Instability, turbulence, and enhanced transport in accretion disks

Steven A. Balbus and John F. Hawley

Virginia Institute of Theoretical Astronomy, Department of Astronomy, University of Virginia, Charlottesville, Virginia 22903-0818

Recent years have witnessed dramatic progress in our understanding of how turbulence arises and transports angular momentum in astrophysical accretion disks. The key conceptual point has its origins in work dating from the 1950s, but its implications have been fully understood only in the last several years: the combination of a subthermal magnetic field (any nonpathological configuration will do) and outwardly decreasing differential rotation rapidly generates magnetohydrodynamic (MHD) turbulence via a remarkably simple linear instability. The result is a greatly enhanced effective viscosity, the origin of which had been a long-standing problem. The MHD nature of disk turbulence has linked two broad domains of magnetized fluid research: accretion theory and dynamos. The understanding that weak magnetic fields are not merely passively acted upon by turbulence, but actively generate it, means that the assumptions of classical dynamo theory break down in disks. Paralleling the new conceptual understanding has been the development of powerful numerical MHD codes. These have taught us that disks truly are turbulent, transporting angular momentum at greatly enhanced rates. We have also learned, however, that not all forms of disk turbulence do this. Purely hydrodynamic turbulence, when it is imposed, simply causes fluctuations without a significant increase in transport. The interplay between numerical simulation and analytic arguments has been particularly fruitful in accretion disk theory and is a major focus of this article. The authors conclude with a summary of what is now known of disk turbulence and mention some knotty outstanding questions (e.g., what is the physics behind nonlinear field saturation?) for which we may soon begin to develop answers. [S0034-6861(98)00501-7]

CONTENTS

I. INTRODUCTION

Nature's affinity for cosmic disk structures was noted in the earliest surveys of the night sky. Chaldean astrologers puzzled over the narrowly confined, erratic paths of the planetary wanderers and the high-arching ribbon of light prominent in the summer sky. The intervening millenia have not witnessed all the progress one might hope for. It is true that the existence of the planes of the solar

ecliptic and galactic disk can now be understood with the help of a few simple, far-reaching concepts: gravitational forces, angular momentum conservation, and dissipational processes. A comprehensive understanding, however, of the origin and evolution of our own solar system and galaxy—or just about any other astrophysical disk system—continues to elude us.

Ignorance of their detailed physical properties has not prevented disks from becoming a pervasive presence in modern astrophysics. In particular, our understanding of star formation and much of high-energy phenomenology is centrally based upon the paradigm of an *accretion disk*. ("Accretion" refers to the accumulation of matter onto a massive central body.) Release of gravitational binding energy is both a dynamic necessity for accretion to occur, as well as a powerful source of luminosity. But in systems with even a small amount of rotation, gravitational contraction is severely limited by a centrifugal barrier. Consider the collapse of a slowly rotating gas cloud. In contrast to internal energy, which is easily radiated, specific angular momentum tends to be conserved and not easily separated from infalling fluid elements. Thus the interim formation of a disk, a repository of the angular momentum, is usually unavoidable. Similar considerations apply to the transfer of matter from one star in a binary to its compact companion, or to the gathering of gas in the cores of active galactic nuclei. The formation of a disk need not halt the infall. But once formed, it is the disk itself that mediates continued accretion, and the physical processes that regulate mass inflow will generally be very different in character from those that may have triggered the initial infall. These processes, which for many years were poorly understood and therefore had a somewhat mysterious, often controversial character, are the focus of this review.

The thread of modern interest in accretion disks can be traced back more or less directly to Kuiper's remarkable paper (1941) on contact binary systems (two stars sharing a common gaseous envelope). Kuiper studied the properties of the streams of gas induced to flow from one star to the other by tidal forces. He was particularly intrigued by the "curious result" (his term) that in the course of its transference, matter seemed to form a ring around the accreting star. These rings were in fact the first hints of structures that would come to be viewed as worthy of study in their own right. Along with neutron stars and black holes, accretion disks are one of a handful of astrophysical objects whose existence was deduced theoretically well in advance of observational guidance.

In the decades following its publication, Kuiper's work, though by no means obscure, seems neither to have been profoundly influential. For example, by the late 1950s, the work of R. P. Kraft and others had established that novae outbursts (which can be recurrent) occurred in binary star systems and involved matter transfer from a normal main-sequence star onto a degenerate white dwarf companion. Several important papers from this era (e.g., Crawford and Kraft, 1956) show an understanding of Kuiper's results and appeal to them as support for what would now be referred to as disk-based accretion. Yet, when compact x-ray sources made their appearance in the 1960s and were identified with binary stars, the realization that a disk must form during the mass transfer process required renewed appreciation (Prendergast and Burbidge, 1968). Thus the detailed study of the physics of accretion really did not begin in earnest until late in this decade, when Lynden-Bell put forth his disk/black-hole model as the central power source for quasars (1969). In turn, this laid the groundwork for the appearance of the seminal papers which have become the basis of a "standard model" of accretion disks: Shakura and Sunyaev (1973) and Lynden-Bell and Pringle (1974).

The central problem of nearly 30 years of accretion disk theory has been to understand how they accrete. In principle, the presence of shear viscosity allows the transfer of angular momentum from one fluid element to another, but this fails badly when actual numbers are used: in almost all cases, the needed accretion rates are orders of magnitude higher than standard microscopic viscosities (see, for example, Spitzer, 1962) could provide. If, on the other hand, the disk were for some reason turbulent, the *effective* viscosity due to interacting eddies could easily be large enough to provide the needed accretion rates. [According to Frisch and Orszag (1990), the idea of an enhanced "eddy viscosity" dates from the early part of this century and originated with Boussinesq and Prandtl.]

What might be the physical cause for the turbulence? Differential rotation by itself is not unstable to linear perturbations unless the angular momentum per unit mass (i.e., the vorticity) decreases outward, a requirement that is not met in astrophysical disks. But when the Reynolds number (defined roughly as the product of a characteristic large-scale flow velocity and its associated length scale divided by the microscopic viscosity) is large, shear flow is notoriously unstable to nonlinear disturbances, an observation going back to Osborne Reynolds himself (1883). Whether this tendency is present in differentially rotating disks has long been a matter of some controversy. For the same reason that the fluid shear viscosity is quite incapable of causing significant accretion, the Reynolds number of astrophysical disks is very high. For some, this meant that disks must be turbulent. Others were less convinced, citing differences between laboratory flow and accretion disk flow, such as supersonic flow velocity and an absence of confining walls. For most, the issue remained undecided.

In recent years, a combination of analytic and numerical work has led to a much better understanding of accretion disk turbulence, and a broad (though not yet universal) consensus has been reached: accretion disks are in fact turbulent, but for reasons having nothing to do with purely hydrodynamic processes. Instead, it is the pervasive presence of magnetic fields that is the culprit.

This is not to say that magnetic fields have been routinely ignored until recently in the study of accretion disk physics. On the contrary, their presence was already noted in the works that would establish the standard accretion disk model. In his pioneering paper, Lynden-

Bell (1969) conceptually separated the stresses due to turbulence and magnetic fields and opined that the latter would exceed the former. Along very similar lines, Shakura and Sunyaev (1973) noted that a magnetic field, tangled by preexisting fluid turbulence, would be an effective source of enhanced transport. A different approach was taken by Eardley and Lightman (1975), who carried through an explicit calculation of a purely magnetic viscosity resulting from Keplerian fluid shear and field-line reconnection. Finally, many authors have pointed to the importance of dynamically significant magnetic fields in extracting angular momentum from disk fluid elements (e.g., Blandford, 1989). The likelihood that magnetic fields played some role in disk transport was a notion that had been about for some time, even among those whose principal interests were hydrodynamic. But in all this, a key idea was missing: weak magnetic fields cause a linear instability (Balbus and Hawley, 1991) that leads directly to disk turbulence (Hawley, Gammie, and Balbus, 1995). Magnetic fields, disk turbulence, and their respective stresses are inseparably linked. This understood, it has now become possible to study accretion disk turbulence and nonlinear transport from first principles, at a level comparable to that of stellar convection.

Broadly, accretion disks can be separated into three categories: (a) protostellar disks, from whence stars and planets are born; (b) disks formed by mass transfer in binary star systems, which are the heart of eruptive novae and compact x-ray sources; and (c) disks in active galactic nuclei (AGN), thought to be the most luminous sources in the universe. The study of protostellar disks has become a very specialized discipline, characterized by highly developed observational and theoretical techniques. The combination of coolish temperatures, large number densities, and the presence of dust leads to a complex nebular chemistry. We shall have little to say directly about protostellar disks in this review; an excellent source is the series *Protostars and Planets* (Gehrels, 1978; Black and Mathews, 1985; Levy and Lunine, 1993). Binary and AGN disks are hot enough to be significantly or fully ionized, and their high conductivity means that currents flow freely. Magnetic fields strongly influence the structure and evolution of these disks. It is not possible to understand the behavior of disks at a fundamental level unless magnetism is included in the analysis from the very beginning.

Although AGN disks have been directly and beautifully imaged by the *Hubble Space Telescope* (see Fig. 1), it is of binary disks that we have the most detailed knowledge. There are several reasons for this. First, there are a large number of such systems within 100 pc;[1] such proximity allows fainter spectral features to be analyzed and more sophisticated techniques (e.g., tomography) to be exploited. Second, the primary star in a binary can sometimes be used as a probe, eclipsing portions of the disk at different times, thereby allowing disk properties to be mapped. Finally, accretion disks in eruptive binary systems are often transitory. The reconstitution of the disk occurs over a time scale convenient for observational monitoring, and much can be learned by studying the evolving spectra. For these reasons, the fiducial physical system referred to in this review is disk accretion about a compact object in a binary. Many of the results, however, will be more broadly applicable.

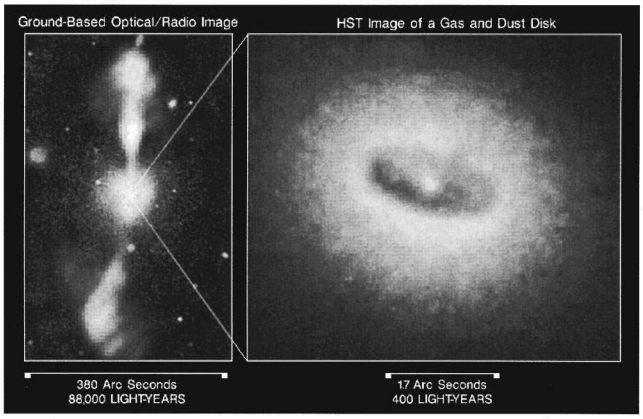

FIG. 1. Hubble Space Telescope (HST) observation of the core of an active galaxy, NGC 4261. The photo on the left combines data from ground-based optical and radio telescopes, showing powerful jets emanating from the core of the galaxy. The right-hand photo is the HST image of the disk in the center of the galaxy. Photo credit: Holland Ford, Johns Hopkins; Walter Jaffe, Leiden Observatory; STScI/NASA.

Our task of reviewing accretion disks has been greatly aided by the large number of excellent reviews that have appeared over the years. This allows us the luxury of concentrating in this review rather exclusively upon the problem of disk turbulence. The article of Novikov and Thorne (1973) still remains an outstanding and authoritative presentation of the astrophysics of black-hole accretion disks. Pringle (1981), much more briefly, offers a clear discussion of phenomenological disk theory. Two recent reviews, those of Papaloizou and Lin (1995, theoretically oriented) and Lin and Papaloizou (1996, observationally oriented), are both very accessible, presenting the reader with a thorough, even-handed treatment of historically contentious issues. Many reviews are geared toward particular classes of objects, but contain discussions of disks and are of more general interest: Petterson (1983), Rees (1984), Katz (1985), Blandford (1989), and Livio (1994) comprise a representative selection. Somewhat dated, but otherwise lucid textbook discussions may be found in Shapiro and Teukolsky (1983), and Frank, King, and Raine (1985).

A. An informal overview

1. Spectral signatures of disks

It is important to understand that far less is understood about astrophysical disks than is understood about

[1]The abbreviation "pc" stands for the astronomical length unit parsec. It is approximately 3.09×10^{18} cm. Typically, the distance between neighboring stars in the disk of our galaxy is a few pc.

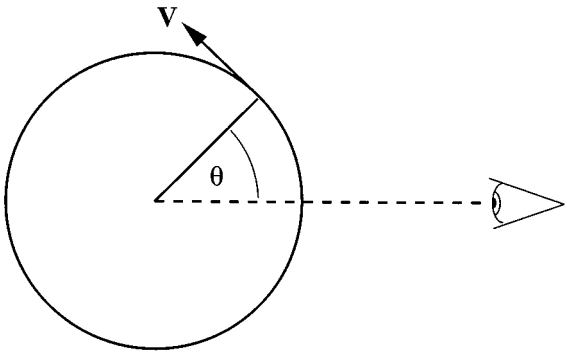

FIG. 2. Geometry for rotating ring viewed edge-on. V is the circular rotation velocity.

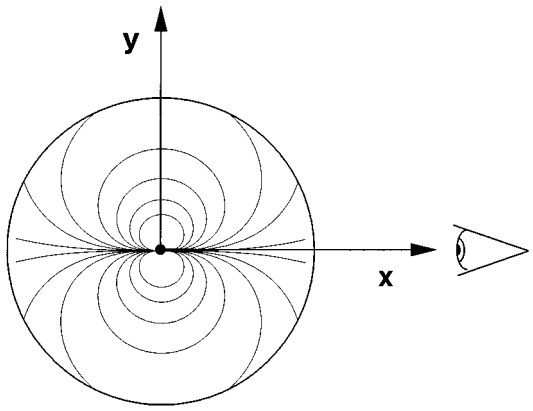

FIG. 3. Loci of constant line-of-sight velocities in a Keplerian disk viewed edge-on along the x axis. They have the same shape as dipole lines of force.

stars. There are very good reasons for this. Not only are disks dynamically much more complicated than stars, they are extremely difficult to detect directly. There are ample and compelling theoretical reasons for believing in the existence of accretion disks; are there equally compelling observations? In the main yes, but details are not easy to extract and are often ambiguous. Binary disks are too small to be directly imaged, and bright disks do not occur in isolation. They occur in systems in which their emission is often overwhelmed by companion stars.

The most obvious disk spectral feature should be a rotational spread in velocities from any discrete line emission (or absorption) present. Consider a simple ring, rotating at velocity V, viewed edge-on (Fig. 2). Ring material subtending an angle θ at the ring center relative to the line of sight has a projected velocity $v=V\sin\theta$. If the ring is emitting uniformly at a well-defined rest frequency corresponding to $v=0$, the emission at velocity v is proportional to $(1-v^2/V^2)^{-1/2}$. This follows because the total emission from velocities less than v is clearly proportional to the angle $\theta=\sin^{-1}(v/V)$ in Fig. 2, and the distribution function is obtained by differentiation. Note not just the spread in velocities, but the double-peaked line structure, which is the hallmark of rotation. Indeed, the emission is formally infinite (but integrable!) at $v=\pm V$.

If we now fill in the ring interior to make a (Keplerian) disk with a sharp outer boundary, emission is obtained from velocities in excess of V. Figure 3 shows contours of constant line-of-sight velocities; for a Keplerian rotation profile these lines are mathematically identical to dipole lines of force. If the emissivity over the face of the disk is assumed for simplicity to be spatially constant, it is not too difficult to work out the form of the resulting emission line profile—a task we leave to the reader. The result is plotted in Fig. 4. The still distinctive double humps are now finite. Emission occurs at all velocities, but declines rapidly ($\sim v^{-5}$) for ($|v|>V$). In real systems, geometrical and thermodynamic effects are considerably more complex, but double-peaked line profiles remain a distinguishing feature of disks.

The combination of a partial disk eclipse and spectral diagnostics provides very strong evidence for the existence of disks in binary star systems. The situation is depicted in Fig. 5. When a companion star partially blocks a disk, only one of the peaks of a spectral line may become visible. Furthermore, since the disk and the binary will be rotating in the same sense, the blueshifted (approaching side) peak ought to be eclipsed first, followed later in the orbit by the redshifted peak. These effects have been unambiguously observed (e.g., Young, Schneider, and Shectman, 1981) and leave no doubt of the basic existence of disks, whether they can be directly imaged or not.

2. Disk luminosity and accretion rates

The bolometric luminosity of an accretion disk is characterized, and to some extent limited, by the so-called *Eddington luminosity* L_E. This quantity is the luminosity at which the radiative momentum flux from a spherically symmetric source is balanced by the gravitational force from a central object. Were the luminosity in a spherical source to exceed L_E, accretion would not be possible. Disks are not spherical, and under some circumstances luminosities in excess of L_E can be produced. Generally, however, most source luminosities

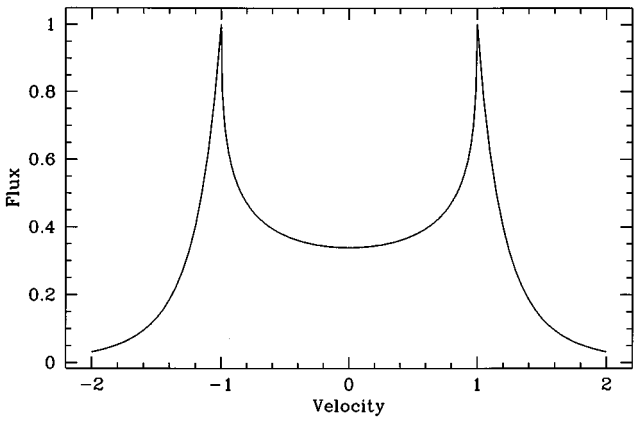

FIG. 4. Emission line from a sharp-edged uniform disk, seen edge-on. Velocity in units of the outermost orbital velocity V; flux units are arbitrary.

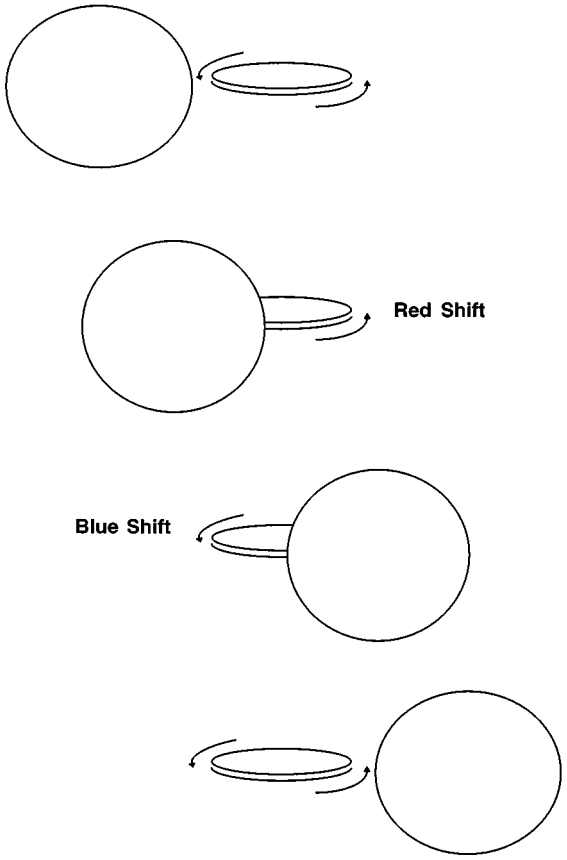

FIG. 5. Partial disk eclipse. As the orbital phase advances, first the blueshifted emission line peak disappears, then the redshifted peak disappears as the blueshifted peak reemerges.

tend to be below their Eddington value, with the more powerful x-ray sources coming within an order of magnitude or so. L_E is thus a useful fiducial benchmark. Because both the radiation flux and the Newtonian force diminish as the inverse square of the distance, L_E is determined solely by the central mass M and fundamental constants of nature:

$$L_E=\frac{4\pi G m_p M c}{\sigma_T}=1.3\times10^{38}\frac{M}{M_\odot}\ \text{ergs s}^{-1}. \tag{1}$$

Here, G is the gravitational constant, m_p the proton mass, c the speed of light, and σ_T the Thompson scattering cross section. The ratio $M/M_\odot$ normalizes the central mass to one solar mass ($M_\odot=1.989\times10^{33}$ g). A binary x-ray source luminosity of 10^{37} ergs s^{-1} is fairly typical. Recall, for comparison, that the total luminosity of the sun is a mere 3.8×10^{33} ergs s^{-1}, relatively little of which is in the form of UV or harder radiation. Thus the discovery of the first compact x-ray source, Sco X-1 (Giacconi *et al.*, 1962), came as a great surprise.

The physical origin of accretion disk luminosity is the accretion process itself. Unlike stars, whose central regions are nuclear furnaces, an accretion disk is hot and radiates as a byproduct of the dynamic activity that accompanies the accretion. The luminosity associated with an accretion rate $\dot{M}$ must, on dimensional grounds, be of the form $GM\dot{M}/R$, where R is some fiducial disk radius. Simple virialization arguments suggest that only half of this energy is actually radiated as a fluid element spirals into its final orbit, the other half being retained as orbital kinetic energy. The latter may itself be radiated in a boundary layer on the surface of the central star, but if a black hole is the accretor, the energy may be permanently lost. In any case, we may define an Eddington accretion rate $\dot{M}_E$ by $L_E\equiv GM\dot{M}_E/R_0$, so that

$$\dot{M}_E=4\pi R_0\frac{m_p c}{\sigma_T}=9.5\times10^{11}\ R_0\text{g s}^{-1} \tag{2}$$

where R_0 is the innermost disk radius (where appropriate, the stellar surface).[2] A typical accretion rate onto a neutron star might thus be $\dot{M}\sim10^{17}$ g s^{-1}. Interestingly, white dwarf accretion sources have comparable values of $\dot{M}$, which implies source luminosities some three orders of magnitude smaller than their neutron star counterparts.

3. Density and temperature scales

The range of densities and temperatures both within a disk and from one disk to another is enormous, but it is possible to make some sensible estimates. We shall be more precise in Sec. III B below, when steady-state disk models are discussed. First, disks occupy the broad density scale gap between interstellar matter, which is at most 10^6 particles cm^{-3} in molecular cloud cores, and stars, whose number density is typically $\sim10^{25}$ cm^{-3}. Disks in binary systems generally have interior densities above 10^{15} cm^{-3}, but well below the stellar regime. Considerable radiation comes from the disk atmosphere, which will typically have a density less than 10^{15} cm^{-3} but well above the molecular cloud core value quoted above.

The innermost regions of an accretion disk can be very hot. If 10^{37} ergs s^{-1} is emerging from a gas disk over a region of radial dimension 10^6 cm (i.e., neutron star dimension) and the gas is emitting as a blackbody, then its temperature will be of order 10^7 K. It will be a plentiful source of keV electrons, as compact x-ray sources indeed are. The surface temperature cools as one moves outward in the disk; the above reasoning suggests that the "local luminosity" of a disk scales as $1/R$, and the radiated flux as $1/R^3$, which implies an $R^{-3/4}$ scaling law for the surface temperature. Thus, on scales of $R\sim10^{10}$ cm, our fiducial disk will have cooled to 10^4 K. Disks around white dwarfs get no hotter than 10^5 K or so in their innermost orbits, and they ought not to be powerful x-ray sources. In general this is the case. However, the physics of the accretion process becomes complex very near the stellar surface, where such phenomena as standing shock waves are possible (see Shapiro

[2]The reader should beware: unlike L_E, there is no standard definition for $\dot{M}_E$. Other authors may define this quantity differently (e.g., Rees, 1984), their choices motivated by the natural scales of the problem of interest (e.g., c^2 instead of GM/R_0).

and Teukolsky, 1983, for a pedagogical discussion), and harder x rays may originate in such processes. A rich variety of eruptive outbursts are associated with white dwarf accretion (Livio 1994; Lin and Papaloizou, 1996); collectively these sources are known as *cataclysmic variables* (CVs). A detailed discussion of their properties would take us too far afield and is left to the reviews cited above.

The interior midplane temperature of a disk can be considerably higher than the surface temperature, a condition similar to, and well understood in, stars: heat flows from hot to cold. Precise temperature estimates depend upon how radiation is vertically transported. It is customary to assume that radiative transport diffuses heat from the disk midplane and then to adopt a model for the radiative opacity, with a degree of complication dictated by one's needs. There is no compelling reason to believe that convective turbulence dominates the radiation transport in binary disks, as it does in the outer layers of low mass stars. There is a question, however, of whether the disk turbulence that is present and doing the job of radial angular momentum and energy transport might also be vertically diffusing passive contaminants such as heat (Stone *et al.*, 1996). The extent to which this is important is not yet known.

If we assume that the emission of radiation from the surface of the disk is driven by radiative diffusion from the interior (conditions of high "optical thickness"), the vertical radiative energy flux is given by (Schwarzschild, 1958):

$$\mathbf{F_{rad}} = -\frac{4\sigma}{3\kappa_r \rho}\frac{\partial}{\partial z}T^4\,\hat{\mathbf{e}}_Z, \tag{3}$$

where σ is the Stefan-Boltzmann radiation constant, $\hat{\mathbf{e}}_\mathbf{Z}$ is a unit vector in the upward vertical direction, and κ_r is the radiative opacity (not to be later confused with the disk epicyclic frequency κ). The vertical *optical depth* τ of the disk is defined by

$$\frac{d\tau}{dz} = -\rho\kappa_r. \tag{4}$$

Transitions between optically thin and optically thick conditions occur when the height-integrated τ goes from $\ll 1$ (in which case the diffusion approximation breaks down) to $\gg 1$. For a very rough orientation, κ_r, which is a radiative cross section per gram of fluid, is likely to be of order unity (in cgs units) at temperatures above 10^4 K. Taking a fiducial ρ of 10^{-8} g cm^{-3} and a disk thickness of some 10% of $R \sim 10^{11}$, we obtain an optical depth of about 100. We therefore expect optically thick conditions to be more pervasive than optically thin, but also that it should be common to find regions in a disk representing both extremes. (Optical depths in protostellar disks are likely to be very much greater than unity, but otherwise very uncertain because of our ignorance of how dust grains modify the opacity law.) Under optically thick conditions, with the vertical flux $F_{rad} \equiv \sigma T_{eff}^4 = \text{const}$, the effective radiating blackbody surface temperature T_{eff} is related to the central midplane disk temperature T_c by

$$T_c^4 = \frac{3}{8}\tau T_{eff}^4 \tag{5}$$

where τ is the total vertical optical depth. This follows simply from the previous two equations. [A somewhat more sophisticated radiative transfer calculation gives instead of τ, the modified sum $\tau + 2/3$ (Shu, 1991), but this refinement will not be necessary here.] Under optically thick conditions, midplane temperatures are thus typically a factor of $\sim 2-4$ higher than surface temperatures, a result that is fairly insensitive to τ. The innermost regions of the disk are likely to be optically thin (see Sec. III below) and therefore more uniform in temperature with height.

II. FORMAL EQUATIONS

A. Dynamics

The fundamental equations of accretion disk structure are

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0, \tag{6a}$$

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + (\rho\mathbf{v}\cdot\nabla)\mathbf{v}\right) = -\nabla\left(P + \frac{B^2}{8\pi}\right) - \rho\nabla\Phi + \left(\frac{\mathbf{B}}{4\pi}\cdot\nabla\right)\mathbf{B} + \eta_V\left(\nabla^2\mathbf{v} + \frac{1}{3}\nabla(\nabla\cdot\mathbf{v})\right), \tag{6b}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla\times(\mathbf{v}\times\mathbf{B} - \eta_B\nabla\times\mathbf{B}). \tag{6c}$$

Equation (6a) is mass conservation, Eq. (6b) is momentum conservation, and Eq. (6c) is the induction equation. Our notation is standard: ρ is the mass density, $\mathbf{v}$ the fluid velocity, P the pressure (plus radiation pressure when important), Φ the gravitational potential, $\mathbf{B}$ the magnetic-field vector, η_V the microscopic kinematic shear viscosity, and η_B the microscopic resistivity. For future reference, let us also define the standard plasma β parameter, which is simply the ratio of gas to magnetic pressure:

$$\beta \equiv \frac{8\pi P}{B^2}. \tag{7}$$

Finally, our treatment is restricted to the domain of nonrelativistic magnetohydrodynamics (MHD) and is therefore not rigorously applicable to flow near neutron star surfaces or black-hole horizons.

A brief digression is appropriate. In writing down the magnetohydrodynamic equations (6b) and (6c), we have assumed that the disk consists of a magnetized plasma of ions and electrons. In such a fluid, Ohm's law takes the form (Jackson, 1975)

$$\mathbf{J} = \sigma\left(\mathbf{E} + \frac{\mathbf{v}}{c}\times\mathbf{B}\right), \tag{8}$$

where σ is the plasma conductivity (taken here to be a scalar), and the current $\mathbf{J}$ and electric field $\mathbf{E}$ are measured in a local rest frame. Using

$$\mathbf{J}=\frac{c}{4\pi}\nabla\times\mathbf{B} \tag{9}$$

(the displacement current is negligible for the nonrelativistic fluid motions that will concern us here), we find that the electric field becomes

$$\mathbf{E}=-\frac{\mathbf{v}}{c}\times\mathbf{B}+\frac{c}{4\pi\sigma}\nabla\times\mathbf{B}. \tag{10}$$

Using this result in Faraday's law,

$$\frac{\partial\mathbf{B}}{\partial t}=-c\nabla\times\mathbf{E}, \tag{11}$$

leads immediately to the induction equation (6c), with

$$\eta_B=\frac{c^2}{4\pi\sigma}.$$

The magnetic terms in Eq. (6b) are simply an expansion of the Lorentz force $\mathbf{J}\times\mathbf{B}/c$ using Eq. (9). The constraint $\nabla\cdot\mathbf{B}=0$ is assured by Eq. (6c), if it is imposed as an initial condition.

The dissipative coefficients in Eqs. (6b) and (6c) are very small, and for simplicity's sake we have ignored the spatial variation of η_V compared with that of $\mathbf{v}$. This simplifies the mathematics while retaining all the important physics. To the extent that η_V depends only upon temperature (as in a collisional Coulomb gas), and that radiative diffusivity is large (keeping small structures nearly isothermal), this approximation is very good. Both the magnetic and velocity fields are rapidly fluctuating in a turbulent fluid, so that the product of second-order spatial gradients with the dissipation coefficients is not, in general, negligible. We have assumed that the bulk viscosity vanishes.

If we integrate Eq. (6c) over an arbitrary volume V in the fluid and use the equivalent of the divergence theorem, we obtain

$$\frac{\partial}{\partial t}\int\mathbf{B}\,dV=\int\mathbf{n}\times(\mathbf{v}\times\mathbf{B}-\eta_B\nabla\times\mathbf{B})\,dS, \tag{12}$$

where S is the surface of V and $\mathbf{n}$ is its outward-pointing unit normal. In other words, the mean value of the magnetic field in a bounded region can be altered only by changing conditions at the region's surface. If the region in question is bounded by an insulator so that the currents are fully contained within, then the integral on the left is simply $8\pi\mathbf{m}/3$, where $\mathbf{m}$ is the dipole moment of the volume (Jackson, 1975). Self-driven MHD turbulence is quadrupolar, and does not act directly on the fluid's dipole properties. Changing a fluid dipole moment or mean magnetic field intimately involves surface dynamics.

An interesting and exact Lagrangian solution to the induction equation is possible when the resistivity is small enough to ignore (Parker, 1979). If we expand Eq. (6c) and use mass conservation to eliminate the term proportional to $\nabla\cdot\mathbf{v}$, we may write the result as

$$\frac{\partial}{\partial t_{\mathcal{L}}}\frac{\mathbf{B}}{\rho}=\frac{\mathbf{B}}{\rho}\cdot\nabla\mathbf{v} \tag{13}$$

where

$$\frac{\partial}{\partial t_{\mathcal{L}}}\equiv\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla\right)$$

is the Lagrangian derivative. Equation (13) has a simple geometrical interpretation. Lay down in the fluid a Lagrangian coordinate system X_i at $t=0$. Imagine now the evolution of an embedded infinitesimal element $d\mathbf{x}$ as the fluid evolves. An elementary construction shows that $d\mathbf{x}$ satisfies

$$\frac{\partial}{\partial t_{\mathcal{L}}}d\mathbf{x}=d\mathbf{x}\cdot\nabla\mathbf{v}. \tag{14}$$

This is exactly Eq. (13). Denoting the initial value of $d\mathbf{x}$ as $d\mathbf{x}(0)$, we must have

$$dx_i=dX_j\frac{\partial x_i}{\partial X_j}=dx_j(0)\frac{\partial x_i}{\partial X_j}.$$

Thus the solution to Eq. (13) is

$$\frac{B_i}{\rho}(\mathbf{x},t)=\frac{B_j}{\rho}(\mathbf{X})\frac{\partial x_i}{\partial X_j}. \tag{15}$$

The indices i,j take on the Cartesian values x,y,z ($\partial_x=\partial/\partial x$, etc.), and we use the convention of summation over repeated subscripts.

Among other things, this solution shows that magnetic-field lines behave as though they were embedded line segments in the fluid ("frozen in"), a classical result which may be obtained in several different ways (Jackson, 1975). We shall make explicit use of this solution to the induction equation when we discuss dynamos in Sec. VI.B.

B. Energy

By way of review, as well as for ease of future reference, we discuss here the energy equations in their various guises. Arriving at the final energy equation is a little tricky. Because we make some nonstandard observations en route, we have chosen to sketch out a few key points in its derivation.

1. Internal energy

The thermodynamic equation of internal energy balance for a gas of adiabatic index 5/3 is

$$\rho\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla\right)\frac{3P}{2\rho}=-P\nabla\cdot\mathbf{v}-\nabla\cdot\mathbf{F}_{\mathbf{rad}}+\Psi \tag{16}$$

where $\mathbf{F}_{\mathbf{rad}}$ the radiative flux and Ψ the volumetric heating rate due to both viscous and ohmic dissipation. If the radiation can freely escape, then $\nabla\cdot\mathbf{F}_{\mathbf{rad}}$ should be replaced by a volumetric radiative loss function, whose precise form depends upon the chemical composition

and thermodynamic state of the gas (Shu, 1991). Detailed modeling of accretion disks in principle requires an explicit formulation and solution of Eq. (16). This is usually very complicated, because of the technical details of handling the radiative transfer, but it will not concern us here, as we shall limit ourselves to more general inferences that can be drawn from the governing energetics.

2. Mechanical energy

The derivation of the equation of mechanical energy conservation first involves taking the dot product of Eq. (6b) with $\mathbf{v}$, using mass conservation [Eq. (6a)], and integrating several terms by parts. In a notation mixing indicial and vectorial representations, we may write the result as

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho v^2+\rho\Phi\right)+\nabla\cdot[\ ]=\left(P+\frac{B^2}{8\pi}\right)\nabla\cdot\mathbf{v}-\frac{B_iB_j}{4\pi}\partial_j v_i - \eta_V\left[\frac{1}{2}(\partial_i v_j)(\partial_i v_j)+\frac{1}{3}(\nabla\cdot\mathbf{v})^2\right] \tag{17}$$

where, within the divergence term $\nabla\cdot[\ ]$, the square brackets represent a dynamic flux,

$$\mathbf{v}\left(\frac{1}{2}\rho v^2+\rho\Phi+P+\frac{B^2}{8\pi}\right)-\frac{\mathbf{v}\cdot\mathbf{B}}{4\pi}\mathbf{B}, \tag{18}$$

plus a viscous flux

$$-\eta_V\left(\nabla\frac{v^2}{2}+\frac{\mathbf{v}}{3}\nabla\cdot\mathbf{v}\right). \tag{19}$$

Transport arising from the viscous flux is generally negligible.

The term proportional to $B_iB_j\partial_j v_i$ represents an interactive coupling between magnetic and mechanical energy. Using the induction equation, we can eliminate it in favor of more readily identifiable magnetic fluxes and ohmic dissipative processes. Dotting Eq. (6c) with $\mathbf{B}$, expanding the right-hand side, and regrouping leads to

$$\frac{\partial}{\partial t}\frac{B^2}{8\pi}+\nabla\cdot\left(\frac{B^2\mathbf{v}}{8\pi}\right)+\frac{B^2}{8\pi}\nabla\cdot\mathbf{v} = \frac{B_iB_j}{4\pi}\partial_j v_i-\nabla\cdot\left[\frac{\eta_B}{4\pi}(\nabla\times\mathbf{B})\times\mathbf{B}\right]-\frac{\eta_B}{4\pi}|\nabla\times\mathbf{B}|^2. \tag{20}$$

Substituting for $B_iB_j\partial_j v_i$ from Eq. (20) into Eq. (17) and simplifying gives us the equation of mechanical energy conservation:

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho v^2+\rho\Phi+\frac{B^2}{8\pi}\right)+\nabla\cdot[\ ]=P\nabla\cdot\mathbf{v}-\eta_V(\partial_i v_j)(\partial_i v_j) - \frac{\eta_V}{3}(\nabla\cdot\mathbf{v})^2-\frac{\eta_B}{4\pi}|\nabla\times\mathbf{B}|^2 \tag{21}$$

where the fluxes in the divergence term $\nabla\cdot[\ ]$ are both dynamic,

$$\mathbf{v}\left(\frac{1}{2}\rho v^2+\rho\Phi+P\right)+\frac{\mathbf{B}}{4\pi}\times(\mathbf{v}\times\mathbf{B}), \tag{22}$$

and viscous,

$$\eta_V\left(\nabla\frac{v^2}{2}+\frac{\mathbf{v}}{3}\nabla\cdot\mathbf{v}\right)-\frac{\eta_B}{4\pi}(\nabla\times\mathbf{B})\times\mathbf{B}. \tag{23}$$

The terms on the right are sources that represent the rate of doing work upon the fluid (the $P\nabla\cdot\mathbf{v}$ term), and the rate at which mechanical energy is lost as heat, corresponding to $-\Psi$ in Eq. (16).

Equation (21) is the main result of this subsection, but it should be noted that what one chooses to call a flux divergence and what one calls a source (or loss) is not always uniquely defined. For example, the positive-definite double sum $(\partial_i v_j)(\partial_i v_j)$ may be recast by using the identities

$$|\nabla\times\mathbf{v}|^2=(\partial_i v_j)(\partial_i v_j)-(\partial_i v_j)(\partial_j v_i), \tag{24}$$

and

$$(\partial_i v_j)(\partial_j v_i)=\nabla\cdot[(\mathbf{v}\cdot\nabla)\mathbf{v}-\mathbf{v}\nabla\cdot\mathbf{v}]+(\nabla\cdot\mathbf{v})^2. \tag{25}$$

If η_V is taken to be a constant, this leads to an alternate formulation of the mechanical energy equation in which the loss terms take on a vector-invariant form:

$$-\eta_V\left[\frac{4}{3}(\nabla\cdot\mathbf{v})^2+\left|\nabla\times\mathbf{v}\right|^2\right]-\frac{\eta_B}{4\pi}|\nabla\times\mathbf{B}|^2. \tag{26}$$

Comparison of these loss terms with those in Eq. (21) shows that they differ by an exact divergence. Because of this loss-term ambiguity, when calculating integrated fluid losses some care must be taken to ensure that the contributions from the fluxes at the outer boundaries have been properly accounted for. Note, for example, that in both Eqs. (21) and (26) the respective expressions for the viscous losses do not vanish for flow in uniform rotation. When the contribution from the divergence term is included, however, the net loss due to viscosity does vanish. In high-Reynolds-number turbulent flow, our main interest here, viscous contributions to the flux divergence are generally negligible. Furthermore, $|\nabla\times\mathbf{v}|^2$ often dominates $(\nabla\cdot\mathbf{v})^2$, a consequence of the near incompressibility of the disturbances. Therefore, from the above expression for the dissipative losses we learn the interesting fact that vorticity and currents directly trace mechanical energy loss in a turbulent gas.

3. Total energy

In the equation for the conservation of the sum of the mechanical and internal energies, dissipation can lead to neither sinks nor sources. Combining Eqs. (16) and (21) leads to

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho v^2+\frac{3}{2}P+\rho\Phi+\frac{B^2}{8\pi}\right)+\nabla\cdot[\ ]=-\nabla\cdot\mathbf{F}_{\mathbf{rad}}, \tag{27}$$

which expresses overall energy conservation. The flux term $\nabla\cdot[\ ]$ consists of the dynamic contribution

$$\mathbf{v}\left(\frac{1}{2}\rho v^2+\rho\Phi+\frac{5}{2}P\right)+\frac{\mathbf{B}}{4\pi}\times(\mathbf{v}\times\mathbf{B}) \tag{28}$$

and a viscous contribution given by Eq. (23). Note the absence of any explicit viscous source terms. In the absence of a wind, energy is lost from an isolated disk only via radiation.

C. Angular momentum

The equation of angular momentum conservation follows from considering the azimuthal component of Eq. (6b). Let (R,ϕ,z) be a standard cylindrical coordinate system centered on the origin. Multiplying the azimuthal equation by R and regrouping terms leads to

$$\frac{\partial}{\partial t}(\rho R v_\phi)+\nabla\cdot R\left[\rho v_\phi\mathbf{v}-\frac{B_\phi}{4\pi}\mathbf{B_p}+\left(P+\frac{B_p^2}{8\pi}\right)\hat{\boldsymbol{e}}_{\boldsymbol{\phi}}\right]$$
$$-\nabla\cdot\left[\frac{R\eta_V}{3}(\nabla\cdot\mathbf{v})\hat{\boldsymbol{e}}_{\boldsymbol{\phi}}+\eta_V R^2\nabla\frac{v_\phi}{R}\right]=0 \tag{29}$$

where $\hat{\boldsymbol{e}}_{\boldsymbol{\phi}}$ is the unit vector in the azimuthal direction, the p subscript refers to a poloidal magnetic-field component (i.e., the R or z component), and $B_p^2=B_R^2+B_z^2$. There are no source terms in this equation; angular momentum may be redistributed in the fluid, but never destroyed. Generally, the direct transport of angular momentum due to microscopic viscosity η_V is negligible, as noted in our discussion of the energy equation (17), and the second line of the above equation may be ignored.

III. FLUCTUATIONS

As we shall see, the combination of differential rotation and a subthermal magnetic field is highly unstable, and there is every reason to believe that magnetized disks are turbulent. But the large-scale properties of the turbulence are tightly constrained. The same dynamic entity—the turbulent stress tensor—must transport angular momentum and energy, while simultaneously extracting free energy from the mean flow to power the rms turbulent fluctuations. These obligations weigh heavily, and in hydrodynamic (unmagnetized) disks, conditions are not conducive to their fulfillment. Indeed, simply by inspecting the governing equations of the velocity fluctuations it is possible to arrive at some important inferences. The conclusions are striking because they call into question some widespread phenomenological approaches to disk turbulence which have for many years held sway.

A. Steady-state Keplerian disks

Our fiducial disk system consists of a point mass potential, situated at the origin of our coordinates at the center of the disk, and an accompanying gaseous flow. The angular velocity of circular orbits is denoted by $\Omega(R)$, i.e.,

$$\Omega^2(R)=\frac{GM}{R^3} \tag{30}$$

where G is the gravitational constant and M is the central mass. We define the *fluctuation velocity* $\mathbf{u}$ to be the difference between the fluid velocity $\mathbf{v}$ and the azimuthal circular velocity $R\Omega$,

$$u_R=v_R,\quad u_\phi=v_\phi-R\Omega,\quad u_z=v_z. \tag{31}$$

When $R\Omega$ much exceeds the isothermal sound speed c_s, as is often the case in astrophysical applications, the disk is thin: the scale height H satisfies $H\ll R$. More explicitly, vertical structure in the disk is governed by hydrostatic balance,

$$\frac{\partial P}{\partial z}=-\frac{GM\rho}{R^3}z=-\rho\Omega^2 z, \tag{32}$$

assuming that $z/R\ll 1$. Thus the density of an isothermal gas satisfies

$$\rho=\rho_0\exp(-\Omega^2z^2/2c_s^2)\equiv\rho_0\exp(-z^2/H^2) \tag{33}$$

where ρ_0 is the midplane density, $c_s^2=P/\rho$ the isothermal sound speed, and H the characteristic scale height $\sqrt{2}c_s/\Omega$. The ratio H/R is clearly a measure of the ratio of the sound speed to the local rotation velocity.

When H/R is a small parameter, one may use perturbative techniques even when the disk turbulence is fully developed. The u velocities will be at most of the order of the thermal velocity (assuming that the magnetic field is not highly suprathermal), since fluctuations well in excess of this will form shocks and become highly dissipative. (Of course, subsonic turbulence is itself dissipative and requires an external source to be maintained.) The practical consequence of $u/R\Omega\ll 1$ is that it is possible to analyze turbulent velocity correlations by retaining only terms through second order.

As an example, let us consider the classical problem of a thin disk under steady-state conditions. The radial flux of angular momentum follows from Eqs. (29) and (31):

$$R\left[\rho u_R(R\Omega+u_\phi)-\frac{B_R B_\phi}{4\pi}\right]. \tag{34}$$

(We have ignored the viscous terms.) We now average this quantity over ϕ, height integrate over z, and average over a narrow range ΔR in R large enough to "smooth over" the rapid fluctuations associated with detailed radial structure. We obtain

$$\Sigma R[R\Omega\langle u_R\rangle_\rho+\langle u_R u_\phi-u_{A\,R}u_{A\,\phi}\rangle_\rho] \tag{35}$$

where the surface density $\Sigma=\int_{-\infty}^{\infty}\rho\,dz$, and for any X

$$\langle X\rangle_\rho=\frac{1}{2\pi\Sigma\Delta R}\int X\rho\,d\phi\,dR\,dz. \tag{36}$$

The notation $u_{A\,R}$, etc. denotes the Alfvén velocity

$$\mathbf{u_A}=\frac{\mathbf{B}}{\sqrt{4\pi\rho}},$$

which is the characteristic velocity of transverse shear waves in a magnetized fluid.

The first term of the angular momentum flux is the direct inflow of angular momentum due to accreting matter in nearly circular orbits (see below). The second term represents an outward component of the flux: the turbulent transport. It arises from the statistical correlations in the Reynolds (velocity) and Maxwell (magnetic) stress tensors. The dominant $R\phi$ component of the stress tensor,

$$W_{R\phi}\equiv\langle u_R u_\phi - u_{A\,R}u_{A\,\phi}\rangle_\rho, \tag{37}$$

occurs often in our analysis. It is convenient, if somewhat inaccurate, to refer to this height-averaged, single component as *the* stress tensor. Traditionally, this has been treated simply as an enhanced viscous flux (Lynden-Bell and Pringle, 1974, Pringle, 1981; Lin and Papaloizou, 1996), but there are advantages to working directly with the fluctuations themselves. Since the intensity of radiative emission lines from accretion disks under some circumstances depends upon the turbulent velocity dispersion along the observer's line of sight (Horne, 1995), it may be possible to learn something of the turbulent Reynolds stress directly from observations.

Under steady conditions, the (height-integrated) angular momentum flux must be divergence free and vary as $1/R$, if there is no wind carrying off fluid from the disk. The constant of proportionality may be evaluated at the inner edge of the disk (R_0, say), where it is assumed that the stress tensor vanishes. The angular momentum flux at the inner edge of the disk is then determined by the inward mass flux alone. The freedom to choose this constraint is simply a boundary condition.

Since the presence of a nonvanishing stress tensor drains the angular momentum from a fluid element, matter spirals inward, giving rise to a mass accretion rate

$$\dot{M}=-2\pi R\Sigma\langle u_R\rangle_\rho. \tag{38}$$

The accretion velocity $\langle u_R\rangle_\rho$ is of the order of the square of the velocity fluctuations divided by the Keplerian orbital velocity and is thus very small compared directly with the velocity fluctuations themselves. The accretion rate must not vary with R under steady conditions. With Ω_0 defined as Ω evaluated at $R=R_0$, vanishing of $W_{R\phi}$ at $R=R_0$ leads to

$$-\frac{\dot{M}}{2\pi}R\Omega+\Sigma R W_{R\phi}=-\frac{\dot{M}}{2\pi R}R_0^2\Omega_0, \tag{39}$$

whence

$$W_{R\phi}=\frac{\dot{M}\Omega}{2\pi\Sigma}\left[1-\left(\frac{R_0}{R}\right)^{1/2}\right]. \tag{40}$$

Thus far our analysis has been exact, because the angular momentum flux involved no correlations higher than second order. The energy flux, however, contains higher-order moments, and it is here where perturbative techniques are the most useful. The largest terms in the radial energy flux [cf. Eq. (22)] are

$$\rho u_R\left(\Phi+\frac{1}{2}R^2\Omega^2+R\Omega u_\phi\right)-\frac{R\Omega}{4\pi}B_R B_\phi. \tag{41}$$

Using the Keplerian virial result $\Phi=-R^2\Omega^2$, we find upon averaging that the flux $\mathcal{F}_E$ becomes

$$\mathcal{F}_E=\frac{\dot{M}R\Omega^2}{4\pi}+\Sigma R\Omega W_{R_\phi}. \tag{42}$$

Substituting for Ω and using Eq. (40) for the stress tensor, this reduces to

$$\mathcal{F}_E=\frac{3GM\dot{M}}{4\pi R^2}\left[1-\frac{2}{3}\left(\frac{R_0}{R}\right)^{1/2}\right]. \tag{43}$$

Unlike its angular momentum counterpart, the energy flux divergence does not vanish for a steady flow. Indeed, the energy deposited by this flux is the origin of the disk's luminosity. Minus the divergence of the flux gives the disk surface emissivity Q. Dividing by a factor of two for each side of the disk yields a final value for the emissivity of

$$Q=\frac{3GM\dot{M}}{8\pi R^3}\left[1-\left(\frac{R_0}{R}\right)^{1/2}\right]. \tag{44}$$

Notice that the stress tensor appears nowhere in this relation. This should not be surprising; the Q-$\dot{M}$ relationship depends only upon local energy conservation and must be recovered whatever the form of the stress tensor (see, for example, Shakura and Sunyaev, 1973; Pringle, 1981). The approach we are using here, however, leads immediately to a relationship between Q and the first-order disk fluctuations. This one cannot be obtained by other methods. Elimination of $\dot{M}$ between Eqs. (40) and (44) yields

$$Q=\frac{3}{4}\Sigma\Omega W_{R\phi}=\frac{3}{4}\Sigma\Omega\langle u_R u_\phi-u_{A\,R}u_{A\,\phi}\rangle_\rho, \tag{45}$$

a sort of fluctuation-dissipation relation for accretion disks (Balbus, Gammie, and Hawley, 1994).

The form of $W_{R\phi}$ along with this tensor's appearance in Eq. (45) makes an important point. It is the correlation of the velocity fluctuation components, not their mere existence, that is responsible for the relatively high levels of outward disk transport and luminosity. The common tacit (in some cases explicit) assumption that turbulence makes "blobs" whose collisions mimic viscosity is often simply just wrong, as experience with vertical convection models has shown (Stone and Balbus, 1996; Cabot, 1996). Turbulent velocity fluctuations are much too strongly coupled to the mean flow to be regarded as bloblike. To quote the text of Tennekes and Lumley (1972): "Momentum is not a passive contaminant; 'mixing' of mean momentum relates to the dynamics of turbulence, not merely to its kinematics." Turbulence is not a fluid attribute; it is a flow attribute.

The energy radiated by the disk is simply obtained, since it amounts to an integration of a pure divergence. Between R_0 and R the total luminosity emitted is

$$L(R_0<R)=2\pi[R_0\mathcal{F}_E(R_0)-R\mathcal{F}_E(R)]$$
$$=\frac{GM\dot{M}}{2R_0}\left[1-3\frac{R_0}{R}+2\left(\frac{R_0}{R}\right)^{3/2}\right]. \quad (46)$$

In the limit $R\to\infty$, $L=GM\dot{M}/2R_0$, which shows that precisely half the binding energy of the innermost orbit has been converted to radiation. The other half is retained as kinetic energy. The fate of this residual energy depends upon the nature of the central accretor, as was noted in Sec. I.A.2. If a stellar surface is present, the remaining energy will be radiated in a boundary layer; if the central object is a black hole, the energy may be swallowed and lost.

Equation (46) makes a very important and general point concerning the luminosity of an accretion disk: it is dominated by the innermost orbits. Within a stellar boundary layer, the Keplerian modeling we have been using will certainly break down. Unfortunately, this occurs precisely in the region that dominates the disk luminosity. To model the details of the emission properties of accretion disks requires both a sophisticated fluid treatment and a careful handling of the radiation physics. Away from the boundary layer, a somewhat simpler treatment is possible. A crude model, but one whose basic scaling properties are instructive, is to assume that the disk is optically thick and locally radiates like a blackbody with an effective surface temperature $T_{eff}(R)$. Then, equating the energy radiated from each side of the disk with the energy generated, we have

$$2\sigma T_{eff}^4=-\frac{1}{R}\frac{d}{dR}(R\mathcal{F}_E)=\frac{3GM\dot{M}}{4\pi R^3}. \quad (47)$$

As previously deduced, the surface temperature is proportional to $R^{-3/4}$. (Interestingly, if the surface heating is instead dominated by a radiative flux from a centralized source, and if all the surface radiation is absorbed and thermalized, then simple disk geometries lead once again to $T_{eff}\sim R^{-3/4}$.)

This simple radial scaling prediction has recently enjoyed some observational support. Eclipse mapping techniques in the ultraviolet (Baptista *et al.*, 1995) and optical (Moreno, Rutten, and Dhillon, 1996) continua of the outer regions of CV accretion disks yield radial temperature profiles in good agreement with the $R^{-3/4}$ law. The assumption of a local blackbody spectrum implies that the emission at radius R peaks at a frequency $\nu \propto T_{eff}(R)$. Combined with the above scaling for the temperature, this tells us that the characteristic radius R, associated with frequency ν, goes as $\nu^{-4/3}$. Not surprisingly, higher frequencies are dominated by the inner regions of the disk.

We may use this result to deduce a continuum spectral signature of our model disk. For frequencies at which photon occupation numbers are large, the disk is a superposition of Rayleigh-Jeans spectra, since each radius produces an energy spectrum proportional to ν^2. For frequencies at which the occupation numbers are $\ll 1$, the emission is dominated by the Wien spectrum at the smallest (and therefore hottest) contributing radii. The intermediate asymptotic regime, with occupation numbers of order unity, is formed by adding up the peaks of emission from each radius. For frequency ν, the energy spectrum is then proportional to the product of the effective area able to contribute, $R^2\sim\nu^{-8/3}$, and to ν^3, the local blackbody scaling at each radius. In other words, the Rayleigh-Jeans and Wien limits of the energy spectrum are joined by a rather flat link proportional to $\nu^{1/3}$ (Fig. 6). This result was first obtained (more rigorously) by Lynden-Bell (1969; see also Pringle, 1981). In practice, the $\nu^{1/3}$ spectrum has rarely been seen, a shortcoming initially considered to be an embarrassment for accretion disk theory. It is no longer perceived as a serious failure. Not only does this intermediate behavior span a rather limited range of frequencies, it is now more clearly understood that several important complications may qualitatively change the appearance of disk spectra. Perhaps the most important point is that the disk surface temperature is extremely sensitive to external irradiation. Accretion disks rarely find themselves in isolation, and radiative scattering from gas above the disk, which may be present in the form of a wind or a structure analogous to the solar corona, is a significant external heat source.

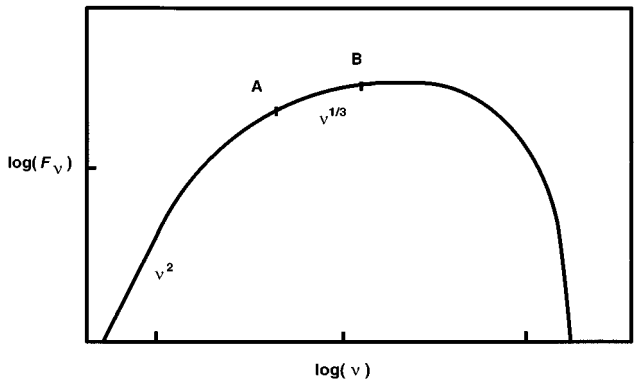

FIG. 6. Model continuum spectrum from a Keplerian disk. The intermediate asymptotic $\nu^{1/3}$ behavior occurs over a limited spectral range, marked off approximately by the points A and B. Adapted from Frank *et al.*, 1985.

B. α disk models

Although the relationship between a steady-state accretion disk's surface emissivity Q and $\dot{M}$ is independent of its detailed transport properties, most other relations involve a dependence on the stress tensor $W_{R\phi}$. Recognizing both the central importance of $W_{R\phi}$ and its computational inaccessibility, Shakura and Sunyaev (1973) suggested that the natural scaling of the stress tensor is $W_{R\phi}\sim c_s^2$ and introduced the now classical α parameter:

$$W_{R\phi}=\alpha c_s^2. \quad (48)$$

The idea is that the turbulent velocities, whose correlation determines $W_{R\phi}$, would be limited by the local sound speed: velocities in excess thereof would develop shocks and quickly drop to below c_s. Furthermore, val-

ues of α within an order of magnitude or so of unity are not *a priori* unreasonable: there is no other obvious or natural limit to which the fluctuation velocities are subject. The presence of magnetic fields would not, as we have seen, be incompatible with this formalism; in the guise of the Maxwell stress they are part of our definition in Eq. (40). (This is not a trivial point: not every external force gives rise to a stress tensor coupling to the mean flow.) The point of view taken by Shakura and Sunyaev (1973) was that disk turbulence would be present, aided perhaps by an entangled magnetic field; magnetically induced turbulence was still some ways off. In fact, at the time of their paper, nonlinear hydrodynamic shear instabilities were considered to be the leading candidate for the origin of turbulent transport.

Of course, the α formalism was invented precisely to finesse the thorny formalism of turbulence with a single dimensionless scaling factor. To the extent that α is taken to be a known quantity, Eq. (48) is a closure relation for the stress tensor. This allows a complete disk model to be constructed, which we shall shortly do.

Another formalism that is sometimes used is to introduce a "turbulent viscosity" ν_t via

$$\nu_t = \alpha c_s H, \tag{49}$$

a scaling based on the form of a microscopic shear viscosity. The role of random-particle velocity is played by c_s, and the scale height H is the effective mean free path. Since both Eqs. (48) and (49) are ultimately a matter of dimensional analysis, their αs must be equal, up to factor of the order of unity. The ν_t formalism is quite popular because it turns a turbulent fluid into a standard Navier-Stokes system, and the equations beckon rigorous analysis. This they should be denied.

While there is a certain sensibility to approaching the gross scaling and evolutionary properties of disks this way (see, for example, Lynden-Bell and Pringle, 1974), turbulence, as we have emphasized, is not reducible to a Stokes viscosity. It may in fact happen that the stress tensor produces the opposite sense of transport from what one would find from a viscous stress. One need not be dealing with pathological flows to encounter this; one need look no further than the solar convection zone (Brummel *et al.*, 1995). A large-Reynolds-number turbulent disk is simply not a laminar disk with a much smaller Reynolds number. If this seems self-evident, it also describes precisely how accretion disks have been (and continue to be) treated innumerable times in the literature. Temptation, and the very real need to do something short of a full-scale simulation, force one's hand. That caution is essential cannot be too strongly emphasized.

Stability analyses are probably the most problematic application of the α formalism, since what comprises the unperturbed state is impossible to define. Furthermore, although they give equivalent *scalings,* different formulations of the turbulence [e.g., along the lines of Eqs. (48) and (49)] do not, in general, lead to equivalent stability conditions. It is tempting to argue that, on sufficiently large length scales and sufficiently long time scales, the presence of turbulence may be adequately modeled as simple viscous dissipation, but the problem is more severe: there is no unique formulation for the stress tensor. Contrast this with the Chapman-Enskog procedure (Chapman and Cowling, 1970), a formalism that bridges the gap between discrete gas particles and fluid continuum, and does lead to well-defined dissipation coefficients.

Having thus criticized the α formalism, we should emphasize that much of observationally driven disk phenomenology is, in fact, based on the stability properties of α-disk models (Livio, 1994). The most robust features of these models have a compelling physical basis which transcends any particular formal implementation scheme for the turbulence. We are not yet at a stage where numerical modeling can answer sensitive stability questions, and this subfield remains more of an art than a science. A good rule of thumb when gauging the plausibility of an α-based model is to trust gross physical properties (e.g., heating and cooling), but tread cautiously in the domain of mathematically provocative details (e.g., WKB waves). For all their foibles, there is little doubt but that α disks will remain part of the landscape for some time.

Let us see how a simple α model works in detail for a standard free-free absorption model of the opacity known as *Kramers' law* (Schwarzschild, 1958; Frank *et al.*, 1985):

$$\kappa_r = 6.6\times 10^{22}\ \rho T^{-7/2}\ \mathrm{cm}^2\ \mathrm{g}^{-1}. \tag{50}$$

An astronomical aside: there is a corresponding bound-free Kramers' opacity with the same ρ and T scaling, but depending linearly upon the relative abundance of elements heavier than helium ("metallicity" in astronomical parlance). Older, lower-mass stars tend to be lower in metals than younger, more massive stars. In a low-metallicity gas, free-free opacity is more important than bound-free. It is simple and physically sensible to examine the case in which the donor star is relatively metal poor. Another contribution to the opacity, electron scattering, becomes important at high temperatures, but need not be included here.

The solution is parametrized by two quantities: α and $\dot{M}$. There are then seven variables and seven relations between them. The variables are ρ_c, the midplane density of the disk; T_c, the midplane temperature of the disk; Q, the heating rate/surface emissivity; Σ, the disk column density; H, the disk scale height; τ, the integrated optical depth; and $W_{R\phi}$, the stress tensor. The seven equations are either self-evident or have been discussed already. They are

$$Q = \frac{3GM\dot{M}}{8\pi R^3}\left[1-\left(\frac{R_0}{R}\right)^{1/2}\right], \tag{51a}$$

$$Q = \frac{3}{4}\Sigma\Omega W_{R\phi}, \tag{51b}$$

$$\sigma T_c^4 = \frac{3}{8}\tau Q, \tag{51c}$$

$$\tau=\kappa_r\Sigma, \tag{51d}$$

$$W_{R\phi}=\alpha c_s^2, \tag{51e}$$

$$H=\sqrt{2}c_s/\Omega, \tag{51f}$$

$$\Sigma=2\rho_c H. \tag{51g}$$

Although the optical depth τ is an integrated quantity and its value depends upon the entire vertical temperature structure, there is little point in striving for high numerical accuracy (hence, for example, our use of "2" in Eq. (51g). We shall evaluate κ_r at the midplane.

The solution to the system (51) is (cf. Frank *et al.*, 1986)

$$T_c=1.4\times10^4(M/M_\odot)^{1/4}\alpha^{-1/5}(\dot{M}_{16}f)^{3/10}R_{10}^{-3/4}, \tag{52a}$$

$$\rho_c=4.7\times10^{-8}(M/M_\odot)^{5/8}\alpha^{-7/10}(\dot{M}_{16}f)^{11/20}R_{10}^{-15/8}, \tag{52b}$$

$$\tau=29\alpha^{-4/5}(\dot{M}_{16}f)^{1/5}, \tag{52c}$$

$$\frac{H}{R}=1.8\times10^{-2}(M/M_\odot)^{3/8}\alpha^{-1/10}(\dot{M}_{16}f)^{3/20}R_{10}^{1/8}, \tag{52d}$$

$$\Sigma=17(M/M_\odot)^{1/2}\alpha^{-4/5}(\dot{M}_{16}f)^{7/10}R_{10}^{-3/4}, \tag{52e}$$

where $M/M_\odot$ is the central mass in units of solar masses, $\dot{M}_{16}$ is the accretion rate in units of 10^{16} g s^{-1}, R_{10} is the position radius in units of 10^{10} cm, and

$$f=1-\left(\frac{R_0}{R}\right)^{1/2}.$$

All dimensional quantities are in cgs.

The solutions of Eqs. (52) should not be regarded as more than a collection of suggestive scalings; as such, however, they have proven to be very useful. The intricate couplings of disk dynamics and radiative diffusion give results that are often remarkably insensitive to astrophysical parameters and in general agreement with the order-of-magnitude arguments given in Sec. I.A. In particular, the midplane temperature and scale height are essentially independent of α. According to the solution of Eqs. (52), thin disks emitting principally in the UV continuum should be a hallmark of CV systems, a prediction that has garnered considerable observational support. In turn, this offers encouragement that the other scalings deduced from these α models are credible.

C. Fluctuation dynamics; the local and Boussinesq approximations

The individual dynamic equations for the u velocities are of interest because of their close connection with nonlinear stability and disk turbulence. In particular, these equations bring out the importance of the fluctuation–mean flow interaction that is critical to understanding the nature of accretion disk transport. In this section we derive these equations and discuss some of their implications.

In addition to establishing a hierarchy of velocity moments, an important simplification is possible when $u\ll R\Omega$: curvature terms in the dynamical equations may be dropped, since they are typically of order $u/R\Omega\ll1$. In its orbital dynamic guise, this procedure goes back at least as far as Hill's (1878) treatment of the Earth-Moon-Sun three-body problem. More recently, it has been used with success in studies of galactic structure (Toomre, 1981), planetary rings (Goldreich and Tremaine, 1978), and hydrodynamic accretion disks (Goldreich, Goodman, and Narayan, 1986). We shall also exploit the result that the local approximation reveals an important connection between constant-angular-momentum disks and Cartesian shear flows. This, in turn, will help us to contrast the hydrodynamic stability properties of disks (more generally) with those of shear layers.

Lastly, we take advantage of the fact that the velocity field is nearly (but not exactly) incompressible for the turbulent flows of interest. This is the *Boussinesq approximation*. Its consistent implementation sometimes causes confusion; one cannot simply put $\nabla\cdot\mathbf{v}=0$ everywhere. Among other difficulties, this overconstrains the flow when one makes use of the energy equation, which is generally incompatible with $d\rho/dt=0$. The approximation is best thought of as a condition on the inverse time scale $t_{char}^{-1}\equiv\nabla\cdot\mathbf{v}=\nabla\cdot\mathbf{u}$. The time t_{char} is taken to be long compared with characteristic turnover time $\sim(\partial u/\partial x)^{-1}$ of the turbulence, but not necessarily long compared with other time scales (e.g., rotational, thermal) of interest. We may, for example, neglect $(\nabla\cdot\mathbf{u})^2$ compared with $|\nabla\times\mathbf{u}|^2$; we may neglect $\nabla(\nabla\cdot\mathbf{u})$ compared with $\nabla^2\mathbf{u}$; but we should *not* neglect $P\nabla\cdot\mathbf{v}$ in the thermal equation (16), although this is often done.

1. Hydrodynamic equations

Let us start with the case in which there are no magnetic effects. This has the virtue of simplicity, but it also serves to accent the qualitative effects of the fields when they are later reintroduced. One may then appreciate the difficulties of trying to understand disk turbulence on a purely hydrodynamic basis. Writing the equations of motion in terms of the $\mathbf{u}$ velocities, using the local and Boussinesq approximations, we obtain

$$\rho\left(\frac{Du_R}{Dt}-2\Omega u_\phi\right)=-\frac{\partial P}{\partial R}+\eta_V\nabla^2u_R, \tag{53}$$

$$\rho\frac{Du_z}{Dt}=-\frac{\partial P}{\partial z}-\rho\frac{\partial\Phi}{\partial z}+\eta_V\nabla^2u_z, \tag{54}$$

$$\rho\left(\frac{Du_\phi}{Dt}+\frac{\kappa^2}{2\Omega}u_R\right)=-\frac{1}{R}\frac{\partial P}{\partial\phi}+\eta_V\nabla^2u_\phi, \tag{55}$$

where

$$\frac{D}{Dt}\equiv\frac{\partial}{\partial t}+\mathbf{u}\cdot\nabla+\Omega\frac{\partial}{\partial\phi} \tag{56}$$

and

$$\kappa^2 \equiv \frac{1}{R^3}\frac{d(R^4\Omega^2)}{dR} \tag{57}$$

is the so-called "epicyclic frequency." The explicit potential Φ appears only in Eq. (54); if the disk is thin it may be regarded as a function only of z at some fixed fiducial R.

If only gravitational forces were important, then a fluid element perturbed from its circular orbit would appear to execute retrograde epicycles at frequency κ, as seen by a comoving observer in the unperturbed orbit. Hence the origin of the name. (It is, of course, a major point of this review that gravitational forces alone do not explain local disk behavior.) The linear Rayleigh criterion for local stability is simply $\kappa^2>0$, i.e., the specific angular momentum should increase outward.[3] The emergence of κ^2 in Eq. (55) is due to a combination of the Coriolis force and shear: both couple directly to u_R. For this reason, Coriolis forces have a profound influence on flow stability. Streamwise momentum fluctuations in nonrotating shear layers couple to the background flow completely differently from the way angular momentum fluctuations couple to the mean large-scale flow in disks. This is true in both the linear and the nonlinear domains.

Next we write each of the Eqs. (53)–(55) in a kinetic-energy form and average them. Averaging is meant to be understood as a full azimuthal average, but a local average in both the radial and vertical directions. Furthermore, there is no implied density weighting; ρ is explicitly present in the appropriate terms, and our angle brackets are not subscripted by ρ. Multiplying Eq. (53) by u_R, regrouping terms as in Sec. II, averaging, and retaining only the dominant terms leaves us with

$$\frac{\partial}{\partial t}\left\langle\frac{\rho u_R^2}{2}\right\rangle+\nabla\cdot\left\langle\frac{1}{2}\rho u_R^2\mathbf{u}\right\rangle=2\Omega\langle\rho u_R u_\phi\rangle-\left\langle u_R\frac{\partial P}{\partial R}\right\rangle-\eta_V\langle|\nabla u_R|^2\rangle. \tag{58}$$

[Terms that have been dropped in obtaining Eq. (58) include the viscous flux divergence.] Following a similar procedure for the azimuthal equation leads to

$$\frac{\partial}{\partial t}\left\langle\frac{\rho u_\phi^2}{2}\right\rangle+\nabla\cdot\left\langle\frac{1}{2}\rho u_\phi^2\mathbf{u}\right\rangle=-\frac{\kappa^2}{2\Omega}\langle\rho u_R u_\phi\rangle-\left\langle\frac{u_\phi}{R}\frac{\partial P}{\partial\phi}\right\rangle-\eta_V\langle|\nabla u_\phi|^2\rangle, \tag{59}$$

while the z equation is simply

$$\frac{\partial}{\partial t}\left\langle\frac{\rho u_Z^2}{2}\right\rangle+\nabla\cdot\left\langle\frac{1}{2}\rho u_Z^2\mathbf{u}\right\rangle=-\left\langle u_Z\frac{\partial P}{\partial z}\right\rangle-\left\langle\rho u_Z\frac{\partial\Phi}{\partial z}\right\rangle-\eta_V\langle|\nabla u_Z|^2\rangle. \tag{60}$$

An energy equation for the fluctuations may be obtained by summing Eqs. (58)–(60), and using the Ψ term in Eq. (16) to eliminate the viscous terms:

$$\frac{\partial}{\partial t}\langle\mathcal{E}\rangle+\nabla\cdot\langle\mathbf{u}\mathcal{E}+\mathbf{u}P+\mathbf{F}_{\mathbf{rad}}\rangle=-\frac{d\Omega}{d\,\ln R}\langle\rho u_R u_\phi\rangle \tag{61}$$

where

$$\mathcal{E}=\frac{1}{2}\rho u^2+\frac{3}{2}P+\rho\Phi. \tag{62}$$

Note that our fluctuation equation has a source term, in contrast with the energy equation (27): fluctuations can exchange energy with the mean flow.

It is well established that Cartesian shear flow is nonlinearly unstable and that the instability leads to fully developed turbulence (Orszag and Kells, 1980; Drazin and Reid, 1981). Since we are interested in the possibility that disks might break down along similar lines, it behooves us to compare Eqs. (58)–(61) with their Cartesian counterparts. The analogous equations for a Cartesian shear flow with background velocity $V(x)\hat{\mathbf{e}}_\mathbf{y}$ are

$$\frac{\partial}{\partial t}\left\langle\frac{\rho u_X^2}{2}\right\rangle+\nabla\cdot\left\langle\frac{1}{2}\rho u_X^2\mathbf{u}\right\rangle=-\left\langle u_X\frac{\partial P}{\partial x}\right\rangle-\eta_V\langle|\nabla u_X|^2\rangle \tag{63}$$

and

$$\frac{\partial}{\partial t}\left\langle\frac{\rho u_Y^2}{2}\right\rangle+\nabla\cdot\left\langle\frac{1}{2}\rho u_Y^2\mathbf{u}\right\rangle=-\frac{dV}{dx}\langle\rho u_X u_Y\rangle-\left\langle u_Y\frac{\partial P}{\partial y}\right\rangle-\eta_V\langle|\nabla u_Y|^2\rangle. \tag{64}$$

The z equation for Cartesian flow is identical to Eq. (60), and the energy equation is the same as Eq. (61) with dV/dx replacing $d\Omega/d\,\ln R$.

Notice that precisely the same correlation tensor $\langle\rho u_R u_\phi\rangle$ [or its shear counterpart $\langle\rho u_X u_Y\rangle$] that appears in all the above equations is just the nonadvective component of the angular momentum flux in a hydrodynamic disk. Whether or not sustained outward turbulent transport is possible depends upon whether a positive value for $\langle\rho u_R u_\phi\rangle$ is compatible with sustaining a source for the fluctuation velocities. Since the coupling between the mean flow gradients and the correlation tensor is manifestly different in the disk and Cartesian systems, we should not be surprised to find that their nonlinear behaviors are likewise manifestly different. For many years, however, this has been a point of controversy.

It is convenient to unify the disk and shear equations by a simple parametrization (Balbus *et al.*, 1996). Consider the system of equations

[3]The Rayleigh criterion strictly applies to axisymmetric disturbances only. It is possible to construct discontinuous or nearly discontinuous flow profiles that are linearly unstable to nonaxisymmetric disturbances (Howard, 1962; Drazin and Reid, 1981). These are in essence Kelvin-Helmholtz instabilities and are irrelevant to our present consideration of smoothly flowing velocity profiles. Interestingly, no general rigorous stability criterion is known (Howard and Gupta, 1962) for simple Couette flow.

$$\frac{\partial}{\partial t}\left\langle \frac{\rho u_X^2}{2}\right\rangle + \nabla\cdot\left\langle \frac{1}{2}\rho u_X^2 \mathbf{u}\right\rangle = A_X\langle \rho u_X u_Y\rangle - \left\langle u_X \frac{\partial P}{\partial x}\right\rangle - \eta_V\langle |\nabla u_X|^2\rangle, \quad (65)$$

$$\frac{\partial}{\partial t}\left\langle \frac{\rho u_Y^2}{2}\right\rangle + \nabla\cdot\left\langle \frac{1}{2}\rho u_Y^2 \mathbf{u}\right\rangle = A_Y\langle \rho u_X u_Y\rangle - \left\langle u_Y \frac{\partial P}{\partial y}\right\rangle - \eta_V\langle |\nabla u_Y|^2\rangle, \quad (66)$$

where A_X and A_Y are constants. Clearly any local disk model or Cartesian flow profile can be formed by picking the appropriate values of A_X and A_Y. Indeed, one of the nonzero A values can always be scaled out of the problem by choosing appropriate units; it is only the A ratio that matters. Rayleigh instability corresponds to a positive value of A_X/A_Y, (linear) stability to a negative value. Furthermore, there is obviously an $X\leftrightarrow Y$ symmetry to this system, so that we expect to find the same stability behavior in Cartesian shear ($A_X=0$) flow and constant-angular-momentum disks ($A_Y=0$).

With this formalism in place, we may understand how turbulence is triggered in both shear flows and rotating disks.

2. Sources of turbulence: high-Reynolds-number shear?

It is a matter of everyday experience that large-Reynolds-number shear flows are unstable to nonlinear disturbances. This has spawned the common wisdom that where there is high-Reynolds-number shear there is turbulence, and where there is turbulence there is greatly enhanced transport. But the fluctuation equations we have just written down, as well as detailed large-scale numerical simulations, suggest neither of these is intrinsic to the hydrodynamics of accretion disks.

To be sure, three-dimensional inviscid solutions to Eqs. (63) and (64) are found, in accord with experience, to be quite unstable even at modest numerical resolutions. Therein lies the importance of these equations: they represent a system that truly *is* turbulent, and their qualitative content is crucial to understand if we are to understand what might or might not happen in disks. While some viewed the existence of shear flow turbulence as compelling evidence that disks were turbulent, the matter was always controversial (Pringle, 1981; Larson, 1989), because of the suspicion that astrophysical disks differed in some important way from laboratory shear flows. This suspicion is justified. The problem is not, however, the oft-cited supersonic motion of the disk (which matters not at all to the local turbulent dynamics on scales less than H), nor is it the absence of laboratory walls. The problem is that astrophysical disks exhibit epicycles (Balbus *et al.*, 1996).

A casual glance at Eqs. (58)–(64) is enough to show that this is likely to be a critical distinction. If one exchanges dR with dx and $R\,d\phi$ with dy, the structural form of the disk and shear equations is seen to be extremely similar, but with crucial dissimilarities in their source terms. These all-important components, which couple the fluctuation amplitudes with the background flow, are qualitatively different. In particular, the shear system has a source term in the y equation, but no sink coupled to the Reynolds stress in either the x or the z equations. By way of contrast, in the disk equations, there *is* such a dynamic sink: the epicyclic term in the azimuthal equation has the opposite sign from its shear counterpart. This follows directly from the fact that angular momentum fluctuations draw upon a background angular momentum gradient that increases outward, even as the angular velocity decreases outward. This is the defining difference between the two systems. In the Cartesian shear system there is only one characteristic gradient, dV/dx; in the disk system there are two, one for the angular velocity, and one for the angular momentum. Both of the latter gradients have great dynamic significance, and they have *opposite signs*. As a consequence, the linear and nonlinear stability properties of disk flow and pure shear flow are qualitatively different. In a Keplerian disk, there is no asymptotic domain for either linear or nonlinear perturbations in which the governing dynamical equations behave locally like Cartesian shear. If $\Omega(R)$ is a simple decreasing power law, the Coriolis parameter 2Ω will generally be of the same magnitude as the angular shear ($d\Omega/d\ln R$). In astrophysical disks, far from being negligible, it will be larger than the shear.

There is nothing obvious in the appearance of the Cartesian shear equations that immediately suggests the presence of an instability. In fact, it is well known that there are no *linear* instabilities in shear flows lacking an inflection point (a result due to Rayleigh; see Drazin and Reid, 1981). Conversely, however, neither is there anything obvious in our equations to prevent a nascent nonlinear instability from growing. The term $\langle \rho u_X u_Y\rangle dV/dx$ represents the interaction between the fluctuations and background shear, and it is an energy source for the azimuthal velocity fluctuations. Energy extraction is possible if $\langle \rho u_X u_Y\rangle$ has the opposite sign from dV/dx, i.e., if the transport is from larger to smaller background velocities. The pressure terms, while not of themselves an energy source couple, redistribute energy from the azimuthal to the other components. Furthermore, if stability is a matter of keeping the velocity profile free of local inflection points, it should be noted that these formally nonlinear defects can occur with only minute changes to $V(x)$ and dV/dx. It is not surprising that laboratory nonlinear shear instabilities have been documented for over a century (Reynolds, 1883).

While many features of turbulent flow are mysterious, the mechanism by which energy is channeled from a shear layer into the turbulence is reasonably well understood and uncontroversial: vortices in the fluid are ensnared and stretched by the shear layer, vorticity conservation causes the circulation velocity around the narrowed vortex tube to rise, and free energy is extracted from the background flow (Tennekes and Lumley, 1972). The mathematical term that embodies vortex stretching is obviously the same $\langle \rho u_X u_Y\rangle dV/dx$ we have

been discussing; there is no other source in the energy equation. Pressure terms reallocate energy between components, but their role in vortex stretching and energy extraction is nil. Vortex stretching is critical to creating a correlation between u_X and u_Y and thus to maintaining an outward flux of angular momentum; it is not the presence of turbulence *per se*. This has been a widely misunderstood point. *Enhanced transport is not an inevitable consequence of turbulence.* Enhanced transport is a consequence of the high degree of correlation between u_X and u_Y. In pure shear flow, the turbulence is fed by vortex stretching, which also happens to correlate u_X and u_Y fluctuations. If there were some other source for the turbulence (e.g., external driving, convection), there would be no reason why u_X and u_Y need be positively correlated everywhere. If they are not, there will not be enhanced transport, regardless of the amplitude attained by the turbulent fluctuations. This becomes particularly important for Keplerian disks, where non-shear sources of disk turbulence are often suggested (Lin and Papaloizou, 1980; Ryu and Goodman, 1992).

Some further understanding emerges from relating the minimum energy states of these simple flows to the transport behavior. Consider first planar flow. The integral $\int \rho v_Y^2/2$, where the integration is over a large but finite volume, is the flow kinetic energy. If we minimize this integral holding the total momentum $\int \rho v_Y$ constant, it is a straightforward matter to show that the flow profile must be one of constant velocity. The mixing term in Eq. (63), $-\langle \rho u_X u_Y \rangle dV/dx$, couples directly to the velocity gradient and redistributes "specific momentum" (i.e., the velocity) towards a state of nonshearing flow. Because the mixing term spontaneously takes the flow from a higher to a lower energy state, the transition is unstable—a nonlinear instability.

Next, consider differential rotation. We seek to minimize $\int \rho v_\phi^2/2$ subject to the constraint that $\int \rho R v_\phi$ be constant. The solution is a profile of constant angular velocity, which is quite distinct from constant specific angular momentum. The mixing term in Eq. (59), however, tries to homogenize the specific angular momentum. This redistribution would not take the flow to a state of lower energy; in fact it tries to do quite the opposite. This is the ultimate reason that flows with epicyclic motions are nonlinearly stable. There is no couple to the angular velocity, but it is angular velocity that needs to be mixed if there is to be a move toward a lower energy state.

The prominent physical role of fluctuation—mean flow interactions—suggests an extremely simple approach to the onset of disk turbulence: all that matters in Eqs. (58) and (59) are the $\langle \rho u_R u_\phi \rangle$ source terms. When $\kappa^2 < 0$ (specific angular momentum decreasing outward), the source terms are both positive and set up an active feedback loop: an increase in u_ϕ leads to an increase in u_R, leading to a further increase in u_ϕ, and so on. An outwardly displaced fluid element has too much angular momentum compared to its orbital surroundings and continues outward; the converse holds for an inwardly displaced element. This is a prescription for a linear instability, the Rayleigh instability.

When $\kappa^2 > 0$, the case of greatest direct astrophysical relevance, the equations offer no route by which energy can be systematically extracted from the background flow. Instead, the epicyclic oscillations are a powerful stabilizing influence, and as long as the fluctuation velocities remain small compared with the $R\Omega$, nonlinearity is of little avail. Hydrodynamic Keplerian disks ought to be stable against the nonlinear disruptions that wreak havoc on Cartesian shear layers.

3. An analogy between shear layers and disks of constant specific angular momentum

The case of $\kappa^2 = 0$ deserves special attention. When the background vorticity vanishes, the disk system (58)–(59) becomes formally identical to the Cartesian system (63)–(64). Even the 2Ω Coriolis term is the same as the disk shear rate $-d\Omega/d \ln R$, so that the coupling between mean flow and stress tensor takes on the precise form of shear flow. The only difference is that the radial (x) and azimuthal (y) variables are interchanged: in a $\kappa^2 = 0$ disk, the turbulence source term appears in the "cross-stream" radial equation, not the "streamwise" azimuthal equation.

One immediate consequence of this observation is that we expect constant-specific-angular-momentum disks to be unstable to nonlinear perturbations, breaking down into enhanced transport turbulence. Simulations in fact bear this out. But the analogy runs deeper, because it suggests why nonlinear disturbances are so important for understanding the behavior of shear layers. The point is that the characteristic local response frequency of an unmagnetized disk to linear disturbances is the epicyclic frequency κ^2 (see Sec. IV.B below). This response frequency is a measure of linear restoring forces on a displaced fluid element—the greater the value of κ^2, the greater the sum of the restoring forces.

If we now imagine taking the limit of vanishing of κ^2 in a sequence of disk models approaching constant specific angular momentum, the linear restoring forces come into ever closer balance. At $\kappa^2 = 0$, they balance precisely, and nonlinear dynamics—restoring or otherwise—must come into play. The κ^2 modes do not disappear as $\kappa^2 \to 0$; they are still present in the disk and easy prey for nonlinear forces.

It can be no less true for Cartesian shear layers. The same zero-frequency modes lie dormant. The classical nonlinear disruption of shear layers should not be thought of as the result of linear forces somehow being overcome. There simply is no linear response. There is nothing to check the nonlinear forces, which will consequently take over. The linearly dormant modes of a shear layer or a constant-specific-angular momentum disk reflect two isolated, singular limits: $A_X = 0$ or $A_Y = 0$. Generic disks, in particular Keplerian disks, behave much differently: they are hydrodynamically stable.

4. Sources of turbulence: convection?

The notion that disks under some circumstances may be convectively unstable, and that convective turbulence will act as an enhanced shear viscosity, has been intensively pursued by protostellar and CV theorists (Cameron, 1978; Lin and Papaloizou, 1980; Canizzo, Ghosh, and Wheeler, 1982; Smak, 1982; Ruden, Papaloizou, and Lin, 1988). In the protostellar case, detailed modeling of the cooling properties of dust grains suggested that the disks would become convectively unstable near their surfaces; CV applications relied on opacity temperature sensitivity to do the same. In both cases the subsequent development of full nonlinear turbulence was assumed to be self-sustaining, with the dissipated energy heating the midplane substantially more than the upper disk layers. Of course, the ultimate source of the convective energy was taken to be the differential rotation somehow—running a convective engine off of its own waste heat *per se* is a thermodynamic impossibility.

This model of turbulent viscosity is clearly derivative of the textbook blob explanation of the Schwarzschild instability (Schwarzschild, 1958). In its more formal guise as mixing-length theory, this method of accounting for convective heat transport in stars has some phenomenological value. But hot and cold gas blobs have a beguiling simplicity which imbues them with a distinct physical reality they do not merit. It is a small step from forming a picture of hot blobs rising and cool blobs sinking to imagining these same blobs colliding and exchanging momenta. Indeed, the macrokinetic phenomenology of colliding gas blobs is the mental "cartoon" of the underlying process of enhanced turbulent viscosity that comes most readily to mind, whatever the source of the turbulence might be. Convection simply offers a plausible mechanism for creating blobs. The transport, one assumes, will take care of itself.

Such reasoning is perilous. The dangers of treating turbulent heat transport and turbulent momentum transport on the same footing have been known for some time (they may be found in textbooks). Heat behaves much more like a passive contaminant than momentum does. Recall as well that convection transports angular momentum *against* the angular velocity gradient in the solar convection zone, so as to maintain the differential rotation. Nevertheless, the first detailed critique of convective angular momentum transport in accretion disks is surprisingly recent (Ryu and Goodman, 1992). An explicit calculation of the angular momentum transport, not an order-of-magnitude mixing-length argument, carried through for the linear stage of the instability, found that the flux had the wrong sign: angular momentum went inward. Ryu and Goodman also argued (but only very briefly) that inward transport may happen in the nonlinear stages of the instability, as well. The idea is that fluid elements on epicyclic oscillations centered on separated radii interact at a common intermediate radius with the outer element moving more rapidly than the inner element. (At the same location in the disk, the outer element has more angular momentum than the inner one.) This process, manifested mathematically in Eq. (59) by the term $-(\kappa^2/2\Omega)\langle\rho u_R u_\phi\rangle$, would tend to transport angular momentum inward. It should be noted, however, that the tendency for sub-Keplerian orbital motion to cause the drift of fluid elements inward (and in the opposite direction for supra-Keplerian motion) tries to transport angular momentum outward. This process is represented by the term $2\Omega\langle\rho u_R u_\phi\rangle$ in Eq. (58). These two opposing drivers prevent the net transport from being very large with either sign. Simply put, turbulence is too coupled to the mean flow to model by some sort of macroscopic gas kinetics; the local mean flow in a nonmagnetic disk is pressure-modified epicycles. The presence of epicyclic oscillations is not conducive to substantial transport.

Limited numerical simulations of convection (heated from below) were carried through by Cabot and Pollack (1992), who investigated artificially small Reynolds numbers in shearing coordinates, and by Kley, Papaloizou, and Lin (1993), who studied axisymmetric convection. Kley *et al.* found inward transport, a result seen to follow immediately from setting $\partial P/\partial\phi=0$ in Eq. (59) and demanding a non-negative source on the right-hand side.

Three-dimensional, inviscid simulations of convectively unstable disks were recently carried out by Stone and Balbus (1996). The basic result was that angular momentum transport was both very small—at least three orders of magnitude below values found for magnetized disks—and directed inward. This holds for any rotation law (assuming, of course, that it is Rayleigh stable), even uniform rotation with no shear at all. On the other hand, when Stone and Balbus drove vertical convection in a Cartesian shear layer flow, the disruption was profound, with transport climbing to levels comparable to those found in a turbulent magnetized disk. But precisely the same disruption is observed in shear layers that are convectively stable (Balbus *et al.*, 1996), so the turbulent transport in this case has little to do with convection, which serves merely as a disruptive trigger. As we have seen, enhanced turbulent transport is intrinsic to the dynamics of Cartesian shear. Convection does not rely on the extraction of free energy from the shear in a disk, so the basis for requiring $\langle\rho u_R u_\phi\rangle>0$ is lost. In a Cartesian shear layer, the free energy of the shear can be tapped, but it comes spontaneously. One does not need convection to extract it.

We may be more explicit. For example, if we consult Eq. (61), there is an apparent inconsistency with the finding of inward transport. Equating the energy source term with the outward radiation losses in steady state gives

$$\langle\rho u_R u_\phi\rangle=-\left(\frac{d\Omega}{d\ln R}\right)^{-1}\nabla\cdot\mathbf{F}_{\mathbf{rad}}>0, \tag{67}$$

which would imply only outward transport. The problem is that Eq. (61) is derived under the assumption that there are no external heat sources, whereas in the numerical simulations this is violated, since the fluid is explicitly heated from below. In the Stone and Balbus

(1996) simulations, the energy equation (16) is modified by adding on the right a term of the form

$$-\rho\chi\nabla^2 T,$$

where T is the gas temperature and χ is an effective thermal conductivity (the usual designation is "κ," but this may be confused with the epicyclic frequency). A similar modification is used by Cabot (1996) and follows the approach of solar convection simulations (Porter and Woodward, 1994). At the bottom of the numerical domain a heat flux enters by way of this term, which then drives unstable thermal convection. The flow is bounded on top and bottom by hard walls, fixed respectively at a cold (top) and hot (bottom) temperature. Discussion of the radial and azimuthal boundary conditions is deferred to the next section, where the properties of the convective stress tensor will be presented.

The reason that the stress tensor in convective turbulence is likely to have very different properties from its shear turbulence counterpart is rooted in energetics. When differential rotation is the source of the fluctuations, rotational mechanical energy must be extracted by a positive stress tensor (outwardly oriented angular momentum flux); there is no other source. Energy is then both radiated locally and transported outward. The energy of the disk's inner regions is lowered in the process. But when thermal energy is introduced continuously into the disk, matters are quite different. Under these circumstances, there is no need for a positive stress tensor (outward transport) to extract energy from the shear to maintain the presence of kinetic-energy fluctuations. Angular momentum fluctuations can only be maintained by two sources: inward transport (direct angular momentum extraction from its background gradient) or large-scale coherence in the azimuthal pressure gradients. The latter is not impossible, but it is a very different type of flow from the mixing of "turbulent blobs" envisioned by the proponents of convective transport, and numerical simulations of local disk sections consistently find inward transport. (The exception to this statement occurs only for runs with unphysically high explicit viscosities. See, for example, the discussion of low-Reynolds-number runs in Cabot and Pollack, 1992.)

The plausibility arguments against hydrodynamic turbulence in disks can be directly tested by three-dimensional numerical integration of the fluctuation equations. This is the topic of the next section.

D. Hydrodynamic numerical analysis

1. Introduction

The set of Eqs. (53)–(55), while extremely useful for understanding qualitative features of disk turbulence, is not the system of choice for a detailed numerical study of local disk dynamics. For this purpose, it is much more convenient to use a set of axes locally corotating with an unperturbed fluid element at fiducial radius R_0; the corresponding angular frequency is $\Omega_0=\Omega(R_0)$. Denoting the velocity in this system as $\mathbf{w}$, we have

$$w_\phi = v_\phi - R\Omega_0. \tag{68}$$

For values of $R-R_0\equiv x\ll R_0$, this implies

$$w_\phi - u_\phi = R(\Omega-\Omega_0)\simeq x\left(R\frac{d\Omega}{dR}\right)_0 \tag{69}$$

which is $-(3/2)\Omega_0 x$ for a Keplerian disk. More generally, we shall take $\Omega(R)\sim R^{-q}$, giving

$$w_\phi - u_\phi = -q\Omega_0 x. \tag{70}$$

With the time derivative in these coordinates given by

$$\frac{\partial}{\partial t_W}=\frac{\partial}{\partial t}+\Omega_0\frac{\partial}{\partial\phi}, \tag{71}$$

Eqs. (53)–(55) transform into the *Hill system*:

$$\rho\left(\frac{\partial}{\partial t_W}+\mathbf{w}\cdot\nabla\right)w_R - 2\rho\Omega w_\phi + \rho x\frac{d\Omega^2}{d\ \ln R} = -\frac{\partial P}{\partial R}+\eta_V\nabla^2 w_R, \tag{72a}$$

$$\rho\left(\frac{\partial}{\partial t_W}+\mathbf{w}\cdot\nabla\right)w_Z = -\frac{\partial P}{\partial z}-\rho\frac{\partial\Phi}{\partial z}+\eta_V\nabla^2 w_Z, \tag{72b}$$

$$\rho\left(\frac{\partial}{\partial t_W}+\mathbf{w}\cdot\nabla\right)w_\phi + 2\rho\Omega w_R = -\frac{1}{R}\frac{\partial P}{\partial\phi}-\eta_V\nabla^2 w_\phi. \tag{72c}$$

We have dropped the 0 subscript on the Ω terms. The Coriolis force is now manifest in the R and ϕ equations, while the x-dependent term of the radial equation represents the local tidal interaction due to the (im)balance between gravity and centrifugal forces. The local Hill approximation thus includes all rotational effects within these three simple terms.

Full-scale numerical simulations of accretion disks, even in the local Hill limit, are a very recent development. There are at least two reasons for this, one practical, one conceptual. The practical difficulty is that nonlinear instabilities in shear flows are three dimensional, and computers capable of carrying out such simulations at high resolution have only recently become widely available. Limited simulations of this type, however, have been possible for nearly two decades. They were not done for a conceptual reason, which, at face value, appears more serious than it actually is. This is the belief that numerical codes have too much numerical viscosity to permit the development of nonlinear shear instabilities that would otherwise be present. The best way to refute this misunderstanding is by explicitly simulating high-Reynolds-number Cartesian shear instabilities at modest numerical resolution.

Of course, the presence of viscosity *can* prevent the development of turbulence. In Cartesian shear layers, low-Reynolds-number flows remain laminar, while high-Reynolds-number flows break down and become turbulent. This (nonlinear) instability is essentially inviscid, however, and would be present in a purely Eulerian fluid. A small viscosity coefficient does not somehow create turbulence. Rather, it limits dissipation to the

point at which inviscid hydrodynamics can generate the turbulence. [Knowledgeable readers will be aware that wall-bounded flows, by way of contrast, may be directly destabilized in the boundary layer by viscosity (Bayly, Orszag, and Herbert, 1988). Such instabilities are not relevant for Keplerian accretion disks, however.]

This line of argument of course applies to both linear and nonlinear instabilities. A case in point is the classical Couette cylinder analysis and the experiments of Taylor (1923), who found a critical minimum Reynolds number for the onset of (linear) instability and turbulence. When both inner and outer cylinder rotated in the same direction at roughly comparable speeds, the instability was observed only in those flows that would have been linearly unstable by the inviscid Rayleigh criterion. It is important to appreciate that in these narrow gap experiments, the basic Rayleigh instability is itself a high-Reynolds-number instability, manifesting itself only when this parameter exceeds $\sim 10^3$ (Drazin and Reid, 1981). Nonlinear instability was later found (Taylor, 1936; Coles, 1965) when the cylinders rotated in opposite directions, or under rather restrictive conditions when the rotation was completely dominated by the outer cylinder. In all of these experiments, viscosity could only prevent the onset of an inviscid instability; it did not cause a new instability to appear.

The long-standing pessimism surrounding the attempted simulation of nonlinear disk instabilities was based upon the expectation that a very-high-Reynolds-number Navier-Stokes code would be required. The mystique attached to high-Reynolds-number nonlinear instabilities has served to obscure the fact that they are as easily simulated in three-dimensional codes as the more familiar (but also large-Reynolds-number) Rayleigh instability. In fact, numerical simulations have been able to recover all known local high-Reynolds-number instabilities, linear and nonlinear, in both shear layers and non-Keplerian disks. Despite this, hydrodynamic Keplerian disks have proven to be extraordinarily stable, well into any sensible nonlinear regime.

2. Numerical viscosity

Let us briefly discuss some properties of numerical diffusion error, often loosely referred to as "numerical viscosity." In fact, diffusion error is quite different from a true physical viscosity. Although, like a true viscosity, its primary effect is to decrease the amplitudes of large-wave-number Fourier components, it is also highly nonlinear, anisotropic, and poorly modeled by a viscous stress tensor. The effective diffusion rate from this process is roughly proportional to both the wave number and a power of the grid spacing, which varies with application. The numerical grid picks out a preferred frame of rest, and numerical diffusion is *not*, in general, Galilean invariant with respect to this frame. When turbulence is present, the small-wave-number region of the spectral energy distribution will be largely unaffected by numerical viscosity, while the high-wave-number region will be strongly damped. Increasing the resolution extends the turbulent damping region to larger wave numbers, but leaves the rest of the wave numbers in the inertial range essentially unaffected (Oran and Boris, 1993).

Since the number of grid zones in a finite-difference simulation is, in fact, quite finite, one might expect the resulting effective numerical Reynolds number to be low. But matters are not so straightforward. The difference between physical and numerical viscosities is brought out nicely by comparing a Navier-Stokes code with an Eulerian (meaning, in this context, "inviscid") code. Porter *et al.* (1990) carried out just such a comparison between the Navier-Stokes and Euler versions of their piecewise parabolic method (PPM) code on a series of test problems. The Eulerian code resolved much more finely detailed structure and shorter-wavelength instabilities than was possible with the highest feasible Reynolds-number Navier-Stokes simulation. This is not difficult to understand. The explicit presence of viscosity forces means that the smaller resolvable grid scales are reserved for dissipative and diffusive flow structure—the dynamic consequences of viscosity. In an Eulerian code, these same small scales become part of the nondissipative turbulent cascade, which is therefore able to resolve finer-scale structure before it is ultimately lost at the grid scale itself. For the PPM code, the effective numerical Reynolds number was found to be roughly proportional to the cube of the number of grid zones. A good Eulerian code can accurately model high-Reynolds-number flows, even at relatively modest resolution.

When is a numerical simulation sufficiently resolved? This depends upon the application one has in mind. Here our interest is in uncovering linear and nonlinear inviscid instabilities. Any local instability operates over some range of wavelengths. Were viscous instabilities our target, the resolved wavelengths would have to be less than the microscopic viscous scale. Then, the required number of grid zones would be of the order of the Reynolds number. But the dynamic instabilities of interest do not depend on a viscous length scale for their existence. The following truth should be self-evident: in a Eulerian code simulation, if there are instabilities present for well-resolved wavelengths, and the growth rate of the instability is greater than the numerical diffusion rate at any of the unstable wavelengths, the instability will be seen. In this case, viscosity (numerical or physical) serves only as the ultimate dissipative sink for high-wave-number structure.

3. Hydrodynamical simulation results

The above considerations suggest that a local simulation of Eulerian accretion dynamics should capture any inviscid, nonlinear instabilities—if they are there to be captured. For studies of instabilities triggered by local shear or differential rotation, it is sufficient to work with the Hill equations. To understand the full stability behavior, however, long-term simulations are required. This is problematic for the local approximation, since "long-term" necessarily implies communication over the

entire computational domain, and boundary conditions must become important. To accommodate the need for long-term evolutions, the numerical simulations use the "shearing box" system.

The shearing box scheme uses a Cartesian grid, with x,y,z corresponding, respectively, to R,ϕ,z. Even though the geometry is simple Cartesian, the local dynamics are not. We have already written down the Hill equations (72); here we discuss their implementation in a finite-difference code. In the shearing box model, the computational domain is a rectangular prism with sides L_X, L_Y, and L_Z. It is assumed to be surrounded by identical prisms, which at $t=0$ form a strictly periodic box lattice. However, a global continuous linear shear flow is present across all the boxes. Thus at later times the central computational box maintains periodic boundary conditions in the y and z directions, while the radial x boundary condition is fixed by the location of the neighboring boxes as they slide past one another due to the background shear. For example, when a fluid element moves off the outer x boundary, it reappears at the inner x boundary at its appropriate sheared position, and with its y velocity compensated for the uniform mean shear across the box. Figure 7 illustrates this.

The boundary conditions can be expressed mathematically for a flow attribute f as

$$f(x,y,z)=f(x+L_X,y-q\Omega L_X t,z) \quad (x \text{ boundary}), \tag{73}$$

$$f(x,y,z)=f(x,y+L_Y,z), \quad (y \text{ boundary}). \tag{74}$$

$$f(x,y,z)=f(x,y,z+L_Z) \quad (z \text{ boundary}). \tag{75}$$

There are two exceptions to these rules. First, in convection simulations, hard wall boundary conditions are used in the z direction. The other exception is for the azimuthal component of the velocity, $w_\phi=w_y$, which must be adjusted to account for the relative shear between neighboring boxes. Specifically, the azimuthal velocity on the radial (x) boundary at time t is

$$w_y(x,y,z)=w_y(x+L_X,y-q\Omega L_X t,z)+q\Omega L_X. \tag{76}$$

At times $t=nL_Y/(q\Omega L_X)$, $n=1,2,3\ldots$ the box becomes strictly periodic again.

The local model has been used extensively in N-body simulations (e.g., Toomre, 1981; Wisdom and Tremaine, 1988). Implementation of the shearing periodic boundary conditions is not as straightforward for a finite-difference scheme as it is for an N-body scheme. The question is how to finite-difference across a radial boundary when in general the grid does not smoothly mesh onto its counterpart in the adjacent sliding "image box." Our approach has been to establish *ghost zones* just outside the computational domain that remain aligned with the grid in the computational domain. The center of each ghost zone corresponds to a point within the computational domain as given by the shearing box boundary conditions; in general, this will not coincide with a computational grid point. Because the ghost zone is fixed to the computational grid, whereas the "image box" is shearing with respect to this grid, the center of the ghost zone will be identified with different locations in the computational domain as time goes by. The image of the ghost zone region will generally straddle two physical zones. Hence, to obtain a quantity f in a ghost zone, it is necessary to interpolate (or remap) f between these two physical zones. Although this is an additional source of truncation error at the boundaries, the size of this error should not be significantly larger in the ghost zones than in the physical zones. It should be noted, however, that *discontinuities* in the truncation error can produce unavoidable numerical artifacts at the boundaries. A description of the specific numerical implementation for the disk problem is given in Hawley, Gammie, and Balbus (1995).

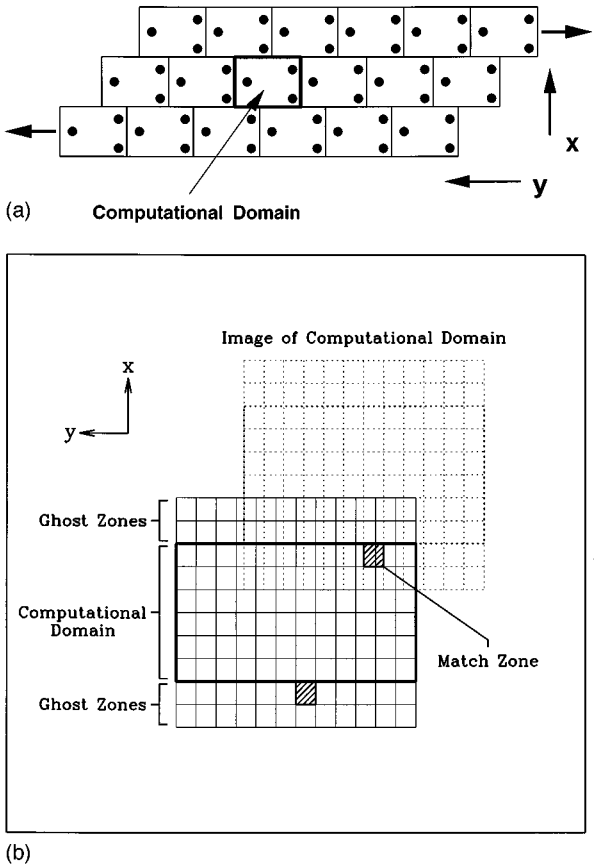

FIG. 7. Shearing box boundary conditions: (a) The computational domain is surrounded by identical domains moving with a fixed relative shear velocity; (b) Implementation of shearing box boundary conditions. The computational domain is surrounded by ghost zones corresponding to neighboring domains. Numerical values in these zones are obtained through interpolation of data from the corresponding matching zones.

We consider first the results of a fiducial run from the convection study of Stone and Balbus (1996), shown in Figs. 8(a) and 8(b). Figure 8(a) displays the hydrodynamic profile of the initial state of the convectively unstable box. Figure 8(b) shows the evolution of the vertical component of the kinetic energy $\langle \rho w_z^2/2 \rangle$ and the normalized stress tensor in the form of the disk α parameter. Although large excursions occur in α, its mean

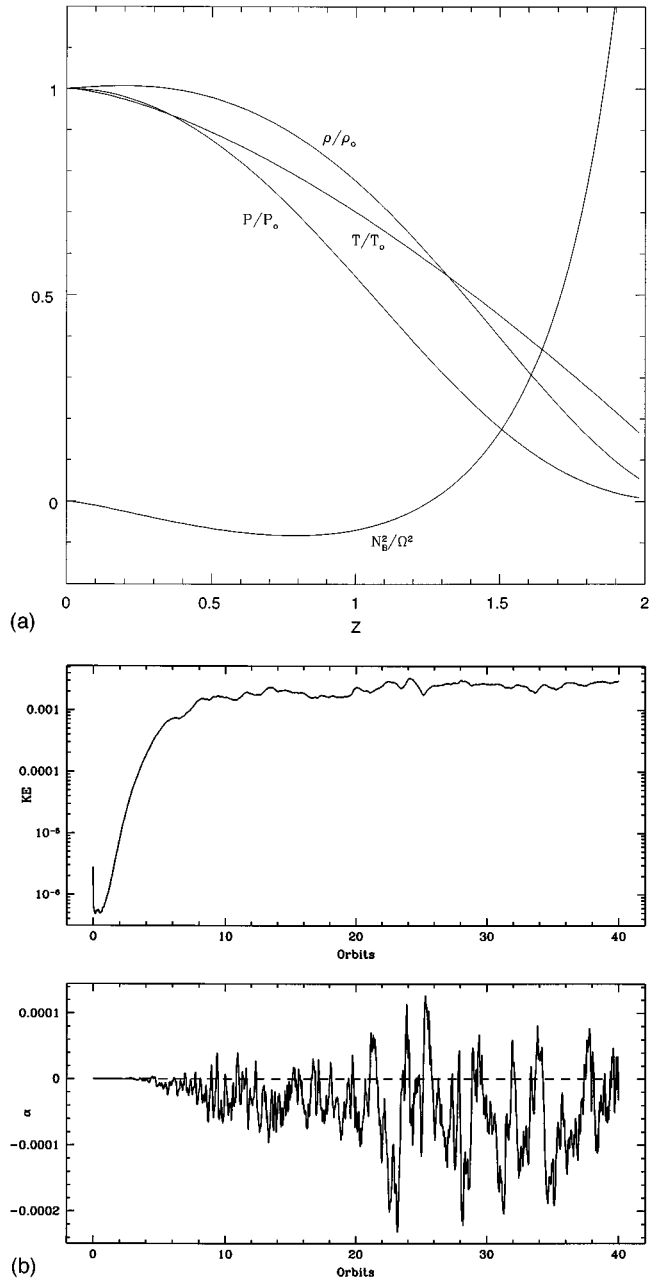

FIG. 8. Simulation of a convectively unstable disk: (a) Profiles of density, temperature, pressure, and Brunt-Väisäla (N_B) frequency in the initial state of a convectively unstable simulation. The model is convectively unstable wherever $N_B^2/\Omega^2<0$. From Stone and Balbus, 1996. (b) Evolution of the vertical component of the kinetic energy (top) and the angular momentum transport parameter α (bottom) in a hydrodynamic disk simulation with external heating applied at the equator. Although considerable vertical convection occurs, the time-averaged α value is negative. From Stone and Balbus, 1996.

value is very small and unambiguously negative ($\alpha\simeq-4.2\times10^{-5}$). A result of similar magnitude (and sign) was found when a local simulation of a uniformly rotating disk was carried out. The absence of shear flow permitted significantly larger azimuthal pressure gradients to be maintained, but even under these circumstances, angular momentum flowed inward in the simulations.

If, instead of a source of external heat, the waste heat from the dissipation of the fluctuations ("viscous heating") were to drive convection, the ultimate energy source would be Keplerian shear. This is, in fact, the standard picture for how convective instability might arise in nature (Lin and Papaloizou, 1980). Then, at least as a point of principle, the sign of α would have to be positive. There are, however, significant potential thermodynamic difficulties with convective dissipation's creating the thermal gradients that drive convective instability. Furthermore, the azimuthal angular momentum equation (59) still must be confronted. If transport is to be outward, it creates a sink for angular momentum fluctuations, which must be overcome by azimuthal pressure gradients. The more effective the transport, the greater the magnitude of the sink. At the very least, convective transport as a generic and effective transport mechanism certainly faces major, and as yet quite unaddressed, difficulties.

For studies of the hydrodynamic stability of differential rotation and shear, we adopt a simple adiabatic equation of state, $P\sim\rho^{5/3}$. It is of course possible to implement other equations of state. As an additional simplification, we may neglect the vertical component of the gravitational field, since it adds nothing to the physics of this problem. (When vertical stratification is of interest, this component of gravity may be added easily.) In the absence of a magnetic field, an exact steady-state solution of the hydrodynamic equations in the shearing box is given by a constant background pressure and density and uniform shear. Assuming that $\Omega(R)\propto R^{-q}$, the local shear velocity is $-q\Omega x\hat{\mathbf{e}}_\mathbf{y}$. We work in length units $L_X=L_Y=L_Z=1$. The initial pressure P_o and the constant mass density ρ_o are chosen so that

$$P_o/\rho_o=(L_X\Omega)^2=\Omega^2=10^{-6}.$$

Onto this initial equilibrium we place a spectrum of random velocity and adiabatic pressure fluctuations.

We have carried out a simulation of the Rayleigh instability on grids as crude as 31^3 zones. Growth rates were found to be in agreement with linear theory, and nonlinear turbulence developed (Balbus *et al.*, 1996). Recall from our earlier discussion that to find such a behavior response already requires an effective Reynolds number in excess of 10^3. Such experiments suggest that if there are local nonlinear hydrodynamic instabilities to be found in Rayleigh-stable flows, three-dimensional simulations should uncover them at accessible resolutions.

The interesting stability issues can be examined by a series of simulations using several Rayleigh-stable background angular velocity distributions, parametrized by q. These linearly stable flows are perturbed at large ($>10\%$) levels over all available wavelengths. The results are summarized in Fig. 9, which shows the time evolution of the kinetic energy in the velocity perturbations. The $q=2.1$ curve is Rayleigh unstable. The contrast between this curve and the Keplerian $q=1.5$ curve

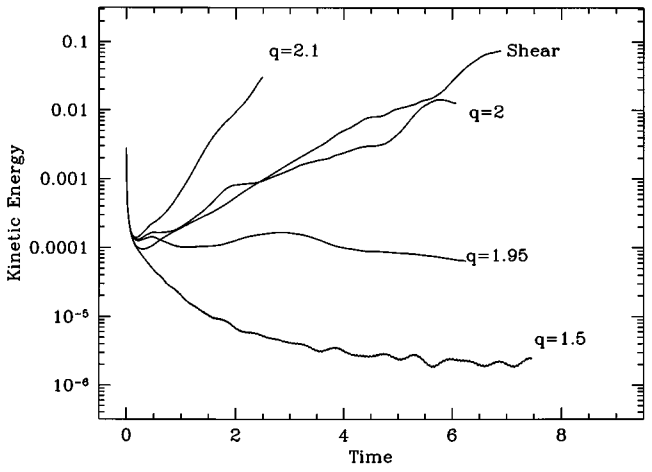

FIG. 9. Evolution of kinetic-energy fluctuations for different rotation profiles. The labels refer to the background angular velocity distribution, as given by $\Omega(R)\sim R^{-q}$. The curve labeled "Shear" has the same shear as the Keplerian $q=1.5$ case, but without the dynamics of differential rotation (Coriolis force and tidal potential). The $q=2$ curve corresponds to constant specific angular momentum, $q=2.1$ to a Rayleigh-unstable profile.

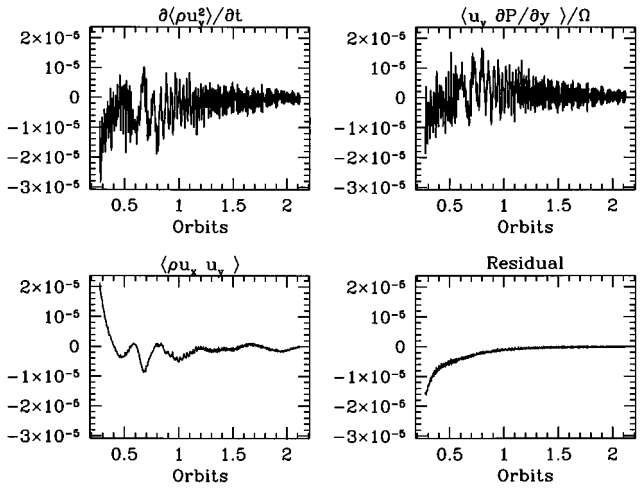

FIG. 10. Evolution of the separate averages of Eq. (77) and their residual sum, for a hydrodynamic simulation of Keplerian differential rotation. The pure divergence term vanishes when averaged over the computational domain. The initial perturbations introduce low-level turbulence that briefly has a positive Reynolds stress, but the disturbances are rapidly damped. From Balbus *et al.*, 1996.

is striking. The finding that Rayleigh-stable disks show no evidence of turbulence cannot be due to the effects of numerical diffusion.

This conclusion is strengthed by the curve labeled "Shear," which corresponds to a Cartesian shear flow without Coriolis or tidal forces. This flow is known to be unstable to finite-amplitude perturbations, and the claim that disks should be nonlinearly hydrodynamically unstable is often made on this basis. The simulation again demonstrates that, whatever the code's effective Reynolds number is, it is large enough to find a nonlinear instability when present. Conversely, the complete absence of any hint of instability in the Keplerian case points to physical, not numerical, causes.

The $q=2$ curve is a linearly stable, constant-angular-momentum disk. This case is particularly interesting because its linear stability is marginal; the epicyclic frequency vanishes. From the discussion in Sec. III.C.3, we see that the absence of the epicyclic term in Eq. (59) means that this flow should share the stablility properties of a shear layer, which also has no dynamic sink. This prediction is confirmed by the simulation. The constant-angular-momentum flow is nonlinearly unstable, with a growth rate indistinguishable from the shear flow.

This effect is further demonstrated by the $q=1.95$ curve. Although the flow has only a very small angular momentum gradient, and hence a small epicyclic frequency, this is enough to stabilize the system completely. The velocity perturbations decline with time, albeit less rapidly than in the Keplerian case. The difference in the nonlinear stability properties of the $q=2$ and $q=1.95$ runs is a clear demonstration that the stabilization is dynamic in origin.

In fact, the entire ensemble of simulations testifies to the stabilizing power of the $-\langle \rho u_R u_\phi \rangle \kappa^2/2\Omega$ term. As q approaches 2, κ^2 vanishes, and with it the rate of decline of the kinetic energy of the fluctuations. However, one need not rely on the appearance of the curves in Fig. 9 to see that this behavior is physical, not numerical, in origin. The precise volume-averaged terms of Eqs. (58) and (59)—or their shear counterparts (63) and (64)—can be computed from the simulations of the Hill system. There is no explicit viscosity term in the implemented equations; that role is assumed by numerical diffusion. In a numerically stable scheme, this will only be a sink for the fluctuation energy. While not directly calculable, it will appear as a residual in the fluctuation energy equation, which in local Cartesian coordinates reads

$$\frac{\partial}{\partial t}\left\langle \frac{\rho u_Y^2}{2} \right\rangle + \frac{\kappa^2}{2\Omega}\langle \rho u_X u_Y \rangle + \left\langle u_Y \frac{\partial P}{\partial y} \right\rangle = \text{Residual}. \tag{77}$$

The contribution to the rate of change of the fluctuation energies due to the epicyclic and pressure gradient terms can be measured directly, and from that the total numerical dissipation may be inferred. (These are the grid scale losses caused by the presence of a turbulent cascade.) In the Keplerian simulation (Fig. 10), the amplitude of each dynamic term is comparable. At the start of the simulation the Reynolds stress term is positive, but the turbulence is dying out. The numerical dissipation residual begins with a comparable negative value, but this term too decays to zero. The pressure term has a comparable amplitude, but oscillates around zero and makes no net time-integrated contribution. At late times only these (mostly) dissipationless pressure waves are present. Tracked by the time derivative term, the pressure term oscillates about zero.

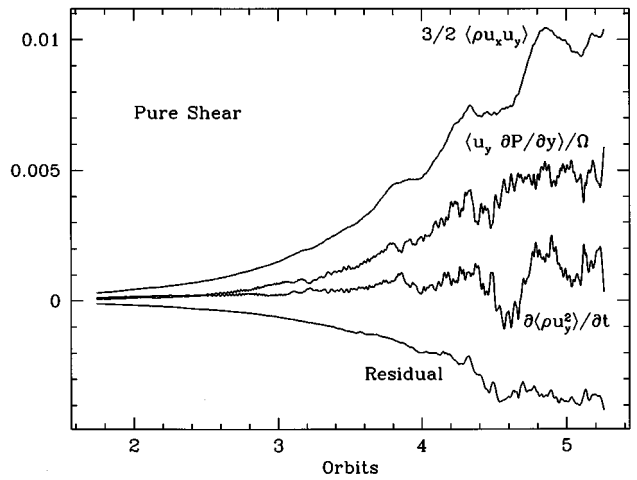

FIG. 11. Evolution of the separate averages of Eq. (77) and their residual sum, for a hydrodynamic simulation of a simple shear layer. The initial perturbations introduce low-level turbulence that is rapidly amplified by hydrodynamic instability, resulting in a growing positive Reynolds stress. (Compare the ordinate scale with that of Fig. 10.) The steadily growing residual loss term represents dissipation and is the hallmark of turbulence. From Balbus *et al.*, 1996.

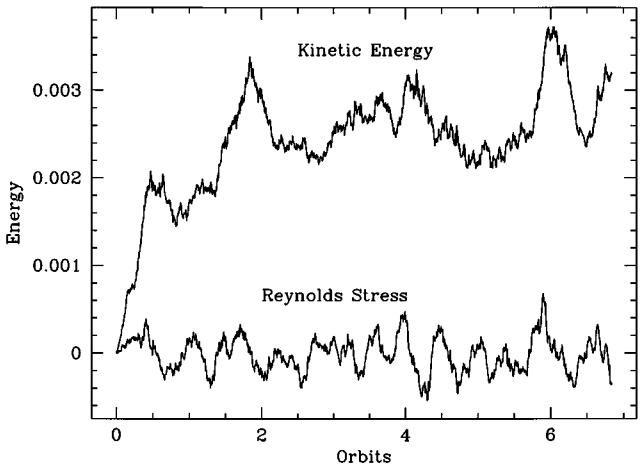

FIG. 12. Kinetic energy and Reynolds stress (in units normalized to the initial gas pressure) for an unstratified, adiabatic, Keplerian shearing box subject to external forcing. Gas is initially isothermal. Although the kinetic energy grows rapidly, the average Reynolds stress vanishes. The presence of turbulence does not necessarily lead to enhanced angular momentum transport in unmagnetized Keplerian disks.

By way of contrast, the nonlinearly unstable shear flow (Fig. 11) has a rising Reynolds stress, matched by an increasing numerical dissipation term. The interpretation of this is straightforward: the shear flow has developed turbulence, which is extracting energy from the background shear velocity gradient. The energy thereby obtained cascades down to high wave numbers, where it is lost to numerical dissipation.

As additional checks, the resolution of the hydrodynamic simulations was both doubled and halved, and larger initial perturbations were applied. None of these variations altered the qualitative results. Nonlinear perturbations grew for the pure shear case and died out in the Keplerian disk case. Despite the fact that high-Reynolds-number instabilities—both linear and nonlinear—were found with relative ease, not a shred of evidence was ever found for nonlinear hydrodynamic instabilities in Keplerian flows.

If turbulence and enhanced transport do not emerge spontaneously from a Keplerian disk, might the latter result from externally driving the former? Our reasoning suggests not, since a positive value for $\langle \rho u_X u_Y \rangle$ is not required to maintain the u fluctuations. Unless the turbulent driving is itself biased, neither outward or inward transport should arise. This is precisely what numerical experiments reveal. In Fig. 12, we show the results of a local Keplerian shearing box subject to an unbiased, noncompressive driving force. (A precise description of the external forcing can be found in the Appendix of Hawley, Gammie, and Balbus, 1996). Note the rapid rise in the turbulent kinetic energy, while the Reynolds stress hovers about zero mean. Thus the presence of turbulence in an accretion disk need not lead to an enhanced turbulent viscosity. We shall encounter the same phenomenon in Sec. III.E.3, where the consequences of tidally induced parametric instability are discussed.

Figure 13 summarizes our results in the A_X-A_Y plane [see Eqs. (65) and (66)].

E. Global hydrodynamic disturbances

The analytic properties of the fluctuation equations, together with numerical simulations, make a very strong

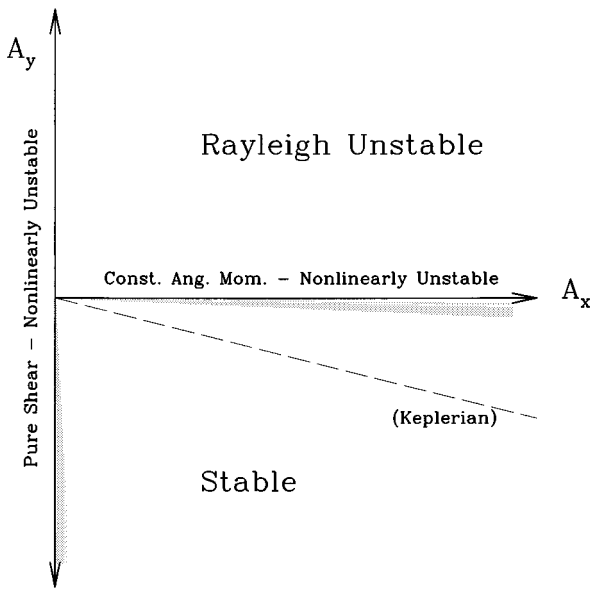

FIG. 13. Summary of hydrodynamic stability in the A_X-A_Y plane. The A_X and A_Y axes correspond, respectively, to shear flow and constant specific-angular-momentum disk flow. Only points along the axes themselves were found to be nonlinearly unstable. (Numerical resolution uncertainty is represented by stippling near the axes.) Keplerian flow was found to be stable, well into any sensible nonlinear regime.

case that local hydrodynamic shear turbulence is not the answer to the question of enhanced disk transport. But this does not mean that angular momentum transport in hydrodynamic disks is impossible. There are other possibilities. We present two, and contrast them with a third mechanism which, though a source of turbulence, is not a source of internal transport.

1. The Papaloizou-Pringle instability

Even if they cannot be generated directly by turbulence, angular momentum fluctuations can be excited by azimuthal pressure gradients. This is, in fact, the *only* way that outward transport occurs in a hydrodynamic disk, since all other "source" terms in Eq. (59) are negative sinks. We have emphasized the need for there to be a good correlation on large scales between azimuthal pressure gradients and velocities (or, equivalently, between pressure and velocity gradients) if angular momentum fluctuations are maintained, and that this is generally accomplished by the formation of trailing spiral waves. The density wave theory of spiral structure (Lin and Shu, 1964; Toomre, 1981) is a classic manifestation of this process in self-gravitating disks. Normally the venue for this is disk galaxies, but related global self-gravitating instabilities have been proposed as a source of outward transport in protostellar disks (Adams, Ruden, and Shu, 1989; Shu *et al.*, 1990). We shall, however, continue to focus on disks in which self-gravity is unimportant. The interested reader will be well served by the review of Papaloizou and Lin (1995), in which an informative summary may be found.

One of the most striking and unexpected results in accretion theory was the discovery (Papaloizou and Pringle, 1984), and subsequent elucidation (Goldreich *et al.*, 1986) of a global nonaxisymmetric instability afflicting an important class of disks. These are accretion tori: bagel-shaped disks with high internal temperatures and well-defined boundaries. If the internal temperature is sufficiently high, pressure gradients can become important enough to rival centrifugal force as the primary source of hydrostatic support for the disk. Under these circumstances, significant departures from a Keplerian rotation curve are inevitable. The extreme Rayleigh-stable limit is $\Omega \sim R^{-2}$, constant specific angular momentum. This requires large pressure gradients for its support, and significant vertical thickening results.

In such a thick accretion torus, the central region is essentially a narrow evacuated vortex along the rotation axis. For a time, these disks were the subject of great astrophysical interest, because it was thought that the empty funnel might collimate outflows. This would explain why well-collimated jets of gas are a ubiquitous feature of the central regions of active galactic nuclei. But the work of Papaloizou and Pringle sharply challenged the viability of such tori: they are dynamically unstable.

The mechanism of the instability is subtle (Goldreich *et al.*, 1986). Ordinarily, when a wave moves through a medium, the mechanical energy is increased by the wave's passage. But if the medium itself is in motion, it is possible for the wave passage to decrease the mechanical energy. Such disturbances are referred to as *negative-energy waves*, in contrast to the more familiar *positive-energy waves*. When a large-scale spiral mode is set up in a disk, the resulting spiral pattern has a fixed angular velocity Ω_p, *the pattern speed.* In some cases, there will be a point in the disk interior, the so-called corotation radius R_c, at which $\Omega(R)=\Omega_p$. For $R>R_c$, $\Omega(R)<\Omega_p$, for $R<R_c$, $\Omega(R)>\Omega_p$. The significance of this is that outside of R_c the wave properties are similar to a passage through a static medium, i.e., the waves are positive energy. Inside R_c, on the other hand, large $\Omega(R)$ leads to a region of negative-energy waves. The corotation point R_c is the only location at which energy exchange between the waves and the background disk medium occurs. Conditions for instability are set up when negative-energy waves lose energy at the corotation radius, energy which is quickly picked up by the positive-energy waves outside of corotation. Losing energy increases the amplitude of a negative-energy wave (it becomes *more* negative in energy), while gaining energy naturally increases the amplitude of a positive-energy wave. To keep the instability going, the negative-energy waves cannot be absorbed at the inner boundary, they must be reflected. When they return to R_c, the process continues, with the positive-energy waves carrying the energy off. Note the critical role played by the reflecting boundary condition; without it the instability loses its feedback mechanism and ceases to be. Given the right conditions, however, the waves can build to a large amplitude and exert considerable torque. The growth rate of the instability is maximal for a constant-angular-momentum torus, decreasing rapidly as one approaches a Keplerian disk (Goldreich *et al.*, 1986).

The nonlinear evolution of the Papaloizou-Pringle instability has been studied via two-dimensional simulations of radially extended thick accretion disks (Blaes and Hawley, 1988) and three-dimensional simulations of tori orbiting around black holes (Hawley, 1991). These simulations show that in radially wide, nearly-constant-angular-momentum tori, the instability saturates in a strong spiral pressure wave, *not* in turbulence (Fig. 14). Moreover, the simulations also confirmed the analysis of Blaes (1987), who suggested that accretion flows through the torus could reduce and even halt the growth of the global instability. Gat and Livio (1992) argued that the onset of accretion would mean the loss of an inner reflecting boundary, and that this was the reason for the sudden drop in activity. More recently, Dwarkadas and Balbus (1996) suggested that even a small amount of radial flow through the corotation radius would disrupt the energy exchange process. Whatever the precise explanation of the abrupt flow stabilization, numerical simulations have taught us that the Papaloizou-Pringle instability is important primarily for nonaccreting constant-angular-momentum tori, and that Keplerian disks are relatively immune to this form of disruption.

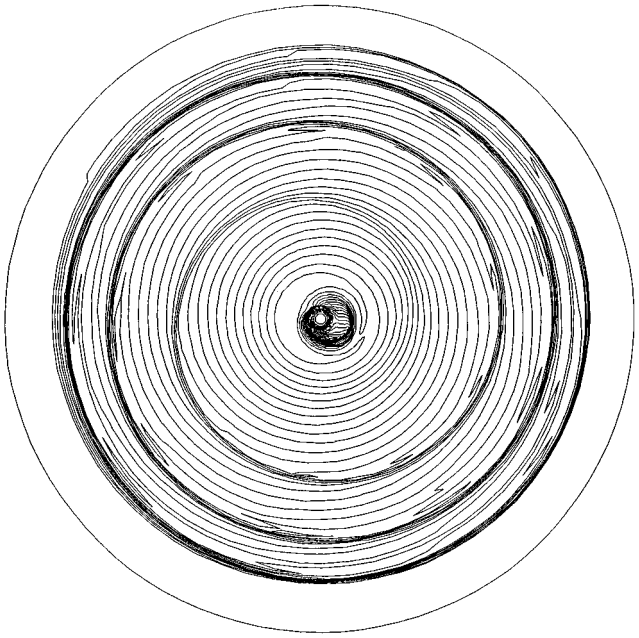

FIG. 14. Contour plot showing density in the equatorial plane from a three-dimensional simulation of a thick, constant-angular-momentum accretion torus. The spiral wave represents the nonlinear saturation of an unstable Papaloizou-Pringle mode. From Hawley, 1991.

2. Spiral shocks

A hydrodynamic disk need not rely on internal instabilities to form spiral structure. In binaries, for example, external forcing from the tidal gravitational field of the companion star can be important (Sawada, 1986; Spruit, 1987; Larson, 1989). The leading-order quadrupole forcing results in the formation of pronounced, two-armed global spiral structure. The peaks of compression in the arms are strong enough to form shock fronts, a feature familiar to astronomers from classical studies of spiral structure in disk galaxies (Roberts, 1969; Shu, Milione, and Roberts, 1973). Tidally induced spiral shocks have been successfully produced in two-dimensional (R,ϕ) hydrodynamic studies by a number of different researchers. These shocks may be significant for understanding disk structure and variability, but their efficiency as an angular transport mechanism appears to be rather low, with $\alpha \sim 10^{-3}$ or less (e.g., Różyczka and Spruit, 1993).

Whether or not shocks contribute noticeably to the α parameter, some degree of spirality is unavoidable in binary systems. For its existence, spiral structure relies on little else but tidal and perhaps mechanical forcing, the latter if an accretion stream impacts upon the disk. Even if (as seems likely) underlying MHD turbulence dominates the transport within the disk, the large-scale properties of compressive shocks are not sensitive to the presence of turbulence or subthermal magnetic fields. (The sound speed and Alfvén velocity couple to the wave number in the same way in a magnetosonic response, discussed in Sec. IV.A.) This stands in sharp contrast to noncompressive disturbances, which are extremely sensitive to the presence of weak magnetic fields—sometimes explosively so.

There are, however, conceptual subtleties associated with spiral shocks, and the models have been criticized in the literature (Lin, Papaloizou, and Savonije, 1990; Lin, 1989). One perceived difficulty is that tidally induced disturbances, which are most pronounced at the outer edge of the disk, must propagate inwards over long radial distances to establish the global two-armed spiral pattern. But, it is argued, these compressive "density waves" (rotationally modified sound waves) will be strongly refracted by the as yet unsimulated vertical pressure gradients and will be diverted sharply upwards before the inner disk regions are affected. It has also been noted that the strength of the spiral structure in the inner regions is sensitive to the disk temperature (Savonije, Papaloizou, and Lin, 1994): in simulations hotter disks show much stronger spiral arms. This is explained by the larger sound speed, which in turn allows for a more rapid propagation rate. Steepening and dissipation of the wave thus occurs farther inward, nearer the center in a hot disk than in a cool disk.

These arguments are of concern for the propagation of driven short-wavelength WKB density waves (it is less of a concern for wavelengths in excess of a vertical scale height), and two-dimensional simulations clearly show disturbances starting from the outer disk edge and moving inwards. On the other hand, these considerations should not obscure the fact that the tidal potential, viewed in a frame corotating with the binary, is static. If the size of the disk is small compared with the binary separation, i.e., if the disk is sensibly Keplerian, it is not difficult to show that the response of the disk will be a well-behaved distortion of its streamlines going gently to zero $\sim R^{5/2}$ as the disk center is approached. However, tidal simulations generally start with a Keplerian disk well out of equilibrium with the external potential. The discrepancy is strongest in the outer regions of the disk, and the ensuing spiral waves generated are clearly part of a transient response. It is far from clear that wave propagation would be a key difficulty in a simulation in which the tides were "turned on" adiabatically.

If the companion star is not far from the outer edge of the disk, then it is possible that resonant wave excitation can occur. An orbiting gas element senses the two-armed tidal field as a periodic disturbance with angular frequency $2(\Omega_{orb}-\Omega)$. (Here, Ω_{orb} is the binary orbital frequency and Ω is the local Keplerian orbital frequency.) If the magnitude of this frequency (either positive or negative quantities are allowed) happens to equal the local epicyclic response frequency κ (which is just Ω for a Keplerian disk), then what is known as a *Lindblad resonance* occurs. The excitation "cross section" is greatly enhanced at a Lindblad resonance, and a detailed analysis (Goldreich and Tremaine, 1979) shows that the coupling between the tidal forcing and the disk is dominated by the resonance. Finite-amplitude waves propagate only inward from this resonance, but angular momentum is transported outward, removed at the resonance by the tidal torque. This represents a viable

means of enhancing the transport by nonturbulent, purely hydrodynamic means. But clearly its generality is limited and, as noted above, it is likely to be much less efficient than fully developed turbulence.

Of course, close binary systems, particularly dwarf novae, often have a mass transfer stream striking the disk, and such a direct impact may be an important nonsteady source of wave excitation. [The case for *inertial wave* excitation has been made by Vishniac and Diamond (1989).] Różyczka and Spruit (1993) have simulated such disks with sound speeds of the order of a few percent of the rotation speed. After an initial transient phase, they found that the disk settled into a quasiequilibrium with α values of order 10^{-3}. Although this is an order of magnitude less than α values typically found in MHD turbulence with self-contained magnetic fields (Hawley *et al.*, 1995), it is large enough to be astrophysically interesting.

Doppler tomography techniques (Marsh and Horne, 1988) should be well-suited to observing spiral structure in accretion disks (Livio, 1994). That such structures are not generally seen is an embarrassment to these models. It may be that only in the outer regions of the largest disks, those whose size approaches the tidal radius, is spirality prominent enough to be marginally detectable. Dwarf novae in outburst may be the most likely objects to develop spiral structure. But even here the role of spiral structure would be secondary, a consequence of whatever leads to eruptive behavior, not the cause thereof.

3. A tidally induced parametric instability: turbulence without internal transport

If the orbits are mildly eccentric, rather than perfectly circular, it is possible to excite a sort of parametric instability in the disk. Originally discovered as an unstable mode of elliptical fluid vortices (Pierrehumbert, 1986; Bayly, 1986), its role in accretion disk physics was first explored by Goodman (1993). In the latter context, elliptical orbits are tidally induced—but gently, or we are back to spiral shock waves. In brief, the Pierrehumbert-Bayly (hereafter PB) instability is present because of a coupling between the noncircular orbital excursions, which are pressure-modified epicycles, and inertial waves, which are also, in essence, pressure-modified epicycles. (Perhaps the coupling is not too surprising.) When the frequency (or a simple harmonic thereof) of the orbital epicycle is close to the inertial wave frequency, orbital energy may be transferred to wave energy. The growth rate is of order $\epsilon\Omega$, where ϵ is the eccentricity of the elliptical orbit. Note that, unlike classical shear instabilities, the free-energy source of the PB instability is not differential rotation: the instability is triggered even in a disk with uniform rotation underlying the ellipses. The ultimate source is the time-dependent gravitational tide, which drives the orbital distortions.[4]

Since the velocity fluctuations of the PB instability do not draw upon $d\Omega/dR$, there is no reason to expect the Reynolds stress $\langle \rho u_R u_\phi \rangle$ associated with the nonlinear saturation to be positive. Even in the linear stages, Goodman (1993) showed that standing, not traveling, waves are excited. Standing waves do not transport energy or angular momentum. Two-dimensional numerical simulations by Ryu and Goodman (1994) confirmed the linear analytic analysis of Goodman (1993), but also showed no significant angular momentum transport within the disk, even when the disk became turbulent. The authors left open the possibility that three-dimensional calculations might produce angular momentum transport within the disk. But the dimensionality of the turbulence does not alter the problem that there is no sense of directionality associated with the disk's internal Reynolds stress for this instability. While angular momentum may be transported from the disk to the tidal source by the action of the instability (Lubow, Pringle, and Kerswell, 1993; Ryu and Goodman, 1994), internal transport at dynamically important levels seems unlikely. Finally, if the disk is magnetized, the magnetorotational instability will certainly dominate the linear phase of growth. Given both the sensitivity of inertial waves to the presence of weak magnetic fields and the strength of MHD turbulence, prospects for the PB instability as the α source in binary accretion disks do not seem propitious.

F. Hydromagnetic fluctuations

The difficulties of extracting the rotational free energy of a disk to power turbulence are at once overcome when magnetic fields are included. The analogs to Eqs. (53)–(54) are

$$\rho\left(\frac{Du_R}{Dt}-2\Omega u_\phi\right)=-\frac{\partial}{\partial R}\left(P+\frac{B^2}{8\pi}\right)+\left(\frac{\mathbf{B}}{4\pi}\cdot\nabla\right)B_R + \eta_V\nabla^2 u_R \tag{78}$$

$$\rho\frac{Du_Z}{Dt}=-\frac{\partial}{\partial z}\left(P+\frac{B^2}{8\pi}\right)-\rho\frac{\partial\Phi}{\partial z}+\left(\frac{\mathbf{B}}{4\pi}\cdot\nabla\right)B_Z + \eta_V\nabla^2 u_Z, \tag{79}$$

$$\rho\left(\frac{Du_\phi}{Dt}+\frac{\kappa^2}{2\Omega}u_R\right)=-\frac{1}{R}\frac{\partial}{\partial\phi}\left(P+\frac{B^2}{8\pi}\right)+\left(\frac{\mathbf{B}}{4\pi}\cdot\nabla\right)B_\phi + \eta_V\nabla^2 u_\phi, \tag{80}$$

where D/Dt is given in Eq. (56).

In the presence of a magnetic field, the route to the analogs of Eqs. (58)–(59) is more lengthy, but the procedure follows the same course as the derivation of the mechanical energy equation. We need the expanded form of the induction equations

$$\frac{DB_R}{Dt}=-B_R\nabla\cdot\mathbf{u}+(\mathbf{B}\cdot\nabla)u_R+\eta_B\nabla^2 B_R, \tag{81}$$

[4]Goodman (1993) argues that it is the orbital distortion itself, rather than the gravitational tide, that should be regarded as the seat of the free energy.

$$\frac{DB_Z}{Dt} = -B_Z \nabla\cdot\mathbf{u} + (\mathbf{B}\cdot\nabla)u_Z + \eta_B \nabla^2 B_Z, \tag{82}$$

$$\frac{DB_\phi}{Dt} - B_R \frac{d\Omega}{dR} = -B_\phi \nabla\cdot\mathbf{u} + (\mathbf{B}\cdot\nabla)u_\phi + \eta_B \nabla^2 B_Z. \tag{83}$$

We have used the local approximation in Eq. (83) to drop the term $u_R B_\phi / R$, and retained only what will prove to be the largest contribution to the resistive terms. Carrying out rather lengthy algebraic manipulations, it is possible to combine the six equations (78)–(83) into three energylike equations,

$$\frac{1}{2}\frac{\partial}{\partial t}\langle \rho(u_R^2 + u_{AR}^2)\rangle + \nabla\cdot\langle\ \rangle = 2\Omega\langle \rho u_R u_\phi\rangle - \left\langle u_R \frac{\partial P_{tot}}{\partial R}\right\rangle - \left\langle \frac{B_R^2}{8\pi}\nabla\cdot\mathbf{u}\right\rangle - \left\langle \eta_V \left|\nabla u_R\right|^2 + \frac{\eta_B}{4\pi}\left|\nabla B_R\right|^2\right\rangle, \tag{84}$$

$$\frac{1}{2}\frac{\partial}{\partial t}\langle \rho(u_Z^2 + u_{Az}^2)\rangle + \nabla\cdot\langle\ \rangle = -\left\langle u_Z \frac{\partial P_{tot}}{\partial z}\right\rangle - \left\langle \rho u_Z \frac{\partial \Phi}{\partial z}\right\rangle - \left\langle \frac{B_Z^2}{8\pi}\nabla\cdot\mathbf{u}\right\rangle - \left\langle \eta_V \left|\nabla u_Z\right|^2 + \frac{\eta_B}{4\pi}\left|\nabla B_Z\right|^2\right\rangle, \tag{85}$$

$$\frac{1}{2}\frac{\partial}{\partial t}\langle \rho(u_\phi^2 + u_{A\phi}^2)\rangle + \nabla\cdot\langle\ \rangle = -2\Omega\langle \rho u_R u_\phi\rangle - T_{R\phi}\frac{d\Omega}{d\ln R} - \left\langle \frac{u_\phi}{R}\frac{\partial P_{tot}}{\partial \phi}\right\rangle - \left\langle \frac{B_\phi^2}{8\pi}\nabla\cdot\mathbf{u}\right\rangle - \left\langle \eta_V \left|\nabla u_\phi\right|^2 + \frac{\eta_B}{4\pi}\left|\nabla B_\phi\right|^2\right\rangle. \tag{86}$$

Once again, the flux divergences have been written $\nabla\cdot\langle\ \rangle$, and in the equation for the ith component, the flux takes the form

$$\frac{1}{2}\rho(u_i^2 + u_{Ai}^2)\mathbf{u} - \rho u_{Ai} u_i \mathbf{u_A} \tag{87}$$

(no summation over repeated i). P_{tot} denotes the sum of the gas plus magnetic pressure,

$$P_{tot} = P + \frac{B^2}{8\pi},$$

and $T_{R\phi}$ is a density-explicit form of the stress tensor,

$$T_{R\phi} = \langle \rho(u_R u_\phi - u_{AR} u_{A\phi})\rangle, \tag{88}$$

closely related to $W_{R\phi}$.

Summation over all three of the above equations, together with some regrouping and integration by parts, leads to a mechanical energy equation for the fluctuations:

$$\frac{1}{2}\frac{\partial}{\partial t}\langle \rho(u^2 + u_A^2 + \Phi)\rangle + \nabla\cdot\langle\ \rangle = -T_{R\phi}\frac{d\Omega}{d\ln R} + \langle P\nabla\cdot\mathbf{u}\rangle - \sum_i \left\langle \eta_V \left|\nabla u_i\right|^2 + \frac{\eta_B}{4\pi}\left|\nabla B_i\right|^2\right\rangle \tag{89}$$

where now only the gas pressure P appears in the equation. The divergence term $\nabla\cdot\langle\ \rangle$ has the flux

$$\left(\frac{1}{2}\rho u^2 + \rho\Phi + P\right)\mathbf{u} + \frac{\mathbf{B}}{4\pi}\times(\mathbf{u}\times\mathbf{B}). \tag{90}$$

Notice how the couplings have changed. The radial energy equation (84) is very similar to its hydrodynamic counterpart (58), except for the unimportant appearance of $B^2/8\pi$ in the pressure term. But the form of the azimuthal equation (86), which was critical to the nonlinear stabilization of a hydrodynamic disk, has been fundamentally altered. The full stress tensor now couples not to the angular momentum gradient, but to the angular *velocity* gradient. There is still a stabilizing couple between $\rho u_R u_\phi$ and the Coriolis force, but it does not involve the full stress tensor. If the angular velocity couple dominates over the Coriolis couple, and we shall presently see that with its magnetic contribution it does, then the entire complexion of the problem has changed. When $d\Omega^2/dR < 0$, then outward transport $\langle \rho u_R u_\phi\rangle$ can set up precisely the same sort of feedback loop that triggers the Rayleigh instability: a positive correlation in the off-diagonal components of the stress tensor maintains hydromagnetic fluctuations and is self-sustaining. The dominant turbulent couple in the all-important azimuthal equation is now "mixing" angular velocity, not angular momentum, and this is the local path to lower energy (cf. Sec. III.C). The resulting instability and dissipative turbulence comes at the expense of the orbital energy, as can be seen in Eq. (89). The ensuing inevitable inspiral of fluid elements is what allows accretion disks to live up to their name.

IV. THE LINEAR STABILITY OF MAGNETIZED ACCRETION DISKS

Having gained some understanding of some of the nonlinear physics of MHD disk turbulence, we are in a better position to appreciate what the linear physics of the instability is trying to do. In particular, we know the most important question to ask: given the difficulties of extracting energy from differential rotation in hydrodynamic disks, precisely how is this overcome by linear magnetized disturbances?

A. A review of MHD waves

Consider a disk medium threaded by a weak magnetic field with an azimuthal component B_ϕ and a vertical

component B_Z. The presence of a radial component would cause a linear time dependence in B_ϕ. In the end, this makes little difference to a weak-field axisymmetric instability (which carries on independently of B_ϕ), but it is simplest to consider the case of vanishing radial field first. The gas is perturbed by local WKB linear disturbances of the form $\exp i(\mathbf{k}\cdot\mathbf{r}-\omega t)$, $kr\gg 1$. Our notation is standard: $\mathbf{k}$ is the wave vector, $\mathbf{r}$ the position vector, ω the angular frequency, and t the time. We denote all linear amplitudes by δ: δP, $\delta\rho$, etc. We restrict our attention for the present to vertical wave numbers, $\mathbf{k}=k\mathbf{e_Z}$. The leading-order local linear equations are

$$-\omega\frac{\delta\rho}{\rho}+k\delta u_Z=0, \tag{91}$$

$$-i\omega\delta u_R-2\Omega\delta u_\phi-i\frac{kB_Z}{4\pi\rho}\delta B_R=0, \tag{92}$$

$$-i\omega\delta u_\phi+\frac{\kappa^2}{2\Omega}\delta u_R-i\frac{kB_Z}{4\pi\rho}\delta B_\phi=0, \tag{93}$$

$$-\omega\ \delta u_Z+k\left(\frac{\delta P}{\rho}+\frac{B_\phi\delta B_\phi}{4\pi\rho}\right)=0, \tag{94}$$

$$-\omega\delta B_R=kB_Z\delta u_R, \tag{95}$$

$$-i\omega\delta B_\phi=\delta B_R\frac{d\Omega}{d\ \ln R}+ikB_Z\delta u_\phi-B_\phi ik\delta u_Z, \tag{96}$$

$$\delta B_Z=0, \tag{97}$$

$$\frac{\delta P}{P}=\frac{5}{3}\frac{\delta\rho}{\rho}. \tag{98}$$

The resulting dispersion formula is

$$[\omega^2-(\mathbf{k}\cdot\mathbf{u_A})^2][\omega^4-k^2\omega^2(a^2+u_A^2)+(\mathbf{k}\cdot\mathbf{u_A})^2k^2a^2]$$
$$-\left[\kappa^2\omega^4-\omega^2\left(k^2\kappa^2(a^2+u_{A\phi}^2)+(\mathbf{k}\cdot\mathbf{u_A})^2\frac{d\Omega^2}{d\ \ln R}\right)\right]$$
$$-k^2a^2(\mathbf{k}\cdot\mathbf{u_A})^2\frac{d\Omega^2}{d\ \ln R}=0 \tag{99}$$

where the adiabatic sound speed and Alfvén speed are, respectively,

$$a^2=\frac{5}{3}\frac{P}{\rho},\quad u_A^2=\frac{B_Z^2+B_\phi^2}{4\pi\rho}.$$

This rather unwieldy equation is best approached piecemeal. Consider first the nonrotating limit:

$$[\omega^2-(\mathbf{k}\cdot\mathbf{u_A})^2][\omega^4-k^2\omega^2(a^2+u_A^2)+(\mathbf{k}\cdot\mathbf{u_A})^2k^2a^2]$$
$$=0. \tag{100}$$

This equation is, in fact, the most general dispersion formula for a magnetic field in a uniform homogeneous medium. No generality is lost by taking the z axis to lie along $\mathbf{k}$. Taking θ to be the angle between $\mathbf{B}$ and $\mathbf{k}$, one root is obviously

$$\omega_A^2=k^2u_A^2\cos^2\theta, \tag{101}$$

corresponding to *Alfvén waves* (Jackson, 1975). These are compressionless disturbances which propagate along the magnetic field, with an effective wave number of $k\cos\theta$. They have been likened to waves on a string, with "magnetic tension" providing the restoring force. The remaining roots of the quartic (quadratic in ω^2) correspond to the fast and slow modes; one root propagates with a phase velocity more rapid than the Alfvén wave, the other less so. The physics of these modes is highlighted most clearly when at least one of three conditions holds: (1) $a\ll u_A$; (2) $u_A\ll a$; (3) $\cos\theta\ll 1$. If any of these are true, then to leading order the fast mode is

$$\omega_+^2=k^2(u_A^2+a^2) \tag{102}$$

and the slow mode is

$$\omega_-^2=\frac{k^2u_A^2a^2\cos^2\theta}{u_A^2+a^2}. \tag{103}$$

Clearly, $\omega_+>\omega_A>\omega_-$. The ω_+ mode represents magnetic and thermal pressure acting in concert. It is sometimes referred to as a *magnetosonic wave*. The ω_- mode represents magnetic tension and gas compression in opposition. When the magnetic field is strong, the result is an ordinary sound wave channeled along the field lines; when the field is weak the slow mode becomes degenerate with the Alfvén mode. This suggests that in the weak-field limit, the slow and Alfvén modes are more closely linked to each other than to the fast mode. Indeed, the Boussinesq approximation may be thought of as the limit $a\to\infty$, which leaves the slow and Alfvén modes completely degenerate.

The effects of Keplerian rotation on the three MHD modes are shown in Fig. 15. We have fixed the values of $\mathbf{k}\cdot\mathbf{u_A}=1$, $ku_{A\phi}=2$, and $ka=5$, i.e., all frequencies are reckoned in $\mathbf{k}\cdot\mathbf{u_A}$ units. The behavior of ω^2 as a function of Ω^2 is then examined. The fast, Alfvén, and slow modes all start at $\Omega=0$ as well-behaved propagating modes, as shown in the figure. But something remarkable happens when Ω^2 reaches 1/3: ω_-^2 turns negative. The slow mode becomes unstable.

This is the heart of accretion disk turbulence.

B. Weak-field shearing instability: a simple treatment

It is desirable to have a less formal, more physical description of this far-reaching instability. We begin by noting that the fast mode is irrelevant here for the same reason that the ensuing turbulence is nearly incompressible. By using the Boussinesq approximation at the very start, we may eliminate magnetosonic modes from the analysis. This greatly simplifies matters.

The simplest fluid system displaying the instability is an axisymmetric gas disk in the presence of a weak vertical magnetic field. The field has no effect on the disk equilibrium, which is a balance of gravitational and rotational forces. If a fluid element is displaced from its circular orbit by an amount $\boldsymbol{\xi}$, with spatial dependence e^{ikz}, it is not difficult to show that the induction ("field-freezing") equation leads to

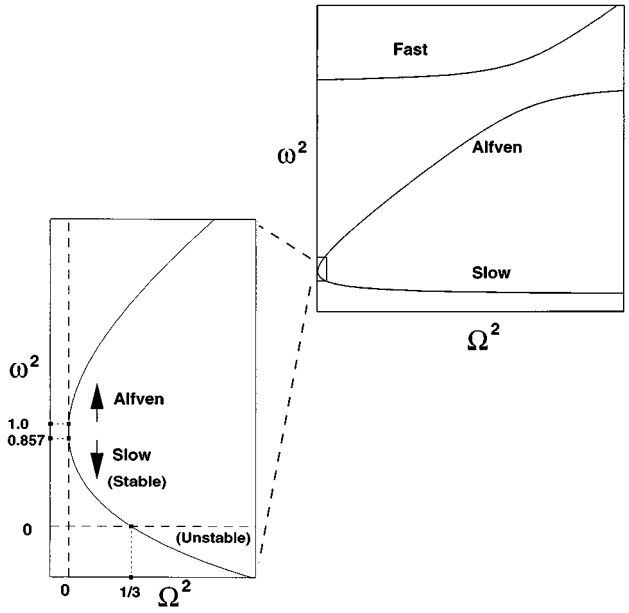

FIG. 15. Plot of ω^2 vs Ω^2 for a Keplerian disk. All frequencies are reckoned in units of $\mathbf{k}\cdot\mathbf{u_A}=1$. Numerical values correspond to the case $ku_{A\,\phi}=2$, $ka=5$. The expanded scale at the left shows the Alfvén and slow-wave branches. For $\Omega^2=0$, their degeneracy is broken only by finite compressibility effects. Note that the onset of instability ($\omega^2=0$) corresponds to $(\mathbf{k}\cdot\mathbf{u_A})^2=-d\Omega^2/d\ln R$.

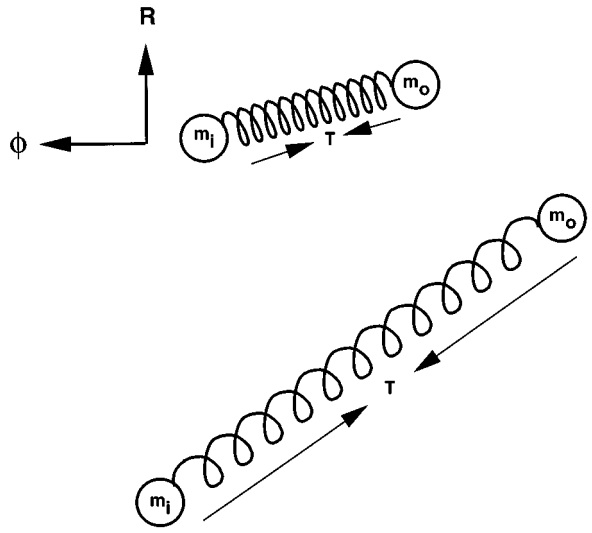

FIG. 16. Two masses in orbit connected by a weak spring. The spring exerts tension force T resulting in a transfer of angular momentum from the inner mass m_i to the outer mass m_o. If the spring is weak, the transfer results in an instability as m_i loses angular momentum, drops through more rapidly rotating inner orbits, and moves further ahead. The outer mass m_o gains angular momentum, moves through slower outer orbits, and drops further behind. The spring tension increases and the process runs away.

$$\delta\mathbf{B}=ikB\boldsymbol{\xi},\quad \delta B_Z=\xi_Z=0. \tag{104}$$

The magnetic tension force is then

$$\frac{ikB}{4\pi\rho}\,\delta\mathbf{B}=-(\mathbf{k}\cdot\mathbf{u_A})^2\boldsymbol{\xi}. \tag{105}$$

Since these incompressible planar displacements are pressure free, the equations of motion take on a very simple form,

$$\ddot{\xi}_R-2\Omega\dot{\xi}_\phi=-\left(\frac{d\Omega^2}{d\ln R}+(\mathbf{k}\cdot\mathbf{u_A})^2\right)\xi_R, \tag{106}$$

$$\ddot{\xi}_\phi+2\Omega\dot{\xi}_R=-(\mathbf{k}\cdot\mathbf{u_A})^2\xi_\phi. \tag{107}$$

As before, the 2Ω and $d\Omega^2/d\ln R$ terms represent Coriolis and tidal forces, respectively. These equations, which are the proper leading-order WKB equations for local fluid displacements in a magnetized disk, also happen to be the equations describing two orbiting mass points connected by a spring with spring constant $(\mathbf{k}\cdot\mathbf{u_A})^2$. Understanding the fate of slow-mode disturbances in weakly magnetized, differentially rotating systems amounts to nothing more than thinking about orbiting mass points connected by a spring (Balbus and Hawley, 1992a).

Consider the situation in Fig. 16. Two mass points, initially at the same orbital location, are displaced to two new orbits close by. Mass point m_i is orbiting at inner radius R_i, while mass point m_o orbits at outer radius R_o. They are connected by a massless spring. Assume, as would be the case for a Keplerian disk, that m_i orbits more rapidly than m_o. The string stretches and builds up tension T. T pulls backwards on m_i and forwards on m_o. Thus m_i continuously loses angular momentum, while m_o continuously gains the same. This means that m_i cannot remain in orbit R_i and must drop down to a yet lower orbit. Similarly, m_o acquires too much angular momentum to stay in orbit R_o and must move outwards. The separation widens, the spring stretches yet more, T goes up, and the process runs away. This is the essence of the weak-field instability in differentially rotating systems. Introducing other magnetic-field components does not change our argument, since by selecting $\mathbf{k}=k\mathbf{e_Z}$ we have ensured that only the vertical component will couple dynamically. For the same reason, it matters little that, if B_R is present, B_ϕ is increasing with time. Ultimately, we shall see that the instability is present for any field geometry regular enough to allow large-wavenumber disturbances.

It is of course crucial that the spring be weak. If the spring constant is strong enough that there are many vibrations in an orbital time, the stretching scenario we have outlined will not work. A glance at the right side of the radial equation (106) suggests the stability requirement

$$(\mathbf{k}\cdot\mathbf{u_A})^2>-\frac{d\Omega^2}{d\ln R} \tag{108}$$

as a condition for the effective spring constant, a result we confirm below [Eq. (111)]. [If we fix $\mathbf{k}\cdot\mathbf{u_A}$ and vary the magnitude of a Keplerian $\Omega(R)$, this agrees precisely with our earlier formal calculation.] But if one is free to choose small enough k, there will always be instability unless

$$\frac{d\Omega^2}{d\ln R}>0. \tag{109}$$

In astrophysical disks, this is almost never true, except possibly for isolated anomalous regions.

Just how large a wavelength is permitted? What restriction would such a maximum scale place on the field strength? In this example, we have chosen to focus on coupling to a vertical field. Thus at least one wavelength needs to fit within a disk thickness of twice the scale height $2H$. The stability requirement is then

$$u_A^2>-\frac{1}{k^2}\frac{d\Omega^2}{d\ln R}\sim-\frac{H^2}{\pi^2}\frac{d\Omega^2}{d\ln R}\sim\frac{6}{\pi^2}c_s^2, \tag{110}$$

assuming a Keplerian profile, and an isothermal Gaussian for the vertical scale height. In other words, the Alfvén speed must significantly exceed the sound speed, if all allowed wavelengths in a disk thickness are to be stable. "Weak field" means subthermal. If the magnetic field had no vertical component, in principle much longer wavelengths would be allowed and instability might be present for much larger field strengths (Papaloizou and Szuszkiewicz, 1992). But once the magnetic-field energy density is comparable to the thermal energy density, the dynamics qualitatively changes, including the nature of the disk equilibrium state. What is interesting is the opposite limit: in the regime in which dissipational length scales are much smaller than the smallest dynamical scales of interest ("ideal MHD"), the magnetic field is *never* too small to be dynamically ignored.

This last point runs deep and calls into question an entire school of MHD analysis, the so-called *kinematic limit*. In this asymptotic domain, thought to coincide with highly subthermal magnetic energy densities, the field is regarded as passive and Lorentz forces ignored. This is implicit in every textbook illustration of a disk wrapping up magnetic field lines and quite explicit in the most developed theories of dynamos (Moffatt, 1978). The problem is that, whereas a weak field is indeed negligible in a magnetosonic disturbance, it can never be negligible for Alfvénic and slow disturbances. These are internal degrees of freedom created by the magnetic field, with no hydrodynamic analog. For these disturbances, there is nothing for the field to be "weak" with respect to—it does not compete with anything. If pressure forces are not involved, the fact that the magnetic energy density is small compared with the thermal energy density is of little consequence.

The dispersion relation resulting from Eqs. (106) and (107) from assuming a $\boldsymbol{\xi}$ time dependence of $\exp(-i\omega t)$ is

$$\omega^4-\omega^2[\kappa^2+2(\mathbf{k}\cdot\mathbf{u_A})^2]+(\mathbf{k}\cdot\mathbf{u_A})^2\left((\mathbf{k}\cdot\mathbf{u_A})^2+\frac{d\Omega^2}{d\ln R}\right)=0, \tag{111}$$

which is precisely the $a^2\to\infty$ Boussinesq limit of Eq. (99). Note also that the stability condition (108) follows directly from Eq. (111).

Equation (111) is a simple quadratic in ω^2, and it is straightforward to show that there is a maximum unstable growth rate

$$|\omega_{max}|=\frac{1}{2}\left|\frac{d\Omega}{d\ln R}\right| \tag{112}$$

which occurs when

$$(\mathbf{k}\cdot\mathbf{u_A})^2_{max}=-\left(\frac{1}{4}+\frac{\kappa^2}{16\Omega^2}\right)\frac{d\Omega^2}{d\ln R}. \tag{113}$$

For a Keplerian rotation profile, these become

$$|\omega_{max}|=\frac{3}{4}\Omega,\quad (\mathbf{k}\cdot\mathbf{u_A})_{max}=\frac{\sqrt{15}}{4}\Omega. \tag{114}$$

This is an enormous growth rate. Unchecked, it would result in a factor of more than 10^4 amplification in energy per orbit. It has been conjectured (Balbus and Hawley, 1992a) that no linear instability feeding off the free energy of differential rotation can grow any faster. This appears to be true even when the disk is near the horizon of a Kerr black hole. In this case, the maximum growth rate is precisely the relativistic generalization of Eq. (112) (Gammie, 1997).

The destabilization of Couette flow by a vertical magnetic field was first analyzed by Velikhov (1959) and later extended to general rotation laws by Chandrasekhar (1960). These authors treated the problem globally and made a point of restricting the analysis to vertical fields only. A purely axial field configuration was deemed to be special because more general poloidal and toroidal field configurations lacked a well-defined equilibrium state and otherwise appeared to greatly complicate the analysis. Vertical fields, on the other hand, allowed for a rigorous global treatment.

While mathematically laudable, this approach had the unfortunate consequence of obscuring the underlying simplicity and generality of the weak-field limit (for which none of the above concerns is an issue) that is captured by Eqs. (106) and (107). As a result, despite the fundamental significance of Velikhov's pioneering study and its subsequent experimental verification by Donnelly and Ozima (1960), almost three and a half decades passed before the paper was first cited in the accretion disk literature.

C. Weak-field shearing instability: full generality

Because of the importance and wide scope of the instability, we now examine its behavior under very general circumstances. Our results will be applicable to pressure-supported rotating stars, non-Keplerian "thick" disks, and rotationally supported thin disks. The equilibrium state may be stratified in both radial and vertical directions. We allow the rotation profile Ω to be a function of both R and z, and displacements need not be confined to the planes of constant z. Our analysis will once again be local, so that leading-order WKB disturbances of the form $\exp i(k_R R+k_Z z-\omega t)$ will be assumed. As in the simple disk treatment given above, the insta-

bility has an evanescent, nonpropagating character when the magnetic field is weak, so that a local treatment makes physical sense. This approach is in complete accord with both numerical simulations, and more detailed global analyses.[5] Like the Rayleigh instability (associated with a decreasing angular momentum gradient), or the Schwarzschild instability (associated with a decreasing entropy gradient), the weak-field instability (associated with a decreasing angular velocity gradient) is insensitive to the global structure in which it is embedded. This, of course, is why all of these instabilities are so widely applicable.

In the Boussinesq limit, the linearized equations of motion are (following Balbus, 1995; see also Papaloizou and Szuszkiewicz, 1992)

$$k_R\delta u_R+k_Z\delta u_Z=0, \tag{115}$$

$$-i\omega\delta u_R+\frac{ik_R}{\rho}\delta P-2\Omega\delta u_\phi-\frac{\delta\rho}{\rho^2}\frac{\partial P}{\partial R}+\frac{ik_R}{4\pi\rho}(B_\phi\delta B_\phi+B_Z\delta B_Z)-\frac{ik_Z}{4\pi\rho}B_Z\delta B_R=0, \tag{116}$$

$$-i\omega\delta u_\phi+\delta u_R\frac{\kappa^2}{2\Omega}+\delta u_Z R\frac{\partial\Omega}{\partial z}-i\mathbf{k}\cdot\mathbf{B}\frac{\delta B_\phi}{4\pi\rho}=0, \tag{117}$$

$$-i\omega\delta u_Z+\frac{ik_Z\delta P}{\rho}-\frac{\delta\rho}{\rho^2}\frac{\partial P}{\partial z}+\frac{ik_Z}{4\pi\rho}\times(B_\phi\delta B_\phi+B_R\delta B_R)-\frac{ik_RB_R}{4\pi\rho}\delta B_Z=0, \tag{118}$$

$$-i\omega\delta B_R-i\mathbf{k}\cdot\mathbf{B}\delta u_R=0, \tag{119}$$

$$-i\omega\delta B_\phi-\delta B_R\frac{\partial\Omega}{\partial\ln R}-\delta B_Z R\frac{\partial\Omega}{\partial z}-i\mathbf{k}\cdot\mathbf{B}\delta u_\phi=0, \tag{120}$$

$$-i\omega\delta B_Z-i\mathbf{k}\cdot\mathbf{B}\delta u_Z=0. \tag{121}$$

To complete our set, for general axisymmetric disturbances we also require the entropy equation in the Boussinesq limit:

$$i\omega\frac{5}{3}\frac{\delta\rho}{\rho}+\delta u_Z\frac{\partial\ln P\rho^{-5/3}}{\partial z}+\delta u_R\frac{\partial\ln P\rho^{-5/3}}{\partial R}=0. \tag{122}$$

The reduction of this set of equations to a fourth-order dispersion relation is a lengthy but straightforward operation. It is convenient to introduce the $\mathcal{D}$ operator,

$$\mathcal{D}=\left(\frac{k_R}{k_Z}\frac{\partial}{\partial z}-\frac{\partial}{\partial R}\right), \tag{123}$$

and to set

$$k^2=k_R^2+k_Z^2,\quad \varpi^2=\omega^2-(\mathbf{k}\cdot\mathbf{u_A})^2. \tag{124}$$

Then the dispersion formula may be written as

$$\varpi^4\frac{k^2}{k_Z^2}+\varpi^2\left(\frac{3}{5\rho}(\mathcal{D}P)\mathcal{D}\ln P\rho^{-5/3}+\frac{1}{R^3}\mathcal{D}(R^4\Omega^4)\right)-4\Omega^2(\mathbf{k}\cdot\mathbf{u_A})^2=0. \tag{125}$$

When the limit $\mathbf{k}\cdot\mathbf{u_A}\to0$ is taken, Eq. (125) reduces to the hydrodynamics problem considered by Goldreich and Schubert (1967), in the adiabatic limit of their equation.

It is an elementary matter to show that Eq. (125) implies that ϖ^2, and hence ω^2, must be real. Stability may then be studied by noting conditions under which $\omega^2\to0$. In this limit, Eq. (125) becomes

$$(\mathbf{k}\cdot\mathbf{u_A})^2=\frac{k_Z^2}{k^2}\left(4\Omega^2+\frac{3}{5\rho}(\mathcal{D}P)\mathcal{D}\ln P\rho^{-5/3}+\frac{1}{R^3}\mathcal{D}(R^4\Omega^2)\right). \tag{126}$$

For stability, Eq. (126) cannot have any real solutions. In other words,

$$\left[4\Omega^2+\frac{3}{5\rho}(\mathcal{D}P)\mathcal{D}\ln P\rho^{-5/3}+\frac{1}{R^3}\mathcal{D}(R^4\Omega^2)\right]<0. \tag{127}$$

We have rather quickly arrived at an interesting juncture: our stability condition for a magnetized disk makes no reference to the magnetic field. In particular, this is *not* the stability condition for a hydrodynamic system, but there is no way to alter Eq. (127) as $B\to0$. This rather subtle point is addressed below.

Returning to our analysis, let us set $x=k_R/k_Z$. Then, replacing the $\mathcal{D}$ operator with its definition (123), simplifying and regrouping, we obtain

$$x^2N_Z^2+x\left[\frac{3}{5\rho}\left(\frac{\partial P}{\partial z}\frac{\partial\ln P\rho^{-5/3}}{\partial R}+\frac{\partial P}{\partial R}\frac{\partial\ln P\rho^{-5/3}}{\partial z}\right)-R\frac{\partial\Omega^2}{\partial z}\right]+N_R^2+\frac{\partial\Omega^2}{\partial\ln R}>0 \tag{128}$$

where

$$N_Z^2=-\frac{3}{5\rho}\frac{\partial P}{\partial z}\frac{\partial\ln P\rho^{-5/3}}{\partial z},\qquad N_R^2=-\frac{3}{5\rho}\frac{\partial P}{\partial R}\frac{\partial\ln P\rho^{-5/3}}{\partial R}. \tag{129}$$

The so-called Brunt-Väisälä oscillation frequency associated with convectively stable buoyant oscillations is

$$N^2=N_Z^2+N_R^2.$$

Two conditions will together ensure that the left-hand side of Eq. (128) is positive. First, it must be positive for some value of x. This is assured if

$$N^2+\frac{\partial\Omega^2}{\partial\ln R}>0, \tag{130}$$

[5]When the Alfvén speed becomes comparable to the rotation speed, the character of the instability can change, but at this field strength the disk is magnetically, not rotationally, driven. See the critique of Knobloch (1992) and the reply of Gammie and Balbus (1994).

since this condition means either very small or very large x^2 will make the quadratic positive. The second condition is that there be no real roots to the quadratic equation, which is simply a matter of restricting its discriminant to negative values. This is, once again, a rather lengthy exercise in algebraic manipulations. The task is simplified somewhat by noting the vorticity relation

$$R\frac{\partial\Omega^2}{\partial z}=\frac{1}{\rho^2}\left(\frac{\partial\rho}{\partial R}\frac{\partial P}{\partial z}-\frac{\partial P}{\partial R}\frac{\partial\rho}{\partial z}\right). \tag{131}$$

(We explicitly use the fact that the magnetic field is too weak to affect the equilibrium state of the disk.) The final result of the calculation is

$$\left(-\frac{\partial P}{\partial z}\right)\left(\frac{\partial\Omega^2}{\partial R}\frac{\partial \ln P\rho^{-5/3}}{\partial z}-\frac{\partial\Omega^2}{\partial z}\frac{\partial \ln P\rho^{-5/3}}{\partial R}\right)>0. \tag{132}$$

Equations (130) and (132) are extremely general, and may be applied to stars as well as to disks. They should be compared with the classical *Høiland criteria* (Tassoul, 1978) that obtain for hydrodynamic stability. With $l^2\equiv R^4\Omega^2$, these are

$$N^2+\frac{1}{R^3}\frac{\partial l^2}{\partial R}>0, \tag{133}$$

$$\left(-\frac{\partial P}{\partial z}\right)\left(\frac{\partial l^2}{\partial R}\frac{\partial \ln P\rho^{-5/3}}{\partial z}-\frac{\partial l^2}{\partial z}\frac{\partial \ln P\rho^{-5/3}}{\partial R}\right)>0. \tag{134}$$

Equations (133) and (134) are very similar to their magnetic counterparts (130) and (132), the only difference being that angular momentum gradients are replaced by angular velocity gradients. This is the signature characteristic of magnetic fields in rotating systems: angular momentum couplings are replaced by angular velocity couplings. This is evident in comparing the nonlinear Eqs. (59) and (86). It is also true if we compare the Rayleigh criterion $\kappa^2>0$ with Eq. (109), and it is once again evinced in the Høiland criteria. It should be noted that the primary application of the classical hydrodynamic Høiland criteria has been to stellar interiors. The presence of any magnetic field at all, however, renders these criteria all but useless. Given this finding and the ubiquity of astrophysical magnetic fields, the domain of applicability of the classical criteria is unclear.

Why do the Høiland criteria change in a mathematically discontinuous way when there is a small and quite continuous physical change in the system? The point is that a weak magnetic field enters the dynamics only via the combination $\mathbf{k}\cdot\mathbf{u_A}$. Just as the presence of any finite viscosity leads (in bounded flows) to a boundary layer inside of which dissipation is an order-of-unity effect, the presence of any finite magnetic field defines a characteristic wave-number scale $k\sim\Omega/u_A$ at which magnetic effects become important. Provided that this scale is not so tiny that microscopic resistivity is dominant, the ensuing magnetic tension forces can always be "tuned" through the proper choice of $\mathbf{k}$ [Eq. (113)] to the maximum growth rate (112). Further discussion of this point may be found in Balbus (1995).

D. Effects of resistivity and viscosity

Resistivity and viscosity dampen the growth of unstable modes. (Thermal diffusivity is less important, since the least stable modes do not have thermal gradients.) In this section we examine these dissipative effects.

Restricting the analysis to vertical wave numbers $\mathbf{k}=k\hat{\mathbf{e}}_\mathbf{z}$ is both physically sensible (since these are the most unstable modes) and mathematically simplest. Let us define, in this section only, $\sigma=-i\omega$, so that we work directly with the growth rate. Including the effects of a kinematic viscosity $\nu=\eta_V/\rho$, we obtain the linearized large-k radial and azimuthal equations of motion,

$$(\sigma+\nu k^2)\delta u_R-2\Omega\,\delta u_\phi-\frac{\delta\rho}{\rho^2}\frac{\partial P}{\partial R}-\frac{i\mathbf{k}\cdot\mathbf{B}}{4\pi\rho}\,\delta B_R=0, \tag{135}$$

$$(\sigma+\nu k^2)\delta u_\phi+\delta u_R\frac{\kappa^2}{2\Omega}-i\mathbf{k}\cdot\mathbf{B}\frac{\delta B_\phi}{4\pi\rho}=0. \tag{136}$$

Including the effects of resistivity, the radial and azimuthal induction equations are

$$(\sigma+\eta_B k^2)\delta B_R-i\mathbf{k}\cdot\mathbf{B}\delta u_R=0, \tag{137}$$

$$(\sigma+\eta_B k^2)\delta B_\phi-\delta B_R\frac{\partial\Omega}{\partial \ln R}-i\mathbf{k}\cdot\mathbf{B}\;\delta u_\phi=0. \tag{138}$$

These four equations lead to the dispersion relation

$$[\sigma^2+(\mathbf{k}\cdot\mathbf{u_A})^2(\eta_B+\nu)\sigma k^2]^2+\kappa^2[\sigma^2+(\mathbf{k}\cdot\mathbf{u_A})^2+2\eta_B\sigma k^2]-4\Omega^2(\mathbf{k}\cdot\mathbf{u_A})^2=0. \tag{139}$$

This reduces to Eq. (111), as it must, when $\eta_B=\nu=0$.

We introduce the dissipation rate γ by defining

$$\gamma=\sigma_0-\sigma, \tag{140}$$

where σ_0 is the value of σ in the absence of resistivity and viscosity. If $|\gamma/\sigma_0|\ll 1$, upon substituting for σ in the dispersion relation and retaining the leading-order terms, we find

$$\gamma=\eta_B k^2\frac{\kappa^2+(1+\mathcal{P})[\sigma_0^2+(\mathbf{k}\cdot\mathbf{u_A})^2]}{\kappa^2+2[\sigma_0^2+(\mathbf{k}\cdot\mathbf{u_A})^2]} \tag{141}$$

where the *magnetic Prandtl number* $\mathcal{P}$ is ν/η_B.[6] Using the Keplerian results for the most rapidly growing mode,

$$\sigma_0^2=\frac{9}{16}\Omega^2,\quad (\mathbf{k}\cdot\mathbf{u_A})^2_{max}=\frac{15}{16}\Omega^2, \tag{142}$$

we find that γ becomes

$$\gamma=\frac{5}{8}\eta_B k^2\left(1+\frac{3}{5}\mathcal{P}\right). \tag{143}$$

If we consult Spitzer (1962) for values of our (collisional) dissipation parameters in an ionized plasma, we find

[6]The term *Prandtl number* without the "magnetic" qualifier refers to the ratio of viscosity to thermal conductivity.

$$\frac{\eta_V}{\rho} \simeq 2.2\times 10^{-16}\frac{T^{5/2}}{\rho}\ \mathrm{cm^2\,s^{-1}}$$

and

$$\eta_B \simeq 5.1\times 10^{12} T^{-3/2}\ \mathrm{cm^2\,s^{-1}},$$

which implies a magnetic Prandtl number of

$$\mathcal{P} \equiv \frac{\eta_V}{\rho\,\eta_B} \simeq \left(\frac{T}{1.4\times 10^5}\right)^4 \left(\frac{10^{16}}{n_e}\right). \tag{144}$$

Note that $\mathcal{P}$ is extremely sensitive to the temperature T, a point to which we shall later return. We focus separately on the cases $\mathcal{P}\gg 1$ and $\mathcal{P}\ll 1$.

When $\mathcal{P}\ll 1$, the criterion for unfettered growth $\gamma/\sigma_0 \ll 1$ for the maximally unstable Keplerian mode becomes

$$\frac{5}{6}\frac{\eta_B k^2}{\Omega} \ll 1.$$

We take k from its value at the wave number of maximal growth, $(\mathbf{k}\cdot\mathbf{u_A})_{max}$. Then, using the Spitzer value for η_B, we may write the above inequality as a constraint on the plasma β parameter,

$$\beta \ll 6\times 10^7 (T_4)^{5/2} (R_{10})^{3/2} \left(\frac{M}{M_\odot}\right)^{-1/2}, \tag{145}$$

where $T_4 = T/10^4 K$, $R_{10} = R/10^{10}$ cm, and $M/M_\odot$ is the central mass in solar units. Thus, the magnetic field must be very weak in any ionized plasma before departures from the ideal MHD limit are important. As a practical matter, only protostellar disks are likely to be affected by finite resistivity (Blaes and Balbus, 1994; Gammie, 1996).

When $\mathcal{P}\gg 1$, viscosity dominates the dissipation in a collisional plasma. The damping rate is now given by $\gamma = (3/8)\nu k^2$. The growth condition $\gamma/\sigma_0 \ll 1$ becomes a constraint on the magnetic field:

$$B \gg 6\times 10^{-4}\ \mathrm{G}\,(T_4)^{5/4} (R_{10})^{-3/4} \left(\frac{M}{M_\odot}\right)^{1/4}. \tag{146}$$

This is a tiny field by either stellar or disk standards.

E. Low-ionization fluids

The weak-field magnetic instability will be important only if the fluid is sufficiently ionized that there is a good coupling between the disk mass and **B**. In some cases of interest, the disk may be predominantly neutral, and it may not be obvious that a magnetic treatment is justifiable. In this section, we quantify this condition under the assumption that the disk has three populations of particles: neutrals, ions, and electrons. Results are taken from Blaes and Balbus (1994).

Neutrals and ions interact predominantly via elastic scattering. The volume specific rate of momentum transfer from ions to neutrals is given by

$$\mathcal{F}_{in} = -\mu_{ni} n_n n_i \langle\sigma v\rangle_{ni} (\mathbf{v_i} - \mathbf{v_n}), \tag{147}$$

where μ_{ni} is the reduced mass of a two-body ion-neutral interaction,

$$\mu_{ni} = \frac{m_n m_i}{m_n + m_i},$$

$\langle\sigma v\rangle_{ni}$ is the momentum rate coefficient (angle brackets indicate averaging over the distribution of cross section×relative velocity), m_n and m_i are the neutral and ion masses, respectively, and n_n and n_i are their respective number densities. The quantity $\mathbf{v_i}-\mathbf{v_n}$ is the relative drift velocity between the two species. The momentum rate coefficient is approximately constant (σ decreases with increasing relative velocity), independent of the kinetic temperature of the populations (Draine, Roberge, and Dalgarno, 1983):

$$\langle\sigma v\rangle_{ni} = 1.9\times 10^{-9}\ \mathrm{cm^3\ s^{-1}}. \tag{148}$$

The main constituents of the neutral population in a cool astrophysical disk will be H_2 molecules and He atoms. Typically the He abundance is 0.1 of the H abundance by number, leading to an effective neutral mass of $m_n = 2.33 m_H$, where m_H is the atomic hydrogen mass. At the threshold of magnetic coupling, trace alkali species, in particular Na and K, are the most numerous ions. Accordingly, we follow Draine *et al.* (1983) and take $m_i \simeq 30 m_H$. The force per unit volume exerted upon the ions by collisions with neutrals is written

$$-\gamma\rho_i\rho_n(\mathbf{v_i}-\mathbf{v_n}) \tag{149}$$

which defines the drag coefficient γ:

$$\gamma = \frac{\langle\sigma v\rangle_{ni}}{m_n + m_i} = 3.5\times 10^{13}\ \mathrm{cm^3\,s^{-1}\,g^{-1}}. \tag{150}$$

Blaes and Balbus (1994) carried through a detailed stability analysis of a two-fluid system including coupling terms of the form of Eq. (149), assuming that ion fluid was perfectly conducting. Only displacements in the plane of the disk were considered, since these are generally the most unstable. In the simplest case, which is all that we shall examine here, the ions and neutrals exist as independently conserved fluids; ionization and recombination processes are ignored. This is sufficient to give us an order of magnitude estimate for the ionization fraction, n_i/n_n, at which magnetic coupling is important.

The two-fluid stability criterion can be cast in a form very reminiscent of Eq. (108), the simple "masses on a spring" criterion. Assuming that $n_n \gg n_i$, Blaes and Balbus show that an ion-neutral system is stable if

$$(\mathbf{k}\cdot\mathbf{u_A}^*)^2 > -\frac{d\Omega^2}{d\,\ln R}, \tag{151}$$

where the effective Alfvén velocity is given by

$$(u_A^*)^2 = \frac{B^2}{4\pi\rho_i}\times\frac{\kappa^2+\gamma^2\rho_i^2}{\kappa^2+\gamma^2\rho_i\rho}. \tag{152}$$

This has a readily interpretable form. The second factor falls from 1 to ρ_i/ρ as collisions become progressively more important. When collisions completely dominate,

the effect is to replace the ion density with the neutral density as the effective inertial mass associated with the Alfvén velocity. Whether collisions dominate or not depends upon the results of comparing the various collision frequencies with the fundamental epicyclic fluid response frequency. The problem is analogous to determining the "effective mass" of an electron traveling though a lattice (Goodstein, 1974); interactions with the background lattice can be subsumed by redefining a new mass parameter. In our problem, when collisions are of no importance, the effective Alfvén velocity is that of the ion fluid alone; when collisions dominate, the effective Alfvén velocity is that associated with the dominant neutral fluid serving as the inertial density. The transition from one asymptotic limit to the other is mediated continuously by varying the ion density.

In all cases of interest,

$$\gamma^2\rho_i\rho \gg \kappa^2.$$

This means that

$$(u_A^*)^2=\frac{B^2}{4\pi\rho_n}(1+\kappa^2/\gamma^2\rho_i^2). \tag{153}$$

Thus the disk is effectively magnetized with the dominant neutrals anchoring the field if

$$\rho_i\gamma \gg \kappa=\Omega, \tag{154}$$

with the latter equality holding for a Keplerian disk. The content of Eq. (154) is simply that the neutral-ion collision rate must greatly exceed the characteristic dynamic response frequency if a weakly ionized disk is to be magnetically coupled.

Very small ionization fractions will often be sufficient to magnetize a disk. To illustrate this, we take our numerical normalization factors from a well-known model of the solar nebula, the primordial disk of the solar system (Hayashi, 1981). Then Eq. (154) may be written as a minimum value for the ionization fraction

$$\frac{n_i}{n_n}=10^{-13}(1+m_n/m_i)\left(\frac{M}{M_\odot}\right)^{1/2}R_{AU}^{-3/2}\left(\frac{n_n}{10^{15}}\right)^{-1}, \tag{155}$$

where R_{AU} is the radius in astronomical units. The limit on the ion density itself, a number independent of the total disk density, is

$$n_i=10^2(1+m_n/m_i)\left(\frac{M}{M_\odot}\right)^{1/2}R_{AU}^{-3/2}. \tag{156}$$

At these low values of the ionization fraction, the electrical conductivity of the gas should also be examined (Umbayashi, 1983). It is, of course, no longer Coulombic. A similar value for the minimal ionization fraction is found (Gammie, 1996). In general, these ionization fractions are sufficiently small that, with the possible important exception of protostellar and protoplanetary disks, all gaseous astrophysical systems are likely to be magnetically well coupled.

F. Global modes and nonaxisymmetric disturbances

1. Global modes

As pointed out at the beginning of the last section, the weak-field instability is present when $d\Omega^2/dR<0$ and is both local and evanescent. In this respect it is similar to a simple convective instability (present when the local entropy gradient decreases outward) or the Rayleigh instability (present when the local angular momentum gradient decreases outward). In all of these cases there is little reason to doubt that the essence of the weak-field destabilization is captured by a local analysis; nevertheless, global analyses can be informative. In particular, they can pick up subtle interactions with boundaries or reveal the presence of potentially powerful instabilities with no local counterpart. The hydrodynamic instability of Papaloizou and Pringle (1984), which chiefly afflicts non-Keplerian disks (cf. Sec. III.E.1), is a well-known example of the latter. But because global analyses can be lengthy and quite technical, and because they tend also to be specialized to particular field configurations and boundary conditions, we shall not attempt to reproduce such calculations here. Rather, we shall concentrate upon and summarize the results of some illustrative cases.

Papaloizou and Szuszkiewicz (1992) approached the problem of magnetized disk stability via a variational global eigenmode method. They emphasized that this treatment was more rigorous than a local analysis. The price to be paid was complexity, and of necessity the approach must be somewhat more restrictive in the class of equilibrium solutions that may be treated. The powerful formalism can also make the underlying dynamics difficult to discern.

Papaloizou and Szuszkiewicz addressed the stability of isorotational field configurations, i.e., those characterized by constant Ω along lines of magnetic force. By generalizing Lagrangian techniques originally developed by Bernstein *et al.* (1958) and Frieman and Rotenberg (1960), Papaloizou and Szuszkiewicz found general agreement with the local analysis of Balbus and Hawley (1991) when the field was weak. When the field was strong, however the domain of instability might extend beyond the saturation level suggested by Eq. (110). Rather than the Alfvén speed's being limited to subthermal values for instability, Papaloizou and Szuszkiewicz concluded that values of u_A up to $\sim\sqrt{R\Omega c_s}$ would be unstable, a velocity considerably in excess of the sound speed.

Gammie and Balbus (1994) carried out an explicit eigenmode analysis of a similar system. These authors treated the radial structure locally (i.e., the only radial structure came from differential rotation), but treated the vertical structure globally, thereby assuring that the desired eigenfunctions would be one dimensional. When boundary conditions corresponding to field lines anchored in an external rigid conductor were applied, a new global instability was found. Instead of having the magnetic field unstably drain angular momentum from inner to outer fluid elements, the unstable mode had the

magnetic field unstably draining angular momentum from interior disk fluid elements to an angular momentum reservoir (the conductor) at large distances from the disk midplane. The magnetic tension term in the stability criterion for this truly global mode involved the product of the curvature of the azimuthal displacement eigenmode and the square of the Alfvén speed (the global analog of $k^2 u_A{}^2$). The eigenmode curvature was appreciably different from zero only within a disk scale $\sim c_s/\Omega$ of the midplane. Away from this region, displacements were nearly linear (because the magnetic fields were approximately forcefree) out to the conducting boundary. Effectively, this introduces a $1/R$ scale in the mode's first derivative, and a $1/RH$ scale in its curvature. With $k^2 \sim 1/RH$, the form of the Papaloizou and Szuszkiewicz global criterion is recovered by invoking the same type of magnetic tension argument used in the local analysis.

We have elected to discuss this particular example in some detail because it nicely illustrates the relation between global and local modes for the magnetized disk problem. When global and local approaches have produced seemingly incompatible results, the influence of the boundaries is generally at the heart of the matter.

Because larger field strengths have larger associated length scales, global modes are most useful for understanding the stability of strong-magnetic-field configurations. Of course, such fields need not be unstable for their influence on disk transport to be felt: $-\langle \rho B_R B_\phi \rangle$ stresses directly provoke an outward angular momentum flux. Furthermore, the initial conditions required of a global mode analysis are of necessity rather artificial: isorotational (always) and forcefree (often). This limits their range of applicability. In addition, it is well to emphasize that if a disk is not born with a strong magnetic field, then it must be grown from seed. Under these circumstances, it is likely that it will be the nonlinear resolution of the weak-field instability that determines the magnetic properties and the disk behavior, not the stability of an ordered large-scale field.

2. Nonaxisymmetric disturbances

The nonaxisymmetric behavior of the weak-field instability is of interest because the only couple to the toroidal field component comes from ϕ-dependent (plane-wave) disturbances. Futhermore, dynamo amplification is not possible under conditions of axisymmetry (Moffatt, 1978), so the behavior of nonaxisymmetric perturbations is important to establish as a first step toward understanding how magnetic fields might be amplified by the nonlinear phase of the instability.

The presence of shear complicates the interpretation of the local behavior of nonaxisymmetric disturbances. This point is a kinematical one and must be faced in both magnetic and nonmagnetic problems alike. When one speaks of a "local" disturbance, this adjective refers to the unperturbed flow, not to the fixed Eulerian space-time coordinates. Plane waves of the form $\exp i(\mathbf{k}\cdot\mathbf{r}-\omega t)$ are no longer local eigenfunctions when $\mathbf{k}$ has a ϕ component; the wave number itself may depend upon time in fixed coordinates. A familiar example of a time-dependent wave number in an axisymmetric problem is the behavior of density perturbations in a homogeneously expanding Friedmann cosmology (Weinberg, 1972). One adopts comoving coordinates, tailored to remain fixed on elements moving with the large-scale Hubble flow. As is well known, the temporal behavior of comoving density instabilities does not, in general, show exponential growth. The precise form depends upon the nature of the fluid and the cosmological model; for a pressureless gas in a critically expanding universe, a simple 2/3 power law is obtained.

The local unperturbed behavior of a disk is not expansion, but shear, and the comoving Lagrangian coordinates are thus somewhat more complicated than their cosmological counterparts (Goldreich and Lynden-Bell, 1965; Balbus and Hawley, 1992b). For example, rather than simply diminishing with time as space expands, a radial wave number k_R behaves as

$$k_R(t) = k_R(0) - mt\,\frac{d\Omega}{dR}\,,$$

where $m = k_\phi R$ is the usual azimuthal wave-number variable. When $m = 0$ the wave number remains fixed, of course. But when m is finite, a local "leading" disturbance is first unwound by the shear ($|k_R|$ decreases and passes through zero) and then wrapped up as k_R increases linearly with time. The evolution of the disturbance can be very different in the large and small wave-number regimes.

The evolutionary equations for local disk perturbations in arbitrary magnetic-field geometries, incorporating the winding of radial wave numbers, were derived and studied by Balbus and Hawley (1992b). The question of local stability may be treated as an initial-value problem. Despite the presence of wave-number winding, the Alfvén coupling parameter $\mathbf{k}\cdot\mathbf{B}$ is constant with time. The effect of the shear is to force B_ϕ to evolve linearly with time if B_R is nonvanishing:

$$B_\phi(t) = B_\phi(0) + tB_R\frac{d\Omega}{d\,\ln R}$$

where $B_\phi(0)$ is the initial azimuthal field component. The shear terms cancel in the dot product $\mathbf{k}\cdot\mathbf{B}$, and the magnetic tension remains constant.

A relatively simple form of the evolutionary equation holds when three conditions are met: the winding time of k_R is long compared with the rotation period, $m \ll k_Z R$, and vertical disk structure is ignored. Then, the evolution of δB_R is given by

$$\left[\frac{k^2}{k_Z^2}D^4 + \kappa^2 D^2 - 4\Omega^2(\mathbf{k}\cdot\mathbf{u_A})^2\right]\delta B_R = 0 \qquad (157)$$

where

$$D^2 = \frac{d^2}{dt^2} + (\mathbf{k}\cdot\mathbf{u_A})^2, \quad k^2 = k_R(t)^2 + m^2/R^2 + k_Z^2 \qquad (158)$$

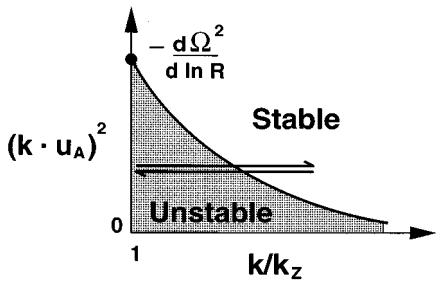

FIG. 17. Region of stability and instability in the $(\mathbf{k}\cdot\mathbf{u}_\mathbf{A})^2, k/k_z$ plane. The quantity $\mathbf{k}\cdot\mathbf{u}_\mathbf{A}$ is constant, even for a shearing wave vector. For a leading disturbance, k/k_z is initially large, and the point defining the wave vector moves leftward in the plane. The minimum value attained by k/k_z is unity, after which the wave vector becomes trailing and retraces its path to the right. If $(\mathbf{k}\cdot\mathbf{u}_\mathbf{A})^2 < -d\Omega^2/d \ln R$, then a finite portion of time is spent in the unstable region, and substantial growth may occur. From Balbus and Hawley, 1992b.

and δB_R is the linear perturbation in the radial field. This differential equation, which is very similar to the axisymmetric dispersion formula (111) (it clearly reduces to it when $m=0$) may be solved using WKB methods. The gist of the result is easily grasped by referring to the $(\mathbf{k}\cdot\mathbf{u}_\mathbf{A})$-$(k/k_Z)$ plane in Fig. 17. Start with a point in this plane whose coordinates $(k/k_Z, \mathbf{k}\cdot\mathbf{u}_\mathbf{A})$ correspond to a "leading" disturbance—one in which the shear causes k/k_Z to decrease with time, reach a value near unity, and then increase again. The effect is that the system evolves from a region of stability into a zone of instability and then once again a stable zone. The demarcation between stable and unstable regions is just the boundary between stability and instability found in an axisymmetric problem.

Balbus and Hawley (1992b) studied the stability behavior for poloidal wave numbers k_p of order Ω/u_A, which in the axisymmetric problem is the domain of maximum growth. In the presence of a poloidal field, rapid growth to enormous amplification factors ($\sim 10^{18}$ and greater) was found. When the radial wave number became sufficiently large, the unstable domain was left and amplitudes remained steady. Though not strictly an instability (because the linear phase is not marked by unconstrained growth), the manifestation of this behavior in numerical simulations (Hawley, Gammie, and Balbus, 1995) makes it indistinguishable from the axisymmetric instability.

Nonaxisymmetric wave numbers additionally allow for the presence of a dynamic Alfvén coupling $(\mathbf{k}\cdot\mathbf{u}_\mathbf{A})$ when there is no poloidal field, only a weak toroidal field. There has been some confusion over characteristic growth rates, however. In the wave-number regime $(k_p \sim \Omega/u_A)$, Balbus and Hawley (1992b) found growth rates of several percent of Ω. Hawley *et al.* (1995) numerically confirmed the instability of a purely toroidal field, but found a more rapid maximal growth rate. A subsequent analytic study by Terquem and Papaloizou (1996), based on what is in essence a large-wave-number WKB approximation, also found growth rates an order of magnitude greater than the original analysis of Balbus and Hawley (1992b).

In fact, the maximum growth rate for a weak toroidal field is $(1/2)|d\Omega/d \ln R|$, precisely its value for axisymmetric disturbances in the presence of a poloidal field. But it is reached only at much larger poloidal wave numbers. This maximum growth rate is reached in the limit of large m, larger $k_R R$, and even larger k_Z. In this limit, to avoid rapid winding, one must have

$$k_Z \gg k_R \sim k_R(0) \gg m \frac{d \ln \Omega}{dR} \sim \frac{\Omega}{u_A} \frac{d \ln \Omega}{d \ln R} \gg \frac{1}{H}.$$

In this wave-number domain, the geometry of the disturbances is essentially axisymmetric; the presence of m would be quite ignorable were it not for the fact that it is the only way the perturbations can couple to the magnetic field. Equation (157) goes over to an axisymmetric form in this limit. When the above inequality is satisfied, and

$$m u_{A\phi} = \sqrt{\frac{15}{4}}\, R\Omega,$$

the resulting growth rate approaches the axisymmetric maximum. One should note, however, that this is a singular field geometry, and everything is extremely sensitive to the assumption of a vanishing poloidal field component. The poloidal wave numbers are huge: m is large (if the field is weak), $Rk_R(0)$ is an asymptotic order larger, and Rk_Z yet an asymptotic order larger than that. A wisp of vertical field will result in enormous tension forces, which completely alter the character of the problem.

It is occasionally argued in the literature that because the toroidal field component is likely to be the largest in a disk, neglect of the poloidal field components is a good approximation. Applied to a magnetosonic wave, this reasoning would be correct; applied to the weak-field instability, it is not. The presence of a subthermal toroidal field, even if it is larger than the poloidal components, is a matter of indifference for the most rapidly growing axisymmetric instabilities. But the presence of very modest poloidal field components would cause a complete breakdown for the most rapidly growing local modes emerging from a toroidal field analysis: the radial and vertical wave numbers are, respectively, one and two asymptotic orders larger than the already large azimuthal wave number. Poloidal Alfvénic couplings would take over even for very small poloidal fields. It is likely, therefore, that direct physical applications of purely toroidal weak-field phenomena are quite limited.

Finally, we may note that the presence of shear means that a true mathematical instability of a nonaxisymmetric disturbance—one whose linear growth is unbounded as $t \to \infty$—must be sought by global methods. Many papers have explored this technical point with some care. For a sample of the variety of approaches and degrees of rigor, the reader may consult Curry and Pudritz (1996), Matsumoto and Tajima (1995), Ogilvie and Pringle (1996), and Terquem and Papaloizou (1996).

V. NUMERICAL SIMULATIONS OF MAGNETIZED ACCRETION DISKS

We have seen from the linear perturbation analysis that differentially rotating systems are unstable to a powerful, local MHD instability. Further progress requires studying the nonlinear consequences of this instability; for this we must appeal to numerical simulations. In this section we shall review some of the findings obtained to date from numerical MHD simulations of accretion disks.

A. Two-dimensional simulations: streaming or decay

Analytic and numerical techniques are sometimes regarded as reigning over nonintersecting domains—"solvable" and "intractable." Numerical stability studies, however, are most compelling when there is contact with analytic results. It is important to understand that both approaches generally involve substantial approximations, and thus each will always have its share of controversies. But if afforded a stature in the theorist's toolbox comparable with its analytic complement, numerical simulations can now provide insights, suggest pathways, and, most importantly, confirm qualitative ideas, all in ways that a few years ago were not possible. Magnetohydrodynamic accretion disk studies have benefited enormously.

In a paper published as a companion to their linear stability analysis, Hawley and Balbus (1991) carried out a series of local simulations of a simple vertical and helical field configuration and compared the results with predictions. Although idealized, these first simulations provided important confirmations: the mode growth rates, dispersion relation, qualitative content (e.g., the indifference of the instability to an azimuthal field), and stability criterion were all in accordance with linear theory. The simulations were also able to treat more complicated loop field geometries. Cylindrical, not local Cartesian, coordinates were used, which included the curvature terms dropped from the linear WKB analysis. Both the local and linear assumptions used in the analysis were in complete agreement with the simulations, which left no question that a combination of Keplerian rotation and weak magnetic fields was extraordinarily unstable.

The 1991 simulations illustrate the behavior of the instability very clearly (see Fig. 18). In fact, the striking clarity and simplicity of this image has created a curious problem: it has been too successful. Its appearance is so compelling, it has led to the notion that the creation of radial field by stretching out a vertical field is the most important thing the instability does. In fact, all geometries are unstable (including those with zero vertical field), and the most general outcome of the weak-field instability is turbulence, not coherent fluid motions. As we shall see, this is true even when nonlinear coherent flow is an exact nonlinear solution to the equations of motion.

Long-term local simulations require using the shearing box technique, as described earlier in the hydrodynamics section, Sec. III.D. The first implementation of this technique was two dimensional and axisymmetric (Hawley and Balbus, 1992). Note that the assumption of axisymmetry removes the time dependence of the radial boundary condition and the associated cumbersome grid interpolation. These simulations examined various initial vertical-field (B_Z) and radial-field (B_x) configurations and were the first to illustrate the evolution of the instability in the absence of a vertical field. If $B_Z=0$, an axisymmetric instability requires the presence of a radial field, and thus an azimuthal field will be generated by the shear. An unambiguous prediction of the linear analysis was that the presence of a (possibly time-dependent) weak B_ϕ component would make no difference to the early onset and growth of the instability. From the general dispersion formula (125) it is straightforward to show that the maximum growth rate is obtained for $k_Z\rightarrow\infty$. As can be seen in Fig. 19, a meridional slice of the disk, this leads to long flat structures, which are snakelike viewed in cross section. Most importantly, the indifference of the linear instability to the presence of B_ϕ was fully confirmed. The two-dimensional shearing box simulations produced a surprising result. When the initial magnetic-field configuration is a uniform vertical field, the resulting evolution produces two exponentially growing inward- and outward-flowing streams on the largest available scale (Fig. 20). In fact, the presence of *any* mean vertical field in the initial conditions, no matter how otherwise irregular the field geometry, always leads eventually to the appearance of these channels. The solutions remain co-

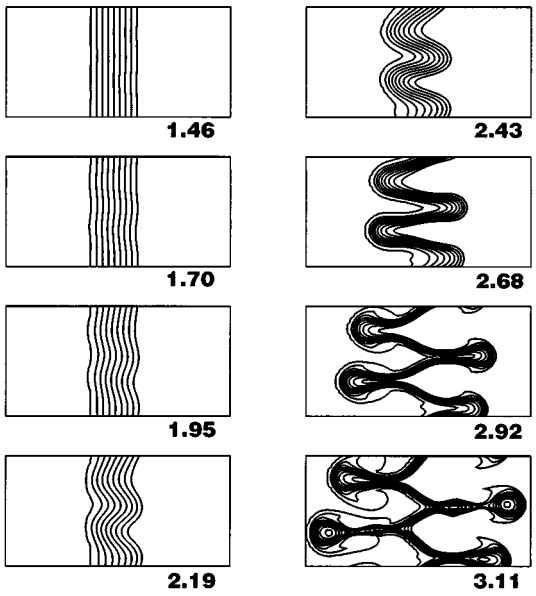

FIG. 18. Evolution of the vertical-field instability. The figure is an (R,z) cross section of field lines in an axisymmetric simulation of Keplerian differential rotation. R increases from left to right; z increases from bottom to top. The numbers give the time in orbits, $2\pi/\Omega$, where Ω is the angular frequency at the center of the computational domain.

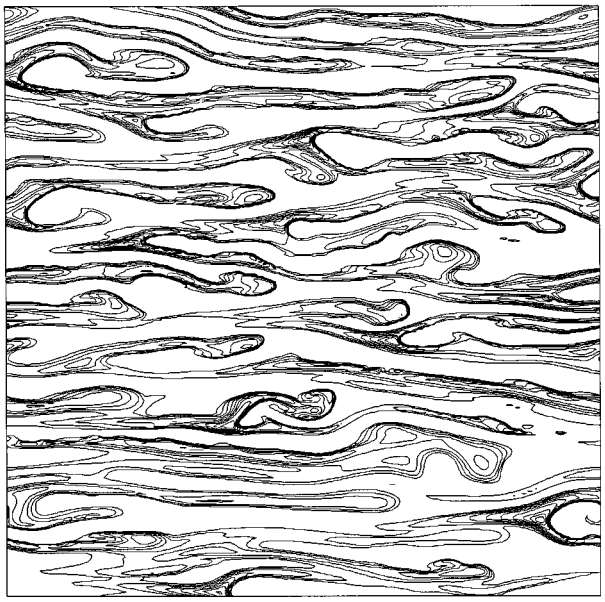

FIG. 19. Contours of angular momentum perturbations in a simulation with an initial radial field, viewed in (R,z) cross section. The axes are oriented as in Fig. 18. The long radial wavelength and short vertical wavelength are characteristic of the most unstable modes of a radial background magnetic field. From Hawley and Balbus 1992.

herent well into the nonlinear regime and look nothing like turbulence. On the other hand, when the average value of the poloidal field over the computational box is zero, the flow quickly breaks down into turbulence. The

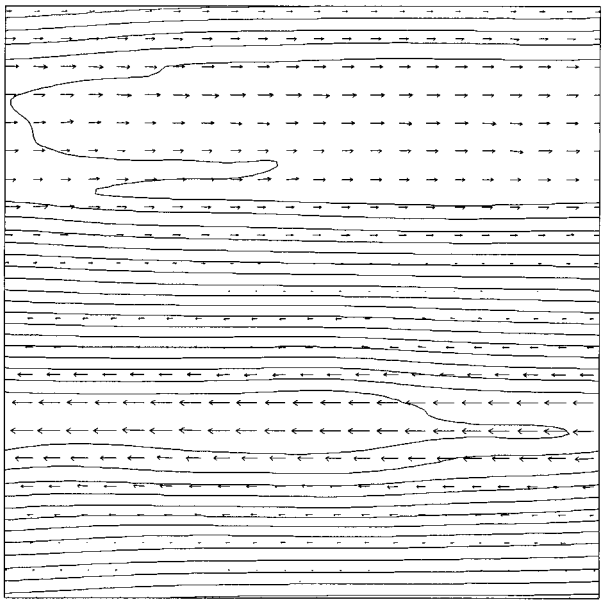

FIG. 20. Magnetic-field lines (solid curves) and velocity vectors (arrows) in a simulation of a uniform initial vertical field, viewed in (R,z) cross section. The axes are oriented as in Fig. 18. The flow evolves to two rapidly flowing channels. In cross section, they appear as oppositely moving radial streams. From Hawley and Balbus, 1992.

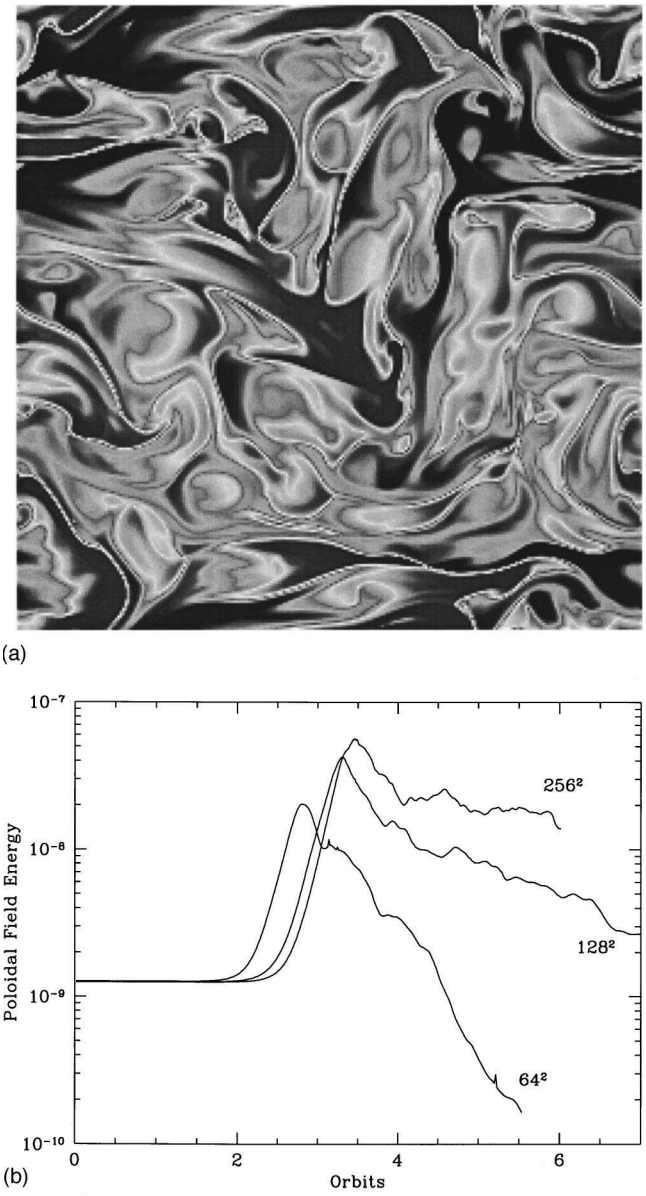

FIG. 21. Two dimensional MHD turbulence simulation: (a) Grey-scale plot of angular momentum perturbations in an axisymmetric simulation of an initial vertical field with $\langle B_Z\rangle=0$, viewed in (R,z) cross section. The axes are oriented as in Fig. 18. This field configuration does not lead to streams (cf. Fig. 20); instead, the flow becomes turbulent. (b) The time evolution of poloidal magnetic-field energy in axisymmetric simulations of an initial vertical field with $\langle B_Z\rangle=0$. Labels correspond to number of grid zones. After an initial period of growth, the magnetic field declines with time at a rate determined by the numerical resolution. This behavior accords with Cowling's anti-dynamo theorem.

turbulence does not persist, however; it decays over a period of many orbits (Fig. 21).

Let us consider the latter result first, as it is more easily understood. It is in fact a consequence of the "anti-dynamo theorem" (Moffatt, 1978; the theorem is due to Cowling). Simply stated, any sort of sustained magnetic-field amplification by axisymmetric turbulence in an isolated dissipative system is impossible. To see

this, note that the poloidal component of the magnetic field in an axisymmetric flow is derivable from the azimuthal component alone of a vector potential—in effect a scalar potential. Denoting this potential as A, its induction equation in the ideal MHD limit is (Moffatt, 1978)

$$\frac{\partial A}{\partial t}+v_i\frac{\partial A}{\partial x_i}=0 \tag{159}$$

where, as usual, the (locally Cartesian) i components are summed over by the repeated index convention. If $\nabla\cdot\mathbf{v}=0$, the divergence theorem implies

$$\frac{\partial}{\partial t}\int A^2\,dV=-\int A^2\mathbf{v}\cdot d\mathbf{S}, \tag{160}$$

where the expression on the left is a volume integral and the expression on the right is a surface integral. The heart of Cowling's anti-dynamo theorem is the observation that the surface integral will become arbitrarily small at large distances or vanish outright for periodic boundary conditions. The volume integral cannot, therefore, change with time; moreover, the presence of any (hitherto neglected) dissipation will cause it to decay to zero.

Simulations characterized by vanishing mean poloidal field components are also characterized by periodic boundary conditions for these quantities. The eventual outcome is therefore in the hands of the anti-dynamo theorem: decay (caused by grid losses in the simulations) is inevitable. Growth in the field energy B^2 is possible only by bringing fluid elements with markedly different values of A (a fluid element label) close together. This generates large A gradients, which are the source of B. This kneading of the fluid would have to continue to finer and finer scales to maintain the growth of B^2, but the dissipation scale is soon reached, and that is the end.

On the other hand, the presence of a nonvanishing vertical B field means that A has a piece proportional to x. Periodic boundary conditions on A now no longer hold and the surface integral need not vanish. Indeed, a steady throughput across the computational box would bring in fluid with ever larger $|A|$, in accord with what is found numerically. The magnetic field grows rapidly with time because the A potential in the box does. There is nothing unphysical about this setup. A real disk might well have a magnetic field threading it, connecting to a central star or to an ambient medium. The question then arises whether such disks really would break apart into outgoing and incoming channels.

The nature of these "channel solutions" was fully clarified by Goodman and Xu (1994), who pointed out a remarkable fact: starting with a vertical field, the linear plane-wave eigensolutions to Eqs. (106) and (107) are in fact exact nonlinear solutions of Eqs. (78)–(80). In the local approximation, the gas really does appear to act like two orbiting masses connected by a spring stretching out indefinitely. But it is well known that the behavior of a magnetized gas in three dimensions can be qualitatively different from its behavior in two dimensions. Will the streams remain stable in three dimensions? Goodman and Xu noted that this ostensibly nonlinear question actually reduces to a *linear* stability problem, but a linear stability problem perturbed about a most unusual equilibrium solution.

The Goodman and Xu analysis is a *tour de force* of technique. The presence of vertically periodic velocity streams renders the problem amenable to Floquet analysis, an approach that will be familiar to readers who have used the Bloch wave formalism in the study of crystal lattices. A detailed presentation of Goodman and Xu's analysis would unfortunately take us too far afield, but the interested and mathematically inclined reader would do well to study the original.

Goodman and Xu concluded that the new equilibrium of streaming motions should be unstable. Streaming instabilities, which feed off of the initial weak-field instability, were dubbed by the authors as "parasitic instabilities." The most important instability is a magnetized Kelvin-Helmholtz mode, which sets in for perturbation radial wavelengths in excess of the streaming equilibrium flow's vertical wavelength. Other parasitic instabilities were found, but their physical interpretation is less clear. To determine the ultimate nonlinear fate of the channels requires therefore a three-dimensional numerical simulation, whose detailed discussion we defer until the next section.

The existence and persistence in two dimensions of coherent disk streams provides an explanation for a recurring puzzling behavior seen in a number of earlier axisymmetric global MHD disk simulations. Uchida and Shibata (1985) and Shibata and Uchida (1986) sought to understand the creation of MHD jets and investigated this problem by threading a disk of gas with a vertical magnetic field. By imparting less than the Keplerian value of angular momentum to the orbiting fluid elements, they sought to mimic the slow radial infall of accretion—but without turbulence. The infall produced radial fields that became wrapped up by differential rotation into strong toroidal fields, whose gradients in turn drove dynamic outflows along the vertical-field lines. These simulations were the first time-dependent demonstration of the efficacy of magnetic fields for jet acceleration and collimation. Some of these simulations, however, began with a Keplerian disk embedded in a vertical magnetic field. Such disks collapsed on a dynamic time scale. At the time, the reason for this was not at all clear.

Nearly a decade later, Stone and Norman (1994) revisited and extended Uchida and Shibata's work. In the interim, the magnetorotational instability had been elucidated, and it became possible to understand the cause of the collapse of the Keplerian disks. Once again, Stone and Norman simulated a fully axisymmetric Keplerian disk penetrated by a weak vertical magnetic field and observed rapid collapse and infall. This configuration is the global analog of the Hawley and Balbus (1992) simulations that led to channel solutions, and the collapse of the disk simply represents their reappearance in a new guise. In general, the largest rapidly growing wave-

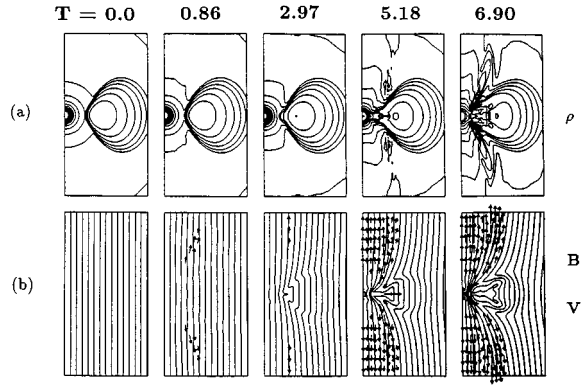

FIG. 22. Global evolution of an axisymmetric thick disk threaded by a uniform vertical field at time $T=0$, viewed in (R,z) cross section. The axes are oriented as in Fig. 18. The top row (a) shows the density; the bottom row (b) shows the magnetic-field lines (solid curves) and velocity vectors (arrows). The columns are labeled by orbital time. The weak-field instability leads promptly to rapid accretion. From Matsumoto *et al.*, 1996.

lengths are to be found in the low-density surface region of a disk. These (top and bottom) surface layers quickly slough inwards, their angular momentum feeding into the high-density (smaller rapidly growing wavelengths) disk interior, which responds by "squirting" outwards. The same behavior was also observed (Matsumoto *et al.*, 1996) in a simulation of a thick torus embedded in a weak vertical magnetic field (Fig. 22).

Stone and Norman (1994) also explored the consequences of strong ($\beta<1$) magnetic fields for disk accretion. They found that such fields also lead to inflow on orbital time scales, via the process of magnetic braking. In this case, the disk angular momentum is siphoned to an external wind (Blandford and Payne, 1982; Wardle and Königl, 1993), not diffused outwards within the disk. Turbulent transport by weak magnetic fields and removal of angular momentum by strong magnetic fields need not be mutually exclusive processes. They may both be present at different evolutionary stages or in different physical regions of the same disk. An obvious way this could emerge is that the high density near the disk midplane leaves the field subdominant (and therefore unstable), while the low-density upper layers are dominated by a relatively strong field and form the base of a wind. It also true that, depending upon the circumstances of their birth, different disks may be entirely dominated by one or the other mechanism. Nature herself is clearly fond of disk jets and winds, for they are a ubiquitous phenomenon (Begelman, Blandford, and Rees, 1984; Lada, 1985; Lin and Papaloizou, 1996). We must, unfortunately, leave this vast and important topic essentially untouched and return to the focus of this article, MHD turbulence.

B. Turbulent MHD transport in three dimensions

1. A brief survey

To pursue the fate of the accretion streams in the channel solution or to study dynamo amplification requires implementation of three-dimensional MHD codes. In recent years, several local shearing box studies have been carried out and a wide variety of models explored. The simplest consists of a homogeneous box, in which only the radial component of the large-scale gravitational field is retained, and the magnetic field is initially uniform (Hawley *et al.*, 1995; Matsumoto and Tajima, 1995). The initial field geometry in these studies had some combination of vertical and toroidal components. A more complicated initial field configuration, important for understanding dynamo activity, is achieved by letting the initial field have a random character with vanishing mean (Hawley, Gammie, and Balbus, 1996). The next level of complexity is to allow the presence of the vertical component of the gravitational field with its attendant density stratification (Brandenburg *et al.*, 1995; Stone *et al.*, 1996; Matsuzaki *et al.*, 1997). This introduces the possibility of magnetic buoyancy, a qualitatively new effect, and by bringing the pseudoscalar quantity $\Omega\cdot\nabla\rho$ into the problem it also breaks chiral symmetry. The resulting handedness of the turbulent flow is potentially important for the development of local mean helicity, a feature upon which much of classical dynamo theory is based (Moffatt, 1978). The initial magnetic-field geometry of these stratified simulations includes radially sinusoidally varying vertical field (Brandenburg *et al.*, 1995; Stone *et al.*, 1996), and pure toroidal field (Stone *et al.*, 1996; Matsuzaki *et al.*, 1997).

2. Numerical technique

To develop a numerical consensus for a problem involving a process as complex as three-dimensional MHD turbulence, it is extremely important that different groups employ codes based on different numerical techniques. Such codes are never straightforward to implement, and each brings with it its own advantages and compromises. While it is not possible here to describe in detail each of the techniques that have been used, some discussion of code particulars is in order. Of necessity, our comments will be brief and contain jargon that may be unfamiliar to readers without numerical experience, and they must be restricted in scope to the techniques used by our own codes.

Our code (Hawley and Stone, 1995) solves the equations of compressible MHD through operator-split, time-explicit, piecewise-linear monotonic finite differencing in a three-dimensional, locally Cartesian accretion disk system. The simulations are single-fluid, use an adiabatic or isothermal equation of state, and assume infinite conductivity; any of these assumptions can be relaxed, if required. The algorithms employed are very similar to those used in the well-known ZEUS-3D astrophysics code (Stone and Norman, 1992a, 1992b). The hydrodynamic techniques have been widely applied in previous studies and are described in Stone and Norman (1992a). The MHD portion of the algorithm uses the constrained transport (CT) approach of Evans and Hawley (1988), which directly evolves the magnetic-field components while preserving the constraint $\nabla\cdot\mathbf{B}=0$. The

CT framework is a prescription for locating fields and their derived electromotive forces on the grid to ensure that the Stokes theorem is satisfied numerically to machine accuracy for all curvilinear coordinate systems. CT permits wide latitude in the way that time-averaged electromotive forces are obtained, and many choices for the electromotive forces lead to highly dispersive (and hence unsatisfactory) transverse (Alfvén) wave evolution. The procedure currently in use takes information propagated along Alfvén characteristics to solve a restricted set of characteristic equations for time-advanced fields and electromotive forces. This is known as the method of characteristics constrained transport (MOCCT) algorithm, and its first implementation is described in detail in Stone and Norman (1992b). Recent algorithm improvements, along with a discussion of MOCCT's strengths and weaknesses, are presented in Hawley and Stone (1995).

There are several sources of dissipation in our code, all of which, unfortunately, tend to be referred to as a type of viscosity. There is "numerical viscosity," better referred to as diffusion, which is not an explicit term in any equation, but is present due to finite gridding effects. Numerical diffusion was briefly discussed in Sec. III.D.2. "Artificial viscosity" is explicitly added to the equations of motion, but it does not take the form of a true, Navier-Stokes viscous stress tensor. Rather, its form is based on that proposed by von Neumann and Richtmyer (1950), which is in essence an anisotropic pressure tensor. But in this viscous pressure, the role of thermal velocity is played by a velocity gradient multiplied by a characteristic length. The artificial viscosity is constructed so that it is unimportant except as a means to resolve shocks. Its properties are closer to bulk viscosity than to shear viscosity, but it is equivalent to neither. Artificial viscosity serves a dual role: it dissipates high-frequency numerical (unphysical) noise and it provides a route for the increase in entropy accompanying irreversible processes (i.e., shocks). A more detailed description of the artificial viscosity term can be found in Stone and Norman (1992a). No other form of viscosity is added to the equations of motion.

Generally, an explicit resistivity can be included in any of the MHD codes commonly used. Even without such a term, however, the magnetic field is subject to dissipation and reconnection through numerical truncation. In large part this is due to the mutual cancellation of oppositely directed magnetic-field components when advected into a single zone. Since there is no associated heating in this "grid resistivity," energy is lost from the system. Because these losses occur at the grid scale, when MHD turbulence plays a significant role in the disk dynamics there can be no formal convergence of the solution with increasing resolution. Such convergence could be achieved by introducing a sufficiently large resistivity and viscosity that the dissipation length scales are fully resolved. However, given the dynamic range limitations of a three-dimensional computation, this approach is generally impractical: large scales in our simulations would be directly affected along with the small scales. Instead, we allow grid-based reconnection, but calibrate its effects through resolution studies and (most importantly) by checking the results against other simulations using different numerical techniques or explicit dissipation.

Code performance has been tested using a variety of problems, including Alfvén wave propagation, MHD solar wind problems, MHD shock tubes, and the evolution of a two-dimensional coronal mass-ejection transient (Stone *et al.*, 1992). Together, these tests verify the ability of the algorithm to model Lorentz force terms in multiple dimensions, to capture MHD shocks, and to propagate all MHD wave families in moving and stationary media. Further diagnostics include direct examination of the error terms and measurements of the rate of convergence to the appropriate analytic solution.

We have carried out both resolution experiments and direct comparisons with dynamo spectral code simulations (as described in Hawley *et al.*, 1995, 1996); these tests indicate that while the present code has numerical diffusion typical of finite-difference schemes, unstable wavelengths are reasonably well described when they extend over a minimum of about six grid zones. Further direct comparisons have been made between our results and those of Brandenburg *et al.* (1995), who use a vector potential MHD code, and Matsumoto and Tajima (1995) and Matsuzaki *et al.* (1997), who employ a modified Lax-Wendroff scheme. Although some of the quantitative details vary from code to code in these simulations, their major conclusions are all mutually consistent.

3. The evolution of an initially vertical field

Let us return to the case of the uniform initial vertical magnetic field, which, in two dimensions, led to streaming. The weak-field instability grows rapidly for wave numbers $kv_{Az}/\Omega \sim 1$ (maximum growth rate $=0.75\Omega$), provided that these wavelengths are adequately resolved by the grid. The growth of this incompressible instability and its saturation level are unaffected by the hydrodynamic pressure in the box. The saturation amplitude is also unaffected by the presence of a subthermal toroidal field, except insofar as instabilities associated with that field add to the overall energy. The fastest growing mode is axisymmetric and produces spiral streaming. In three-dimensional studies (Hawley *et al.*, 1995), if the computational box is large enough to allow an unstable radial wavelength, the streaming is disrupted within a few orbits, as the parasitic instabilities of Goodman and Xu (1994) lead to fluid turbulence. This turbulence is characterized by significant outward angular momentum transport. Transport results because the turbulence is inherently anisotropic; perturbations in the x and y components of the magnetic field and the velocity are highly correlated. Further, power spectra of the simulations are sharply peaked to the lowest wave numbers, indicating that there are significant fluctuations on scales comparable to the domain size. These large-scale fluctuations contain most of the magnetic energy and contribute the bulk of the stress.

Both Reynolds $\langle \rho u_R u_\phi \rangle$ and Maxwell $\langle -\rho u_{AR} u_{A\phi} \rangle$ stresses transport angular momentum, with the Maxwell stress typically a factor of 4 larger than the Reynolds stress. The ratio of their sum $T_{R\phi}$ to the average magnetic pressure in the computational box is about 0.6, a number that is fairly constant across a wide variety of vertical-field simulations. The magnetic pressure remains subthermal, and a typical value of the α parameter $T_{R\phi}/\rho c_s^2$ in these runs is $\alpha \sim 0.2$. But the numerical value of α so obtained depends upon the initial field strength.

Since the total stress is directly proportional to the magnetic energy density, it is important to understand what causes the latter to saturate. For vertical-field simulations, Hawley *et al.* (1995) found empirically that the average magnetic-field strength at saturation is proportional to $\langle B_Z L_Z \Omega \rangle^{1/2}$. This is the geometric mean of the (unchanging) box-averaged vertical field and the largest possible field having an unstable vertical wavelength $< L_Z$. This field is

$$(B_Z)_{max} \simeq \Omega L_Z \sqrt{\rho/\pi}.$$

The simplest guess for the saturated field strength would have been this latter value. (In a physical setting, this same reasoning suggests saturated Alfvén velocities of the order of the sound speed.) The memory of the initial vertical field is, however, never fully lost in the saturated state, a result clearly due in some part to the choice of boundary conditions, which ensure that $\langle B_Z \rangle$ remains constant. The effects of grid resolution were found to be minor, provided that the characteristic rapidly growing wave numbers were resolved. As we shall see, initially sinusoidal vertical fields lead to saturated states that are independent of initial conditions. But significantly stronger (factor of ~ 10) angular momentum transport can be produced for the same computational domain and initial field strength when $\langle B_Z \rangle \neq 0$. The astrophysical consequences of qualitative changes in α arising from changes in field topology are intriguing and potentially important; at the time of this writing, they remain largely unexplored.

4. The evolution of an initially toroidal field

Of course, a vertical field is hardly essential for the weak-field instability: when the initial field is toroidal, the instability is still present. Although certain mathematical issues arise as to when a local nonaxisymmetric instability is well posed (Ogilvie and Pringle, 1996), numerical simulations are quite unambiguous: growth rates are large enough that fully nonlinear turbulence sets in almost as rapidly as it does in the axisymmetric vertical-field case. The maximum growth rate for a toroidal configuration in a Keplerian disk is 0.75Ω, the same value as for a vertical field. However, as we have seen, extremely large values of k_Z are required to achieve this limit. In numerical simulations (Hawley *et al.*, 1995; Matsumoto and Tajima, 1995) the observed growth rates are accordingly lower than the vertical-field runs, but not dramatically so. Short-wavelength nonaxisymmetric modes are the first to appear, but long wavelengths soon dominate.

Once again, turbulence develops and angular momentum is transported outward. However, the turbulent magnetic energy densities that result are an order of magnitude smaller than those seen in the vertical-field simulations. In the runs described in Hawley *et al.* (1995) there is an increase of about one order of magnitude in the field energy before saturation. The magnitude of α is typically 10^{-2}, but, in common with their vertical-field counterparts, the toroidal saturated states retain a memory of their initial field strength. Once again the saturated field strength is proportional to $\langle B_Y L_Y \Omega \rangle^{1/2}$, the geometric average of the background field strength and the largest field with unstable wavelengths able to fit azimuthally in the box. The total stress remains proportional to the magnetic pressure, with a constant of proportionality ~ 0.5. And, as before, the Maxwell stress is greater than the Reynolds stress by a factor of 3 to 4.

In Sec. IV.F.2, it was emphasized that the dominance of toroidal fields, both in the numerical simulations and (to the extent it may be inferred) in nature, does not mean that neglecting poloidal fields is a good approximation. Hawley *et al.* (1995) gave an unambiguous demonstration of this. In one case, after an initially toroidal $\beta = 100$ field evolved for a little over ten orbits, a $\beta = 400$ uniform vertical field was introduced "by hand." The effect was immediate and dramatic. An eruptive increase in the value of $T_{R\phi}$ took place. When a new saturation level was reached, the magnetic energy was a factor of ~ 5 larger than the level reached without the vertical field. In another experiment, a $\beta = 3200$ uniform vertical field was started off with a $\beta = 100$ uniform toroidal field. Even though the initial vertical-field energy was only some 3% of the toroidal field energy in this run, the development was qualitatively similar to a pure vertical-field evolution. At the conclusion of the simulation, the magnetic energy density was a factor of ~ 4 larger than the purely toroidal initial field simulation. A difference of 3% in the initial field energy going into a poloidal component made a difference of 400% when saturation was reached.

5. The evolution of an initially random field

The initial conditions with the most general applicability are probably those for which $\langle \mathbf{B} \rangle = 0$ (Hawley, Gammie, and Balbus, 1996), specifically, a random initial field with zero net value within the computational domain. Once again the instability is found to lead to rapid field growth, turbulence, and enhanced transport. The most important difference between these simulations and those with $\langle \mathbf{B} \rangle \neq 0$ is that there appears to be no dependence on the initial magnetic-field strength, provided of course that the characteristic unstable wavelengths are resolved. Figure 23 shows the evolution of the magnetic energy in several runs that began with different initial values of β, ranging from 50 to 1600.

Although the magnetic energies show considerable variation with time, their time averages are $\beta \sim 50$. The

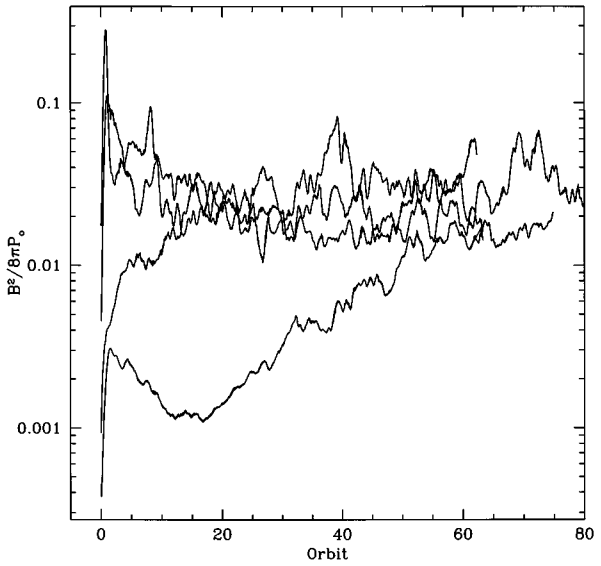

FIG. 23. Plots of normalized magnetic energy (β^{-1}) as a function of time for a series of three-dimensional shearing box simulations. All simulations begin with a random magnetic field ($\langle \mathbf{B} \rangle = 0$), but varying initial magnetic energy $\langle B^2 \rangle$. All initial conditions lead to turbulence with comparable magnetic energies. From Hawley *et al.*, 1996.

toroidal field dominates; its energy is typically almost 10 times that of the poloidal field. The magnetic energy values in these simulations were consistently less than those for which net fields passed through the simulation domain. Evidently, the presence of a net mean field (and especially a net vertical field) serves as an effective source term, which endures even as turbulence develops. Absent from such a source, fluid motions must be self-sustaining. In two-dimensional runs, we have seen this most dramatically: the presence of a net vertical field leads to an exponential runaway; its absence leads to a complete decay of the instability.

Aside from the differences in magnetic energies, the zero and net mean-field simulations are similar. As before, the turbulence is highly anisotropic, transporting angular momentum outward through Maxwell and Reynolds stresses, and the former exceeds the latter by a factor of 3. The total stress remains proportional to the magnetic energy, which here means that a typical value for α is 10^{-2}. In all simulations, including these random-initial-field models, the toroidal field energy is the largest, followed by the radial, and then the vertical field energy.

In $\langle \mathbf{B} \rangle = 0$ simulations, what sets the level of the magnetic energy in the saturated state? Although only a weak dependence was found on the grid resolution, Hawley *et al.* (1996) noted a significant dependence of the final magnetic-field energy on the level of artificial viscosity: the larger the viscosity, the higher the final field strength. Why should increasing a dissipative parameter lead to larger field growth? Certainly for the linear development of the instability, it has the opposite effect.

The point is that magnetic reconnection of field lines, both in numerical simulations and in the real world, often involves large velocity gradients. Increasing the viscosity while maintaining constant resistivity causes thermalization of kinetic energy on larger scales, but leaves the small grid scale for the thermalization of magnetic energy untouched. If the viscosity gets large, fluid elements are enjoined from getting close enough to allow field-line reconnection. Whereas strong (suprathermal) magnetic fields may overwhelm viscous stresses in a reconnection front, the subthermal magnetic fields of interest here are much less dominant. Driven to their extreme, viscous stresses saturate at about the level of thermal gas pressure, and this is more than enough to overcome subthermal magnetic stresses. This suggests that the ratio of the microscopic viscosity to resistivity, i.e., the *magnetic Prandtl number*, is likely to be an important parameter controlling the level of magnetic-field saturation. Suprathermal fields are less likely to be sensitive to the magnetic Prandtl number.

In simulations, the viscosity and resistivity have diffusive lengths set by the size of the grid zones. Of necessity, these must be considerably greater than those appropriate to real disks. However, it is still possible to vary their ratio, and in the Hawley *et al.* (1996) simulations, a large viscous diffusion relative to resistive diffusion led to proportionately larger final mean B-field values. In the same spirit, including an explicit resistivity term leads to smaller mean B-field values. While this must still be regarded as a preliminary result, it is potentially very important: for an ionized plasma the Prandtl number is extremely temperature sensitive. Collisional mean free paths increase sharply with temperature, raising the viscosity while lowering the resistivity. This is reflected in the steep T^4 temperature dependence of $\mathcal{P}$ in Eq. (144). All of this suggests that disk midplane temperatures in excess of 10^5 K may be associated with significantly higher values of α. For the present, this is beyond the range of easy numerical verification. Phenomenological applications of this behavior may be quite interesting. In the temporal domain, increases in α are associated with eruptive disk phenomena (Livio, 1994); spatial changes in α are needed in some important disk models which become hot and thick in their inner regions (Abramowicz, 1996; Narayan, 1996).

6. Shear vs vorticity

The role of the background rotational shear flow was examined by Abramowicz, Brandenburg, and Lasota (1996), who carried out a series of simulations allowing the disk rotation law to vary from Keplerian. Such an exercise is of more than academic interest; disks orbiting Schwarschild or Kerr black holes are expected to have rotation profiles that differ from Keplerian near the horizon. Further, disks with significant internal pressure, the so-called thick disks, also have non-Keplerian angular momentum distributions. Abramowicz *et al.* (1996) examined the turbulent saturation levels for the local shearing box system with a variety of background shear

index values q. They found an increase in α (and of course the field energy) roughly in proportion to the background-shear-to-background-vorticity ratio, as $q \to 2$. That is, α was found to be proportional to $q(2-q)^{-1}$, as long as q did not get too close to 2. (In the published runs, the maximum value of the ratio was 9.)

The notion of a precise linearity of the proportionality should not be taken too far, since α and all other natural fluid quantities presumably remain quite finite at $q=2$, when the constructed shear-to-vorticity ratio becomes infinite. But the general trend exhibited by these simulations—vorticity limits disk turbulence, while shear promotes it—is an interesting lesson, which is easily understood. It follows quite naturally from the fluctuation equation (86). Focus on the turbulent source terms, and note

$$2\Omega\langle\rho u_R u_\phi\rangle + \frac{d\Omega}{d\ln R}\langle\rho(u_R u_\phi - u_{AR}u_{A\phi})\rangle = \frac{1}{R}\frac{dR^2\Omega}{dR}\langle\rho u_R u_\phi\rangle - \frac{d\Omega}{d\ln R}\langle\rho u_{AR}u_{A\phi}\rangle. \quad (161)$$

In other words, the Reynolds stress couples to the large-scale vorticity, and the Maxwell stress couples to the shear. In all the runs reported by Abramowicz *et al.* (1996), the shear and vorticity had opposite signs, i.e., the disks were Rayleigh stable. Hence the above formulation shows that outward (vorticity-coupled) Reynolds transport is a dynamic sink for angular momentum fluctuations, whereas outward (shear-coupled) Maxwell transport is a dynamic source. Their net sum is a dynamic source, but without the Maxwell stress there is, of course, no turbulence at all. Increasing the shear while maintaining or lowering the vorticity should strengthen the sources of turbulence, and this is precisely what the Abramowicz *et al.* (1996) simulations show.

7. Density stratification

Of the numerical MHD work done to date, simulations including vertical gravity come the closest to modeling a full accretion disk. Brandenburg *et al.* (1995), Stone *et al.* (1996), and Matsuzaki *et al.* (1997) have carried out MHD simulations of local, vertically stratified shearing boxes. Although all groups report some numerical dependencies in the precise numbers resulting from the simulations, comparisons between the various results and the use of several grid resolutions helps to separate the physics from the numerical artifacts. Higher resolution results in larger magnetic-field levels, stronger turbulence, and greater transport. For example, by doubling the resolution in all three dimensions, Stone *et al.* (1996) found an increase in magnetic energy of about 25% in one comparison simulation.

Somewhat surprisingly, the effects of stratification for the evolution of a weak magnetic field, at least as revealed in the simulations, are rather weak. Homogeneous and stratified boxes evolve remarkably similarly. Of course, the weak-field linear instability is unaffected, since the most unstable displacements lie in planes normal to the stratification. However, even the nonlinear development of the instability, which might have been strongly quenched by magnetic buoyancy, continues much as it does in the homogeneous simulations. There appears to be nothing about stratification that alters the basic picture of MHD turbulent angular momentum transport. The initial instability grows rapidly and generates self-sustaining MHD turbulence. The turbulence is characterized by a Kolmogorov-like power-law spectral distribution; most of the turbulent energy is found in the smallest wave numbers (Brandenburg *et al.*, 1995; Stone *et al.*, 1996). Correlated fluctuations in the magnetic and velocity fields produce Maxwell and Reynolds stresses, which transport angular momentum outward. The total stress is once again proportional to the magnetic pressure, which saturates at a subthermal level. In essence, a stratified disk simulation is dominated by the dynamics within a scale height of its midplane, where it strongly resembles a homogeneous disk. Values of α range from 0.005–0.01 in Stone *et al.* (1996) and Matsuzaki *et al.* (1997) to 0.001–0.005 in Brandenburg *et al.* (1995). No single value of α should be raised to canonical status. One would be more justified at this point in regarding these values as lower limits for the stresses in weakly magnetized disks.

With hindsight, the lack of qualitative change introduced by density stratification could have been foreseen. The point is that the characteristic growth rate of the instability is Ω, which should be compared with the rate v_b/H, the rate at which a buoyant velocity v_b crosses a disk scale height. Unless $v_b \sim c_s$, the instability will win, and magnetic losses will be dominated by resistivity. Numerical resistivity occurs naturally at grid scale in the simulations, but the statement is no less true for physical resistivities as well.

Stone *et al.* (1996) found that effects of different equations of state (e.g., adiabatic vs isothermal) were largely inconsequential for the development of the instability, but there was some influence on the overall structure of the disk in the long term. In one simulation with an adiabatic equation of state, Stone *et al.* (1996) found that, as the disk heated due to (artificial) viscous dissipation, the increase in the disk temperature led to a commensurate increase in the scale height, producing a similar increase in magnetic energy. This result is consistent with the general α viscosity picture, namely, that total stress (which is directly proportional to magnetic pressure) should scale on average with total disk pressure (gas plus magnetic), which determines the disk scale height.

Strong turbulence and resulting significant vertical motions may themselves have important consequences for the vertical disk structure and energy transport. For example, the toroidal field simulation of Stone *et al.* (1996) began with the field confined below one scale height, but after five orbits the field was distributed throughout the computational domain. In fact, at late times the magnetic energy density increased with height by a modest amount. A little field goes a long way in a

low-density atmosphere. Since the gas pressure drops with height, there is a corresponding decrease in β. Thus, while buoyant losses, as noted, do not dominate the local magnetic energy budget, the existence of vertical transport in both magnetic fields and passive disk contaminants strongly hints that accretion disks will be surrounded by a hot magnetized corona. The existence of a very hot gas in accretion disk systems has long been suspected on the basis of their hard x-ray spectra. Whether the explanation has its origins in disk turbulence and the strength of the disk magnetic field (Mineshige, Kusnose, and Matsumoto, 1995) or in some form of non-Keplerian accretion (Narayan, Yi, and Mahadevan, 1995) is an open and vigorously contested question at this time.

C. MHD simulations: a summary

What, then, have we learned from local numerical simulations? First and foremost, they have confirmed the linear stability analysis, namely, that weak magnetic fields produce a powerful, local instability in differentially rotating systems. They have demonstrated that a weak magnetic field of any topology leads directly and naturally to turbulence and outward angular momentum transport. The contrast with purely hydrodynamic simulations is dramatic. As recently as the beginning of this decade no turbulent angular momentum transport mechanism had been explicitly demonstrated to exist which would allow accretion disks to accrete at anywhere near the inferred observational rates. Now such a mechanism has been identified, and no special global conditions, external torques, exotic physics, or other contrivances are required, only a subthermal field, well coupled to the gas, and outwardly decreasing differential rotation.

Shakura and Sunyaev (1973) originally postulated that angular momentum transport was accomplished by turbulence and that the resulting stress tensor could be scaled by the disk pressure, $T_{R\phi}=\alpha P$. The α viscosity approach has been successful in describing mean disk properties, and since the early 1970s it has been clear that there is *something* correct about this phenomenology. Yet, despite the insights such models have provided investigators (e.g., in the study of dwarf nova eruptions), time-dependent models relying upon adjustable viscosity parameters ultimately lack predictive power. The lack of knowledge about α's underlying physical behavior is a serious limitation for such issues as long-term dynamic stability and detailed disk structure and evolution. To date, the primary contribution of numerical disk simulations has been to put the basis of the α formalism on a firmer physical footing. But because these simulations have of necessity been limited to dynamics, important qualitative phenomena are still unexplored: What might cause α suddenly to jump an order of magnitude? What is the role of the microscopic diffusion coefficients and opacity effects in determining α? Numerical formulations of the answers to these questions are still some way off.

Because the turbulence is fundamentally MHD in origin and character, the stress scales directly with the total *magnetic* pressure. The simulations have found a constant of proportionality of about 0.5 between the stress and the magnetic pressure. The relationship between the magnetic pressure and the total pressure (which may in some cases also include a contribution from radiation) is less direct and the subject of continuing active research. When radiation pressure is important, for example, magnetic pressure may not scale with total pressure, but with gas pressure alone (Vishniac, 1995). However, the simulations to date have shown that it is easily possible to obtain numerical values of α in accord with the wide range inferred from observations: from as high as 0.6 when there is a net vertical field (Hawley *et al.*, 1995) to $\sim 5\times 10^{-3}$ for random, zero-mean-field models (Brandenburg *et al.*, 1995).

The elucidation of the underlying physical mechanism regulating accretion disk evolution gives us the expectation that detailed, first-principles disk models will eventually be possible. Several future lines of inquiry can easily be imagined, although carrying them through will, of course, be less than straightforward. One approach is to improve the physics within a small local region of a disk. Examples include radiation transport, partial ionization, buoyancy, reconnection, and improved equations of state. Although there remain several fundamental questions that can be addressed by such local simulations, ultimately fully global three-dimensional MHD disk simulations will be required. To date, only one such simulation has been published, that of Matsumoto and Shibata (1997). These authors have simulated the evolution of an initially constant-angular-momentum torus embedded in a vertical magnetic field, extending to three dimensions the analysis of an earlier two-dimensional simulation (Matsumoto *et al.*, 1996). In this case, the added dimension does not appear to lead to qualitatively new dynamics. In both the two- and the three-dimensional calculations, surface accretion was observed to occur as a consequence of the rapid onset of magnetorotational instability, leading to spiraling infall and the beginnings of jetted outflow. The daunting technical requirements of such simulations, plus the (often underappreciated) difficulties of visualizing, diagnosing, interpreting, and simply managing the massive data stream associated with such complex dynamic flows, means that such global three-dimensional modeling will remain a significant challenge for some time to come.

VI. ACCRETION DISK DYNAMOS

A. The dynamo-electric machine

The first 19th-century current-generating machines consisted of a set of rotating coils in the field of permanent steel magnets. In 1866, the German industrialist and engineer Werner Siemens had a clever idea: instead of permanent magnets, install electromagnets. Then take the generated current itself, wind it around the electromagnets, and thereby greatly strengthen the mag-

netic field through which the coils turn. Siemens named his device (which helped make him a very wealthy man) a *dynamo-electric* machine. Its dynamic character distinguished it from the older, permanent-magnet, *magneto-electric* device.

The British physicist and science popularizer Silvanus Thompson could scarcely contain his enthusiasm: "The cost of producing electric currents of any required power is now simply the cost of ... a dynamo-machine and a steam-engine, and of the coal and labour necessary to supply and attend them." [7] While Thompson seems to have had an unusual notion of what constitutes a bargain, he knew a good idea when he saw it, and he suggested that these machines, soon to be in common usage, be referred to succinctly as *dynamos*.

So efficient seemed the Victorian dynamo machine, it has remained to this day the conceptual prototype of all large-scale cosmic magnetic fields: planetary, stellar, and galactic. None of these fields has well understood origins. According to classical dynamo theory, the time-varying magnetic flux caused by bulk fluid motions in a continuum, conducting, magnetized fluid induces currents in the fluid. If conditions are right, these currents reenforce the original source currents which gave rise to the flux in the first place. It is, however, no simple matter to decide what the "right conditions" are in all circumstances, and rather than remain wedded to the idea of self-reenforcing inductive currents, we follow the lucid text of Moffatt (1978) and adopt an operational definition of a dynamo: if the motions of an isolated, magnetized, resistive fluid maintain a finite magnetic energy density as $t\rightarrow\infty$, then the velocity field acts as a dynamo. Clearly, by this definition, the instability in a weakly magnetized disk leads to a dynamo. But, as we shall see, accretion disk dynamos do not operate by principles that would be familiar to Herr Siemens.

B. A brief review of mean-field dynamo theory

The fundamental equation for the evolution of a magnetic field in a conducting fluid is the induction equation (6c)

$$\frac{\partial \mathbf{B}}{\partial t}=\nabla\times(\mathbf{v}\times\mathbf{B}-\eta_B\nabla\times\mathbf{B}), \tag{162}$$

which is a combination of the laws of Ampère, Faraday, and Ohm. The last of these "laws" is phenomenological, not fundamental, and excludes such effects as inertial and pressure forces on the charge carriers (Spitzer, 1962). The result is the absence of a source term in the induction equation and the need for a seed field to be inserted by hand. For a binary disk this is not an issue, since the gas presumably retains its stellar magnetization, but it is very much an issue for the galaxy (Parker, 1979; Kulsrud and Anderson, 1992), a point we shall touch on very briefly. "Battery forces" (Biermann, 1950) and radiative electron drag are strongly limited by self-induction, and the resulting fields have associated Larmor radii comparable to the macroscopic scale of the system (Balbus, 1993). For the galaxy this is a disappointing field strength of some 10^{-19} G. Amplification by pre-galactic turbulence has recently been suggested by Kulsrud *et al.* (1997), while Rees (1993) has argued that stellar dynamos may be the galaxy's ultimate source of magnetism. The problem remains open.

For the problem at hand, we take as our starting point a weakly magnetized accretion disk. For the turbulent velocity and magnetic fields, it is customary to decompose these quantities into mean plus fluctuating components. Averages are taken on scales over which the local mean is treated as a constant, very much in the style of Sec. III.A, where steady-state accretion was discussed. There, as here, the large scale is defined by the disk radius R, while the smaller averaging size is of order the vertical scale height H or less. In our earlier treatment, however, we did not distinguish between mean magnetic fields and fluctuation fields, and we decomposed the velocity field into a circular and noncircular component, not a mean plus fluctuation. We shall nevertheless continue to use the notation $\mathbf{u}$ in this section for the fluctuation velocity. The small ($|\langle\mathbf{u}\rangle|\ll\langle u^2\rangle^{1/2}$) nonazimuthal drift velocities, which formally ought to be absorbed into our definition for the mean velocity, are immaterial to our present purposes.

We therefore have

$$\mathbf{B}=\langle\mathbf{B}\rangle+\mathbf{b}, \tag{163}$$

$$\mathbf{v}=\langle\mathbf{v}\rangle+\mathbf{u}. \tag{164}$$

The magnetic and velocity fluctuations are, respectively, $\mathbf{b}$ and $\mathbf{u}$. If, following Moffatt (1978), we insert these expressions into the induction equation and sort out the resulting mean and fluctuating components, we find

$$\frac{\partial\langle\mathbf{B}\rangle}{\partial t}=\nabla\times(\langle\mathbf{v}\rangle\times\langle\mathbf{B}\rangle+\langle\mathbf{u}\times\mathbf{b}\rangle-\eta_B\nabla\times\langle\mathbf{B}\rangle), \tag{165}$$

$$\frac{\partial\mathbf{b}}{\partial t}=\nabla\times(\mathbf{u}\times\langle\mathbf{B}\rangle+\langle\mathbf{v}\rangle\times\mathbf{b}+\mathbf{G}-\eta_B\nabla\times\mathbf{b}), \tag{166}$$

where

$$\mathbf{G}=\mathbf{u}\times\mathbf{B}-\langle\mathbf{u}\times\mathbf{B}\rangle. \tag{167}$$

In classical dynamo theory, we are invited to view the turbulent velocity field $\mathbf{v}$ as a prescribed function of position and time, at least in a statistical sense. The main point is that when the field is "weak," Lorentz forces are taken to be unimportant to the statistical properties of the turbulence. Fluid forces are hydrodynamic, not magnetohydrodynamic. Conceptually, the hydrodynamic turbulence molds the current to mimic the coil winding of the dynamo machine, allowing the field to be self-strengthening. Lorentz forces should not interfere with how the coils are wound. Mathematically, the development of dynamo theory begins with the argument that $\langle\mathbf{u}\times\mathbf{b}\rangle$ scales linearly with the local mean field $\langle\mathbf{B}\rangle$. This

[7] The quotation is from the *dynamo* entry in the *Oxford English Dictionary*.

would follow if $\mathbf{u}$ had no dependence upon $\mathbf{B}$, for then Eq. (166) would assure us of a linear scaling of $\mathbf{b}$ with $\langle\mathbf{B}\rangle$.

This conclusion, that the mean electromotive force $\langle\mathbf{u}\times\mathbf{b}\rangle$ scales linearly with the mean magnetic field $\langle\mathbf{B}\rangle$, is the cornerstone of *kinematic* dynamo theory. It has been suggested (Moffatt, 1978) that it is the absence of a corresponding relationship between $\langle\mathbf{v}\rangle$ and $\langle u_i u_j\rangle$ that has prevented shear turbulence theory from enjoying a measure of success comparable to that ascribed to dynamo theory.

It is only fair to forewarn the reader that this formalism does not fare well as an approach to accretion disk dynamos. It is nevertheless important, for several reasons, to have a sense of where this approach leads. First, it is a benchmark with many years of literature behind it, and it is inevitable that other turbulent fluid field amplification schemes will be compared with it. Second, venues other than disks are sites of field amplification, and here kinematic dynamo theory may do better. Its domain of applicability is very much a lively and vigorously contested issue (Cattaneo and Vainshtein, 1991; Kulsrud and Anderson, 1992). Finally, even if kinematic dynamo theory fails quantitatively, it has important lessons to teach us with respect to how field ordering is influenced by the symmetry of the turbulence.

The linearity *ansatz* suggests a relationship of the form

$$\langle\mathbf{u}\times\mathbf{b}\rangle_i=\alpha_{ij}\langle B_j\rangle+\beta_{ijk}\frac{\partial\langle B_j\rangle}{\partial x_k}+\cdots \tag{168}$$

which is expected to converge rapidly with increasing order of the derivatives. The β-parameter term leads to an effective turbulent dissipation of the mean field [because of the presence of higher-order derivatives in Eq. (165)], and we shall not pursue its consequences here. The α term is the heart of the classical dynamo.

If the fluid may be approximated as isotropic on the scales of interest, then

$$\alpha_{ij}\langle B_j\rangle=\alpha B_i,$$

where now α is a simple scalar. Clearly the possibility for mean-field growth is at hand. If the mean velocity and resistivity may be neglected, then the induction equation is

$$\frac{\partial\mathbf{B}}{\partial t}=\nabla\times(\alpha\mathbf{B}). \tag{169}$$

Solutions to this equation based on, for example, force-free configurations ($\nabla\times\mathbf{B}=C\mathbf{B}$ with C constant) grow exponentially while remaining force free. To develop a feel for what properties of the fluid go into determining α, start with

$$\langle\mathbf{u}\times\mathbf{b}\rangle_k=\langle\mathbf{u}\times\mathbf{B}\rangle_k=\varepsilon_{ijk}\left\langle u_i B_m\frac{\partial x_j}{\partial X_m}\right\rangle, \tag{170}$$

where we have used the Lagrangian solution (15) and ignored density variations. It is customary to argue that correlations in the velocity u_i and Lagrangian Jacobian $\partial x_j/\partial X_m$ will be stronger with each other than with the magnetic field, whence

$$\varepsilon_{ijk}\left\langle u_i B_m\frac{\partial x_j}{\partial X_m}\right\rangle\simeq\varepsilon_{ijk}\left\langle u_i\frac{\partial x_j}{\partial X_m}\right\rangle\langle B_m\rangle,$$

so that the α tensor takes the form

$$\alpha_{km}\equiv\varepsilon_{ijk}\left\langle u_i\frac{\partial x_j}{\partial X_m}\right\rangle. \tag{171}$$

We may express this in terms of the velocity by noting

$$\frac{\partial x_j}{\partial X_m}=\delta_{jm}+\int_0^t\frac{\partial u_j}{\partial X_m}d\tau \tag{172}$$

so that

$$\alpha_{km}=\varepsilon_{ijk}\int_0^t\left\langle u_i\frac{\partial u_j}{\partial X_m}\right\rangle d\tau. \tag{173}$$

Returning once again to the assumption of isotropy, we find

$$\alpha=\frac{1}{3}\alpha_{kk}=\frac{1}{3}\int_0^t\varepsilon_{ijk}\left\langle u_i\frac{\partial u_j}{\partial X_k}\right\rangle d\tau=-\frac{1}{3}\int_0^t\langle\mathbf{u}\cdot\nabla_{\mathbf{x}}\times\mathbf{u}\rangle d\tau, \tag{174}$$

where the $\nabla_{\mathbf{x}}$ notation indicates Lagrangian coordinates. When the turbulence is isotropic, dynamo amplification relies on the presence of mean helicity ("coil windings") in the velocity field.

We shall not comment further here on the assumptions and approximations that go into the derivation of Eq. (174). They are often subtle, and mathematical rigor is difficult to establish. Our point here is to arrive at the salient physical features of classical mean-field dynamo theory. To summarize, they are as follows:

(i) When the magnetic field is weak, it is assumed that turbulence in a magnetized fluid can affect $\mathbf{B}$, but the back reaction on the turbulence is statistically unimportant.

(ii) Velocity and velocity gradient fluctuations are significantly more correlated with each other than with magnetic-field fluctuations.

(iii) The primary field amplification method is a sort of self-inductive bootstrap, analogous to the dynamo-electric machine described in Sec. VI.A.

(iv) The dynamics of the flow are not directly important; flow kinematics determine the nature of field amplification. Some "handedness" in the velocity field is necessary to produce amplification by self-induction.

C. Mean-field theory and nonlinear evolution of the magnetorotational instability

It is difficult to reconcile several of the items highlighted above with the behavior of weak magnetic fields in accretion disks. The fundamental problem is the assumption that the velocity field $\mathbf{u}(\mathbf{r})$ may be prescribed independently of $\mathbf{B}$. The essence of the magnetorotational instability is precisely the opposite: any wisp of a

magnetic field is enough to drive and ultimately to maintain a turbulent velocity field. Lorentz forces, assumed to be negligible while the field is "weak," are in fact never negligible. The field is never really weak because there is nothing for it to be weak with respect to: Alfvénic and slow-mode disturbances are indifferent to the pressure, and the pressure is the only nonmagnetic characteristic fluid stress that, in this context, sets an intrinsic scale. As noted in the discussion in Sec. IV.B, the presence of a magnetic field introduces degrees of freedom to the fluid which have no purely hydrodynamic counterpart. Since the u field is not defined independently of **B**, the arguments leading to the conclusion that $\langle \mathbf{u}\times\mathbf{b}\rangle$ scales linearly with $\langle \mathbf{B}\rangle$ break down. That lost, the basis for kinematic mean-field dynamo theory is gone.

The nature of dynamo activity in accretion disks has been investigated in two numerical studies: Brandenburg *et al.* (1995) and Hawley, Gammie, and Balbus (1996). The first was a vector potential code, the second used the MOCCT algorithm described in Sec. V.B.2. Both studies simulated local shearing boxes, both used the same boundary conditions in radius (quasiperiodic, see Sec. III.D.3), and azimuth (periodic), but the vertical boundary conditions were handled differently. Hawley *et al.* (1996) used periodic boundary conditions in z. This allowed flow through the vertical bounding surfaces, but did not allow the mean-field value $\langle \mathbf{B}\rangle$ to evolve. Brandenburg *et al.* (1995) bounded the flow vertically by enforcing what amount to reflecting wall boundary conditions on the top and bottom boundaries. This proscribed vertical outflow, but allowed the mean azimuthal magnetic-field component to change with time. Over the course of their simulation, Brandenburg *et al.* (1995) noted changes in $\langle \mathbf{B}\rangle$, including several field reversals.

The two simulations, notwithstanding differences in technique and boundary conditions, are in agreement on their most important conclusion: Accretion disks are the site of self-sustaining magnetic-field amplification, which occurs despite the presence of significant grid loss, and this amplification lies outside the domain of kinematic dynamo theory. The broad significance of this finding is that the conceptual picture of turbulence amplifying a passive magnetic field,

$$\text{turbulence}\rightarrow\text{magnetic field}$$

is replaced by the synergy of MHD turbulence:

$$\text{turbulence}\leftrightarrow\text{magnetic field}.$$

Magnetic fields are able not only to induce self-reenforcing currents, but also to produce the turbulent velocity field itself. It is simply not possible to prescribe the statistical properties of the velocity field *a priori*.

In one respect, the underlying physics of an accretion disk dynamo is simpler than that of a classical dynamo. The disk dynamo operates by a linear process that feeds off the differential rotation and is in essence no more complicated than the stretching of a spring connecting orbiting point masses (see Sec. IV.B). Stretching of magnetic-field lines increases their energy density at the expense of differential rotation, and the process is ultimately controlled, at least in simulations, by the field lines' pinching off and reconnecting. The energy effectively lost in this reconnection is replenished by the ongoing field-line stretching, and a balance is struck. Because the fluid displacements and field-line stretching are predominantly in the disk plane—a behavior whose roots go back to the linear instability—by the time saturation is reached, the azimuthal (B_ϕ) and radial (B_R) components of the magnetic field are significantly larger than the vertical (B_Z) component (Hawley *et al.*, 1996). Moreover, B_ϕ is fed both directly by shear and by the instability, B_R only by the instability, while any growth in B_Z is due to secondary agitation of the fluid. In what Hawley *et al.* designated their "fiducial run"—a random initial magnetic field with $\beta=800$ in a box with $31\times63\times31$ (x,y,z) grid zones—88% of the magnetic energy was in the B_ϕ^2 component, 9% in B_R^2, and 3% in B_Z^2 at saturation. Accretion disks are likely to be dominated by their azimuthal magnetic-field component, a trait they would share with the interstellar medium in the disk of our own galaxy.

The mean helicity of the fiducial run was found to be small: in nondimensionalized units (velocity as a fraction of sound speed, length normalized to the vertical box size),

$$-\langle \mathbf{v}\cdot\nabla\times\mathbf{v}\rangle\sim4\times10^{-3}.$$

In the same units, the kinetic-energy fluctuations were an order of magnitude larger. Though small, the finite size of the helicity would be of interest if it were really a true time-steady value, since there is no obvious pseudoscalar symmetry built into the initial conditions. Shearing box simulations that do have such symmetry [e.g., the vertically stratified boxes of Stone *et al.* (1996) with finite $\Omega\cdot\nabla\rho$] show mean helicity at saturation of about the same size. Since fluctuations in the helicity are large, and since a nonzero value for its mean is difficult to understand, the statistical significance of finite $\langle \mathbf{v}\cdot\nabla\times\mathbf{v}\rangle$ needs to be more firmly established. What is clear however, is that the field amplification in accretion disk dynamos does not depend upon mean finite helicity for its existence.

Hawley *et al.* (1996) compared their work with earlier simulations by Meneguzzi, Frisch, and Pouquet (1981). The latter used a spectral code to simulate the three-dimensional dynamo activity in an enclosed box resulting from explicit external driving forces. Both helical and nonhelical driving were studied, and both types of driving led to dynamo amplification. While the helical stirring produced fairly coherent large-scale fields, the nonhelical forces gave rise to a more intermittent field structure. The accretion disk dynamo was qualitatively very similar to the nonhelical dynamo of Menneguzzi *et al.*, as one might expect.

The finding that helicity in the turbulent velocity field is not a prerequisite for dynamo amplification may lead

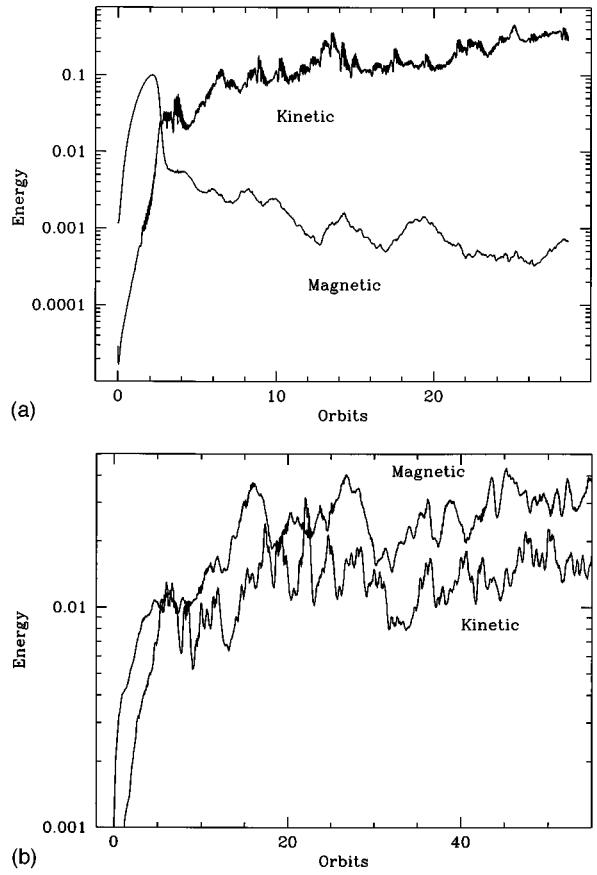

FIG. 24. Evolution of turbulent magnetic and kinetic energies: (a) Simulation of a simple shear flow. The kinetic energies grow due to hydrodynamic instability, but despite the presence of shear and turbulence, the magnetic energy declines after an initial rise. From Hawley *et al.*, 1996. (b) A dynamo simulation driven by the weak-field MHD instability. The behavior of the magnetic energy density relative to the kinetic energy is in sharp contrast to the shear turbulence result of Fig. 24(a).

one to the conclusion that the simple combination of turbulence plus magnetic field invariably causes field amplification. That this is not the case is clearly illustrated by the simulations of Hawley *et al.* (1996) of magnetized Cartesian shear flow. As discussed in Sec. III.D.3, such flows are nonlinearly hydrodynamically unstable and break down into turbulence. The addition of a weak magnetic field does not alter this behavior qualitatively, and it affords the possibility of following the evolution of a magnetic field in self-produced non-MHD turbulence.

The results of the Hawley *et al.* shear layer simulation are compared with a disk dynamo simulation in Figs. 24(a) and 24(b). As these figures show, the sharp difference in the relative behavior of the magnetic and kinetic energies is striking. In the shear layer [Fig. 24(a)], the kinetic-energy fluctuations grow by feeding directly off the background velocity gradient. But in the absence of true differential rotation, the magnetic-field fluctuations do not grow. Instead, the ever present large-scale shear drives magnetic structure to ever larger radial wave numbers, where it is lost to grid dissipation. The role of the large-scale shear is important, as nonhelical randomly forced turbulence does not drive down magnetic energy and seems in fact to amplify it slowly (Meneguzzi *et al.*, 1981). By way of contrast, the kinetic and magnetic energies both rise to roughly equal levels in the disk dynamo results of Fig. 24(b).

The long-term fate of the magnetic field in shear turbulence is unclear (does it level off or vanish as $t\rightarrow\infty$?), and the grid dissipation surely enhances losses well beyond what is to be expected in an astrophysical plasma, but the qualitatively different behavior of shear layers and local disk patches is dynamic and real. Whereas a shear layer is characterized by a long-term kinetic energy well in excess of the magnetic energy density, the magnetic energy is comparable to or larger than the kinetic in a local disk box. What must be appreciated here is that amplification of magnetic fields by hydrodynamic turbulence cannot be taken for granted. Each problem is special.

D. Saturation

Magnetic fields grow in accretion disks through the action of the weak-field instability. What limits the growth? There are two principal mechanisms: buoyancy and dissipation. We consider each in turn.

When the magnetic pressure becomes comparable to the thermal pressure in a gas, it is difficult to keep it confined in a thin disk. Relative to its surroundings, strongly magnetized gas has a smaller thermal pressure, which often translates to a smaller density, and tends to rise out of the disk because it is buoyant. Even static equilibria are not immune to field expulsion because of the magnetic analog of convective instabilities (Newcomb, 1961; Parker, 1966). To cross one vertical scale height of a disk in a time $1/\Omega$ requires a velocity on the order of c_s— roughly speaking, the sound speed. Buoyant velocities are likely to be a small fraction of this, and as a mechanism to continuously rid the disk of generated magnetic field, buoyancy is too inefficient. One could imagine, however, the field building up to a level at which the magnetic instability was suppressed. Without this source, even sluggish vertical motions would be able to clear the field, leaving behind a weak remnant which would retrigger the instability. The cycle would then repeat. This picture is the basis of a simple disk dynamo scenario outlined by Tout and Pringle (1992).

The vertically stratified simulations reveal a more complex picture, however. Power is injected at the wave-number scale $k\sim\Omega/u_A$ and quickly spreads to both larger and smaller scales. Resistive dissipation occurs at small grid scales at a rate which continuously balances the injection. Velocities and magnetic fields are dominated by their radial and azimuthal components. Dissipation, not vertical advection, determines the level of field saturation—at least in the numerical simulations.

Granting that the simulations vastly overestimate the scale at which resistive losses occur, they may yet be telling us something important. First, Hawley *et al.* (1996) carried out runs at several different resolutions

and found that the saturation field levels were statistically identical. These simulations had no vertical background gradients, but the Stone *et al.* (1996) simulations did, and they showed very similar statistical behavior. In particular, although regions of high magnetic-field strength were seen to rise over periods of tens of orbits, buoyancy was never the dominant field limitation mechanism. Even regions of relatively large field rose too slowly to outpace the instability.

It appears likely, therefore, that there is a dynamo regime that is characterized by unstable growth continuously balancing dissipation scale losses. This leads to subthermal magnetic fields and a dimensionless stress tensor α (the α of Shakura and Sunyaev, not the helicity parameter) of order 10^{-2}. Whether there are other modes of dynamo operation that arise naturally in accretion disks—at different magnetic Prandtl numbers, for example—is a fascinating and completely open question.

VII. SUMMARY

It does not seem possible to understand any facet of turbulence-enhanced accretion disk transport without the active participation of magnetic fields. Conversely, when a subthermal magnetic field is present, turbulence is well nigh unavoidable; when the field is suprathermal, its Maxwell stress will be dynamically significant, and once again enhanced transport will be hard to avoid. There is a very good possibility that the decades old mystery of the origin of accretion disk turbulence has been solved.

At one level, matters are very simple, which lends credence to this claim. The origin of disk turbulence is a local linear instability, akin to the classical Rayleigh problem. But instead of inertial waves turning into evanescent instabilities when the specific angular momentum decreases outward, this fate befalls slow MHD waves when the angular *velocity* decreases outward. The formal result of all of this is that angular velocity gradients—not angular momentum gradients—are the critical stability discriminants in a weakly magnetized gas. There is a very simple algorithm for correcting rotational hydrodynamic stability criteria for the presence of a weak magnetic field: replace all angular momentum gradients with angular velocity gradients in the relevant criterion. The new forms, for example, of the classical Høiland criteria are given by Eqs. (130) and (132). Since it is hard to imagine the usual stellar interior application of these criteria to be devoid of *any* magnetic field, in their original form they are of very limited validity.

There is an exact mechanical analog of the instability when applied to disks: two orbiting point masses connected by a weak spring (one whose oscillation frequency is less than the orbital frequency) will drift apart at an exponentially increasing rate. Alfvénic tension acts precisely as such a spring, tethering neighboring fluid elements. Ultimately, understanding the origins of accretion disk turbulence is no more difficult than simple orbital mechanics.

What is known of the turbulence itself? Here, of course, matters become rapidly more complex. Still, there are some reassuring patterns that emerge. It is clear, for example, that the outcome of the instability is outwardly directed enhanced angular momentum transport. This is the hallmark of a true shearing instability, as seen in both planar and Rayleigh-unstable Couette flow. Externally driven disk turbulence need not lead to outward angular momentum transport, as experience with convection and tidal forcing has shown. The coupling between the stress tensor $T_{R\phi}$ and differential rotation is the energy source of the fluctuations if the turbulence is self-generated [see Eq. (89)]. Outward flow of angular momentum is a prerequisite for this coupling to act as a source, not as a sink. Three-dimensional MHD simulations verify this conclusion and produce values of the phenomenological α parameter ranging from $\sim 10^{-2}$ to 0.6, depending upon the magnetic-field geometry. The largest values of α occur when the magnetic field has a nonvanishing mean vertical component. These magnitudes roughly bracket the range of α inferred from observations (Lin and Papaloizou, 1996). How reliable are the numerical α values? We cannot yet be sure.

The central problem in accretion disk theory is no longer to find out what turns turbulence on. It is to find out what tries to turn it off. We do not yet know the role of global disk geometry, for the problem of understanding accretion disk turbulence is intimately linked to understanding the dynamo behavior of the fluid. We do not know how the relative sizes of the microscopic viscosity and resistivity influence the final field strength, but there is just a hint that it is important: raising the artificial viscosity parameter in numerical simulations enhances the final magnetic-field strength at saturation. Astrophysically, the viscosity-to-resistivity ratio (magnetic Prandtl number) is likely to be very temperature sensitive. What are the phenomenological consequences? Might there be a secondary instability taking the disk from one turbulent saturated state to another, as suggested by dwarf novae eruptions? How does any of this relate to formation of diagnostic emission lines or the maintenance of an x-ray corona? Again and again we are ignorant.

The good news is that, for the first time, it appears that we know in which directions we should be looking to begin to find answers to questions like these.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the advice and encouragement of many colleagues as we have struggled over the past few years to understand the dynamics of accretion disks. We must, however, single out our collaborators O. Blaes, C. Gammie, and J. Stone, whose shared insights have helped to shape much of the work presented here. We are pleased to acknowledge helpful suggestions from M. Livio and E. S. Phinney. Finally, C. Gammie, P. Goldreich, and C. Terquem all read an earlier draft of this paper with great care, and their comments have been invaluable. We are particularly grateful

to these individuals. We acknowledge the NSF and NASA for their continuing support through grants AST-9423187, NAG-5-3058, NAGW-4431. Most of the numerical calculations discussed in this paper were carried out at NCSA and PSC, with support from the NSF.

REFERENCES

Abramowicz, M., 1996, in *Physics of Accretion Disks*, edited by S. Kato, S. Inagaki, S. Mineshige, and J. Fukue (Gordon and Breach, Amsterdam), p. 1.

Abramowicz, M., A. Brandenburg, and J.-P. Lasota, 1996, Mon. Not. R. Astron. Soc. **281**, L21.

Adams, F. C., S. P. Ruden, and F. H. Shu, 1989, Astrophys. J. **347**, 959.

Balbus, S. A., 1993, Astrophys. J. Lett. **413**, L137.

Balbus, S. A., 1995, Astrophys. J. **453**, 380.

Balbus, S. A., C. F. Gammie, and J. F. Hawley, 1994, Mon. Not. R. Astron. Soc. **271**, 197.

Balbus, S. A., and J. F. Hawley, 1991, Astrophys. J. **376**, 214.

Balbus, S. A., and J. F. Hawley, 1992a, Astrophys. J. **392**, 662.

Balbus, S. A., and J. F. Hawley, 1992b, Astrophys. J. **400**, 610.

Balbus, S. A., J. F. Hawley, and J. M. Stone, 1996, Astrophys. J. **467**, 76.

Baptista, R., K. Horne, R. W. Hilditch, K. O. Mason, and J. E. Drew, 1995, Astrophys. J. **448**, 395.

Bayly, B. J., 1986, Phys. Rev. Lett. **57**, 2160.

Bayly, B. J., S. A. Orszag, and T. Herbert, 1988, Annu. Rev. Fluid Mech. **20**, 359.

Begelman, M. C., R. D. Blandford, and M. J. Rees, 1984, Rev. Mod. Phys. **56**, 255.

Bernstein, I. B., E. A. Frieman, M. D. Kruskal, and R. M. Kulsrud, 1958, Proc. R. Soc. London, Ser. A **244**, 17.

Biermann, L., 1950, Z. Naturforsch. **52**, 65.

Black, D. C., and M. S. Mathews, 1985, Eds., *Protostars and Planets II* (University of Arizona Press, Tucson).

Blaes, O. M., 1987, Mon. Not. R. Astron. Soc. **227**, 975.

Blaes, O. M., and S. A. Balbus, 1994, Astrophys. J. **421**, 163.

Blaes, O. M., and J. F. Hawley, 1988, Astrophys. J. **326**, 277.

Blandford, R. D., 1989, in *Theory of Accretion Disks*, edited by F. Meyer, W. J. Duschl, J. Frank, and E. Meyer-Hofmeister (Kluwer, Dordrecht), p. 35.

Blandford, R. D., and D. Payne, 1982, Mon. Not. R. Astron. Soc. **199**, 883.

Brandenburg, A., Å Nordlund, R. F. Stein, and U. Torkelsson, 1995, Astrophys. J. **446**, 741.

Brummel, N., F. Cattaneo, and J. Toomre, 1995, Science **269**, 5229; **269**, 1370.

Cabot, W., 1996, Astrophys. J. **465**, 874.

Cabot, W., and J. B. Pollack, 1992, Geophys. Astrophys. Fluid Dyn. **64**, 97.

Cameron, A. G. W., 1978, Moon Planets **18**, 5.

Canizzo, J. K., P. Ghosh, and J. C. Wheeler, 1982, Astrophys. J. Lett. **260**, L83.

Cattaneo, F., and S. I. Vainshtein, 1991, Astrophys. J. Lett. **376**, L21.

Chandrasekhar, S., 1960, Proc. Natl. Acad. Sci. USA **46**, 253.

Chapman, S., and T. G. Cowling, 1970, *Mathematical Theory of Nonuniform Gases* (Cambridge University, Cambridge, England).

Coles, D., 1965, J. Fluid Mech. **21**, 385.

Crawford, J. A., and R. P. Kraft, 1956, Astrophys. J. **123**, 44.

Curry, C., and R. Pudritz, 1996, Mon. Not. R. Astron. Soc. **281**, 119.

Donnelly, R. J., and M. Ozima, 1960, Phys. Rev. Lett. **4**, 497.

Draine, B. T., W. G. Roberge, and A. Dalgarno, 1983, Astrophys. J. **264**, 485.

Drazin, P. G., and W. H. Reid, 1981, *Hydrodynamic Stability* (Cambridge University, Cambridge, England).

Dwarkadas, V. V., and S. A. Balbus, 1996, Astrophys. J. **467**, 87.

Eardley, D. M., and A. P. Lightman, 1975, Astrophys. J. **200**, 187.

Evans, C. R., and J. F. Hawley, 1988, Astrophys. J. **332**, 659.

Frank, J., A. R. King, and D. J. Raine, 1985, *Accretion Power in Astrophysics* (Cambridge University, Cambridge, England).

Frieman, E., and M. Rotenberg, 1960, Rev. Mod. Phys. **32**, 898.

Frisch, U., and S. A. Orszag, 1990, Phys. Today **43**, 1,24.

Gammie, C. F., 1996, Astrophys. J. **457**, 355.

Gammie, C. F., 1997, private communication.

Gammie, C. F., and S. A. Balbus, 1994, Mon. Not. R. Astron. Soc. **270**, 138.

Gat, O., and M. Livio,, 1992, Astrophys. J. **396**, 542.

Gehrels, T., 1978, Ed., *Protostars and Planets* (University of Arizona, Tucson).

Giacconi, R., H. Gursky, F. R. Paolini, and B. B. Rossi, 1962, Phys. Rev. Lett. **9**, 439.

Goldreich, P., J. Goodman, and R. Narayan, 1986, Mon. Not. R. Astron. Soc. **221**, 339.

Goldreich, P., and D. Lynden-Bell, 1965, Mon. Not. R. Astron. Soc. **130**, 125.

Goldreich, P., and G. Schubert, 1967, Astrophys. J. **150**, 571.

Goldreich, P., and S. D. Tremaine, 1978, Astrophys. J. **222**, 850.

Goldreich, P., and S. D. Tremaine, 1979, Astrophys. J. **233**, 857.

Goodman, J., 1993, Astrophys. J. **406**, 596.

Goodman, J., and G. Xu, 1994, Astrophys. J. **432**, 213.

Goodstein, D., 1974, *States of Matter* (Prentice-Hall, Englewood Cliffs, NJ).

Hawley, J. F., 1991, Astrophys. J. **381**, 496.

Hawley, J. F., and S. A. Balbus, 1991, Astrophys. J. **376**, 223.

Hawley, J. F., and S. A. Balbus, 1992, Astrophys. J. **400**, 595.

Hawley, J. F., C. F. Gammie, and S. A. Balbus, 1995, Astrophys. J. **440**, 742.

Hawley, J. F., C. F. Gammie, and S. A. Balbus, 1996, Astrophys. J. **464**, 690.

Hawley, J. F., and J. M. Stone, 1995, Comput. Phys. Commun. **89**, 127.

Hayashi, C., 1981, Prog. Theor. Phys. Suppl. **70**, 35.

Hill, G. W., 1878, Am. J. Math. **1**, 5.

Horne, K., 1995, Astron. Astrophys. **297**, 273.

Howard, L. N., 1962, J. Fluid Mech. **13**, 158.

Howard, L. N., and A. S. Gupta, 1962, J. Fluid Mech. **14**, 463.

Jackson, J. D., 1975, *Classical Electrodynamics*, 2nd ed. (Wiley, New York).

Katz, J. I., 1985, in *Cataclysmic Variables and Low-mass X-Ray Binaries*, edited by D. Q. Lamb and J. Patterson (Reidel, Dordrecht), p. 359.

Kley, W., J. C. B. Papaloizou, and D. N. C. Lin, 1993, Astrophys. J. **416**, 679.

Knobloch, E., 1992, Mon. Not. R. Astron. Soc. **255**, 25p.

Kuiper, G., 1941, Astrophys. J. **93**, 133.

Kulsrud, R. M., and S. W. Anderson, 1992, Astrophys. J. **396**, 606.

Kulsrud, R. M., R. Cen, J. P. Ostriker, and D. Ryu, 1997, Astrophys. J. **480**, 481.

Lada, C. J., 1985, Annu. Rev. Astron. Astrophys. **23**, 267.
Larson, R. B., 1989, in *The Formation and Evolution of Planetary Systems*, Space Telescope Institute Symposium Series 3, edited by H. A. Weaver and L. Danly (Cambridge University, Cambridge England), p. 31.
Levy, E. H., and J. I. Lunine, 1993, *Protostars and Planets III* (University of Arizona Press, Tucson, Arizona).
Lin, C. C., and F. H. Shu, 1964, Astrophys. J. **140**, 646.
Lin, D. N. C., 1989, in *Theory of Accretion Disks*, edited by F. Meyer, W. J. Duschl, J. Frank, and E. Meyer-Hofmeister (Kluwer, Dordrecht), p. 89.
Lin, D. N. C., and J. C. B. Papaloizou, 1980, Mon. Not. R. Astron. Soc. **191**, 37.
Lin, D. N. C., and J. C. B. Papaloizou, 1996, Annu. Rev. Astron. Astrophys. **34**, 703.
Lin, D. N. C., J. C. B. Papaloizou, and G. Savonije, 1990, Astrophys. J. **364**, 326.
Livio, M., 1994, in *Interacting Binaries*, Saas–Fee Advanced Course 22, edited by H. Nussbaumer and A. Orr (Springer, Berlin), p. 142.
Lubow, S. H., J. E. Pringle, and R. R. Kerswell, 1993, Astrophys. J. **419**, 758.
Lynden-Bell, D., 1969, Nature (London) **223**, 690.
Lynden-Bell, D., and J. E. Pringle, 1974, Mon. Not. R. Astron. Soc. **168**, 603.
Marsh, T. R., and K. Horne, 1988, Mon. Not. R. Astron. Soc. **235**, 269.
Matsumoto, R., and K. Shibata, 1997, in *Accretion Phenomena and Related Outflows*, edited by D. Wickramsinghe, L. Ferrario, and G. Bicknell (Astronomical Society of the Pacific, San Francisco), p. 443.
Matsumoto, R., and T. Tajima, 1995, Astrophys. J. **445**, 767.
Matsumoto, R., Y. Uchida, S. Hirose, K. Shibata, M. R. Hayashi, A. Ferrari, G. Bodo, and C. Norman, 1996, Astrophys. J. **461**, 115.
Matsuzaki, T., R. Matsumoto, T. Tajima, and K. Shibata, 1997, in *Accretion Phenomena and Related Outflows*, edited by D. Wickramsinghe, L. Ferrario, and G. Bicknell (Astronomical Society of the Pacific, San Francisco), p. 766.
Meneguzzi, M., U. Frisch, and A. Pouquet, 1981, Phys. Rev. Lett. **47**, 1060.
Mineshige, S., M. Kusnose, and R. Matsumoto, 1995, Astrophys. J. **445**, 43.
Moffatt, K., 1978, *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge University, Cambridge, England).
Moreno, C., R. G. M. Rutten, and V. S. Dhillon, 1996, in *Cataclysmic Variables and Related Objects*, edited by A. Evans and J. H. Woods (Kluwer, Dordrecht), p. 15.
Narayan, R., 1996, in *Physics of Accretion Disks*, edited by S. Kato, S. Inagaki, S. Mineshige, and J. Fukue (Gordon and Breach, Amsterdam), p. 15.
Narayan, R., I. Yi, and R. Mahadevan, 1995, Nature (London) **374**, 623.
Newcomb, W., 1961, Phys. Fluids **4**, 391.
Novikov, I. D., and K. S. Thorne, 1973, in *Black Holes—Les Astres Occlus*, edited by C. De Witt (Gordon and Breach, New York), p. 346.
Ogilvie, C., and J. E. Pringle, 1996, Mon. Not. R. Astron. Soc. **279**, 152.
Oran, E. S., and J. P. Boris, 1993, Comput. Phys. **7**, 523.
Orszag, S. A., L. C. Kells, 1980, J. Fluid Mech. **96**, 161.
Papaloizou, J. C. B., and D. N. C. Lin, 1995, Annu. Rev. Astron. Astrophys. **33**, 505.
Papaloizou, J. C. B., and J. E. Pringle, 1984, Mon. Not. R. Astron. Soc. **208**, 721.
Papaloizou, J. C. B., and E. Szuszkiewicz, 1992, Geophys. Astrophys. Fluid Dyn. **66**, 223.
Parker, E. N., 1966, Astrophys. J. **145**, 811.
Parker, E. N., 1979, *Cosmical Magnetic Fields* (Oxford University Press, Oxford).
Petterson, J. A., 1983, in *Accretion Driven X-ray Sources*, edited by W. H. G. Lewin and E. P. J. van den Heuvel (Cambridge University, Cambridge, England), p. 367.
Pierrehumbert, R. T., 1986, Phys. Rev. Lett. **57**, 2157.
Prendergast, K. H., and G. R. Burbidge, 1968, Astrophys. J. Lett. **151**, L83.
Pringle, J. E., 1981, Annu. Rev. Astron. Astrophys. **19**, 137.
Porter, D. H., and P. R. Woodward, 1994, Astrophys. J., Suppl. Ser. **93**, 309.
Porter, D. H., P. R. Woodward, W. Yang, and Q. Mei, 1990, Ann. (N.Y.) Acad. Sci. **617**, 234.
Rees, M. J., 1984, Annu. Rev. Astron. Astrophys. **22**, 471.
Rees, M. J., 1993, in *Cosmical Magnetism*, edited by D. Lynden-Bell (Kluwer, Dordrecht), p. 155.
Reynolds, O., 1883, Philos. Trans. R. Soc. London **174**, 935.
Różyczka, M., and H. C. Spruit, 1993, Astrophys. J. **417**, 677; **417**, videotape, Segment 3.
Roberts, W. W., 1969, Astrophys. J. **158**, 123.
Ruden, S., J. C. B. Papaloizou, and D. N. C. Lin, 1988, Astrophys. J. **329**, 739.
Ryu, D., and J. Goodman, 1992, Astrophys. J. **388**, 438.
Ryu, D., and J. Goodman, 1994, Astrophys. J. **422**, 269.
Savonije, G., J. C. B. Papaloizou, and D. N. C. Lin, 1994, Mon. Not. R. Astron. Soc. **268**, 13.
Sawada, K., T. Matsuda, and I. Huchisa, 1986, Mon. Not. R. Astron. Soc. **219**, 75.
Schwarzschild, M., 1958, *Structure and Evolution of the Stars* (Dover, New York).
Shakura, N. I., and R. A. Sunyaev, 1973, Astron. Astrophys. **24**, 337.
Shapiro, S., and S. Teukolsky, 1983, *The Physics of Compact Objects* (Wiley, New York).
Shibata, K., and Y. Uchida, 1986, Publ. Astron. Soc. Jpn. **38**, 631.
Shu, F. H., 1991, *The Physics of Astrophysics I. Radiation* (University Science, Mill Valley).
Shu, F. H., V. Milione, and W. W. Roberts, 1973, Astrophys. J. **183**, 819.
Shu, F. H., S. Tremaine, F. C. Adams, and S. P. Ruden, 1990, Astrophys. J. **358**, 495.
Smak, J., 1982, Acta Astron. **32**, 199.
Spitzer, L., 1962, *Physics of Fully Ionized Gases* (Wiley Interscience, New York).
Spruit, H. C., 1987, Astron. Astrophys. **184**, 173.
Stone, J. M., and S. A. Balbus, 1996, Astrophys. J. **464**, 364.
Stone, J. M., and M. L. Norman, 1992a, Astrophys. J., Suppl. Ser. **80**, 753.
Stone, J. M., and M. L. Norman, 1992b, Astrophys. J., Suppl. Ser. **80**, 791.
Stone, J. M., and M. L. Norman, 1994, Astrophys. J. **433**, 746.
Stone, J. M., J. F. Hawley, C. R. Evans, and M. L. Norman, 1992, Astrophys. J. **388**, 415.
Stone, J. M., J. F. Hawley, C. F. Gammie, and S. A. Balbus, 1996, Astrophys. J. **463**, 656.

Tassoul, J.-L., 1978, *Theory of Rotating Stars* (Princeton University, Princeton).
Taylor, G. I., 1923, Philos. Trans. R. Soc. London, Ser. A **223**, 289.
Taylor, G. I., 1936, Proc. R. Soc. London, Ser. A **157**, 546.
Tennekes, H., and J. L. Lumley, 1972, *A First Course in Turbulence* (MIT Press, Cambridge, Mass.).
Terquem, C., and J. C. B Papaloizou, 1996, Mon. Not. R. Astron. Soc. **279**, 767.
Toomre, A., 1981, in *The Structure and Evolution of Normal Galaxies*, edited by S. M. Fall and D. Lynden-Bell (Cambridge University, Cambridge), p. 111.
Tout, C. A., and J. E. Pringle, 1992, Mon. Not. R. Astron. Soc. **259**, 604.
Uchida, Y., and K. Shibata, 1985, Publ. Astron. Soc. Jpn. **37**, 515.
Umbayashi, T., 1983, Prog. Theor. Phys. **69**, 480.
Velikhov, E. P., 1959, Sov. Phys. JETP **36**, 995.
Vishniac, E. T., 1995, Astrophys. J. **451**, 816.
Vishniac, E. T., and P. Diamond, 1989, Astrophys. J. **347**, 435.
von Neumann, J., and R. D. Richtmyer, 1950, J. Appl. Phys. **21**, 232.
Wardle, M., and A. Königl, 1993, Astrophys. J. **410**, 218.
Weinberg, S., 1972, *Gravitation and Cosmology* (Wiley, New York).
Wisdom, J., and S. Tremaine, 1988, Astron. J. **95**, 925.
Young, P., D. P. Schneider, and S. A. Shectman, 1981, Astrophys. J. **244**, 259.