pressure equations of motion with an explicit viscous term

$$\nu\nabla^2\boldsymbol{u} \quad (15)$$

added to Equation (8). (In L07, we did not include this term because we only considered inviscid flows.) The equations are solved in Fourier space by decomposing fields into spatial Fourier modes whose wavevectors are advected by the background flow $-qx\hat{\boldsymbol{y}}$. As a result, the boundary conditions are periodic in the y and z dimensions and "shearing periodic" in x. Most of our techniques are standard (e.g., Maron & Goldreich 2001; Rogallo 1981; Barranco & Marcus 2006). One exception is our method for remapping highly trailing wavevectors into highly leading ones, which is both simpler and more accurate than the usual method. In addition, our remapping does not introduce power into leading modes, because a mode's amplitude has always been set to zero before the remap. The code was extensively tested on two-dimensional flows in L07. A number of rather stringent three-dimensional tests are performed in this paper. We shall show that the code correctly reproduces the linear evolution of three-dimensional modes (Section 4), as well as the nonlinear coupling between them (Section 5). We also show in the present section that it tracks the various contributions to the energy budget, and that the sum of the contributions vanishes to high accuracy.

Figures 1–4 show results from two pseudospectral simulations. One simulation illustrates the formation and survival of a vortex, and the other illustrates vortex destruction. In the first (the "short box"), the number of Fourier modes used is $n_x \times n_y \times n_z = 64 \times 64 \times 32$, and the simulation box has dimensions $L_x = \frac{1}{15}$, $L_y = 1$, and $L_z = \frac{1}{2}$. In the second (the "tall box"), the setup is identical, except that it has $L_z = 2$ instead of $1/2$. Both simulations are initialized by setting

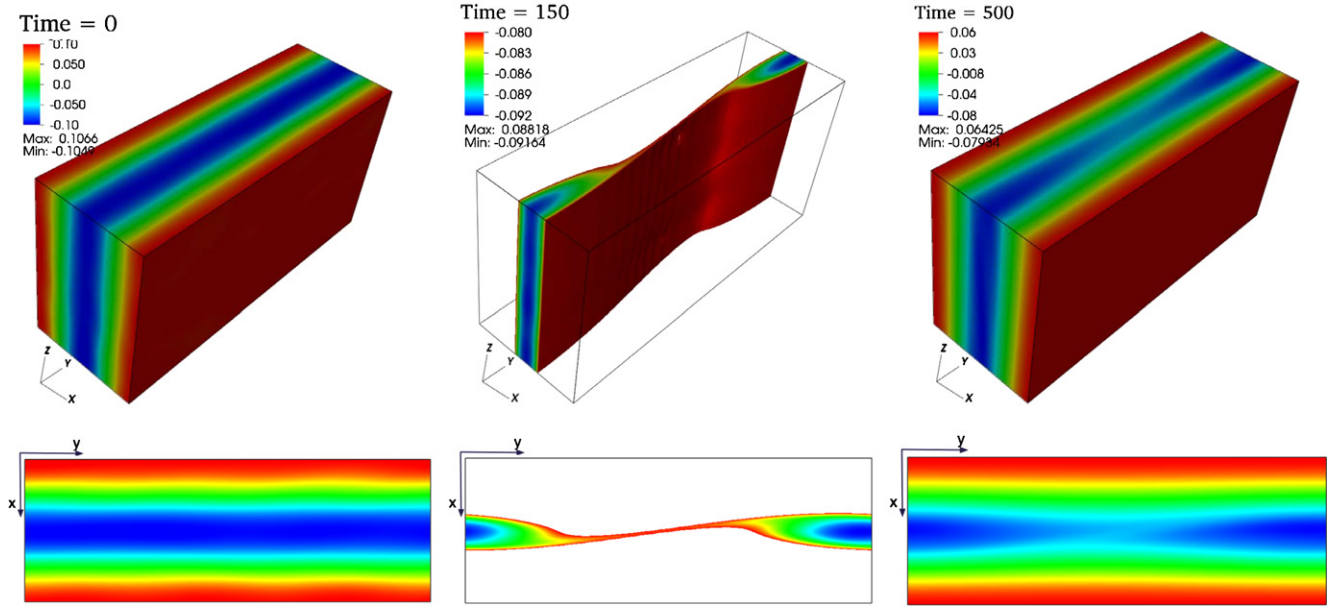$$\left.\omega_z\right|_{t=0} = -0.1\cos\left(\frac{2\pi}{L_x}x\right). \quad (16)$$

Figure 1. Vortex formation and survival in a short box. Color depicts ω_z. The initial state is unstable to vertically symmetric ($k_z = 0$) perturbations, and forms into a vortex. But it is stable to three-dimensional ($k_z \neq 0$) perturbations, and the evolution remains two dimensional. The bottom panels show horizontal slices through the boxes in the upper panels, midway through the boxes. At time = 150, the vortex has already formed. Only fluid with $\omega_z < -0.08$ is shown in the middle panels to highlight the vortex, and to illustrate that surfaces of constant ω_z remain purely vertical. At time = 500, the vortex still survives. Its amplitude slowly decays by viscosity, which acts on timescale = 1130. We set $\Omega = 1$ and $q = 3/2$. The number of modes in the simulation is $n_x \times n_y \times n_z = 64 \times 64 \times 32$, and the size of the simulation box is $(L_x, L_y, L_z) = (\frac{1}{15}, 1, \frac{1}{2})$. In this figure, L_z is to scale relative to L_y, but L_x has been expanded by a factor of 5 for clarity.

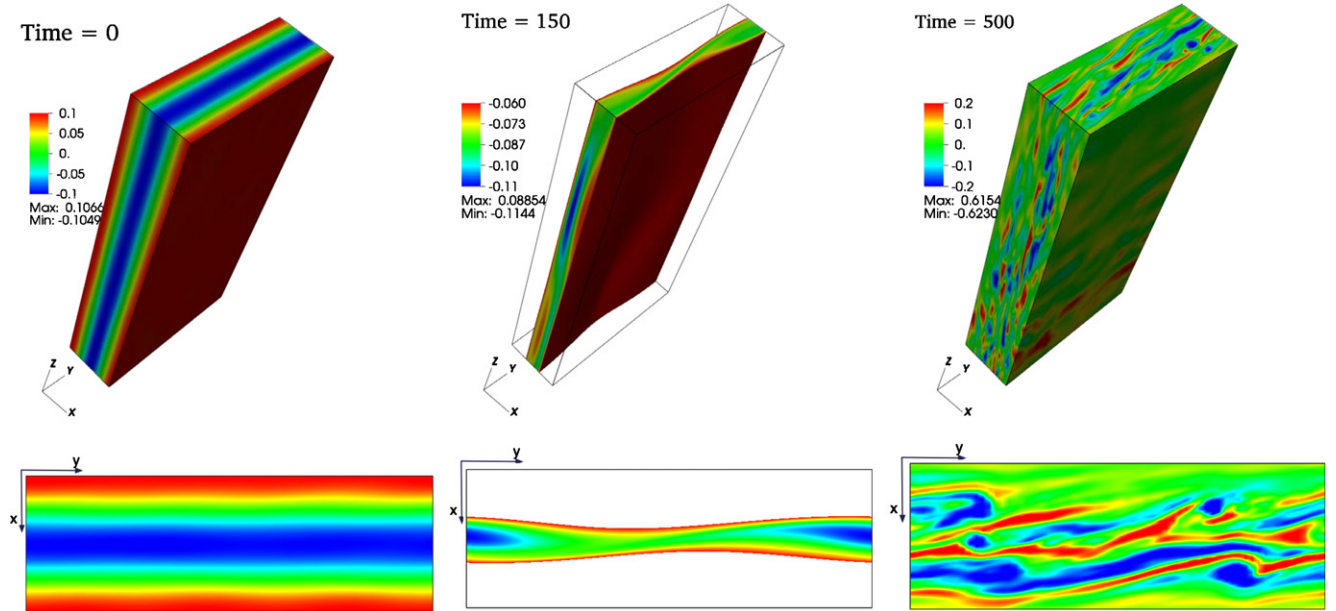(A color version of this figure is available in the online journal.)

Figure 2. Vortex destruction in a tall box. The setup is identical to the short-box simulation of Figure 1, except that the height L_z has been increased by a factor of 4, so that it now exceeds L_y. The resulting evolution is dramatically different. The initial state is now unstable not only to two-dimensional perturbations, but to three-dimensional ones as well. In the middle panels, surfaces of constant ω_z are warped, and the evolution is no longer vertically symmetric. In the right panels, the flow looks turbulent.

(A color version of this figure is available in the online journal.)

In addition, small perturbations are added to long-wavelength modes. Specifically, labeling the wavevectors as

$$(k_x, k_y, k_z) = 2\pi \left(\frac{j_x}{L_x}, \frac{j_y}{L_y}, \frac{j_z}{L_z} \right), \tag{17}$$

with integers (j_x, j_y, j_z), we select all modes that satisfy $|j_x| \leqslant 3$, $|j_y| \leqslant 3$, and $|j_z| \leqslant 3$, and set the Fourier amplitude of their ω_z to $10^{-4} e^{i\phi}$, where ϕ is a random

phase. But we exclude the $(j_x, j_y, j_z) = (0, 0, 0)$ mode, as well as $(j_x, j_y, j_z) = (\pm 1, 0, 0)$, which is given by Equation (16). Finally, we set $\Omega = 1$, $q = 3/2$, $\nu = 10^{-7}$, and integration time step $dt = 1/30$.

With our chosen initial conditions, the mode given by Equation (16) is nonlinearly unstable to vertically symmetric ($k_z = 0$) perturbations, and hence it tends to wrap up into a vortex. From the approximate criterion for instability
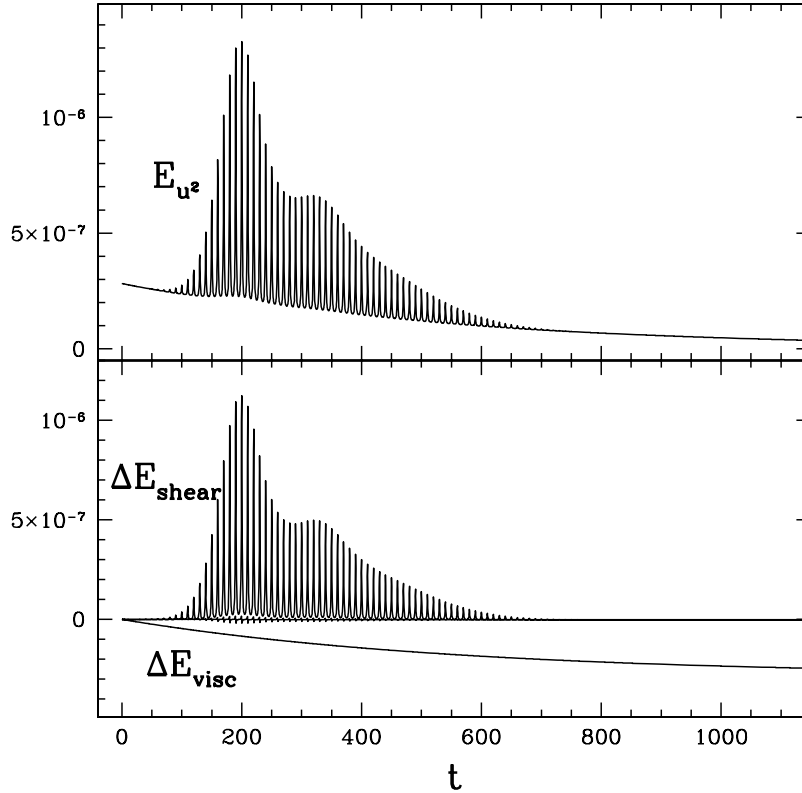
Figure 3. Energy in the short box. The three contributions to the energy budget, E_{u^2}, ΔE_{shear}, and ΔE_{visc}, are defined in Equations (20)–(22). E_{u^2} initially decays, and then rises to a peak near $t \sim 200$ as nonaxisymmetric perturbations turn the axisymmetric mode into a vortex. Subsequently, the vortex decays due to viscosity. The spikiness of the evolution is due to the boundary conditions, as explained in the text. Also shown in the bottom panel is the error due to numerical effects, ΔE_{error} (Equation (24)). It is unlabeled because it is mostly obscured by ΔE_{shear}. But it is nearly equal to zero everywhere, showing that the code accurately tracks the components of the energy budget.

(A color version of this figure is available in the online journal.)

(Equation (1)), we see that to illustrate the wrapping up into a vortex of a mode with a small amplitude, one must make the simulation box elongated in the y-direction relative to the x-scale of the mode in equation (16).

In the short box (Figure 1), the evolution proceeds just as it would in two dimensions. The initial mode indeed wraps up into a vortex, and the evolution remains vertically symmetric throughout. Once formed, the vortex can live for ever in the absence of viscosity. But in our simulation, there is a slow viscous decay. The timescale for viscous decay across the width of the vortex is $\sim 1/\nu k_x^2 = 1130$, taking $k_x = 2\pi/L_x$.

In the tall box (Figure 2), the evolution is dramatically different. In this case, the initial state is unstable not only to two-dimensional perturbations, but to three-dimensional ($k_z \neq 0$) ones as well. In the middle panel of that figure, we see that instead of forming a vertically symmetric vortex as in the short box, surfaces of constant ω_z are warped. By the third panel, the flow looks turbulent.

Figures 3 and 4 show the evolution of the energy in these simulations. Projecting $\boldsymbol{u}$ onto the Navier–Stokes equation (Equation (8) with viscosity included), and spatially averaging, we arrive at the energy equation

$$\frac{d}{dt}\frac{\langle u^2 \rangle}{2} = q\langle u_x u_y \rangle + \nu \langle \boldsymbol{u}\cdot\nabla^2\boldsymbol{u} \rangle, \tag{18}$$

after applying the shearing-box boundary conditions, where angled brackets denote a spatial average. The time integral of this equation is

$$E_{u^2} - E_{u^2}|_{t=0} = \Delta E_{shear} + \Delta E_{visc}, \tag{19}$$

where

$$E_{u^2} \equiv \frac{\langle u^2 \rangle}{2} \tag{20}$$

$$\Delta E_{shear} \equiv q \int_0^t \langle u_x u_y \rangle dt' \tag{21}$$

$$\Delta E_{visc} \equiv \nu \int_0^t \langle \boldsymbol{u}\cdot\nabla^2\boldsymbol{u} \rangle dt'. \tag{22}$$

The pseudospectral code records each of these terms, and Figure 3 shows the result in the short box simulation. At very early times, E_{u^2} decays from its initial value due to viscosity. At the same time, the small vertically symmetric perturbations grow exponentially, and they start to give order-unity perturbations by $t \sim 150$, by which time a vortex has been formed (Figure 1). As time evolves, E_{u^2} gradually decays due to viscosity on the viscous timescale = 1130. The evolution is very spiky. We defer a discussion of this spikiness to the end of this section.

Figure 4 shows the result in the tall box. The early evolution of E_{u^2} is similar to that seen in the short box. Both start with the same E_{u^2}, and an initial period of viscous decay is interrupted by exponentially growing perturbations. But in the tall box, not only are vertically symmetric modes growing, but modes with $k_z \neq 0$ are growing as well. By $t \sim 150$, there is a distorted vortex that subsequently decays into a turbulent-like state. The energy E_{u^2} rises to a value significantly larger than its initial one, and it continues to rise until $t \sim 600$, when it starts to decay. Throughout the time interval $300 \lesssim t \lesssim 600$, ΔE_{shear} rises nearly linearly in time, showing that $\langle u_x u_y \rangle$ is positive and nearly constant.
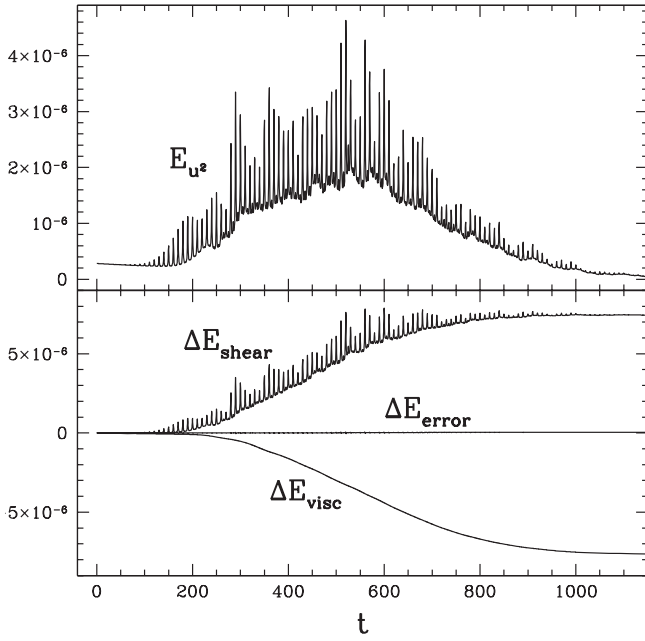
Figure 4. Energy in the tall box. The initial evolution is almost the same as that seen in the short box (Figure 3). But three-dimensional perturbations are unstable and force the destruction of the vortex. In the time interval $300 \lesssim t \lesssim 600$, while the initial axisymmetric disturbance decays in a turbulent-like state, the value of E_{u^2} is significantly larger than its initial value, and ΔE_{shear} rises nearly linearly in time, corresponding to nearly constant outward transport of angular momentum in a disk. The contribution of numerical errors to the time-integrated energy budget, ΔE_{error} (Equation (24)), remains small throughout.

It is intriguing that $\langle u_x u_y \rangle$ is positive for hundreds of orbits, because it suggests that decaying vortices might transport angular momentum outward in disks and, hence, drive accretion. Understanding the level of the turbulence, its lifetime, and its nature are topics for future work. Here, we merely address the sign of $\langle u_x u_y \rangle$. The quantity $\langle u_x u_y \rangle$ is the flux of y-momentum in the $+x$-direction (per unit mass and spatially averaged). It corresponds to the flux of angular momentum in a disk. A positive $\langle u_x u_y \rangle$ implies an outward flux of angular momentum, as is required to drive matter inward in an accretion disk. (Even though the shearing box cannot distinguish inward from outward, the sign of the angular momentum within a box depends on which side of the shearing box one calls inward. Therefore, outward transport of (positive) angular momentum is well defined in a shearing box.) In the shearing box, any force that tends to diminish the background shear flow $-qx\hat{y}$ necessarily transports y-momentum in the $+x$-direction. Hence, the fact that $\langle u_x u_y \rangle > 0$ in Figure 4 shows that the turbulence exerts forces that resist the background shear, as one might expect on physical grounds. One can also understand why $\langle u_x u_y \rangle > 0$ from energy considerations. Since $\Delta E_{\text{visc}} < 0$, as may be seen explicitly by an integration by parts, that is, $\langle \boldsymbol{u} \cdot \nabla^2 \boldsymbol{u} \rangle = -\sum_{i,j} \langle (\partial_j u_i)^2 \rangle$, Equation (19) may be rearranged to read as

$$\Delta E_{\text{shear}} = |\Delta E_{\text{visc}}| + E_{u^2} - E_{u^2}|_{t=0}. \qquad (23)$$

If the turbulence reaches a steady state—as it approximately does in Figure 4 during the time interval $300 \lesssim t \lesssim 600$—then the last two terms in the above equation are nearly constant, whereas $|\Delta E_{\text{visc}}|$ increases linearly with time. Hence, ΔE_{shear} must also increase. It is a general property of accretion disks that energy dissipation implies outward transport of angular momentum (e.g., Lynden-Bell & Pringle 1974). Since

turbulence always dissipates energy, it must also transport angular momentum outward. However, this argument can be violated if an external energy source drives the turbulence, in which case one would have to add this energy to the left-hand side of Equation (23). For example, the simulations of Stone & Balbus (1996) showed that convective disks can transport angular momentum inward when an externally imposed heat source drives the convection.

Also shown in the bottom panels of Figures 3 and 4 is the integrated energy error

$$\Delta E_{\text{error}} \equiv \Delta E_{\text{shear}} + \Delta E_{\text{visc}} + E_{u^2}|_{t=0} - E_{u^2} \qquad (24)$$

due to numerical effects, which is seen to be small. (In Figure 3, ΔE_{error} is not labeled because the curve is mostly obscured by ΔE_{shear}; it can be seen near $t \sim 200$, and is everywhere very nearly equal to zero.) The fact that ΔE_{error} nearly vanishes throughout the simulations is not guaranteed by the pseudospectral algorithm. Rather, we have chosen ν to be large enough that the algorithm introduces negligible error into the energy budget. To be more precise, at each time step in the pseudospectral code, modes that have $|j_x| > n_x/3$ or $|j_y| > n_y/3$ or $|j_z| > n_z/3$, where $j_{x,y,z}$ are defined via Equation (17), have their amplitudes set to zero ("dealiased"). This introduces an error that is analogous to grid error in grid-based codes. By choosing ν to be sufficiently large, it is the explicit viscosity that forces modes with large k to have small amplitudes, in which case the dealiasing procedure has little effect on the dynamics. Increasing the resolution $n_x \times n_y \times n_z$ would allow a smaller ν to be chosen—implying a larger effective Reynolds number—while keeping the energy error small.

The curves of E_{u^2} show sharp narrow spikes every time interval $\Delta t = 10$, with width ~ 1. Similar spikes have been seen in other simulations (Umurhan & Regev 2004; Shen et al. 2006), but they are stronger and narrower in our simulations because our simulation box is elongated. These spikes are due to the shearing-periodic boundary conditions. It is perhaps simplest to understand them by following the evolution in k-space, as we shall do in Section 5 (see also L07). But for now, we explain their origin in real space. By the nature of shearing-periodic boundary conditions, associated with the simulation box centered at $x = 0$ are "imaginary boxes" centered at $x = jL_x$ with integer $j = \pm 1, \pm 2, \ldots$. These imaginary boxes completely tile the x–y plane, and each contains a virtual copy of the conditions inside the simulation box. The boxes move relative to the simulation box in the y-direction, with the speed of the mean shear at the center of each box, $-qjL_x$. Therefore, in the time interval $\Delta t = L_y/(qL_x) = 10$, all the boxes line up. When this happens, the velocity field $\boldsymbol{u}$ that is induced by the vorticity within all the boxes (via Equation (12)) becomes large, because all the boxes reinforce each other, and therefore E_{u^2} exhibits a spike. Even though the shearing-periodic boundary conditions that we use are somewhat artificial, we are confident that using more realistic open boundary conditions would not affect the main results of this paper—and particularly not the stability of axisymmetric modes to three-dimensional perturbations. In L07, where we considered two-dimensional dynamics, we investigated both open and shearing-periodic boundary conditions, and showed explicitly that both give similar results. We also feel that the boundary conditions likely do not affect the level and persistence of the "turbulence" seen in Figure 4. However, this is less certain. Future investigations should more carefully address the role of boundary conditions.
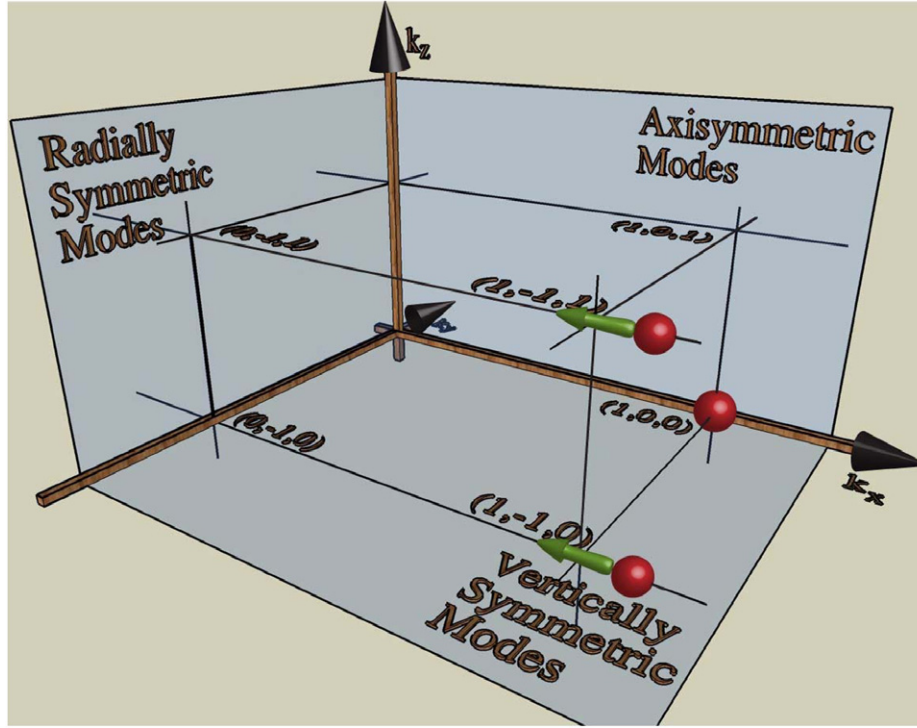
Figure 5. Evolution of wavevectors. Modes have constant k_y and k_z, and $k_x = qtk_y+$const. The three spheres depict modes that play important roles in nonlinear instability. The mode at $(1, 0, 0)$ does not move in k-space. The other two modes are swinging modes that are depicted in the leading phase of their swing. They will become trailing after crossing through the radially symmetric plane. The mode crossing through $(1, -1, 0)$ is responsible for the two-dimensional instability that forms vortices. The one crossing through $(1, -1, 1)$ is responsible for the three-dimensional instability that destroys vortices.

(A color version of this figure is available in the online journal.)

4. LINEAR EVOLUTION

In the remainder of this paper, we develop a theory explaining the stability of vortices seen in the above numerical simulations. We first consider the linear evolution of individual modes, and then proceed to show how nonlinear coupling between linear modes can explain vortex stability.

The linear evolution has been considered previously (Afshordi et al. 2005; Johnson & Gammie 2005; Balbus & Hawley 2006). Only two aspects of our treatment are new. First, we give the solution in terms of variables that allow the simple reconstruction of the full vectors $\boldsymbol{\omega}$ and $\boldsymbol{u}$. And second, we give the analytic expression for matching a leading mode onto a trailing mode that is valid for all k_y and k_z,

The linearized equation of motion is, from Equation (10),

$$(\partial_t - qx\partial_y)\boldsymbol{\omega} = -q\,\hat{\boldsymbol{y}}\omega_x + (2\Omega - q)\partial_z\boldsymbol{u}. \tag{25}$$

A single mode may be written as

$$\boldsymbol{\omega}(\boldsymbol{x}, t) = \hat{\boldsymbol{\omega}}(\boldsymbol{k_0}, t)e^{i[\boldsymbol{k}(\boldsymbol{k_0},t)]\cdot\boldsymbol{x}}, \tag{26}$$

where $\boldsymbol{k_0}$ is a constant vector that denotes the wavevector at time $t = 0$, and the wavevector $\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{k_0}, t)$ has components

$$k_y = k_{0y} = \text{const} \tag{27}$$

$$k_z = k_{0z} = \text{const} \tag{28}$$

$$k_x = k_{0x} + qtk_y \neq \text{const}, \tag{29}$$

so that upon insertion into Equation (25), the time derivative of the exponential cancels the term $-qx\partial_y\boldsymbol{\omega}$. The velocity field induced by such a mode is (Equation (12))

$$\boldsymbol{u}(\boldsymbol{x}, t) = \hat{\boldsymbol{u}}(\boldsymbol{k_0}, t)e^{i\boldsymbol{k}[(\boldsymbol{k_0},t)]\cdot\boldsymbol{x}}, \tag{30}$$

where

$$\hat{\boldsymbol{u}} = i\frac{\boldsymbol{k} \times \hat{\boldsymbol{\omega}}}{k^2}. \tag{31}$$

Figure 5 sketches the evolution of wavevectors. Axisymmetric modes ($k_y = 0$) do not move in k-space, as depicted by the sphere at $(1, 0, 0)$ in Figure 5. "Swinging modes" have $k_y \neq 0$, and their k_x is time dependent. Their fronts of constant phase are advected by the background shear. Swinging modes with $k_x/k_y < 0$, as depicted by the two spheres near $(1, -1, 0)$ and $(1, -1, 1)$ in Figure 5, have phasefronts tilted into the background shear, that is, they are leading modes. As time evolves, the shear first swings their k_x through $k_x = 0$, at which point their phasefronts are radially symmetric. Subsequently, they become trailing modes ($k_x/k_y > 0$) and approach alignment with the azimuthal direction ($k_x/k_y \to \infty$).

We now turn to the evolution of the Fourier amplitudes. In the remainder of this paper, we drop the hats

$$\hat{\boldsymbol{\omega}} \to \boldsymbol{\omega}, \quad \hat{\boldsymbol{u}} \to \boldsymbol{u}. \tag{32}$$

To distinguish real-space fields, we shall explicitly write their spatial dependence, for example, $\boldsymbol{\omega}(\boldsymbol{x})$.

Because $\boldsymbol{\omega}(\boldsymbol{x})$ is divergenceless, $\boldsymbol{\omega}$ only has 2 degrees of freedom, which we select to be ω_x and

$$\omega_{yz} \equiv \frac{\hat{\boldsymbol{x}}\cdot(\boldsymbol{k} \times \boldsymbol{\omega})}{k_{yz}} = \begin{cases} -\omega_y, & \text{if} \quad k_y = 0, \\ \omega_z, & \text{if} \quad k_z = 0, \end{cases} \tag{33}$$

where

$$k_{yz} \equiv \sqrt{k_y^2 + k_z^2}. \tag{34}$$

Our variable ω_{yz} is proportional to the variable U of Balbus & Hawley (2006). Adopting ω_x as the second degree of freedom

enables the full vectors to be reconstructed as

$$\boldsymbol{\omega} = -\omega_x \frac{\boldsymbol{k} \times (\boldsymbol{k} \times \hat{\boldsymbol{x}})}{k_{yz}^2} - \omega_{yz} \frac{\boldsymbol{k} \times \hat{\boldsymbol{x}}}{k_{yz}} \qquad (35)$$

$$\boldsymbol{u} = -i\omega_{yz} \frac{\boldsymbol{k} \times (\boldsymbol{k} \times \hat{\boldsymbol{x}})}{k^2 k_{yz}} + i\omega_x \frac{\boldsymbol{k} \times \hat{\boldsymbol{x}}}{k_{yz}^2}. \qquad (36)$$

The linearized Equation (25) is expressed in terms of these degrees of freedom as

$$\frac{k_{yz}}{qk_y} \frac{d}{dt} \begin{pmatrix} \omega_x \\ \omega_{yz} \end{pmatrix} = \beta \frac{\Omega}{\kappa} \begin{pmatrix} 0 & -\frac{1}{2} \frac{\kappa^2}{\Omega^2} \frac{1}{1+\tau^2} \\ 2 & 0 \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_{yz} \end{pmatrix}, \qquad (37)$$

after introducing the epicyclic frequency,

$$\kappa \equiv \sqrt{2\Omega(2\Omega - q)}, \qquad (38)$$

with $\kappa = \Omega$ in a Keplerian disk, and

$$\tau \equiv \frac{k_x}{k_{yz}} \qquad (39)$$

$$\beta \equiv \frac{\kappa}{q} \frac{k_z}{k_y}. \qquad (40)$$

As long as $k_y \neq 0$, τ varies in time through its dependence on $k_x = k_{0x} + qtk_y$.

For axisymmetric modes ($k_y = 0$), τ is constant and

$$\frac{d^2}{dt^2} \omega_{yz} + \kappa^2 \frac{k_z^2}{k_x^2 + k_z^2} \omega_{yz} = 0, \qquad (41)$$

the solution of which is sinusoidal with frequency $\kappa k_z/\sqrt{k_x^2 + k_z^2}$. Axisymmetric modes with phasefronts aligned with the plane of the disk ($k_x = k_y = 0$) have in-plane fluid velocities, and they oscillate at the epicyclic frequency of a free test particle, κ. But axisymmetric modes with tilted phasefronts have slower frequencies, because fluid pressure causes deviations from free epicycles. In the limit of vertical axisymmetric phasefronts ($k_z = k_y = 0$), the effects of rotation disappear entirely, and this zero-frequency mode merely alters the mean shear flow's velocity profile.

For swinging modes ($k_y \neq 0$), it is convenient to employ τ as the time variable. Since

$$\frac{k_{yz}}{qk_y} \frac{d}{dt} = \frac{d}{d\tau}, \qquad (42)$$

we have

$$\frac{d^2}{d\tau^2} \omega_{yz} + \frac{\beta^2}{1+\tau^2} \omega_{yz} = 0. \qquad (43)$$

(Balbus & Hawley 2006). Figure 6 plots numerical solutions of this equation, and shows that it matches the output from the pseudospectral code, as well as the analytic theory described below. Given ω_{yz}, it is trivial to construct $\boldsymbol{\omega}$ and $\boldsymbol{u}$ from

$$\omega_x = \frac{\kappa}{2\beta\Omega} \frac{d\omega_{yz}}{d\tau} \qquad (44)$$

and Equations (35) and (36).

For highly leading or trailing modes ($|\tau| \gg 1$), Equation (43) has simple power-law solutions,

$$\omega_{yz} = \omega_A |\tau|^{\frac{1-\delta}{2}} + \omega_B |\tau|^{\frac{1+\delta}{2}}, \quad |\tau| \gg 1 \qquad (45)$$
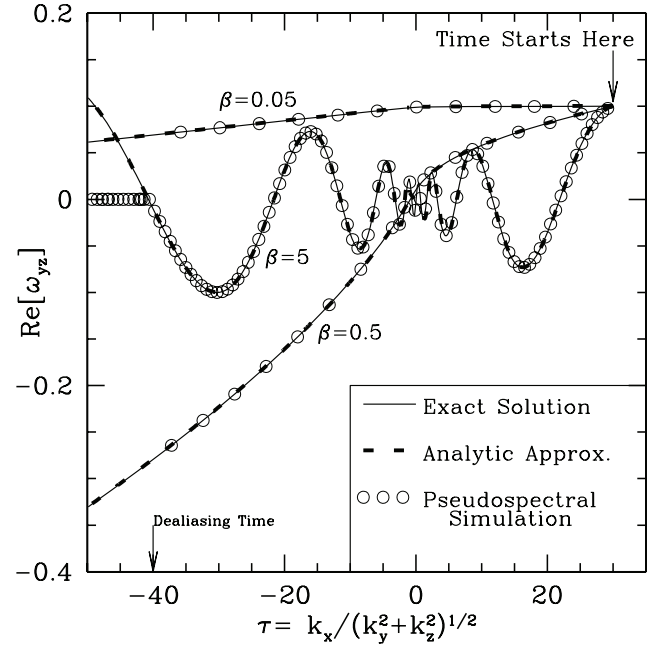

Figure 6. Linear evolution of mode amplitudes for three values of β. Time runs from right to left. Solid curves show the exact, numerically integrated solution of Equation (43). The initial value of $d\omega_{yz}/d\tau$ was chosen so that $\omega_B = 0$ initially (Equation (45)). Dashed lines show the analytic solution (Equation (45)) with constant ω_A and $\omega_B = 0$ for $\tau > 0$, while for $\tau < 0$, the normal mode amplitudes are set to different constants that are given by Equation (47). We exclude the domain $|\tau| < 1$ from the dashed curve, because the analytic approximation does not apply there. Circles show output from the pseudospectral code, integrated with a time step $dt = 1/15$ and with the viscosity set to zero.

(A color version of this figure is available in the online journal.)

(Balbus & Hawley 2006), where ω_A and ω_B are constants that we shall call the "normal-mode" amplitudes, and

$$\delta \equiv \sqrt{1 - 4\beta^2}, \qquad (46)$$

which is imaginary for $|\beta| > 1/2$. As a mode's wavevector evolves along a line in k-space, its amplitude is oscillatory if this line is much closer to the k_z-axis than to the k_y-one, and nonoscillatory if the converse is true. The transition occurs at $|\beta| = 1/2$. This behavior may be understood as a competition between shear and epicyclic oscillations. The timescale for k_x to change by an order-unity factor due to the shear is $t_{\rm shear} \sim |k_x/\dot{k}_x| = |k_x/qk_y|$, and the timescale for epicyclic oscillations of axisymmetric modes is $t_{\rm epi} \sim \kappa^{-1}|k_x/k_z|$ for $|k_x| \gg |k_z|$. Therefore, $|\beta| \sim t_{\rm shear}/t_{\rm epi}$, and when $|\beta| \gg 1$, the epicyclic time is shorter and so the mode's amplitude oscillates as its wavevector is slowly advected by the shear. But when $|\beta| \ll 1$, the shear changes the wavevector faster than the amplitude can oscillate.

Solution (45) breaks down in mid-swing. As a swinging wave changes from leading to trailing, its "normal-mode amplitudes" change abruptly on the timescale that τ changes from ± 1 to ∓ 1 via

$$\begin{pmatrix} \omega_A \\ \omega_B \end{pmatrix}_{\rm trail} = \begin{pmatrix} T_{AA} & T_{AB} \\ T_{BA} & T_{BB} \end{pmatrix} \begin{pmatrix} \omega_A \\ \omega_B \end{pmatrix}_{\rm lead}, \qquad (47)$$

where the transition matrix has components

$$T_{AA} = -T_{BB} = \csc(\delta\pi/2) \qquad (48)$$

$$T_{BA} = -\frac{\cot(\delta\pi/2)^2}{T_{AB}}$$

$$= -2^{\delta+1} \frac{1}{\delta} \frac{1-\delta}{1+\delta} \frac{\Gamma(1+\delta/2)^2}{\Gamma(1/2+\delta/2)^2} \qquad (49)$$

and determinant $= -1$, and hence is its own inverse. The components are complex when $|\beta| > 1/2$. To derive these components, we took advantage of the fact that Equation (43) has hypergeometric solutions (Johnson & Gammie 2005; Balbus & Hawley 2006), and matched these onto the normal-mode solution given above. We omit the unenlightening details.

5. NONLINEAR EVOLUTION: FORMATION AND DESTRUCTION OF VORTICES

5.1. Qualitative Description

The instability that destroys vortices is a generalization of the one that forms them. We review here how vortices form, before describing the instability that destroys them. In Section 5.2, we make this description quantitative.

Vortices form out of a nonlinear instability that involves vertically symmetric ($k_z = 0$) modes. (See L07 for more details of the two-dimensional dynamics than are presented here.) Consider the two vertically symmetric modes shown in Figure 5: the "mother" mode at $(1, 0, 0)$ and the "father" mode that is depicted crossing through $(1, -1, 0)$. Triplets of integers (j_x, j_y, j_z) label values of wavevectors (k_x, k_y, k_z) (e.g., via [17]). The mother is both axisymmetric and vertically symmetric and the father is a leading swinging mode.

As the father swings through radial symmetry, that is, as it crosses through the point $(0, -1, 0)$, its velocity field is strongly amplified by the background shear. This can be seen from Section 4, which shows that swinging modes with $k_z = 0$ have ω_{yz}=const, and hence $u_x = i(\omega_{yz}/k_{yz})/(1 + \tau^2)$, which becomes largest when τ crosses through 0. When the father is near the peak of its transient amplification ($|\tau| \lesssim 1$), it couples most strongly with the mother, and they produce a "son" near $(1, -1, 0) = (1, 0, 0) + (0, -1, 0)$. The son will then swing through radial symmetry where it will couple (oedipally) with the mother to produce a grandson near $(1, -1, 0)$, which can repeat the cycle. We summarize this two-dimensional instability feedback loop as

$$\text{linear amplification:} \quad (1, -1, 0) \rightarrow (0, -1, 0)$$
$$\text{nonlinear coupling:} \quad (0, -1, 0) + (1, 0, 0) \rightarrow (1, -1, 0).$$

The criterion for instability is simply that the amplitude of the son's ω_{yz} be larger than that of the father. As shown in L07, if instability is triggered, its nonlinear outcome in two dimensions is a long-lived vortex.

The three-dimensional instability that is responsible for destroying vortices is a straightforward generalization. The mother mode is still at $(1, 0, 0)$, but now the father mode starts near $(1, -1, 1)$. Symbolically, the feedback loop is

$$\text{linear amplification:} \quad (1, -1, 1) \rightarrow (0, -1, 1)$$
$$\text{nonlinear coupling:} \quad (0, -1, 1) + (1, 0, 0) \rightarrow (1, -1, 1).$$

The two-dimensional instability described above is just a special case of this three-dimensional one in the limit that $k_z = 0$. In general, the stability of a mother mode at $(1, 0, 0)$ with given $k_x = \bar{k}_x$ and $\omega = \bar{\omega}$ depends on k_y and k_z of the father-mode perturbations (as well as on the parameters q and Ω). Which k_y and k_z are accessible in turn depends on the dimensions $L_y \times L_z$ of the simulation box or, equivalently, on the circumferential

distance around a disk and the scale height. In Section 5.2, we quantitatively map out the region in the k_y–k_z plane that leads to instability. For now, it suffices to note that the unstable region has $|k_y| \lesssim |\bar{k}_x \bar{\omega}|/q$ and $|k_z| \lesssim |k_y|$. We conclude that a given mother mode suffers one of three possible fates, depending on L_y and L_z.

1. If L_y is less than a critical value ($\sim q/|\bar{\omega}\bar{k}_x|$), then the mother mode is stable to all perturbations.
2. If L_y is larger than this critical value, then the mother mode is unstable to vertically symmetric ($k_z = 0$) perturbations; if in addition L_z is sufficiently small that all modes with $k_z \neq 0$ are stable, then the mother mode turns into a long-lived vortex (Figure 1).
3. If both L_y and L_z are sufficiently large, the mother mode is unstable both to vertically symmetric and to three-dimensional perturbations. When this happens, the mother starts to form a vortex, but this vortex is unstable to three-dimensional perturbations. The result is turbulence (Figure 2).

There is also a possibility that is intermediate between numbers 2 and 3: if the conditions described in number 2 hold, the essentially two-dimensional dynamics that results can nonlinearly produce new mother modes that are unstable to three-dimensional perturbations. In this paper, we shall not consider this possibility further, since it did not occur in the pseudospectral simulations of Section 3. We merely note that in our simulations of this possibility (not presented in this paper), we found that when the new mother modes decayed, they also destroyed the original mother mode.

5.2. The Stability Criterion

To quantify the previous discussion, we choose an initial state as in Figure 5, with the mother mode at $(1, 0, 0)$ and the father a leading mode crossing through $(1, -1, 1)$. The son mode, not depicted in the figure, initially crosses through the point $(2, -1, 1)$. We set its initial vorticity—as well as the initial vorticity of all modes other than the mother and father—to zero.[4] The father's wavevector and Fourier amplitude are labeled as in Section 4, and the mother's and son's are labeled with bars and primes:

$$\text{father:} \quad \boldsymbol{k}, \ \boldsymbol{\omega}$$
$$\text{mother:} \quad \bar{\boldsymbol{k}} \equiv \bar{k}\hat{\boldsymbol{x}}, \bar{\boldsymbol{\omega}} \equiv \bar{\omega}\hat{\boldsymbol{z}}$$
$$\text{son:} \quad \boldsymbol{k}' \equiv \bar{k}\hat{\boldsymbol{x}} + \boldsymbol{k}, \ \boldsymbol{\omega}'$$

Note that $\bar{k}$=const, and $\bar{\omega}_x = 0$ because the vorticity must be transverse to the wavevector. We also set $\bar{\omega}_y = 0$; otherwise $\bar{u}_z \neq 0$, which corresponds to a mean flow out the top of the box and in through the bottom.

At early times, the father mode swings through the point $(0, -1, 1)$. Since the only other nonvanishing mode at this time

---

[4] We ignore the complex conjugate modes for simplicity. Since $\boldsymbol{\omega}(\boldsymbol{x})$ is real valued, each mode with wavevector and amplitude $(\boldsymbol{k}, \boldsymbol{\omega})$ is accompanied by a conjugate mode that has $(-\boldsymbol{k}, \boldsymbol{\omega}^*)$. In our initial state, there are really four modes with nonzero amplitudes: the mother at $(1, 0, 0)$ and its conjugate at $(-1, 0, 0)$, and the father and its conjugate. We may ignore the conjugate modes because they do not affect the instability described here. As shown in L07 for the two-dimensional case, their main effect is that when the son swings through $(0, -1, 1)$, not only does it couple with the mother at $(1, 0, 0)$ to produce a grandson at $(1, -1, 1)$, but it also couples with the conjugate mother at $(-1, 0, 0)$ to partially kill its father, which is then at $(-1, -1, 1)$ (bringing to mind the story of Oedipus). But since the father is a trailing mode at this time, it no longer participates in the instability. Nonetheless, the conjugate modes do play a role in the nonlinear outcome of the instability.

is the mother, there are no mode couplings that can nonlinearly change the father's amplitude. Therefore its amplitude is governed by the linear Equation (37), which we reproduce here as

$$\frac{d}{d\tau}\begin{pmatrix} \omega_x \\ \omega_{yz} \end{pmatrix} = \beta\frac{\Omega}{\kappa}\begin{pmatrix} 0 & -\frac{1}{2}\frac{\kappa^2}{\Omega^2}\frac{1}{1+\tau^2} \\ 2 & 0 \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_{yz} \end{pmatrix}. \quad (50)$$

During its swing, it couples with the mother to change the amplitude of the son. The linear part of the son's evolution is given by the above equation with primed vorticity and wavevector in place of unprimed. The nonlinear part is given by

$$\frac{d}{dt}\boldsymbol{\omega}'\Big|_{\text{nonlin}} = i\boldsymbol{k}' \times (\bar{\boldsymbol{u}}\times\boldsymbol{\omega} + \boldsymbol{u}\times\bar{\boldsymbol{\omega}}) \quad (51)$$

(Equation (10)), where $\bar{\boldsymbol{u}} = -i(\bar{\omega}/\bar{k})\hat{\boldsymbol{y}}$ and $\boldsymbol{u} = i\boldsymbol{k}\times\boldsymbol{\omega}/k^2$ (Equation (31)). Adding the linear and nonlinear parts, and re-expressing in terms of our chosen degrees of freedom, we find

$$\frac{d}{d\tau}\begin{pmatrix} \omega'_x \\ \omega'_{yz} \end{pmatrix} = \beta\frac{\Omega}{\kappa}\begin{pmatrix} 0 & -\frac{1}{2}\frac{\kappa^2}{\Omega^2}\frac{1}{1+(\tau+\bar{\tau})^2} \\ 2 & 0 \end{pmatrix}\begin{pmatrix} \omega'_x \\ \omega'_{yz} \end{pmatrix}$$
$$ - \frac{\bar{\omega}}{q}\begin{pmatrix} \frac{1}{\tau} & \frac{\beta q}{\kappa}\frac{1}{1+\tau^2} \\ 0 & \frac{1}{\tau}-\frac{\bar{\tau}}{1+\tau^2} \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_{yz} \end{pmatrix}, \quad (52)$$

where the dimensionless constant

$$\bar{\tau} \equiv \frac{\bar{k}_x}{k_{yz}} \quad (53)$$

depends on both the mother's and the father's wavevectors. It is the father's $\tau \equiv k_x/k_{yz}$ that is being used as the time coordinate for evolving the son's amplitude. The grandson's equation is the obvious extension: denoting the grandson's amplitudes with double primes, one need to only make the following replacements in Equation (52): $\boldsymbol{\omega}' \rightarrow \boldsymbol{\omega}''$, $\boldsymbol{\omega} \rightarrow \boldsymbol{\omega}'$, and $\tau \rightarrow \tau + \bar{\tau}$. Subsequent generations evolve analogously.

The father's Equation (50) is easily solved, as shown in Section 4. Inserting this solution into Equation (52) produces a linear inhomogeneous equation for the son's amplitude, and similarly for the grandson's. Figure 7 plots numerical solutions of these equations. Also shown as circles are output from a pseudospectral simulation, showing excellent agreement.

In the appendix, we solve Equation (52) analytically to derive the amplification factor χ, which is the ratio of the son's amplitude at any point in its evolution (e.g., when it is radially symmetric) to the father's amplitude at the same point in its evolution. We find

$$\chi = -\frac{\bar{\omega}}{q}\bar{\tau}^\delta\sqrt{\pi}\frac{1+\delta}{\delta^2}\left(1 + \frac{q\Omega}{\kappa^2}(1-\delta)\right)\frac{\Gamma(1+\delta/2)}{\Gamma(1/2+\delta/2)}, \quad (54)$$

where $\delta = \sqrt{1-4\beta^2}$. Equation (54) is applicable in the limit $|\bar{\omega}|/q \ll 1$. For two-dimensional modes ($\beta = 0 \Rightarrow \delta = 1$), it recovers Equation (42) of L07 (see also Equation (1) of this paper):

$$\chi_{2D} = -\pi\frac{\bar{\omega}}{q}\frac{\bar{k}}{k_y}. \quad (55)$$

Marginally stable modes have $|\chi| = 1$. Figure 8 plots curves of marginal stability. The left panel is for the case $\bar{\omega} = 0.005$, as in Figure 7, and the right panel is for $\bar{\omega} = 0.05$, as in the pseudospectral simulations presented at the outset of this paper (Equation (16); Figures 1 and 2). The left panel shows that
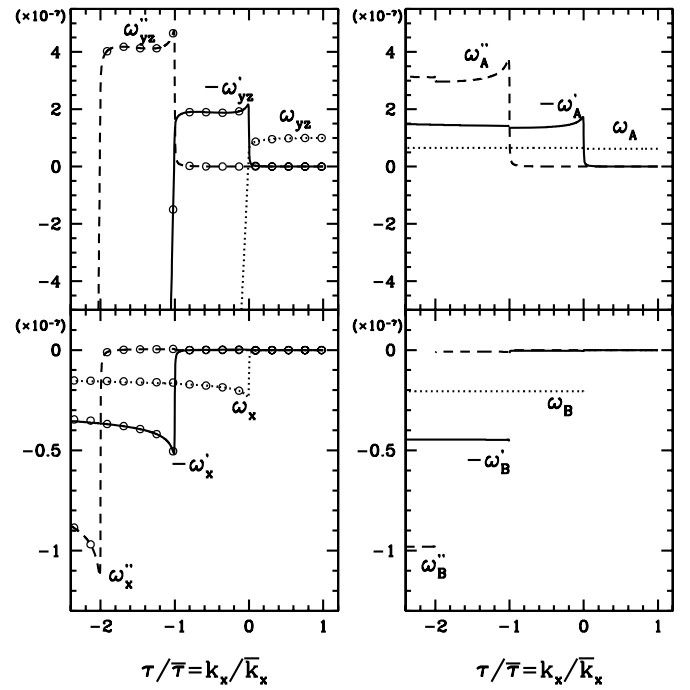

Figure 7. Nonlinear evolution of three-dimensional instability. Time runs from right to left. In the two left panels, lines show numerical solutions of Equations (50) and (52), as well as the grandson's equation. Also shown as circles are the output from a pseudospectral simulation, showing excellent agreement with the "exact" solutions. The following parameters have been chosen: $\Omega = 1$, $q = 3/2$, $\bar{\omega} = 0.005$, $\bar{k}_x = 2\pi\cdot 15$, $k_y = -2\pi/30$, $k_z = 0.45|k_y| \Rightarrow \beta = -0.3$. The small disagreement between pseudospectral and exact solutions for ω'_x at $\tau < -2\bar{\tau}$ is due to the conjugate modes that, for simplicity, we have not included in equations (50) and (52); see footnote 4. The two right panels show the mode amplitudes, defined via Equation (A1) for the son, and similarly for the father and grandson. Although these two right panels contain the same information as the left ones, they are helpful in constructing the analytic form of the growth factor χ (see the appendix). With the parameters chosen for this figure, Equation (54) predicts $\chi = -2.2$ for the amplification factor between successive generations, in agreement with that seen in the figure.

Equation (54) gives a fair reproduction of the exact curve. We do not show Equation (54) in the right panel, because it gives poorer agreement there (since $|\bar{\omega}|/q$ is too large). In the right panel, we also plot X for the values of the smallest nonvanishing three-dimensional wavenumbers in the simulations of Figures 1 and 2. In the short-box simulation, all three-dimensional modes lie in the stable zone. Therefore, the dynamics remains two dimensional. But in the three-dimensional box, there is a three-dimensional mode in the unstable zone that destroys the vortex and gives rise to turbulent-like behavior.

It is interesting to briefly consider how the instability described here connects with the Rayleigh-unstable case, which occurs when $\kappa^2 < 0$. At small $|k_y|$, the marginally stable curves in Figure 8 are given by $|\beta| = 1/2$, where $\beta = (\kappa/q)(k_z/k_y)$. Hence, if one decreases κ from its Keplerian value Ω, the marginally stable curve becomes steeper in the k_z–k_y plane, and an increasing number of three-dimensional modes become unstable. As $\kappa \rightarrow 0$, if a two-dimensional mode with some k_y is unstable, then so are all three-dimensional modes with the same k_y. Therefore, any two-dimensional-unstable state is also three-dimensional unstable, and any forming vortex would decay into turbulence.

6. CONCLUSIONS

Our main result follows from Figure 8, which maps out the stability of a "mother mode" (i.e., a mode with wavevector
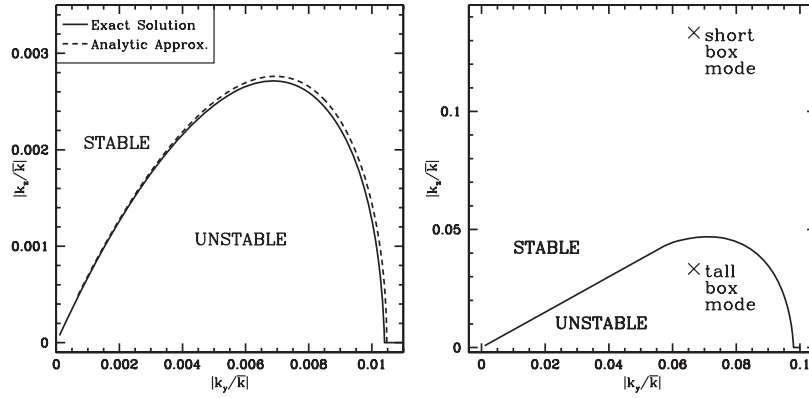
Figure 8. Curves of marginal stability for a mother mode with $\bar{\omega} = 0.005$ (left panel) and $\bar{\omega} = 0.05$ (right panel). The left panel corresponds to Figure 7 and the right panel corresponds to the pseudospectral simulations of Figures 1 and 2. We set $\Omega = 1$ and $q = 3/2$. To make the solid lines in these plots (the "exact solutions"), we repeated the integrations that produced the lines in Figure 7, but varying k_z for each k_y until perturbations neither grew nor decayed. The dashed line in the left panel shows that the analytic approximation of Equation (54) agrees reasonably well with the exact solution. We do not show Equation (54) in the right panel because the agreement is poorer there. The right panel shows two X's for the values of the smallest nonzero $|k_y|$ and $|k_z|$ in the simulations of Figures 1–2, that is, $|k_y/\bar{k}| = L_x/L_y = 0.067$ for both simulations and $|k_z/\bar{k}| = L_x/L_z = 0.13$ for the short box and $= 0.033$ for the tall box. The tall box contains a three-dimensional-unstable mode that leads to the destruction of the vortex into a turbulent-like state. The short box contains no such mode, and is stable to three-dimensional perturbations.

$\bar{k}\hat{x}$ and amplitude $\bar{\omega}$) to nonaxisymmetric three-dimensional perturbations. A mother mode is unstable provided that k_y and k_z of the nonaxisymmetric perturbations satisfy both $|k_y| \lesssim \bar{k}\bar{\omega}/q$ and $|k_z| \lesssim |k_y|$, dropping order-unity constants. Based on this result, we may understand the formation, survival, and destruction of vortices. Vortices form out of mother modes that are unstable to two-dimensional ($k_z = 0$) perturbations. Mother modes that are unstable to two-dimensional modes but stable to three-dimensional ($k_z \neq 0$) ones form into long-lived vortices. Mother modes that are unstable to both two-dimensional and three-dimensional modes are destroyed. Therefore, a mother mode with given $\bar{k}$ and $\bar{\omega}$ will form into a vortex if the disk has a sufficiently large circumferential extent and a sufficiently small scale height, that is, if $r \gtrsim \bar{k}^{-1}q/\bar{\omega}$ and $h \lesssim \bar{k}^{-1}q/\bar{\omega}$, where r is the distance to the center of the disk and h is the scale height. Alternatively, the mother mode will be destroyed in a turbulent-like state if both r and h are sufficiently large ($r \gtrsim \bar{k}^{-1}q/\bar{\omega}$ and $h \gtrsim \bar{k}^{-1}q/\bar{\omega}$).

Our result has a number of astrophysical consequences. In protoplanetary disks that do not contain any vortices, solid particles drift inward. Gas disks orbit at sub-Keplerian speeds, $v_{\rm gas} \sim \Omega r(1 - \eta)$, where Ωr is the Keplerian speed and $\eta \sim (c_s/\Omega r)^2$, with c_s being the sound speed. Since solid particles would orbit at the Keplerian speed in the absence of gas, the mismatch of speeds between solids and gas produces a drag on the solid particles, removing their angular momentum and causing them to fall into the star. For example, in the minimum mass solar nebula, meter-sized particles fall in from 1 AU in around a hundred years. This rapid infall presents a serious problem for theories of planet formation, since it is difficult to produce planets out of dust in under a hundred years. Vortices can solve this problem (Barge & Sommeria 1995). A vortex that has excess vorticity $-\bar{\omega}$ and radial width $1/\bar{k}$ can halt the infall of particles provided that $\bar{\omega}/\bar{k} \gtrsim (\Omega r)\eta$, because the gas speed induced by such a vortex more than compensates for the sub-Keplerian speed induced by gas pressure.[5] Previous simulations

implied that three-dimensional vortices rapidly decay, and so cannot prevent the rapid infall of solid particles (Barranco & Marcus 2005; Shen et al. 2006). Our result shows that vortices can survive within disks, and so restores the viability of vortices as a solution to the infall problem.

A more important—and more speculative—application of our result is to the transport of angular momentum within neutral accretion disks. In our simulation of a vortex in a tall box, we found that as the vortex decayed, it transported angular momentum outward at a nearly constant rate for hundreds of orbital times. If decaying vortices transport a significant amount of angular momentum in disks, they would resolve one of the most important outstanding questions in astrophysics today: what causes hydrodynamical accretion disks to accrete? To make this speculation more concrete, one must understand the amplitude and duration of the "turbulence" that results from decaying vortices. This is a topic for future research.

In this paper, we considered only the effects of rotation and shear on the stability of vortices, while we neglected the effect of vertical gravity. There has been a lot of research in the geophysical community on the dynamics of fluids in the presence of vertical gravity, since stably stratified fluids are very common on Earth—in the atmosphere, oceans, and lakes. In numerical and laboratory experiments of strongly stratified flows, thin horizontal "pancake vortices" often form, and fully developed turbulence is characterized by thin horizontal layers. (e.g., Brethouwer et al. 2007). Pancake vortices are stabilized by vertical gravity, in contrast to the vortices studied in this paper, which are stabilized by rotation. Gravity inhibits vertical motions because of buoyancy: it costs gravitational energy for fluid to move vertically. The resulting quasi-two-dimensional flow can form into a vortex.[6] We may speculate that in an

---

[5] We implicitly assume here that the stopping time of the particle due to gas drag is comparable with the orbital time, which is true for meter-sized particles at 1 AU in the minimum mass solar nebula. A more careful treatment shows that a vortex can stop a particle with stopping time t_s provided that $\bar{\omega}/\bar{k} \gtrsim (\Omega r)(\Omega t_s)\eta$ (Youdin 2008).

[6] Billant & Chomaz (2000) showed that a vertically uniform vortex column in a stratified (and nonrotating and nonshearing) fluid suffers an instability (the "zigzag instability") that is characterized by a typical vertical lengthscale $\lambda_z \sim U/N$, where U is the horizontal speed induced by the vortex, N is the Brunt–Väisälä frequency, and the horizontal lengthscale of the vortex L_h is assumed to be much greater that λ_z (hence, the pancake structure). We may understand Billant & Chomaz's result in a crude fashion with an argument similar to that employed in the introduction to explain the destruction of rotation-stabilized vortices: since the frequency of buoyancy waves is Nk_x/k_z (when $|k_x| \ll |k_z|$), and since the frequency at which fluid circulates around a vortex is $U/L_h \sim k_x U$, there is a resonance between these two frequencies for a vertical lengthscale $1/k_z \sim U/N$.

astrophysical disk, vertical gravity provides an additional means to stabilize vortices, in addition to rotation. But to make this speculation concrete, the theory presented in this paper should be extended to include vertical gravity.

In this paper, we have not addressed the origin of the axisymmetric structure (the mother modes) that give rise to surviving or decaying vortices. One possibility is that decaying vortices can produce more axisymmetric structure and, therefore, they can lead to self-sustaining turbulence. This seems to us unlikely. We have not seen evidence of it in our simulations, but this could be because of the modest resolution of our simulations. Other possibilities for the generation of axisymmetric structure include thermal instabilities, such as the baroclinic instability, or convection, or stirring by planets. This, too, is a topic for future research.

APPENDIX

ANALYTIC EXPRESSION FOR GROWTH FACTOR χ

In this appendix, we derive Equation (54) by analytically integrating Equation (52) for the son's vorticity, given the father's vorticity as a function of time (Section 4), and taking the mother's vorticity $\bar{\omega}$ to be constant, which is valid when the father's amplitude is small relative to the mother's. The numerical integral of Equation (52) is shown in Figure 7. Recall that initially $\tau = \bar{\tau} > 0$ and τ decreases in time, and typically $\bar{\tau} \gg 1$. It simplifies the analysis to work with the son's "normal-mode" amplitudes ω'_A and ω'_B, defined from ω'_x and ω'_{yz} via (Equations (44) and (45))

$$\begin{pmatrix} \omega'_x \\ \omega'_{yz} \end{pmatrix} = \begin{pmatrix} \frac{\kappa}{2\beta\Omega} \frac{1-\delta}{2\tau'} |\tau'|^{\frac{1-\delta}{2}} & \frac{\kappa}{2\beta\Omega} \frac{1+\delta}{2\tau'} |\tau'|^{\frac{1+\delta}{2}} \\ |\tau'|^{\frac{1-\delta}{2}} & |\tau'|^{\frac{1+\delta}{2}} \end{pmatrix} \begin{pmatrix} \omega'_A \\ \omega'_B \end{pmatrix}, \quad (A1)$$

where

$$\tau' \equiv \tau + \bar{\tau}. \quad (A2)$$

Substituting this into Equation (52), the time derivative of the above matrix cancels the homogeneous term in that equation if we approximate $1 + \tau'^2 \simeq \tau'^2$, which holds until just before the time that $\tau = -\bar{\tau}$. The inhomogeneous term produces

$$\frac{d}{d\tau}\omega'_A = \frac{\bar{\omega}}{q}\frac{1}{\delta}|\tau'|^{\frac{1+\delta}{2}}\left(\omega_{yz}\left[\frac{1}{1+\tau^2}\left(\frac{2\beta^2 q\Omega}{\kappa^2} + \frac{1+\delta}{2}\frac{\bar{\tau}}{\tau'}\right)\right.\right.$$
$$\left.\left. - \frac{1+\delta}{2}\frac{1}{\bar{\tau}\tau'}\right] + \omega_x\frac{2\beta\Omega}{\kappa}\frac{1}{\bar{\tau}}\right), \quad \tau \gtrsim -\bar{\tau}. \quad (A3)$$

Since ω_x and ω_{yz} are known (Section 4), a straightforward integration yields ω'_A just before $\tau = -\bar{\tau}$. To perform this integral, we resort to some approximations, guided by the solution shown in Figure 7. For the first term, we need

$$\int_{\bar{\tau}}^{-\bar{\tau}} \omega_{yz}\frac{(\tau+\bar{\tau})^{\frac{1+\delta}{2}}}{1+\tau^2}d\tau \approx -\frac{1}{\beta^2}\int_{\bar{\tau}}^{-\epsilon\bar{\tau}}\frac{d^2\omega_{yz}}{d\tau^2}(\tau+\bar{\tau})^{\frac{1+\delta}{2}}d\tau$$
$$+ \omega_B\int_{-\epsilon\bar{\tau}}^{-\bar{\tau}}|\tau|^{\frac{1+\delta}{2}-2}(\tau+\bar{\tau})^{\frac{1+\delta}{2}}d\tau \quad (A4)$$

$$\approx \omega_B\bar{\tau}^\delta\frac{1+\delta}{2\beta^2}\epsilon^{\frac{-1+\delta}{2}} - \omega_B\bar{\tau}^\delta\int_\epsilon^1 s^{\frac{-3+\delta}{2}}(1-s)^{\frac{1+\delta}{2}}ds \quad (A5)$$

$$\approx \omega_B\bar{\tau}^\delta 2^{-\delta}\sqrt{\pi}\frac{1+\delta}{1-\delta}\frac{\Gamma(1/2+\delta/2)}{\Gamma(1+\delta/2)}, \quad (A6)$$

where ϵ is a parameter that satisfies $1 \gg \epsilon \gg 1/\bar{\tau}$. In the first line, we used Equation (43), and we discarded the ω_A mode from the second integral because the ω_B mode increases faster with increasing $|\tau|$. From Figure 7, ω'_A nearly vanishes until $\tau \simeq 0$. Therefore, in the second line, we approximated the first integral as $-(\bar{\tau}^{\frac{1+\delta}{2}}/\beta^2)d\omega_{yz}/d\tau|_{-\epsilon}$. The third line holds in the limit of small ϵ. The other three terms in Equation (A3) are integrated similarly, yielding

$$\omega'_A = \omega_B\frac{\bar{\omega}}{q}\frac{1}{\delta}\bar{\tau}^\delta 2^{-\delta-1}\sqrt{\pi}\frac{\Gamma(1/2+\delta/2)}{\Gamma(1+\delta/2)}$$
$$\times \left(1 + \frac{q\Omega}{\kappa^2}(1-\delta)\right)\frac{(1+\delta)^2}{1-\delta} \quad (A7)$$

just before the time when $\tau = -\bar{\tau}$, that is, just before the son is radially symmetric. At this time, Figure 7 shows that ω'_B very nearly vanishes. Therefore, just after the son is radially symmetric, it will have $\omega'_B = T_{BA}\omega'_A$ (Equation (47)), with ω'_A given by Equation (A7). This gives the amplification factor $\chi \equiv \omega'_B/\omega_B$ that is displayed in Equation (54).

REFERENCES

Afshordi, N., Mukhopadhyay, B., & Narayan, R. 2005, ApJ, 629, 373
Balbus, S. A., & Hawley, J. F. 1998, Rev. Mod. Phys., 70, 1
Balbus, S. A., & Hawley, J. F. 2006, ApJ, 652, 1020
Barge, P., & Sommeria, J. 1995, A&A, 295, L1
Barranco, J. A., & Marcus, P. S. 2005, ApJ, 623, 1157
Barranco, J. A., & Marcus, P. S. 2006, J. Comput. Phys., 219, 21
Billant, P., & Chomaz, J.-M. 2000, J. Fluid Mech., 419, 29
Brethouwer, G., Billant, P., Lindborg, E., & Chomaz, J.-M. 2007, J. Fluid Mech., 585, 343
Chagelishvili, G. D., Zahn, J.-P., Tevzadze, A. G., & Lominadze, J. G. 2003, A&A, 402, 401
Drazin, P. G., & Reid, W. H. 2004, Hydrodynamic Stability, ed. P. G. Drazin & W. H. Reid (Cambridge: Cambridge Univ. Press), 626
Gammie, C. F., & Menou, K. 1998, ApJ, 492, L75
Gill, A. E. 1965, J. Fluid Mech., 21, 503
Godon, P., & Livio, M. 1999, ApJ, 523, 350
Johnson, B. M., & Gammie, C. F. 2005, ApJ, 635, 149
Lerner, J., & Knobloch, E. 1988, J. Fluid Mech., 189, 117
Lithwick, Y. 2007, ApJ, 670, 789
Lovelace, R. V. E., Li, H., Colgate, S. A., & Nelson, A. F. 1999, ApJ, 513, 805
Lynden Bell, D., & Pringle, J. E. 1974, MNRAS, 168, 603
Marcus, P. S. 1993, ARA&A, 31, 523
Maron, J., & Goldreich, P. 2001, ApJ, 554, 1175
Rogallo, R. S. 1981, NASA STI/Recon Technical Report N, 81, 31508
Saffman, P. G. 1995, in Vortex Dynamics, ed. P. G. Saffman (Cambridge: Cambridge Univ. Press), 325
Sano, T., Miyama, S. M., Umebayashi, T., & Nakano, T. 2000, ApJ, 543, 486
Shen, Y., Stone, J. M., & Gardiner, T. A. 2006, ApJ, 653, 513
Stone, J. M., & Balbus, S. A. 1996, ApJ, 464, 364
Umurhan, O. M., & Regev, O. 2004, A&A, 427, 855
Yecko, P. A. 2004, A&A, 425, 385
Youdin, A. 2008, in Proc. Les Houches Winter School "Physics and Astrophysics of Planetary Systems" (EDP Sciences: EAS Publications Series) in press (arXiv:0807.1114v2)