GLOBAL INSTABILITIES OF DISKS*

J. M. BARDEEN

Yale University, New Haven, Conn., U.S.A.

Abstract. Current understanding of the stability of gas and stellar disks suggests very strongly that local stability to axisymmetric modes is not sufficient for global stability. A global instability to a bar mode will develop unless the rotational kinetic energy is sufficiently small compared with the random kinetic energy for the system as a whole. A disk as cool as the galactic disk near the Sun can survive only if most of the mass of the Galaxy is in a 'hot' component, such as a central bulge and/or an extended halo. We review the theoretical evidence for this conclusion coming from analytic results for simple gas and stellar disks, from numerical simulations of stellar disks, and from numerical calculations of the stability of gas disks. Some new results on the precise form of dynamic bar instabilities of gas disks with and without halos are reported.

1. Introduction

An understanding of the structure and dynamics of spiral galaxies depends in an important way on questions of global stability. Conventional fits to rotation curves, based on the observed light distribution, assume all the mass outside the 'central bulge' type of spheroidal component lies in a 'cool' disk with only enough velocity dispersion to satisfy the Toomre criterion for local stability against axisymmetric ring modes. However, attempts to simulate the stellar dynamics of these cool disks have consistently found strong instabilities to bar modes. The non-linear effect of the bar is to increase the velocity dispersion until it is comparable with the circular velocity over much of the disk. The final state is typically characterized by a ratio t of kinetic energy of rotation to gravitational potential energy of only 0.14 or so, compared with the value 0.5 for a zero velocity dispersion disk (Ostriker and Peebles, 1973). Ostriker and Peebles argue that a cool disk is stable against bar modes only if a substantial part of the mass of the galaxy is in the form of a more or less spherical halo, so that for the system as a whole t has an upper limit in the range 0.1–0.2.

No direct observational evidence for such halos has been found (Freeman, 1975). They must be composed of objects with very large mass to light ratios, such as extreme M dwarfs or black holes. Ostriker *et al.* (1974) have assembled some indirect evidence for very massive halos extending to many times the radius of the visible galaxy. Of course, only the mass within the outer edge of the disk is relevant to our discussion of stability. One should also keep in mind that the observational evidence for disks being cool is rather meagre; in the vicinity of the Sun the galactic disk is at best only marginally stable according to Toomre's criterion (see Schmidt, 1975), but in other galaxies one must appeal to the apparent small ratio of thickness to radius or argue that spiral density waves are only possible in rather cool disks.

My discussion will focus on the theoretical arguments and results of numerical simulations as they affect questions of global stability. The most directly relevant

* Supported in part by National Science Foundation Grant GP-36317

Hayli (ed.), Dynamics of Stellar Systems, 297–320.

calculations are the numerical simulations of stellar disks carried out by Miller *et al.* (1970), Hohl (1971, 1975); and Ostriker and Peebles (1973). I will only mention some of the general implications of Hohl's calculations, since he will describe them in detail in his review paper. The analytic results of Kalnajs (1972) on stability are of great help in checking and interpreting the simulations, but so far they have not been extended to realistic galaxy models.

Bar instabilities seem to be a general property of rapidly rotating, self-gravitating systems, with many qualitative and some quantitative similarities between stellar disks and gaseous disks. Furthermore, the only obvious way for a stellar disk to form is from a gas disk at least as thin as the final stellar disk. Therefore, this review will include some discussion of gas disks; in particular, I will present the results of some of my calculations of the instability of two-dimensional gas disks to bar and two-armed spiral modes. The linearized equations for dynamical perturbations are integrated in time until, if the disk is unstable, one mode dominates and gives a steady, uniform pattern speed and growth rate.

I will begin by reviewing the known analytic results on the stability of gas and stellar disks, both for global and local instability and then discuss the numerical calculations.

2. Analytic Results

Only for very special configurations can one solve analytically for the gravitational potential, both for the axisymmetric equilibrium and for general perturbations. The classic example is the sequence of Maclaurin spheroids, uniform density and uniformly rotating spheroids of incompressible perfect fluid, along with the associated sequences of ellipsoids which have various combinations of internal motion and uniform rotation (see Chandrasekhar, 1969). For a given density and mass the eccentricity of a Maclaurin spheroid goes form zero to one and the ratio t of kinetic energy of rotation to gravitational potential energy goes from zero to one-half as the angular momentum increases from zero to infinity.

In the gas case one distinguishes between secular instability and dynamic instability. In the absence of any dissipative mechanism such as viscosity both the total angular momentum and the circulation ($\int \mathbf{v}\cdot d\mathbf{l}$ around a closed curve comoving with the fluid) are conserved in any dynamic process. Furthermore, an axisymmetric perturbation conserves the angular momentum of each ring of matter.

A perturbation in which one of the dynamic constraints is violated is a secular perturbation. For instance, with only viscosity present the circulation need not be conserved and angular momentum can be transferred between neighboring rings of matter even when the perturbation is axisymmetric, but the total angular momentum still remains constant. On the other hand, gravitational radiation reaction by itself preserves the circulation but allows the total angular momentum to change. The time scale for growth of a secular perturbation is governed by the amount of dissipation present and often is large compared with the characteristic dynamic time scale of the system. The concept of secular instability is meaningful only if the original

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

equilibrium configuration is not affected by the dissipation. One expects secular instability to set in before dynamic instability, since a neighboring lower energy configuration may not be reachable by a perturbation consistent with all the dynamical constraints.

The first instability along the Maclaurin sequence is a secular instability to a bar mode at $t \geqslant 0.1376$. As excited by viscosity the marginally unstable mode is an ellipsoidal deformation which corotates with the matter (a Jacobi ellipsoid). As excited by gravitational radiation reaction the ellipsoidal shape remains fixed in space, a Dedekind ellipsoid (Chandrasekhar, 1970). The non-linear evolution of the secular instability for t a little greater than 0.1376 has been solved by Press and Teukolsky (1973) for viscosity and by Miller (1973) for gravitational radiation.

Only for $t > 0.2738$ are the Maclaurin spheroids dynamically unstable. The marginally unstable mode is a bar which rotates at one-half the angular velocity of the matter. At $t = 0.3589$ the spheroid first becomes secularly unstable to an axisymmetric redistribution of angular momentum and matter, either in the direction of a 'central bulge' surrounded by a thin disk or in the direction of a ring-like distribution of matter. Finally, the first onset of an axisymmetric dynamic instability is at $t = 0.4574$ (see Bardeen, 1971).

A class of stellar dynamic analogues to Maclaurin spheroids has been studied by Kalnajs (1972), who was able to separate the normal modes and solve for their characteristic frequencies. The models are infinitesimally thin disks with the same surface density as a function of radius as the Maclaurin spheroids and with a range of possible velocity distributions in the plane of the disk which are obtained by superimposing a particular type of distribution function with only one free parameter, the mean angular velocity of rotation Ω. The stability to the two-armed bar mode analogous to the mode that dominates the non-axisymmetric instability of the Maclaurin spheroids, Kalnajs shows, depends only on the weighted average of Ω in a composite model. The bar mode is unstable if $\langle \Omega \rangle$ exceeds $(125/486)^{1/2} = 0.507$, in units such that the angular velocity Ω_0 of the zero velocity dispersion disk with the same mass and radius is one. Since $\langle \Omega \rangle$ is the mean angular velocity of the matter in the composite model, the parameter t at marginal instability has the unique value $t = 0.1286$. Thus $t \leqslant 0.1286$ is a necessary condition for overall stability of this class of models. Unfortunately, the velocity distribution has some peculiar properties, and even aside from the question of the importance of shear, one should be cautious in extrapolating this result to more realistic stellar disks.

Kalnajs also considers how a halo which supplies a rigid component to the potential modifies the stability of the disk. For instance his model B_1 without a halo has $t = 0.1735$ and is unstable to three modes; the most rapidly growing instability is to the (2, 2) bar mode. With a halo corresponding to a uniform density sphere equal in radius to the disk, the model which has the same distribution of stellar orbits as B_1 is marginally stable when the halo mass is 0.199 of the disk mass or 0.166 of the total mass. For the system as a whole $t = 0.145$. His model B_3, which without a halo has $t = 0.271$, is stabilized by a halo mass equal to $\frac{2}{3}$ the mass of the disk or 40% of the

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

total mass. The effect of the halo is to multiply the response by a factor equal to the fraction of the equilibrium gravitational force due to the disk.

There is no distinction between secular and dynamic instability in a collisionless stellar system. The conservation of density in phase space is a constraint on the evolution of the system, but one which can allow a considerable amount of relaxation and conversion of bulk motion into random motion to occur (see Lynden-Bell, 1967). There are no macroscopic constraints analogous to conservation of vorticity of a perfect fluid. It is perhaps not surprising, then, that the bar instability for the stellar disk occurs at about the same value of t as the secular instability of the Maclaurin spheroid. The pattern angular velocity of the marginally unstable mode is 0.4564, roughly equal to the mean angular velocity of the stars, so the marginally unstable mode of the stellar disk resembles the Jacobi mode of the Maclaurin spheroid.

Typical galactic disks are highly centrally condensed and are far from uniform rotation. The only generally applicable analytic stability criteria are the 'local' criteria which apply to axisymmetric ring modes or tightly wound spiral modes in which the radial wavelength of the perturbation is small compared with the scale of radial inhomogeneities. In this limit one treats a local region of the disk as an infinite plane sheet.

The Toomre criterion (Toomre, 1964) was derived for infinitesimally thin stellar disks with a Gaussian radial velocity dispersion c_r. It predicts instability when c_r falls below

$$(c_r)_{\text{crit}} = \frac{3.36\, G\sigma}{\kappa},$$

where κ is the local epicyclic frequency and σ is the local surface density. The wavelength of the marginally unstable mode is

$$\begin{aligned}\lambda_{\text{crit}} &= 3.5(c_r)^2_{\text{crit}}/G\sigma \\ &= 40\, G\sigma/\kappa^2.\end{aligned}$$

For a completely self-gravitating disk λ_{crit} is comparable with the radius of the disk unless the surface density in the region being considered is small compared with the average surface density. Therefore, the conditions necessary for the local criterion to be used with confidence will not be satisfied unless the disk contains a relatively small part of the total mass of the system. The bulk of the mass could either be in a 'central bulge' or 'spheroidal' component or in an extended halo component. This point has also been made by Vandervoort (1970) in connection with a study of non-linear density waves.

A local stability criterion for gas disks was derived by Goldreich and Lynden-Bell (1965). If a measure of the mean density in the disk,

$$\bar{\varrho} = \int \varrho^2\, dz \Big/ \int \varrho\, dz,$$

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

exceeds a numerical coefficient times κ^2/G the disk is unstable. The numerical coefficient ranges from 0.23 for an isothermal equation of state to 0.56 for an incompressible fluid equation of state. The critical wavelength at marginal instability is a few times the mean thickness of the disk $W=\sigma/\bar{\varrho}$. The ratio λ_c/W ranges from 4.5 for an isothermal equation of state to 10.4 for an incompressible fluid equation of state.

The exact result for marginal instability of a Maclaurin spheroid to axisymmetric modes is $\varrho=0.61\ \kappa^2/G$ when W is 0.0465 times the diameter of the disk. However, it seems likely that the local criterion will work less well when σ and κ^2 vary strongly with radius.

Even when such a local criterion does apply it gives at most a necessary condition for stability. It says nothing about the stability of modes which are not tightly wound, particularly the bar modes which are the dominant instabilities of the uniformly rotating disks, both gas and stellar.

Nevertheless, it is the local type of analysis which has formed the basis for the density wave theory of spiral structure developed by Lin and Shu (1964, 1966). They and coworkers have obtained dispersion relations and included effects of resonances. Results of Toomre (1969) on the group velocity of density waves have raised serious question as to whether they can be considered persistant normal modes of the stellar system. Further discussion of density waves as such seems outside the scope of this review.

3. Numerical Simulations of Stellar Disks

Direct N-body calculations are impractical for N larger than 500. It is questionable whether 500 stars are adequate to represent a galaxy, since the two-body relaxation time is not too much larger than the dynamical time scale. Certainly 500 stars are not enough to represent any detailed structure in the galaxy. These considerations have prompted the development of approximate schemes which allow calculations with much larger numbers of stars.

For instance, Miller and Prendergast (1968) devised an approximation based on a discrete phase space in which stars jump between integer values of position and velocity. A fast Fourier transform technique was used to solve for the gravitational potential. Miller, Prendergast, and Quirk (1970) applied this method to infinitesimally thin disks made up both of 'stars' and 'gas', with about 10^5 total particles. The gas component was kept cool by reducing the relative velocities of the gas particles or 'clouds' at each point in configuration space at every time step. Without 'gas' present the stars formed a 'hot' pressure-supported system. With gas some moderately persistent spiral patterns developed which were much less prominent in the stars than in the gas. In some cases bars developed as transient phases in the evolution of the system. No attempt was made to look in detail at the conditions necessary to stabilize the disk, but typically there was a substantial background of hot stars once things settled down to a steady state. Ostriker and Peebles (1973) quote a range of 0.130–0.135 for the final value of t in the MPQ simulations.

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

More elaborate and more accurate simulations of infinitesimally thin two-dimensional purely stellar disks have been carried out by Hohl over the last several years (Hohl, 1971, 1975). The configuration space is broken up into cells and the mass density averaged over each cell before the potential is calculated. However, unlike MPQ the equations of motion for the individual stars are solved exactly for an infinitesimally thin disk with velocity dispersion only in the plane. The simulations have been tested against the Kalnajs analytic models. Most of the models calculated have contained 10^5 stars, which is large enough for two-body encounters to have a negligible effect (Hohl, 1973). Hohl found that an initially uniformly rotating disk with just enough velocity dispersion to satisfy the Toomre stability criterion was violently unstable to a bar. Outside the corotation radius, where the pattern speed of the bar equaled the circular velocity of the stars, the stars were pushed outward; inside the corotation radius the stars drifted inward. In the final steady state the surface density was a steep exponential function of radius inside the corotation radius; outside the corotation radius the disk was also exponential, but with a considerably larger scale length. The ratio Q of the radial velocity dispersion to the Toomre critical velocity dispersion in the final state varied from about 2 near the center to 5 or 6 in the outer part of the disk. Ostriker and Peebles (1973) quote a final value of $t \simeq 0.14$. Any transient spiral patterns quickly decayed as the velocity dispersion increased, but a bar did persist indefinitely.

In an attempt to generate a cooler stable disk Hohl symmetrized this final steady state to remove the bar and then tried artificially reducing the velocity dispersion of a certain fraction of the stars peroidically. However, this caused a new bar instability. With steady cooling an open spiral pattern did persist in the outer part of the disk surrounding a central bar. Heating of the disk by Landau damping of the spiral pattern compensated for the cooling and kept the disk rather hot.

The most striking feature of these calculations (Hohl, 1971) was the relaxation to an exponential type of disk structure, somewhat like that observed in many spiral galaxies (Freeman, 1970). This suggests that the exponential surface density reflects an epoch of dynamic relaxation in the disk, rather than the initial angular momentum distribution in the gas cloud out of which the disk formed. A phase of bar instability may also greatly increase the central condensation of the disk and assist in the formation of a condensed nucleus of the galaxy.

More recently Hohl has run experiments with rigid background potentials representing 'halo' populations superimposed on the self-gravity of the disk (Hohl, 1975). A halo potential corresponding to the mass distribution in the Schmidt (1965) model of the Galaxy seemed to suppress the bar instability if the halo mass was comparable to that of the disk, but substantial heating of the disk still took place if the initial Q was one, suggesting that less prominent instabilities were still present. The steady-state Q was in the range 2–3. A less centrally condensed halo of comparable mass allowed a bar to develop, but the heating was less with a final Q of 1.5–2. At present these results are only suggestive, but combined with my results on gas disks reported in Section 4, one can conclude that a halo less centrally condensed than the disk seems to be more effective in stabilizing the disk than a halo corresponding to a massive 'central bulge' component.

Hopefully in the near future it will be possible to make more realistic simulations of this type which allow for a finite thickness of the disk and the presence of a genuine gas component. However, it seems unlikely that the results on global stability will be altered significantly.

An attempt to model a galactic disk with a finite thickness has been carried out by Ostriker and Peebles (1973). Their N-body calculation has N in the range 150–500 and usually $N=300$. The forces are calculated by summing over particle pairs; force is cut off at an interparticle distance about 0.05 times R, the initial radius of the disk. The initial conditions typically have all the stars in the equatorial plane evenly divided between intervals of $0.1R$ in radius and distributed randomly in each interval. The stars are given an initial radial velocity dispersion scaled to satisfy the Toomre criterion by a margin of 20% everywhere except near the centre. The initial z-velocity dispersion was set equal to the initial axial-velocity dispersion obtained from the epicyclic approximation. Finally, the total velocity, including the circular velocity, was rescaled to satisfy the equilibrium virial theorem.

Such a system developed a strong bar instability. The velocity dispersion increased rapidly in the plane, but after an initial adjustment the vertical velocity dispersion and the root mean square value of z for the disk as a whole increased more slowly. The parameter t, initially about 0.35, seemed to approach an asymptotic constant value of about 0.14.

Some runs included a rigid halo potential corresponding to a spherical mass distribution with about the same central concentration as the disk,

$$\begin{aligned} M_H(r) &= (1.1)^2 \frac{r^3 M_H}{R(r+0.1R)^2} \qquad && 0 \leqslant r \leqslant R \\ &= M_H && r > R. \end{aligned}$$

A halo mass M_H greater than or equal to the mass of the disk was sufficient to remove the violent bar instability, but the velocity dispersion still did increase slowly. For the system as a whole the parameter t seemed to level off at about 0.17 for $M_H/M_D = 0.5$ and 1.0. Once the halo mass dominates that of the disk t becomes very insensitive to the velocity dispersion of the disk; a better measure of the velocity dispersion necessary for stability would be the Toomre parameter Q. It does seem that with $M_H/M_D \gtrsim 2$ or 3 the initial disk is basically stable.

Ostriker and Peebles claim their results are insensitive to the value of N in the range 150–500 and therefore that two-body relaxation effects are not important. While this probably is true for the violent bar instability of the cool disk without a halo, the precise point at which the halo stabilizes the initial disk is unclear because a slow increase of velocity dispersion could be due either to a mild instability or to two-body relaxation. The cut-off of the halo at the initial radius of the disk is rather artificial, particularly since the radius of the disk increases substantially during the dynamical evolution. The value of t is very sensitive to the amount of mass in the

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

outer part of the halo, but the stability is not likely to be sensitive to the halo mass outside the great bulk of the mass of the disk. For this reason, Ostriker and Peebles' claim that $t \simeq 0.14$ is a universal number for marginal bar instability, even with a halo, should not be taken too seriously.

Taken together the numerical stimulations and the Kalnajs analytic results constitute a very strong, even overwhelming case that without a massive halo a stable stellar disk must have a much higher velocity dispersion in the plane than that required for local axisymmetric stability. The numerical simulations generally have not used self-consistent stationary solutions to the Liouville equation as initial conditions, so they are not true stability calculations. However, the stationary final states presumable do correspond to stationary solutions of the Liouville equation and attempts to gradually cool these differentially rotating disks did not succeed in reducing the total velocity dispersion by a large amount. Therefore, it seems unlikely that single-component stellar disks can be globally stable with t much greater than 0.14 or a mass-weighted average of Q^{-1} greater than 0.3 or so. However, it is still very important to extend the exact stability analysis to at least some differentially rotating disks and to construct good equilibrium stellar dynamical models of disks with various degrees of central condensation to serve as initial conditions for numerical simulations. The one possibility that has not been ruled out by the numerical simulations to date is that a very hot 'central bulge' in which Q is much larger than one could allow a surrounding disk containing the bulk of the mass (as observed from rotation curves) to be stable with $Q \simeq 1$. My experiments with gas disks make this alternative seem unlikely.

A less realistic version of a stellar disk, but one which is susceptible to a fairly complete stability analysis, uses a modified gravitation interaction potential,

$$\phi_{ij} = -\frac{m_i m_j}{(r_{ij}^2 + a^2)^{1/2}}, \tag{1}$$

to soften the gravitational interaction. The stars move in circular orbits in the equatorial plane of the unperturbed disk. Miller (1971) first noticed that the modified gravity, which with small values of a is often used in N-body calculations to prevent large accelerations in close encounters, can by itself stabilize the axisymmetric ring modes of a zero-velocity-dispersion disk. The WKB dispersion relation relating frequency ω and radial wave number k is

$$\omega^2 = \kappa^2 - 2\pi G\sigma k e^{-ak}. \tag{2}$$

This is qualitatively similar to the dispersion relation of Lin and Shu (1966) in that $\omega^2 \to \kappa^2$ in the high wave number limit as well as at zero wave number. The dispersion relation (2) predicts local axisymmetric stability if

$$a > a_{\text{crit}} = \frac{2\pi G\sigma}{e\kappa^2}. \tag{3}$$

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

Miller (1974) and Erickson (1974) have found this WKB estimate agrees rather well with more exact numerical stability calculations for the axisymmetric modes of some simple disk models. Erickson also looks in detail at the whole spectrum of axisymmetric modes. However, the modified gravity fails to stabilize against some non-axisymmetric modes even when a is comparable with effective radius of the disk (Erickson, 1974); global stability does seem to require a real velocity dispersion.

4. Global Instabilities of Gas Disks

Gas disks are simpler to work with than stellar disks in that fewer dynamic variables describe the response of the matter. As mentioned in the introduction, they also have a direct relevance to the existence of disk galaxies. A stellar system cannot relax into a disk. Therefore, the disk stars must have formed from gas that was already in a thin disk. The original gas disk must not have been too unstable or it would not have lasted in a quiescent state long enough to form stars with small velocity dispersion.

A completely realistic gas disk would still require an elaborate computer simulation to follow its dynamical evolution, since it would involve two or three-dimensional hydrodynamics depending whether the calculation is restricted to small perturbations of an equilibrium configuration or follows the full non-linear development of instabilities.

A relatively simple type of approximate stability analysis is based on the tensor virial equations developed by Chandrasekhar and Lebovitz (1962) and adapted for use on rapidly rotating stellar models by Tassoul and Ostriker (1968). The fluid displacements are constrained to be linear functions of the Cartesian spatial coordinates,

$$\xi_i = A_{ij}(t)\, x_j. \tag{4}$$

Assuming a harmonic time dependence, one can derive a set of algebraic equations for the coefficients A_{ij} which has solutions for certain characteristic frequencies ω_k. The modes which have an angular dependence corresponding to a $l=2$, $m=\pm 2$ spherical harmonic are the 'bar' modes and in fact are the exact bar modes of a Maclaurin spheroid.

The tensor virial equations predict dynamical instability when the characteristic frequencies of the bar modes become complex. Since in general one does not expect the exact eigenfunction for a differentially rotating, centrally condensed star to have the form of equation (4), one does not expect the tensor virial estimate of, say, the value of t at the point of marginal dynamical instability to be exact. There is no minimum principle for non-axisymmetric modes (Lynden-Bell and Ostriker, 1967), so the sign of the error is not known.

As far as secular stability goes, Ostriker and Tassoul (1969) argue that any stationary, non-axisymmetric, differentially rotating configuration with triplanar symmetry must satisfy virial relations which in the limit of axisymmetry make one of the bar mode frequencies of the axisymmetric star calculated with the tensor virial equations equal to zero. In other words, there must be a zero frequency tensor virial mode at any point of bifurcation to a sequence of triplanar configurations along a sequence of axisymmetric rotating stars. One can show from the variational principle for differentially rotating stars of Lynden-Bell and Ostriker (1967) that secular instability will set in when a certain energy integral first becomes negative for some trial displacement and furthermore (see Friedman and Schutz, 1975) that a zero-frequency Dedekind-type mode does exist when some non-trivial trial displacement corresponding to an axial eigenvalue $m=2$ minimizes the energy integral at a value of zero. There is no guarantee that a Jacobi-type mode with a non-zero pattern angular velocity exists at this point. Taken together, these arguments imply that the tensor virial method locates the point of marginal secular instability exactly.

The tensor virial stability analysis has been applied by Ostriker and various collaborators to a variety of rapidly rotating stellar models calculated by the self-consistent field method of Ostriker and Mark (1968). The most systematic study has been carried out for the differentially rotating polytrope models of Ostriker and Bodenheimer (1973). At the point of marginal secular instability they find $t \simeq 0.138$ for all their models, while dynamical instability seems to set in at $t \simeq 0.26$. The models cover a fairly wide range of central condensation and angular momentum distribution.

None of the Ostriker-Bodenheimer models are as thin as typical galactic disks and the general validity of the tensor virial estimate of $t \simeq 0.26$ for dynamical instability is open to question. Also, the effect of halos on the instability of gas disks deserves attention. Therefore, I have developed a simplified treatment of the dynamics of gas disks perturbed away from axisymmetry in which the vertical structure of the disk is suppressed, either by assuming the disk is infinitesimally thin and the 'pressure' only acts in the plane of the disk or by solving for the vertical structure of the disk analytically and inserting this into the horizontal equilibrium and dynamics.

In the case of the infinitesimally thin disk we further simplify things by assuming that the horizontal stress per unit length p is a universal function of the surface density σ both in the equilibrium disk and in the dynamics of a given fluid element, $p=p(\sigma)$. Then the dynamical equations governing motions in the plane of the disk can be written.

$$\frac{d\mathbf{v}}{dt}=\nabla(U-\psi(\sigma)), \tag{5}$$

where

$$\psi(\sigma)=\int_0^\sigma \sigma^{-1}(dp/d\sigma)\,d\sigma \tag{6}$$

and U is the (positive) gravitational potential in the plane of the disk obtained by solving the Laplace equation outside the disk with boundary condition

$$\left.\frac{\partial U}{\partial z}\right|_{z=0^+} = -2\pi G\sigma. \tag{7}$$

Equation (5), linearized about the equilibrium disk, has been used by several authors as a simple context in which to discuss local density waves (e.g. Hunter, 1972, 1973; Feldman and Lin, 1973).

To first order in the thickness of an ordinary isotropic-pressure gas disk in which there is a polytropic relation between pressure and volume gas density the equation governing the horizontal dynamics has the same form as Equation (5), if in Equation (5) $U=U_0$, the potential in the plane of an infinitesimally thin disk with the same surface density as the finite thickness disk. To zeroth order in the thickness the vertical acceleration is small compared with the vertical gravitational potential gradient,

$$\ddot{z}=\mathcal{O}\left[(GM/R^3)\,z\right]=\mathcal{O}(z/R)\,G\sigma.$$

Assume vertical equilibrium, then, and also assume that the thickness is small compared with the scale of horizontal variations in the surface density. The vertical equilibrium gives

$$U(r,\phi,z)=\int_0^p \frac{\mathrm{d}p}{\varrho(p)}+U_1(r,\phi), \tag{8}$$

where U_1 is the value of U at the surface of the disk, $z=z_1(r,\phi)$. In the equation governing horizontal dynamics, the net horizontal gravitational and pressure force per unit mass is $\nabla U_1(r,\sigma)$, independent of z. To first order in z the potential at $z=z_1$ is the same as the potential at $z=z_1$ outside the zero thickness disk,

$$\begin{aligned}U_1(r,\phi)&=U_0(r,\phi)+\frac{\partial U}{\partial z}\,z_1=\\&=U_0(r,\phi)-2\pi G\sigma z_1.\end{aligned} \tag{9}$$

For a polytrope of the form

$$\varrho=K\,p^{\frac{n}{n+1}}, \tag{10}$$

Harrison and Lake (1972) find that z_1 for an infinite plane sheet is

$$z_1=\frac{\sigma}{2K}\left(\frac{\pi G}{2}\,\sigma^2\right)^{-n/(n+1)}\zeta_1, \tag{11}$$

where ζ_1 is a numerical coefficient depending on n. The equation for motions in the plane then has the form of Equation (5) with $\psi\propto\sigma^{2/(n+1)}$. The pressure forces and the reduction in U due to the finite thickness contribute to ψ in comparable amounts.

The correction to the horizontal dynamics second order in z requires taking into account vertical accelerations as well as the second order corrections to $U_1(r,\phi)$ which depend on the radial variation of U. Equation (5) can represent a thin isotropic pressure gas disk only as long as the scale of radial variation of both the equilibrium and perturbed surface density is large compared with the thickness. A local stability analysis based on Equation (5) will give only roughly the same result as the

Goldreich-Lynden-Bell local axisymmetric stability criterion, since the wavelength at marginal stability is only a few times the thickness.

My numerical analysis of global instabilities has been applied to disks whose horizontal dynamics is described by Equation (5). Only some of the results will be discussed here; the details will be published elsewhere.

From now on all equations and quantities will be written in units such that the outer radius of the disk $R=1$, the mass of the disk $M_D=1$, and the gravitational constant $G=1$.

Only first-order perturbations of equilibrium models will be considered. The models discussed in most detail will have a $\psi(\sigma)$ of the form

$$\psi=\beta(\pi\sigma)^{\alpha}. \tag{12}$$

The potential U in the plane of the disk is found by separating variables in oblate spheroidal coordinates (Hunter, 1963, 1965), so it is convenient to make the square of the angular velocity of the zero-pressure equilibrium model, Ω_0^2, a polynomial in η^2, where

$$\eta=(1-r^2)^{1/2}. \tag{13}$$

The corresponding potential U_D in the plane of the disk is found by integrating

$$\Omega_0^2=\frac{1}{\eta}\frac{\mathrm{d}U_D}{\mathrm{d}\eta}; \tag{14}$$

the surface density derived from U_D is a polynomial containing only odd powers of η. The Ω_0^2 polynomial is adjusted to make $\sigma\propto\eta^3$ as $\eta\to 0$ at the rim. Then the angular velocity of the 'warm' disk Ω, given by

$$\Omega^2=\frac{1}{\eta}\frac{\mathrm{d}}{\mathrm{d}\eta}(U-\psi(\sigma)), \tag{15}$$

is a regular function of η^2 at the rim if α in Equation (12) equals $\frac{2}{3}$ or $\frac{4}{3}$, corresponding to a finite thickness disk with polytropic index $n=2$ or $\frac{1}{2}$, respectively. The potential U in (15) may include a contribution U_H from a 'halo',

$$U=U_D+U_H. \tag{16}$$

If the halo is spherical, the halo mass inside radius r is

$$M_H(r)=r^3\Omega_H^2=r^3\left[\frac{1}{\eta}\frac{\mathrm{d}U_H}{\mathrm{d}\eta}\right]. \tag{17}$$

A remarkable property of the choice $\alpha=\frac{2}{3}$ in Equation (12) is that for reasonable disk models in which σ is roughly exponential over a substantial fraction of the radius, the fractional correction to Ω_0^2 produced by the finite pressure is remarkably uniform as the surface density changes by a couple of orders of magnitude. This allows a choice of β which gives a substantial average reduction in Ω^2 from Ω_0^2 without Ω^2 becoming negative near the center or near the rim. Only for $\alpha = \frac{2}{3}$ was I able to increase the pressure enough to stabilize the disk without a halo.

The structure of one standard disk model, Model C, is shown in Figure 1. The central surface density is 16.61 times the average surface density. The disk is roughly an exponential disk with a scale height of 0.16 in radius. The velocities of rotation for the zero-pressure disk and for the disk with $\alpha = \frac{2}{3}$ and $\beta = 0.4601$ are also plotted.

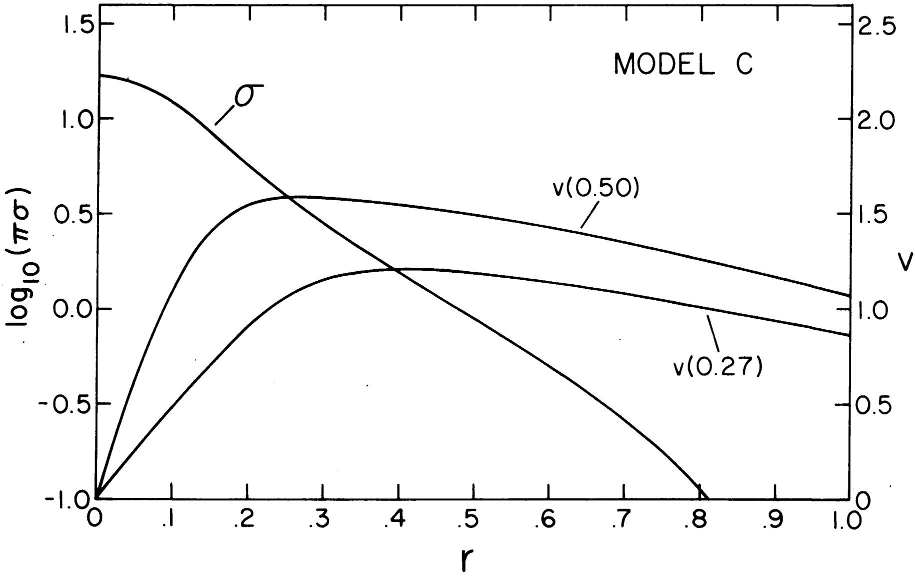


Fig. 1. The surface density σ and rotational velocity of Disk C are plotted. The rotational velocity $v(0.50)$ is for the zero pressure disk and the rotational velocity $v(0.27)$ is for a disk almost dynamical stable to the bar mode.

The ratio of kinetic energy of rotation to gravitational potential energy, t, labels the velocity curves; it is the value estimated for an isotropic pressure, finite thickness disk. The infinitesimally thin disk with $\beta = 0.4601$ has $t = 0.2341$.

The equations governing dynamic perturbations are obtained by expanding Equation (5) to first order in the deviation from equilibrium. It is instructive to consider harmonic time dependence and, of course, a particular axial harmonic; so the perturbed quantities all contain a factor $e^{im\phi - i\omega t}$. Then the perturbations in the velocity are algebraically related to the perturbations in the force per unit mass,

$$[\kappa^2 - (\omega - m\Omega)^2]\,\delta v_r = -i(\omega - m\Omega)\frac{\mathrm{d}}{\mathrm{d}r}(\delta U_{\mathrm{eff}}) + \frac{2im\Omega}{r}(\delta U_{\mathrm{eff}}) \tag{18}$$

and

$$[\kappa^2 - (\omega - m\Omega)^2]\,\delta v_\phi = -\frac{\kappa^2}{2\Omega}\frac{\mathrm{d}}{\mathrm{d}r}(\delta U_{\mathrm{eff}}) + \frac{m(\omega - m\Omega)}{r}(\delta U_{\mathrm{eff}}), \tag{19}$$

where

$$\delta U_{\mathrm{eff}} = \delta U_0 - \frac{\mathrm{d}\psi}{\mathrm{d}\sigma}\,\delta\sigma. \tag{20}$$

These equations are supplemented by the conservation of mass equation

$$i(\omega - m\Omega)\,\delta\sigma = \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}(\sigma r\,\delta v_r) + \frac{im\sigma}{r}\,\delta v_\phi. \tag{21}$$

In Equations (18) and (19) the 'epicyclic frequency' κ is defined by

$$\kappa^2 = 2\Omega\left(2\Omega + r\frac{\mathrm{d}\Omega}{\mathrm{d}r}\right); \tag{22}$$

it is not the epicyclic frequency for a circular test particle orbit.

A special cancellation on the right-hand side of one of the above equations is necessary to avoid a singularity in the eigenfunction at a radius where the pattern angular velocity of the perturbation $\Omega_p \equiv \omega/m$ equals the local angular velocity of the gas Ω (corotation resonance) or at a radius where

$$\omega/m = \Omega_p = \Omega + \kappa/m \qquad \text{(outer Lindblad resonance)}$$

or

$$\omega/m = \Omega - \kappa/m \qquad \text{(inner Lindblad resonance).}$$

These resonances act like extra boundary conditions on the eigenfunction. One boundary condition is already used up in requiring regularity at the center (the rim is a free boundary), so more than one resonance makes it difficult for any *regular* real-frequency modes to exist.

The non-local relation between the surface density perturbation $\delta\sigma$ and the potential perturbation δU_0 prevents the direct integration of Equations (18)–(20). However, if the radial wavelength of the perturbation is small compared with the radial scaleheight of surface density and small compared with r/m (a tightly wound spiral), there is an approximate local relation

$$\delta U_0 \simeq 2\pi\,\delta\sigma/k, \tag{23}$$

where k is the radial wavenumber. In this limit Equations (18)–(21) combine to give a dispersion relation (see Hunter, 1972)

$$(\omega - m\Omega)^2 = \kappa^2 - 2\pi\sigma k + \sigma\frac{\mathrm{d}\psi}{\mathrm{d}\sigma}k^2. \tag{24}$$

The minimum of $(\omega - m\Omega)^2$ is at a wavenumber

$$k_c = [\mathrm{d}\psi/\mathrm{d}(\pi\sigma)]^{-1} \tag{25}$$

or a wavelength

$$\lambda_c = 2\,\mathrm{d}\psi/\mathrm{d}\sigma. \tag{26}$$

The minimum value is less than zero, implying local instability, if

$$Q^2 \equiv \frac{\kappa^2}{\pi\sigma}\frac{\mathrm{d}\psi}{\mathrm{d}(\pi\sigma)} < 1. \tag{27}$$

The parameter Q is the analogue of Toomre's local stability parameter for stellar disks.

The actual method used to study the global stability of these disks was to solve numerically the initial value problem. For a particular angular eigenvalue m ($m=2$ for all cases discussed here) the radial dependence of the perturbations are represented by sums over associated Legendre functions $P_l^m(\eta)$, in such a way that the boundary conditions at the center and at the rim are automatically satisfied. The dynamic equations then reduce to an infinite set of ordinary differential equations for the coefficients C_l of the Legendre functions,

$$\frac{dC_l}{dt}=\sum_{l'=m}^{\infty} M_{ll'}C_{l'}. \tag{28}$$

The coefficients of the matrix $M_{ll'}$ depend only on the equilibrium structure of the disk. In practice the Legendre expansion was truncated at about 20–30 terms (counting only even values of $l-m$), enough to represent a moderately complicated eigenfunction accurately.

Starting from arbitrary initial conditions one mode, the most rapidly growing unstable mode, will eventually dominate if the disk is in fact unstable. By first choosing a small value of β, so the growth rate is large, then using the eigenfunction obtained as the initial condition for a model with a somewhat larger value of β, and so on, one can find the value of β required for marginal instability and the marginally unstable eigenfunction without an inordinate expenditure of computer time.

Table I shows the approach to stability for the Disk C whose structure is shown in Figure 1 and for a Disk B which has the same radial variation of surface density

TABLE I

Unstable modes of two disks without halos

Model	β	t	$\langle\Omega\rangle$	Ω_p	Growth rate	r_{corot}	r_{Lind}
C	0.3336	0.34	2.804	2.24	0.82	0.56	0.75
	0.3708	0.32	2.722	1.81	0.55	0.65	0.83
	0.3891	0.31	2.680	1.58	0.44	0.71	0.88
	0.4072	0.30	2.638	1.42	0.44	0.76	0.92
	0.4250	0.29	2.595	1.32	0.40	0.79	0.96
	0.4427	0.28	2.552	1.24	0.31	0.81	1.00
	0.4514	0.275	2.530	1.21	0.217	0.82	–
	0.4601	0.270	2.508	1.17	0.08	0.84	–
	0.4636	0.268	2.500	1.13?	0	0.86?	–
B	0.4352	0.30	–	1.272	0.463	0.80	0.96
	0.4543	0.29	–	1.20	0.32	0.83	1.00
	0.4619	0.286	2.338	1.169	0.229	0.84	–
	0.4657	0.284	2.329	1.154	0.158	0.85	–
	0.4694	0.282	2.320	1.134	0.076	0.85	–
	0.4732	0.280	2.312	1.06?	0	0.89?	–

in the outer half of the disk but a central surface density only 9.26 times the average surface density. Both disks have an exponent $\alpha=\frac{2}{3}$, corresponding to a polytropic index $n=2$ in the finite thickness interpretation. The parameter t is calculated assuming a finite thickness. The quantity $\langle\Omega\rangle$ is the angular momentum divided by the moment of inertia, the tensor virial prediction for the frequency of the marginally stable bar mode. The pattern angular velocity Ω_p is one-half the real part of the frequency (since $m=2$) and the growth rate is the imaginary part of the frequency. The radii of the corotation resonance and the (outer) Lindblad resonance are in the last two columns.

Both disks become stable at roughly the point predicted by the tensor virial method. The form of the dominant unstable mode is a moderate trailing spiral which straightens into a bar in the limit of marginal stability. The displacements at marginal stability are close to linear functions of the Cartesian coordinates out to $r\simeq 0.6$, the part of disk containing the bulk of the mass, but deviate by a large amount in the vicinity of the corotation radius. The Eulerian perturbations $\delta\sigma$, δv_r, δv_ϕ vary smoothly through the corotation radius near marginal stability. For Disk C there is an indication that the existence of a Lindblad resonance tends to keep the disk unstable. This effect might not be expected to carry over to stellar disks, since there one expects strong Landau damping near a Lindblad resonance.

The influence of a halo on the stability and on the form of the dominant unstable mode has been studied most thoroughly for Disk C. One interesting point is the degree of central condensation of the halo which is most effective in stabilizing the disk. Models C1–C5 combine Disk C with halos all of same Mass $M_H=1.433$. The distribution of mass in each of the halos and in the disk is shown in Figure 2. The properties of the dominant unstable modes are listed in Table II for certain values of β. In particular, note that the growth rate compared at the same value of β, $\beta=$ $=0.2605$, is smallest for a halo considerably less centrally condensed than the disk, Model C2.

Some additional properties of Model C2 with $\beta=0.2605$ are shown in Figure 3. The local stability parameter Q is substantially greater than one everywhere. It is smallest in the center of the disk, partially due a substantial reduction of Ω^2 below $\Omega_0^2+\Omega_H^2$ there. The wavelength at the minimum of the local dispersion relation, λ_c, increases strongly outward, and even at the center of the disk it is somewhat larger than the characteristic scale of radial variation of the surface density.

The angular patterns of the dominant unstable modes of Models C2 and C3 at $\beta=0.2605$ are depicted in Figure 4. The trailing spiral rather clearly will persist at marginal stability, since the growth rates are so small, as will the presence of a Lindblad resonance. The amplitude of the *fractional* perturbation in the surface density has a large peak at $r\simeq 0.21$, well within the central bar. Model C2 has a secondary maximum in $\delta\sigma/\sigma$ near the corotation radius, down by a factor of 1.7. This probably becomes a pole at marginal stability, associated with the sharp break in the pattern there in Figure 4. The mode is strongly dominated by the central bar, and is not even qualitatively similar to a local density wave.

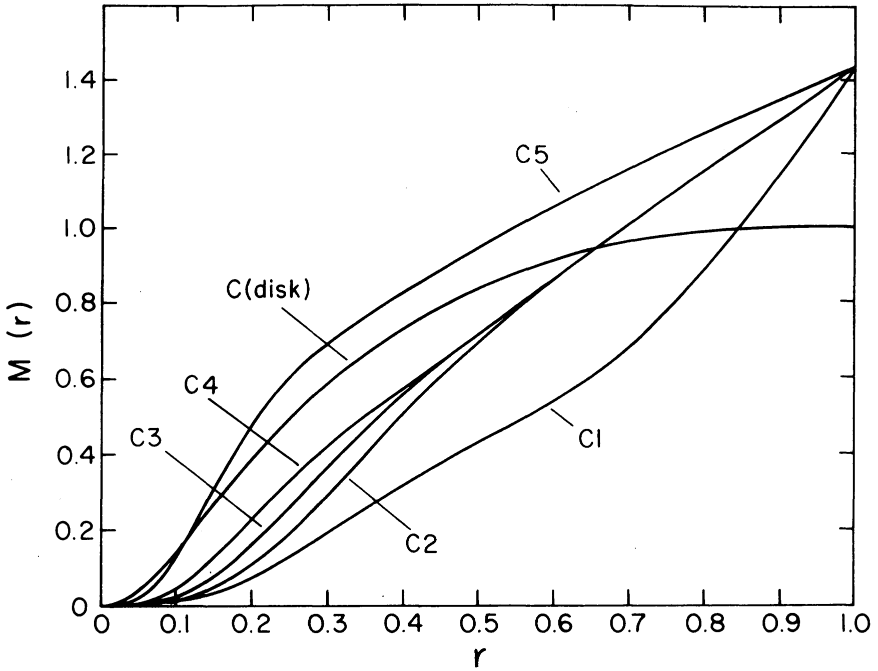


Fig. 2. The mass distribution of the halo as a function of radius is shown for Models C1–C5 along with the mass distribution of Disk C.

TABLE II

Unstable modes of $\alpha = \frac{2}{3}$ Disk C with halo

Model	β	t	$\langle \Omega \rangle$	Ω_p	Growth rate	r_{corot}	r_{Lind}
C1	0.2605	0.2076	3.569	3.03	0.42	0.53	0.86
C2	0.2605	0.2121	3.839	2.75	0.09	0.63	0.92
C3	0.2605	0.2179	3.921	2.68	0.16	0.64	0.93
C4	0.2605	0.2252	4.003	2.862	0.266	0.61	0.88
C5	0.111	0.2658	4.640	5.7	0.8	0.36	0.55
	0.1760	–	4.557	4.3	0.5	0.46	0.67
	0.2605	0.2437	4.446	3.365	0.34	0.56	0.79

The nature of the dominant instability changes substantially if the distribution of pressure in the disk is changed to make Q and λ_c larger in the center of the disk than in the middle part of the disk. This is accomplished by changing the exponent α in Equation (12) to $\frac{4}{3}$ from $\frac{2}{3}$. Now at least some halo is necessary to approach stability while Ω^2 is still positive at the center. Models C11 and C12 still have the zero-pressure

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

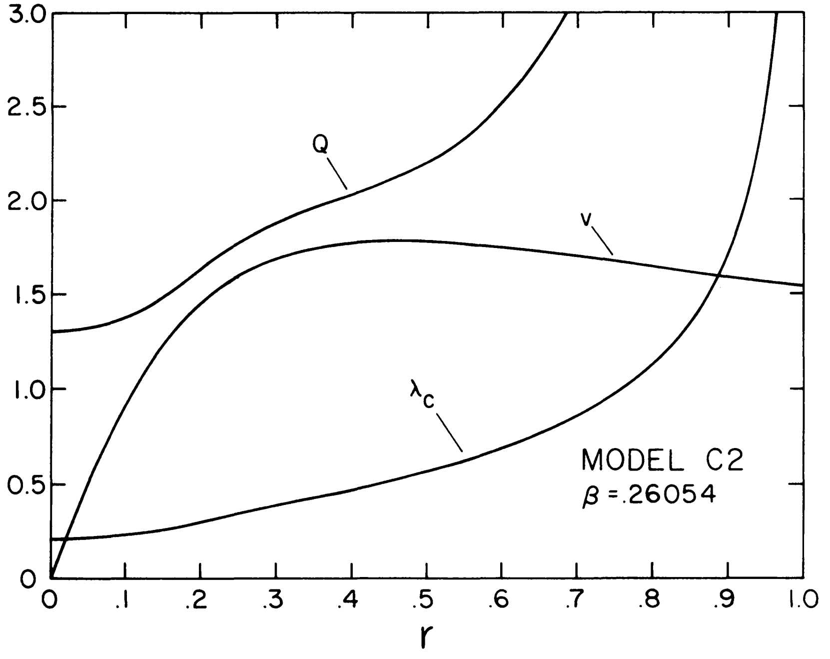


Fig. 3. The local stability parameter Q, the rotational velocity v, and the local characteristic wavelength λ_c for Model C2 at $\beta=0.2605$.

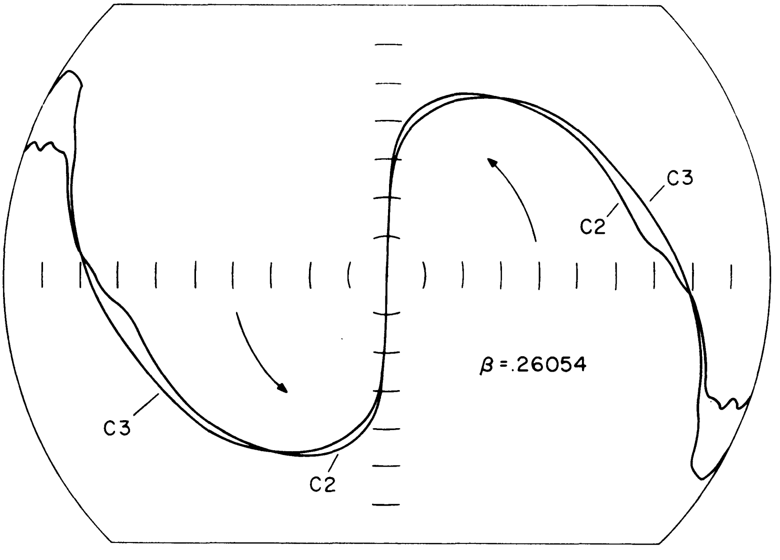


Fig. 4. The pattern of the maximum in surface density as a function of radius for Models C2 and C3 close to stability.

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

structure of Cisk C, but $\alpha=\frac{4}{3}$. The halo mass of C11 is the same as Models C1–C5 and is distributed in radius like Model C2 in the outer part of the disk. A strongly centrally condensed component of the halo is adjusted to keep Ω^2 reasonably large at the center. Some properties of Model C11 are shown in Figure 5. Model C12 has a halo mass $M_H=2.149$, but except for a smaller centrally condensed component to the halo is otherwise identical to Model C11.

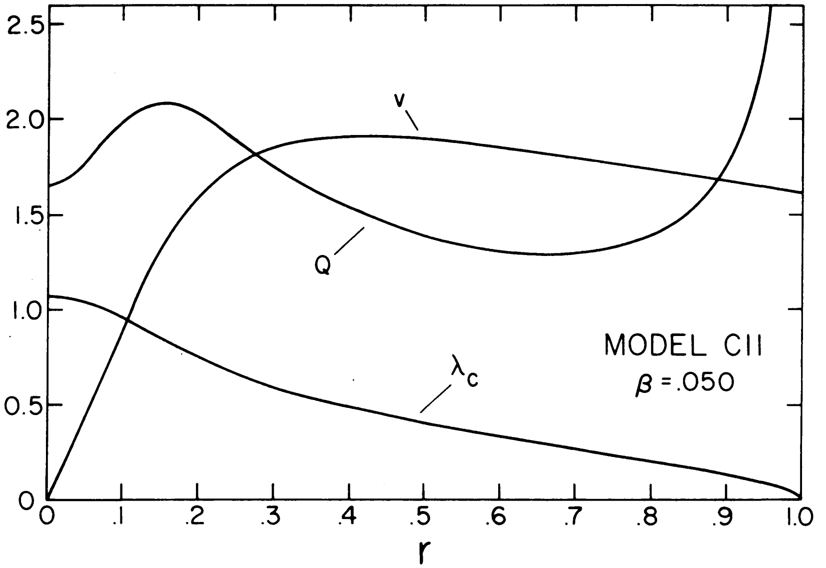


Fig. 5. The local stability parameter Q, the rotational velocity v, and the local characteristic wavelength λ_c for Model C11 at $\beta=0.50$.

Table III lists some of the properties of the dominant unstable modes of these models. Again, t is estimated from the finite thickness interpretation of the disk and includes the halo, which is truncated at the outer edge of the disk. The minimum value of Q for each value of β is given in the third column. In all the models the radius of the minimum Q is $r=0.66$. My numerical methods were stretched to or a little beyond their limit in the case of Model C12. A spurious numerical instability

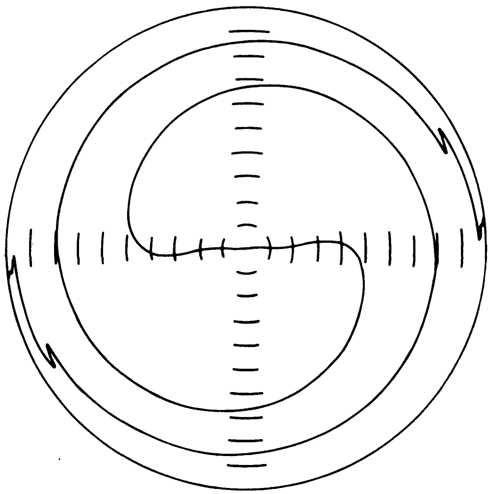
TABLE III

Unstable modes of $\alpha=\frac{4}{3}$ Disk C with halo

Model	β	t	$Q_{\min}$	Ω_p	Growth rate	r_{corot}	r_{Lind}
C11	0.030	0.2408	0.994	3.8	0.7	0.50	0.74
	0.040	0.2327	1.150	3.2	0.5	0.58	0.84
	0.050	0.2225	1.289	2.74	0.328	0.66	0.94
C12	0.023	0.1839	0.984	3.3	$\lesssim 0.3$	0.60	0.90

of very short wavelength was present, but did not obscure the overall properties of the pattern, which maintained itself with little change in the region where the amplitude was largest for a couple of rotation periods.

The amplitude of the dominant mode, as measured by the fractional perturbation in the surface density, has a broad maximum in the vicinity of the corotation radius in these models and is very small in the central bar and outside the Lindblad resonance. The absolute value of $\delta\sigma$ is largest near the inner edge of the spiral pattern. Therefore, these modes are genuine spiral modes. The pattern for Model C11 is shown in Figure 6 and that for Model C12 in Figure 7. The mode at $\beta=0.023$ for



MODEL C11 β =.050

Fig. 6. The pattern of the surface density perturbation for Model C11 at $\beta=0.050$.

Model C12 corresponds roughly to a local density wave. The spacing of the spiral pattern is roughly comparable with λ_c (0.68 times the λ_c plotted in Figure 4), and the fairly smooth behavior near the corotation radius is consistent with the density wave prediction, since $Q_{\min}\sim 1$ there. The inner radius of the spiral pattern is roughly where λ_c becomes comparable with the radius.

The relatively large value of Q near the center does make it easier to get a spiral pattern, but a substantial halo is still required to damp the global instability of the disk to the point that a local density wave analysis has some validity. The spiral pattern in Models C11 and C12 is outside most of the mass in the disk, so in effect these models rely on a 'central bulge' as well as the halo to stabilize the part of the disk with Q near one.

Models with Q and λ_c roughly independent of radius over most of the disk have also been studied, with results intermediate between the extremes quoted here. The fractional surface density perturbation has comparable amplitudes in a central bar

https://doi.org/10.1017/S007418090001562X Published online by Cambridge University Press

and an outer spiral pattern; the latter at comparable halo masses is more open than the spiral pattern of Models C11 and C12 near marginal stability.

The spiral patterns indicate that the antispiral theorem of Lynden-Bell and Ostriker (1967), which does apply to these disks, is not a severe restriction in practice. With a substantial halo it takes only a small imaginary part in the frequency to produce a

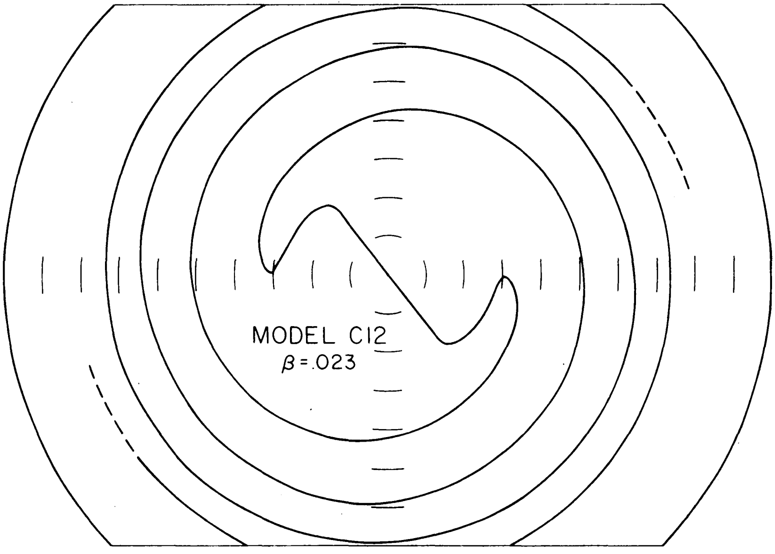


Fig. 7. The pattern of the surface density perturbation for Model C12 at $\beta = 0.023$. The pattern is not well defined inside $r = 0.3$ and outside $r = 0.8$.

fairly tightly wound spiral without a noticeable singularity in the eigenfunction. In a stellar disk Landau damping associated with the spiral pattern can be expected to limit the growth of the instability to small amplitudes.

5. Summary and Conclusion

All lines of theoretical evidence lead to the same conclusion. Any disk which remotely resembles the disk of a spiral galaxy as represented by the neighborhood of the Sun or as contemplated in density wave theories of spiral structure will be globally unstable unless the disk contains only a rather small fraction of the total mass within its outer radius.

In the absence of a halo there seems to be a rather close correspondance between the dynamic instability of a stellar disk and the secular instability of a gas disk. Both are marginally stable when the ratio of kinetic energy to gravitational potential energy t is about 0.14. Even the small difference between the critical value of $t = 0.1376$ for the Maclaurin spheroids and $t = 0.1286$ for the Kalnajs (1972) uniformly rotating disk is probably largely due to the difference in thickness. A more precise gas analogue to the Kalnajs disk is an infinitesimally thin disk with isotropic stress only in the plane; a choice of $\psi = \beta(\pi\sigma)^2$ and a Maclaurin surface density distribution produces a zero frequency mode, corresponding to marginal secular instability, at $t = 0.1250$.

More realistic differentially rotating stellar dynamic configurations have not yet been tested directly for stability against small perturbations, but the numerical simulations do provide quite strong evidence that shear and central condensation do not change t at marginal stability very much. However, the numerical simulations should be carried out with more realistic initial conditions, particularly initial conditions which correspond to self-consistent stellar dynamic equilibrium models with a large velocity dispersion and a large Q near the center relative to that in the outer part of the disk.

A stellar disk presumably began as a gas disk. The gas disk, considering a probable large effective viscosity from turbulence and perhaps magnetic fields at least during its formation, must have been secularly stable. Therefore, it is certainly relevant that Ostriker and Bodenheimer (1973) find t at marginal secular instability is about 0.138. Also, the dynamic instability calculations for thin or infinitesimally thin gas disks reported in Section 4 of this paper do provide necessary conditions for stability which already exclude disks like those commonly used to model spiral galaxies without halos.

Just how much halo is necessary to stabilize a disk with a local stability parameter Q close to one is not yet well established. The numerical simulations have so far been carried out for only a few special cases. The indications are that a disk which does not contain a massive hot 'spheroidal component' in its center requires a halo of perhaps 3 times the mass of the disk or more. With a fairly massive halo it seems likely that the distinction between secular and dynamic instability for a gas disk becomes less important, and that the dynamic instability calculations reported in Section 4 have a more direct relevance to galactic disks. The central condensation of the halo should be somewhat less than that of the disk for an optimum overall stabilizing effect, though a more centrally condensed halo does reduce the amplitude of the central bar component of the overall unstable mode relative to the spiral component.

All in all, I do not think the halo should be considered an extension of a central spheroidal component. The observational evidence is that the halo does not contain anything remotely like a normal stellar component. It seems to me that it is much easier to conceive of a halo containing black holes than a halo containing extreme M-dwarfs, since the black holes would fit in better with current ideas on star formation in metal-deficient interstellar gas (Larson and Starrfield, 1971) and on the chemical evolution of galaxies (Truran and Cameron, 1971). In fact, there are indications some elliptical galaxies may have extended halos with large M/L ratios, perhaps largely composed of black holes (Wolfe and Burbidge, 1970).

The calculations in Section 4 and some of the numerical simulations suggest that at least the more open spiral patterns in galactic disks may be understandable as slightly unstable global modes of the disk. They also indicate that conventional density wave theory may not be applicable. In particular, the local relation between surface density perturbation and potential perturbation breaks down badly when the spiral pattern is fairly open in a centrally condensed disk. The potential perturbation in the outer part of the disk is dominated by the inner part of the spiral pattern, and does not follow the surface density perturbation. Also, if a massive halo is present the density wave fits to observed spiral patterns based on conventional galaxy models, which ignore any extended halo in obtaining the surface density of the disk from a rotation curve (see Lin *et al.*, 1969; Tully, 1974), are invalid.

The detailed study of global instabilities of galactic disks and their possible relation to spiral structure is just beginning. Direct observational confirmation of the existence of halos will be virtually impossible if the halos are made up of black holes, so observational tests of the broad implications of the global stability analysis may well have to rely on indirect evidence, such as M/L variations obtained from rotation curves (see Roberts, 1975).

Acknowledgements

I would like to thank F. Hohl, A. Kalnajs, J. P. Ostriker and A. Toomre for discussions which have greatly increased my understanding of the questions discussed in this paper. I would also like to recognize valuable contributions of Bernard F. Schutz, Jr. to early stages of the dynamic stability analysis of thin gas disks.

References

Bardeen, J. M.: 1971, *Astrophys. J.* **167**, 425.
Chandrasekhar, S.: 1969, *Ellipsoidal Figures of Equilibrium*, Yale University Press, New Haven.
Chandrasekhar, S.: 1970, *Astrophys. J.* **181**, 497.
Chandrasekhar, S. and Lebovitz, N. R.: 1962, *Astrophys. J.* **135**, 248.
Erickson, S. A.: 1974, 'Vibrations and Instabilities of a Disk Galaxy with Modified Gravity', Massachusetts Institute of Technology (Ph.D. Thesis).
Feldman, S. I. and Lin, C. C.: 1973, *Studies in Applied Math.* **7**, 1.
Freeman, K. C.: 1970, *Astrophys. J.* **160**, 811.
Freeman, K. C.: 1975, this volume, p. 367.
Friedman, J. L. and Schutz, B. F.: 1975, 'On the Stability of Relativistic Systems', *Astrophys. J.* (to be published).
Goldreich, P. and Lynden-Bell, D.: 1965, *Monthly Notices Roy. Astron. Soc.* **130**, 7.
Hohl, F.: 1971, *Astrophys. J.* **168**, 343.
Hohl, F.: 1973, *Astrophys. J.* **184**, 353.
Hohl, F.: 1975, this volume, p. 349.
Hunter, C.: 1963, *Monthly Notices Roy. Astron. Sov.* **126**, 299.
Hunter, C.: 1965, *Monthly Notices Roy. Astron. Soc.* **129**, 311.
Hunter, C.: 1972, *Annual Reviews of Fluid Mechanics*, Annual Reviews, Palo Alto.
Hunter, C.: 1973, *Astrophys. J.* **181**, 685.
Kalnajs, A. J.: 1972, *Astrophys. J.* **175**, 63.
Larson, R. and Starrfield, S.: 1971, *Astron. Astrophys.* **13**, 190.
Lin, C. C. and Shu, F. H.: 1964, *Astrophys. J.* **140**, 646.
Lin, C. C. and Shu, F. H.: 1966, *Proc. Nat. Acad. Sci.* **55**, 229.
Lin, C. C., Yuan, C. and Shu, F. H.: 1969, *Astrophys. J.* **155**, 721.
Lynden-Bell, D.: 1967, *Monthly Notices Roy. Astron. Soc.* **136**, 101.
Lynden-Bell, D. and Ostriker, J. P.: 1967, *Monthly Notices Roy. Astron. Soc.* **136**, 293.
Miller, B. D.: 1973, *Astrophys. J.* **181**, 497.

Miller, R. H.: 1971, *Astrophys. Space Sci.* **14**, 73.
Miller, R. H.: 1974, *Astrophys. J.* **190**, 539.
Miller, R. H. and Prendergast, K. H.: 1968, *Astrophys. J.* **151**, 699.
Miller, R. H., Prendergast, K. H., and Quirk, W. J.: 1970, *Astrophys. J.* **161**, 903.
Ostriker, J. P. and Bodenheimer, P.: 1973, *Astrophys. J.* **180**, 171.
Ostriker, J. P. and Mark, J. W-K.: *Astrophys. J.* **151**, 1075.
Ostriker, J. P. and Peebles, P. J. E.: 1973, *Astrophys. J.* **186**, 467.
Ostriker, J. P., Peebles, P. J. E., and Yahil, A.: 1974, *Astrophys. J. Letters* **193**, 1.
Ostriker, J. P. and Tassoul, J. L.: 1969, *Astrophys. J.* **155**, 987.
Press, W. H. and Teukolsky, S. A.: 1973, *Astrophys. J.* **181**, 513.
Roberts, M.: 1975, this volume, p. 331.
Schmidt, M.: 1965, *Stars and Stellar Systems* **5**, 513.
Schmidt, M.: 1975, this volume, p. 325.
Tassoul, J. L. and Ostriker, J. P.: 1968, *Astrophys. J.* **154**, 613.
Toomre, A.: 1964, *Astrophys. J.* **139**, 1217.
Toomre, A.: 1969, *Astrophys. J.* **158**, 899.
Truran, J. W. and Cameron, A. G. W.: 1971, *Astrophys. Space Sci.* **14**, 179.
Tully, R. B.: 1974, *Astrophys. J. Suppl.* **27**, 449.
Vandervoort, P.: 1970, *Astrophys. J.* **162**, 453.
Wolfe, A. and Burbidge, G. R.: 1970, *Astrophys. J.* **161**, 419.

DISCUSSION

Hohl: Why is a central core/halo not effective in stabilizing the bar instability?

Bardeen: At least for these gas disks, I find that a halo as centrally condensed or more centrally condensed than the disk is less effective (for a given total mass) than a less centrally condensed halo in stabilizing the disk as a whole. The central condensation does help damp the central bar, but the pattern angular velocity is increased, which means that the outer Lindblad radius moves in and has a stronger destabilizing effect. This behavior is somewhat in conflict with what is expected from the local criterion, since the central condensation increases Q substantially at the center.

Miller: Have you allowed modes other than $m=2$ yet?

Bardeen: My program is set up to calculate the stability of modes with any value of m, but so far all my calculation have been for $m=2$.

Miller: An interesting problem would result if the halo could respond to the disk, rather than being rigid. Do you see a way to allow for this?

Bardeen: A crude way would be to superimpose a hot disk and a cool disk, which interact only gravitationally.

Pişmiş: Is there any restriction to the extent of the halo required for stability?

Bardeen: All the stability requires is a halo mass somewhat larger than that of the disk within the radius of the disk. The halo mass at larger radii whether it exists or not, has no effect on the dynamics of the disk.