JOURNAL OF COMPUTATIONAL PHYSICS 21, 400–437 (1976)

Validity of Disk Galaxy Simulations

R. H. MILLER*

NASA–Ames Research Center, Moffett Feld, California 94035

Received October 28, 1975; revised March 5, 1976

Several n-body calculations have simulated disk galaxies by means of a large number of self-gravitating particles moving on a plane, and have indicated many features in agreement with observation, but also some major disagreements. Some gaps remain in the arguments that results of these simulations represent physically valid consequences of a model in which a disk galaxy is composed entirely of stars. A new formulation of the n-body calculation was designed to complete the arguments. The entire calculation is done in polar coordinates, and runs as fast as equivalent programs in cartesian coordinates.

Possible difficulties with disk galaxy simulations are reviewed, and the gaps indicated. The polar program is described in detail, and results obtained with it are used to complete the arguments that the disagreement with observation represents a deficiency in the physical model. The physical process of formation of barlike or two-armed spiral structures from a variety of initial conditions is described; the mechanism is not an $m = 2$ instability in an initially axisymmetric configuration.

1. INTRODUCTION

The large gravitational n-body calculations in two dimensions that have produced long-lived spiral patterns [1–4] have made use of a cartesian grid of points over which potentials or forces were computed, and have used a leapfrog difference scheme in cartesian coordinates. Typical computations use 50,000 to 200,000 point particles that move on a plane under $1/r^2$ forces of self-gravitation; the calculations are similar in spirit to several large plasma computations [5]. These gravitational systems usually form barlike density concentrations that shear into trailing spirals by differential rotation, and finally settle down into the spiral density waves that make the computer experiments look successful. But some difficulties remain. The velocity dispersions of particles are typically much greater than the corresponding velocity dispersions in real spiral galaxies. The disagreement is not small; velocity

* Permanent address, Department of Astronomy, and Institute for Computer Research, The University of Chicago, Chicago, Ill. 60637.

dispersions in computed systems scale to be 2 to 5 times as large as the velocity dispersion of stars in our Galaxy, as observed in the solar neighborhood.

These results have important implications if they properly reflect the physics of flat galaxies. Something must be present to stabilize the Galaxy—possibly some mass that is stable and that produces a gravitational potential within the observed stars move. Otherwise, the velocity dispersion in the Galaxy is too small for stability. This has ramifications far beyond the immediate question: If typical galaxies have much more mass than has been thought, the universe may be fairly close to the critical density for closure.

There have been some gaps in the arguments that n-body simulations correctly indicate properties of galaxies composed entirely of stars. The polar program was designed to close those gaps, and runs with it have been successful in doing so. The completion of the arguments is a problem in computational physics; in this paper, results obtained with the polar program are combined with earlier results to establish the validity of the simulations. The goal is to produce a definitive statement that the simulations represent the physics of the model, which implies that any disagreement with observation represents a deficiency in the model. This is done by means of a critical examination of the arguments.

We adopt the viewpoint that computation must be regarded as numerical experimentation. This viewpoint is essential with astronomical systems which do not admit laboratory experiments, but it is useful in other areas of physics as well. An experimenter in a laboratory must be certain that his results reflect the physics of the system under test, rather than some form of experimental error. Experimental error might arise from defective experimental design, it might be an unavoidable systematic error, or it might be some combination of these effects. The computer program is analogous to the experimental apparatus. The usual program verification checks are equivalent to a laboratory experimenter's equipment checks. There is a peculiar tendency to accept computed results uncritically; the computational physicist must be even more careful than a laboratory experimenter to satisfy himself that computational results reflect the physics of the problem under investigation.

An additional danger is that important physical effects may have been omitted in the original design. In this respect, computational physics is closer in spirit to theory, which is usually based on abstractions designed to leave out "unimportant" details in order to produce a manageable system. But the abstraction may be so severe that there is little resemblance between the physical system and the model. The computational physicist can include more effects than can a theoretician, so he can get closer to the true physics; but he still deals with oversimplified systems. Effects can be switched on and off in computer experiments to help find the important physical mechanisms at work in any perticular situation. Computer experiments are most convincing when they can be used to identify those physical

processes which are most important in producing the effects observed in nature.

Features of galaxy simulations that possibly could lead to fallacious results or interpretations are surveyed in Section 2. While some of these possibilities have been eliminated by earlier experiments, checks with the polar-coordinate program have eliminated those that remain.

The polar program is described in Section 3. The potential calculation, in particular, is useful for other astronomical calculations; it has already been used for self-consistent calculations of self-gravitating hydrodynamic flows on a plane. Finally, in Section 4, we describe results obtained with this program, and use these results to complete the argument that the simulations are physically valid.

The polar program proved to be more powerful as a tool in the study of galactic dynamics than had been hoped at the time it was designed. It provides a fine degree of control over the experiments. It has been possible for the first time to construct nontrivial static self-consistent initial conditions that represent true equilibria [6]. Most of these are unstable. These static self-consistent starting conditions are similar in spirit to the "quiet start" of plasma computations [7], but are substantially more difficult to construct.

Further, the polar program has been useful as an aid in the identification of the physical processes responsible for certain observed phenomena, such as the overwhelming preponderance of two-armed systems among spiral galaxies. More detailed expositions of the astronomical results will be published elsewhere.

2. Possible Difficulties

The polar calculation, like its predecessors, is an initial value calculation that undertakes the integration of the Newtonian equations of motion for a large number of particles constrained to move on a plane under $1/r^2$ forces. A galaxy is pictured as a large number of such mass-points, in dynamical equilibrium under their forces of self-gravitation. Large numbers of particles can be handled because forces are computed only at a restricted set of points—if needed, forces are obtained for intermediate locations by interpolation rules. Forces are not computed between particle pairs, which makes the amount of computation necessary to obtain the forces independent of the number of particles. The main computational difficulty in conventional gravitational n-body problems comes from the need to handle close encounters; the possible formation of binary systems. In these calculations, the forces are cut off for close encounters, sidestepping that problem. The range of the cutoff is on the order of the spacing between the grid points at which force-values are tabulated. The justification for this cutoff lies in the observation that stellar dynamical calculations can be corrected for divergences at close encounters, while the computations are designed to explore long-range effects—a matter that

has been left unresolved by theory. The computer models approximate particle dynamics; they may also include additional features such as crude hydrodynamics of some gas or effects of magnetic fields, but these will not be discussed in this paper. In keeping with the crudeness of the force (or potential) calculation, the equations of motion are handled by a time-centered leapfrog scheme. Computational details may be found in [8] and [3, 9, 10].

The most obvious points that might compromise the validity of numerical experiments with a system of mass-points can be listed as follows. These points range from purely technical matters that arise in the specific form of the computations (analogous to experimental technique), through practical matters of whether certain phenomena should be detectable in the experiments, to theoretical results that have been invoked in the interpretation of the experiments.

1. Crude low-order integration scheme (time-centered leapfrog).

2. Roundoff. This is typically more severe than with most computations, either to permit packing several particle attributes into a single machine word, or to assure a Liouville theorem and reversibility [1, 8]. In the velocities, roundoff should look like an additional velocity dispersion, and should help to stabilize systems; however, in the forces, roundoff is expected to have a destabilizing effect [6].

3. Effective change of the force law because of the near cutoff and from the tabulation of potential values on a grid.

4. Calculation in cartesian coordinates may "pump" $m = 2$ or $m = 4$ angular Fourier components (densities with 2φ or 4φ angular dependence; those with $m = 2$ are barlike). In addition, the corners of a square field can fill with particles and further excite density features with $m = 4$ symmetries. Calculation in cartesian coordinates could be responsible for the predominance of barlike ($m = 2$) features.

5. Restriction of the particles to a plane places an artificial constraint that could result in larger velocity dispersions.

6. There might not be enough particles to allow reasonable relaxation times, which should be long compared to the time modeled in an integration.

7. Possible failure to have constructed a static self-consistent model, such that changes in the system are a normal response to whatever initial condition was provided.

8. Posssible failure to recognize an instability (or stability, if we had it).

9. It has been customary to invoke the Toomre stability criterion [11] in discussions of the numerical experiments. This criterion refers to axisymmetric disturbances in axisymmetric systems. A stability threshold should be present if

this criterion is applicable; however, the experiments fail to show any stability threshold. This is the reason for the concerns expressed in points 7 and 8. But, in addition, attempts to determine the stability of systems not constrained to remain axisymmetric by the application of Toomre's criterion may be inappropriate. This affects the interpretation, rather than being a matter of experimental technique.

Several of these have previously been eliminated as possible sources of trouble. Unfortunately, these points cannot be discussed in order because of the nature of the evidence used; a given experiment often provides results that can be applied to several points.

Points 1 and 2 arise in any computations of initial-value problems. While they were disposed of in an earlier paper [8], there is now additional empirical evidence in support of this conclusion.

Point 2, concerning roundoff, has been eliminated by comparison of runs with different amounts of rounding. The calculations of Hohl and Hockney [3, 4] used much finer roundoff than those of Miller and Prendergast [1, 2], but behaved similarly with regard to the problem of large velocity dispersions and with respect to the formation of barlike or $m = 2$ features.

Point 3, modification to the effective force law, has been eliminated as a possible source of trouble by Hohl [12], who showed that the tendency to large velocity dispersions cannot result from the force cutoff for near encounters. If anything, calculations with such "softened" forces underestimate the true effect. Gravitational systems behave differently from plasma simulations in this respect. Langdon [13] demonstrated an instability in plasma simulations if the Debye length is too small (on the order of the grid spacing). But the analog of a Debye length with gravitational systems is the system dimension. Hohl's result [12] also provides experimental confirmation of the conclusion that galaxy simulations are safe with respect to the Langdon condition.

Point 6, concerning relaxation time, has been eliminated by Hohl by means of a study of the rate at which energy was transferred among particles of different masses in a computed system [14]: the experimental relaxation time is more than 500 rotation periods. This is much longer than most integrations, which seldom run more than 20 rotation periods.

A numerical "heating" process, in which particle kinetic energies increase systematically involves an interplay of points 1, 2, 3, and 6. In plasma, the "heating" rate is measured in terms of the "collision time" [15]. In the gravitational problem, the relaxation time provides the measure; numerical "heating" should be negligible. In fact, "heating" has not been detectable in gravitational experiments; Hohl's measurements on the "stable axisymmetrical" models [4] show no perceptible heating, and Hohl had to use a mixture of particle masses to produce a measurable relaxation effect [14].

The Toomre criterion (point 9) results from a theoretical discussion of axisymmetric responses of an axisymmetric disk galaxy to axisymmetric disturbances [11]. Toomre showed that an axisymmetric disk with no velocity dispersion would be unstable to axisymmetric disturbances, but that such a disk could be stable if it had sufficient velocity dispersion. The required velocity dispersion is different in different places within a disk galaxy; we refer to a system in which the velocity dispersion is sufficient everywhere as satisfying the Toomre criterion. Point 9 was eliminated as a possible source of trouble by an experiment with concentric rings as pseudo-particles [16], which forced all distrubances and responses to be axisymmetric. A stability threshold was found and the Toomre criterion was quantitatively satisfied, leading to the conclusion that computed systems indeed obey the Toomre criterion if they are forced to remain axisymmetric, and that the failure of disk galaxies to show any stability threshold, much less one near that of Toomre's criterion, must be ascribed to the admissibility of nonaxisymmetric disturbances and responses. Hohl, in another series of experiments, showed that axisymmetric systems could be maintained in the computer with velocity dispersions only slightly larger than those required by the Toomre criterion when the force field is kept axisymmetric [4], in contrast to the situation when nonaxisymmetric forces are permitted. Hohl's result is consistent with the conclusion stated about the Toomre criterion. An interesting feature of the Toomre criterion is that the velocity dispersion for stars in the solar neighborhood or our Galaxy is near that required by the Toomre criterion; this coincidence is responsible for the importance attached to the criterion under circumstances in which the criterion is evidently not applicable.

Since the experiments concerning the Toomre criterion were run with the time-centered leapfrog difference-scheme, that possible source of difficulty is eliminated as well (point 1). Similarly, since the experiments with rings to check the Toomre criterion showed a sharp stability threshold, the question concerning our ability to recognize stability, if present, in a computed model is also eliminated (point 8).

The numerical experiments are on safe grounds with respect to point 5, restriction to a plane. Real spiral galaxies are quite thin; the thickness is typically about 1/20 of the radius. Thus, while real galaxies have nonzero thickness, approximation by point particles that move on a plane is reasonable. Theoretical estimates, based on attempts to extend the Toomre criterion to disk galaxies with nonzero thickness [17], indicate that such systems should be stable with the dispersion in one component of the velocity 20–30% lower. With this reduction, the kinetic energy of the practicles over a unit area is about the same for planar systems and in three dimensions. Hockey and Brownrigg [18] reported a three-dimensional calculation of a thin disk. They found that three-dimensional systems can exist in the computer with somewhat lower velocity dispersions than can the corresponding plane system. But the pictures presented by Hockney and Brownrigg show rather

small differences, such as might accompany velocity dispersions lower by 20–30% (for one component), but not the large differences expected for a reduction of velocity dispersions by the required factor of 2 to 5. The experimental results confirm the theoretical expectations.

Point 7, possible failure to have generated a static self-consistent model, is largely a matter of experimental technique related to the initial conditions from which the numerical experiments are started. Starting conditions that are intended to mimic simple theoretical galaxy models are seldom correct in all details. Calculations immediately respond to the nonequilibrium state, and begin a general rearrangement. It is possible that the large velocity dispersions result from this rearrangement, and that systems could exist in the computer with lower velocity dispersions. The statement that experimental systems have large velocity dispersions means that no one has yet found a system that would remain substantially unchanged in the computer with small velocity dispersion. This is by no means a proof that no such systems exist, or equivalently, that any equilibrium model with small velocity dispersions is unstable. But enough systems have been tried to make such a conclusion plausible.

Disk galaxies are kept from collapsing under their forces of self-gravitation by a mixture of centrifugal and pressure (velocity dispersion) forces. It has not been possible to work out the proper correction to the rotational velocities in models with sufficient velocity dispersion to meet the Toomre criterion. The correction to the mean rotational velocity is calculated according to the equations of stellar hydrodynamics [19], which make use of the first and second moments of the local stellar velocity distribution. Apparently these moments do not provide an adequate description with such a large velocity dispersion. Models with initial conditions generated by these rules are not close enough to equilibrium.

Self-consistent starting conditions obtained by using a late stage of some other calculation, once that other calculation has settled down, invariably have large velocity dispersions. Starting conditions obtained through modifications of these late-stage models fail to be equilibrium models once the mechanism that generates the modification is turned off. The modification usually consists of a reduction of the velocity dispersion, possibly with readjustment of the mean rotational velocity, and has been referred to as "cooling."

The most satisfactory solution has been to use complete self-consistent equilibrium theoretical models, which properly take all of the features of mean rotational velocity and pressure support into account. A stability analysis is essential as well. So far, only the models of Kalnajs [20] meet these conditions. Computer models based on the Kalnajs models behave significantly better than the models used previously [10]. Even so, linear combinations of the Kalnajs "omega models" must be used to achieve a model that appears to be stable under a linear stability analysis [20]. These more elaborate models are fairly difficult to mimic in the

computer. The disk galaxy experiments reported in this paper all start from a pure Kalnajs omega model.

Note the distinction between point 7, possible failure to have generated a static self-consistent (equilibrium) model for use as a starting condition, and point 8, which relates to stability. Equilibrium is required before it makes sense to inquire into stability, but an equilibrium may be either stable or unstable. From an operational point of view, it can be very difficult to distinguish between instability and failure to have achieved equilibrium. The work reported in this paper has helped to clarify this situation. Examples of both stable and unstable equilibria have been obtained with the polar program. Some equilibrium models have been produced which are stable or unstable according to the value of a parameter. These were not disk galaxy models. A rather sharp stability threshold appeared as the parameter was varied. In this case, instability is easy to distinguish from lack of equilibrium.

Disk galaxy models are surprisingly intolerant of any failure to meet the self-consistency conditions. It makes no difference whether the intolerance results from instability or from lack of equilibrium; the result is the same whatever the cause.

Point 7 clearly requires separate discussion for each experiment. Before the present set of experiments was undertaken, the strongest results that bear on point 7 were those of Hohl [10] with the Kalnajs models [20]. Hohl's results suggested that unchanging self-consistent models could be constructed, but that the models have large velocity dispersions. The present work provides some stronger results. The issues raised by point 7 are understood and are under control; they have been eliminated as a possible cause of the large velocity dispersions and of the preference for barlike structures. This point will be discussed further in Section 4.

All points except number 4, the effect of a cartesian grid, have been eliminated. While it is unlikely that integration on a cartesian grid could be responsible for the distressingly large velocity dispersions in the experimental galaxies, it could account for the preference for barlike structures. The polar program was undertaken to study the effect of point 4. A clear-cut experimental answer is provided by running the calculation with a different coordinate system. Experiments run with the polar program have yielded results with which point 4 can be eliminated, and which strengthen the arguments on many of the remaining points as well.

With the help of results obtained with the polar program, we conclude that all the obvious points by which numerical experiments might lead to results that do not reflect conditions set by the physical model have been systemetically eliminated, as carefully as it is possible to do so. These points relate both to experimental technique and to interpretation. The physical model pictures a galaxy as an isolated assembly of point particles that interact under inverse-square forces of self-

gravitation. To the extent that the numerical results might disagree with observation, the deficiency rests with the physical model, and not with computations designed to study the implications of that model.

3. Details of the Polar Program

A. *The Grid*

There are advantages to a grid obtainable from a cartesian grid by conformal transformation. The integration is carried out in the variables that describe the cartesian coordinates, (u, v), with the mapping $r = Le^{\alpha u}$, $\varphi = \alpha v$ for some convenient value, α, which determines the number of azimuthal grid points along a ring at $u =$ constant. Arbitrarily, $\alpha = 2\pi/36$ has been chosen, leading to $N_a = 36$ azimuthal values. A choice, $M = 24$, makes the radius run from L (at $u = 0$) to about $65L$ at $u = 24$. The radial interval ranges from $0.191L$ at $u = 0$ to $12.6L$ at $u = 24$. (L is a dimensioned length that sets the length scale for the problem.) The potentials are required over a 24×36 grid of points.

With the choice, $a = 3L$, for the "softening parameter" in the potential (defined at Eq. (8), below), the polar calculation can describe as much detail as a 128×128 grid in cartesian coordinates. Several advantages to the polar reprosentation are apparent. First, central regions, where a lot of detail is desirable, are described by a fine grid, while the outer regions, which are typically more sparsely populated and do not show as fine a structure, are described by an appreciably coarser grid. So this 24×36 grid describes a disk galaxy as well as the old 128×128 or 256×256 cartesian grids. This produces a substantial speedup of the force-calculation, a second advantage. A third advantage is that it is easy to restrict the force field to a subset of angular harmonics—for example, only 0 and 2, etc. This facilitates experiments to determine sensitivity of the systems to various angular driving terms. The polar description is more naturally suited to the problem.

An inner curoff at $u = 0$ ($r = L$) is small enough that few particles spill off the inner edge. Those that do spill at this boundary are carried across the center by force-free rectilinear motion. Particles that spill over the outer boundary ($u = 24$) have simply been removed from the calculation, but they could be treated by other rules.

The nonuniform grid cannot be used directly in an integer arithmetic mode equivalent to that used with the cartesian program [8]. Unit velocity in the outer regions corresponds to a very large velocity in the inner regions. Either a floating-point or a finely divided integer representation should be used. As a matter of taste, we prefer an integer representation that corresponds to about 2^{-8} of the potential

grid. Unequal amounts of subdivision could be used at different radii, but the resulting program complexity more than outweighs any advantages.

Potentials can be interpolated in (u, v) by the usual rules for interpolation with a cartesian grid of points, with v interpreted modulo 36. There is little advantage to storing force components over potentials; linear interpolation on forces requires fetching 8 force-values, while the equivalent quadratic interpolation in potentials can be based on 6 or 9 (we prefer the 9-point interpolation rule). Potentials are stored over the grid.

B. *The Difference-Scheme*

The Hamiltonian for the system is

$$\mathscr{H} = \frac{e^{-2\alpha u}}{2\mu\alpha^2 L^2}(\mathbb{P}^2 + \mathbb{J}^2) + \mathscr{V}(u, v), \tag{1}$$

where $\mathbb{P}$ and $\mathbb{J}$ are the momenta conjugate to u and v, respectively, and μ is the mass of a particle.

A simple translation to a time-centered leapfrog scheme yields

$$J^{(n+(1/2))} = J^{(n-(1/2))} + G^{(n)}, \tag{2}$$

$$P^{(n+1/2)} = P^{(n-1/2)} + \frac{\exp(-2\alpha u^{(n)})}{\alpha}(P_a^{(n)2} + J_a^{(n)2}) + F^{(n)}, \tag{3}$$

where dimensioned units $(\mu L^2/\Delta t)$ have been divided out throughout, so $P = \mathbb{P}(\Delta t/\mu L^2)$, $J = \mathbb{J}(\Delta t/\mu L^2)$, $F = (-\partial\mathscr{V}/\partial u)(\Delta t^2/\mu L^2)$, and $G = (-\partial\mathscr{V}/\partial v)(\Delta t^2/\mu L^2)$ are dimensionless. The values, $P_a^{(n)} = \frac{1}{2}(P^{(n+(1/2))} + P^{(n-(1/2))})$, and $J_a^{(n)} = \frac{1}{2}(J^{(n+1/2)} + J^{(n-(1/2))})$, are time-centered for the difference scheme. The equation for $P^{(n+(1/2))}$ is implicit. Similarly, the equations for $\dot{u}, \dot{v}$ become

$$u^{(n+1)} = u^{(n)} + \frac{\exp(-2\alpha u^{(n+1/2)})}{\alpha^2} P^{(n+1/2)}, \tag{4}$$

$$v^{(n+1)} = v^{(n)} + \frac{\exp(-2\alpha u^{(n+1/2)})}{\alpha^2} J^{(n+1/2)}. \tag{5}$$

The time-centered value for $u^{(n+(1/2))}$ inside the exponential is treated like the $P_a^{(n)}$ and $J_a^{(n)}$ terms; that equation, too, is implicit. The relations for $P^{(n+(1/2))}$ and $u^{(n+1)}$ are unpleasant, but not troublesome. The resulting scheme has remainder terms $O(\Delta t^3)$, like the cartesian leapfrog scheme.

The Hamiltonian formulation, with storage of momenta and coordinates, assures a Liouville theorem. The time-step is determined through selection of the

scaling coefficient in the potential calculation. Transition from the dimensionless units of the calculation to the dimensioned units of the modeled physical system is through μ, L, and Δt, the mass of an individual particle, the length scale chosen, and the integration time-step.

Particle attributes are packed into one 60-bit word on a CDC 7600, which allows 14 bits each for P and J and 15 each for u and v. The remaining bits are used for flags. Momenta are scaled so the largest magnitude representable is four times the angular momentum of a particle on the periphery of the initial disk. No momentum components have ever gone out of the allowable range.

Particle-pushing requires about 3 sec per integration step for 60 000 particles. This is about 1.6 times the time required for integration in cartesian coordinates, a pleasant surprise since we had expected that a polar program might require significantly more time. The particle-pushing programs were written without any particular attention to producing fast-running-object codes. Since the polar programs have turned out to be so useful, some attention to careful programming to improve running speed in this part of the program would be appropriate.

An alternative reversible difference-scheme for (u, v) coordinates has been proposed by Buneman [21], in which the $(P_a^{(n)})^2$ term in Eq. (3) is represented by the geometric mean of $P^{(n+1/2)}$ and $P^{(n-1/2)}$. This permits a simpler (and faster) solution for $P^{(n+1/2)}$. Tests were run with Buneman's difference-scheme; the entire integration ran about 10% faster. The small difference results from the usual feature that such a large part of the total computation time is spent in computational overhead; packing and unpacking particle data, fetching and interpolating potentials, bounds-checking, and so on, that a speed-up in the integration algorithm yields little overall reduction in running-time.

Like the present scheme, the leading error term in Buneman's difference-scheme is $O(\Delta t^3)$, typical of these reversible first-order schemes. However, the coefficient of (Δt^3) is $\frac{1}{6}$ for Buneman's method against $\frac{1}{24}$ for the method used. The difference results from the change to updated forces halfway through the momentum step in the present method, while Buneman's method uses the same force-value for the entire step. The shorter time-step required to compensate for the larger truncation error more than outweighs the speed-up gained by the faster method. This is the usual argument in favor of more elaborate integration methods. We continued to use the method set forth in Eqs. (2) to (5), rather than changing to Buneman's method.

C. *The Potential Calculation*

The potential calculation proceeds through Fourier transformation of the density in the angular coordinate: $\sigma(u, v) \rightarrow \hat{\sigma}_n(u)$, where n is the index of the angular Fourier component. The potential energy of interaction between two such Fourier components is zero unless the indices match, but the calculation does not seem to

admit further useful decomposition. There is a matrix $\hat{D}_n(r_i, r_j)$ for each azimuthal index, n, such that

$$\hat{\mathscr{V}}_n(r_i) = \sum_{r_j} \hat{D}_n(r_i, r_j)\, \hat{\sigma}_n(r_j), \tag{6}$$

and such that the azimuthal Fourier transform of the set of potentials at the grid points on the ring at r_i is a linear combination of the transformed mass densities on all rings. With the "softened" potential, the self-potential of a ring is finite, so there is no need for special treatment for r_i in the sum.

The potential energy of interaction between two mass-rings with unit masses in Fourier components m and n is

$$\hat{\mathscr{D}}_n(r_i, r_j) = -\delta_{mn} \frac{G}{(r_i r_j)^{1/2}} \frac{\Gamma(n + 1/2)}{2\pi^{1/2} n!} z^{(n+1/2)} F\left(\frac{1}{2}, n + \frac{1}{2}; n + 1; z^2\right), \tag{7}$$

where

$$z = \frac{(r_i^2 + r_j^2 + a^2) - [(r_i^2 + r_j^2 + a^2)^2 - 4r_i^2 r_j^2]^{1/2}}{2r_i r_j}. \tag{8}$$

Here, a is the "softening parameter" in the potential: the potential energy of interaction between two point particles separated by $(\Delta x, \Delta y)$ is $-G\mu_1\mu_2(\Delta x^2 + \Delta y^2 + a^2)^{-1/2}$. The function, F, is the hypergeometric function. The $n = 0$ case reduces to the complete elliptic integral $(2/\pi)\, K(z)$; as n gets large, the function approaches $(1 - z^2)^{-1/2}$, while intermediate values of n lead to a monotone increasing sequence at fixed z. The argument, z, is positive and $z \leqslant \min(r_i/r_j, r_j/r_i)$ with equality for $a = 0$.

Proper expression of the interaction potential, in the nearest gridpoint (NGP) approximation, requires Fourier components to all orders, since the transformed density is periodic with period N_a (for N_a points equally spaced around the circle) in the index, n. The hypergeometric function is not well suited to numerical computation because of slow convergence, especially for z near unity.

A practical computation that properly sums these functions to all orders and that incorporates the appropriate boundary condition that the potentials go to zero at infinite distance is straightforward to construct. The potential energy of interaction between two particles of equal mass at (r_1, φ_1) and (r_2, φ_2) is

$$D(r_1, \varphi_1; r_2, \varphi_2) = [r_1^2 + r_2^2 + a^2 - 2r_1 r_2 \cos(\varphi_2 - \varphi_1)]^{-1/2}, \tag{9}$$

where $G\mu^2 = 1$. The set of matrices $\hat{D}_n(r_1, r_2)$ is the discrete Fourier transform (technically the complex conjugate of the DFT, but the DFT is real) of the set of these D's for all N_a values of $(\varphi_2 - \varphi_1)$ with r_1 and r_2 fixed.

In the calculation, the potential is tabulated on 26 values of the radial variable

and 39 values of the angular variable. The extra values facilitate interpolation near the boundaries of the allowed region. Because of the symmetries in the potentials, it is sufficient to tabulate $[(N_a/2) + 1]$ matrices $\hat{D}_n(r_1, r_2)$(for N_a even). For the 24×36 calculation, there are 19 matrices of dimension (26×24). The potential calculation consists of the following numbers of operations: (1) twenty-four Fourier transformations of the density. Each transform is of length 36, for a total of 8640 operations. (2) Thirty-six multiplications of 24-element vectors by 26×24 matrices, which yields 22 464 operations. (3) Twenty-six Fourier transforms to yield the potential values at each of the (26×36) points. Each transformation is of length 36, for 9360 operations. The total of 40 464 operations may be compared to 1 048 576 operations for an equivalent 128×128 grid in cartesian coordinates. Storage of the $\hat{D}_n$ requires 11 856 locations, compared to the 65 536 for convolution coefficients in the cartesian calculation. An additional symmetry, $\hat{D}_n(r_1, r_2) = \hat{D}_n(r_2, r_1)$, which could be used to reduce the storage requirement to 6631 locations, has not been utilized. There is a substantial computational advantage to the potential calculation in polar coordinates.

There is no feature analogous to the problem of periodic replication in the radial direction with this formulation; hence there is no need to surround the active region with a fictitious empty region. The angular coordinate is periodic; the DFT fits naturally. A substantial fraction of the computational saving results because empty neighboring regions need not be included.

The potential calculation can be restricted to a subset of angular harmonics by omitting the calculation of the corresponding $\hat{\mathscr{V}}_n$. The subroutine that carries out the potential calculation has been arranged to permit the Fourier amplitudes to be printed for both the density, $|\hat{\sigma}_n(r_i)|$, and the potential, $|\hat{\mathscr{V}}_n(r_i)|$.

The potential calculation is much faster than had been hoped at the time a polar program was planned. The factor 25 had not been anticipated, and only became evident as the details were worked out. The speed advantage results from two features: (1) an adequate description of the potential field requires many fewer points, and (2) it is not necessary to include "empty" regions to avoid troubles from the periodicity of the DFT, as it is with cartesian coordinates [8, 9]. The potential calculation, written without attention to details that would help with running speed, and compiled without concern for the quality of the object code, runs in less than 0.1 sec on a CDC 7600. Three-dimensional calculations based on cylindrical coordinate systems should achieve similar speed advantages.

Several alternative formulations are available for potential calculations in polar coordinates [22]. These fall into two classes: Bessel-function transformations (Toomre and Clutton-Brock) and eigenfunctions of the Laplacian in oblate spherioidal coordinates (Barbanis and Prendergast). The Bessel function formulations present some difficulty if restricted to a finite set of grid points or of "frequencies," or if the mass is restricted to a finite region: all of these are necessary

for practical computation. The spheroidal formulation automatically restricts itself to the interior of a circle on the plane and uses discrete "frequencies" (order and degree of associated Legendre polynomials). However, termination of the expansions at a finite number of terms produces effects analogous to band-limiting in Fourier expansions, which can cause incorrect values to be returned for the potentials. The present formulation was preferred, with its admitted shortcomings, because it returns correct values for the potentials in a uniform and precisely defined way over the configuration.

D. *Starting Conditions*

Experiments have been started from two different basic initial conditions. Each has the property that the model obtained by setting the velocity dispersion to zero is an equilibrium model that remains unchanged after an integration step.

The first condition generates the set of Kalnajs [20] "omega models" also studied by Hohl [10]. These models contain a (dimensionless) parameter, $0 \leqslant \Omega \leqslant 1$ such that models with $\Omega = 1$ have no random motions (no velocity dispersion; all particles move on exactly circular orbits), while $\Omega = 0$ produces models with no mean rotation. Kalnajs [20] showed that these models are all unstable under a linear stability analysis. However, models with $\Omega < 0.811$ are stable against axisymmetric disturbances; this is the analog of Toomre stability for these models.

Particles are loaded into the configuration space on a sequence of concentric rings spaced to produce the desired radial density variation. Thirty-six particles are equally spaced around each ring. The number of rings is chosen to make the total number of particles loaded come out near 60 000 (1665 rings for 59 940 particles). Next, the force field is determined by the potential calculation and the angular momentum of the particles on each ring set such that the centrifugal acceleration just balances the radial force, to produce a purely circular motion ($J = J_c$). This step is essential because the small difference between the actual force field and the analytic force field is too great for a model to remain in equilibrium if the circular motion is calculated from the analytic form. The differences arise principally from the softened forces and from the unequal numbers of rings in each radial interval. With the angular momenta set according to the force field and the radial momenta set to zero, a "cold" disk, or $\Omega = 1$ model, is complete.

Models with $\Omega \neq 1$ have mean angular momentum $J_0 = \Omega J_c$ and mean radial momentum, $P_0 = 0$; "peculiar momenta" are chosen for each particle by sampling from the desired distribution in a Monte Carlo sense, and then are added to the mean momenta. This procedure assures that models generated with $\Omega = 1$ reduce to the appropriate "cold" disk. The "peculiar" velocities in these models are quite large. At the center, the maximum peculiar velocity is $(1 - \Omega^2)^{1/2}$ times the circular velocity of a particle on the periphery.

The quiet disks with $\Omega = 1$ have no apparent change. With 36 particles in each ring, each time one particle moves out of a cell of the grid used for the potential calculation, another particle moves into that cell. The potentials remain constant throughout the motion. While the pure Kalnajs model with $\Omega = 1$ represents a rigidly rotating disk, the disk actually obtained has differential rotations; the angular velocity is slightly different at different radii. Even so, the disk with $\Omega = 1$ is an equilibrium model.

A ring of particles in orbit about a fixed central mass is the second starting condition. For this case, all particles have the same radius (the ring is infinitesimally thin), and there are 6 particles placed on each of 36×256 equally spaced locations in angle around the ring, for a total of 55 296 particles. The mass at the center is Q times the total mass in the ring (55 296 Q times the mass of an individual particle in the ring). Again, by means of the potential calculation, the angular momentum of each particle is set to the circular value, J_c, so the ring is in equilibrium with each particle on (the same) circular orbit centered on the central mass. The quiet ring configuration ($P = 0$) forms another equilibrium model that circulates in the computer without change.

More interesting models, with a nonzero velocity dispersion, are generated by giving each particle a radial momentum, $P = qJ_cU$, where U is pseudorandom in $(-1, 1)$. Models with $q = 0$ reduce to quiet equilibrium models. Particles with different amplitudes of oscillation about the equilibrium orbit oscillate at different frequencies, so once started, the configuration never again collapses to a thin ring. This model was designed for studies of the stability of ring galaxies, but it also has evident application to the problem of Saturn's rings.

The integration time step is determined by setting a coefficient in the potential calculation. This coefficient has been determined by an approximate relation to make the rotational period in the initial configuration to be a specified number of integration steps, usually 64, 128, or 256. Most experiments have been run with 128 integration steps per rotation of the initial configuration. Because these experiments were designed to study properties of the integration and responses to different initial conditions, most runs were rather short; none was carried beyond 4 rotations of the initial configuration.

The Kalnajs "omega models" are disks whose density goes to zero beyond a certain limiting radius. Disks have been loaded with maximum values of u at 22 ($r/L = 46.5$) and 19 ($r/L = 27.6$); these are about 0.42 and 0.7 of the outermost radius allowed in the configuration. Rings were loaded at $u = 19$.

E. *Cartesian Potentials*

An experimental study of spurious effects produced in cartesian coordinates (point 4 of Section 2) is facilitated by providing a way of mimicking a cartesian calculation in the polar program. Simulations can be run from identical starting

conditions, with and without the cartesian perturbation. Comparisons can then be made in which the only difference is that due to the perturbation. Clearly this simulation is approximate. The most important features are identified and the others ignored.

Especially in the calculations of Hohl [3, 10], particle positions and velocities are represented on sufficiently fine divisions that there are effectively no restrictions to the directions in space that can be represented. Our cartesian calculations [8] made use of coarser number representations, but since similar effects are obtained with both calculations, we conclude that cartesian representation of particle positions and velocities cannot contribute any significant effects.

Particles in the corners of a square field present no problem. Hohl eliminates their effect by using only particles interior to a circle inscribed inside the square for the potential calculation.

The process of computing potentials over a cartesian grid can have an important effect. Particle positions are represented on a fairly coarse grid as input to the potential calculation, and the potential values are represented on the same grid. Because the potential values returned are correct for a mass distribution that is a set of point masses at each of the grid-points, the only effect that can arise from the tabulation of potentials on the grid is a short-wavelength effect like that in the Langdon computational instability [13], which has been argued to be unimportant in gravitational problems. The dominant effect with a cartesian grid is the angular perturbation from assigning particles to the square grid, which is straightforward to simulate.

The principal effect of the cartesian grid is felt very near the beginning of the calculation, when the exact potential field is axisymmetric but the cartesian representation introduces small fictitious potentials in the fourth, eighth, etc., Fourier components. If the effect is important, the system responds with even greater amplitudes in those components. In the experiments, the growth can be masked by "noise" that results from the normal development of the system from an initial condition that contains some randomness in the velocities.

The perturbations arise from assigning particles to locations on a cartesian grid in the initial condition of the Kalnajs disk. The grid has spacing L. The non-axisymmetric part of the potential computed from this cartesian particle assignment was inserted as an external potential that remained fixed throughout the calculation. The external potential is added to the potential produced by the particles to yield the total potential. The potential produced by the particles is computed anew at each integration step, while a fixed external potential is reused at each step. The fixed cartesian potential can be made stronger or weaker to vary the strength of the perturbation. At unit strength, it represents a 128×128 active cartesian grid; at four times normal strength, it approximates a 64×64 active grid.

"Cartesian" potentials are nonzero only in the fourth, eighth, etc., angular

Fourier components. Those of order 8, 12, and 16 are nearly as strong as that of order 4, but they are more restricted in their radial range. At unit strength, the fourth angular component attains an amplitude nearly 1/1000 of the axisymmetric potential component over the range from $u = 12$ to 22 ($r/L = 8$ to 47).

F. *Program Verification*

A major part of design and programming effort goes into verification of the programs. Modular structure facilitates checking. This, in turn, involves design of the programs around a basic data structure, which can be very simple in a calculation such as this.

The potential calculation was checked by comparison with values obtained by straightforward "brute force" calculation for a single particle placed arbitrarily on the grid. Some 50 to 60 values of the brute-force calculation were checked by hand. Further checks involved a pair of particles and sampled versions of continuous mass distributions that admit analytic solutions.

The particle-pusher, or integrator, was checked by integrating a few exactly calculable orbits. Straight-line orbits in a constant potential are one good test for a polar program. Harmonic oscillator orbits were also checked. Angular momentum conservation in axisymmetric potentials is not a good check: it follows automatically (Eq. (2)). The particle-pusher neither gains nor loses particles.

It is more difficult to check proper handling of limiting or end-point situations. An example arises in the solution of Eq. (4) for $u^{(n+1)}$, which typically involves a division. Smooth transitions to limiting cases of small divisors are necessary. During checkout, the programs were made to issue a warning whenever the small divisor branch was taken, and in the production form the number of occurrences is counted. There are several places in the particle-pusher where similar situations arise. Even so, it is difficult to be certain that all potential trouble-points have been identified.

The most convincing checks come through overall program performance in test cases. Two of these are unusually sensitive to a variety of possible troubles.

The first is that static self-consistent (equilibrium) models can be generated through the use of rules developed analytically. Such models have been run for several (64) steps and shown to be true equilibria. All equilibrium models found so far are axisymmetric.

Another, more severe, test makes use of models with a certain angular symmetry. For example, if the initial condition is exactly three fold symmetric, a system should retain exact threefold symmetry. The order of the symmetry must divide the number of azimuthal cells used in the density and potential calculations. One such model was run in which an equilibrium "cold" disk was perturbed by giving small radial momenta to three symmetrically disposed particles. When the calculation was terminated after 80 integration steps (1 ¼ rotations), it had developed very

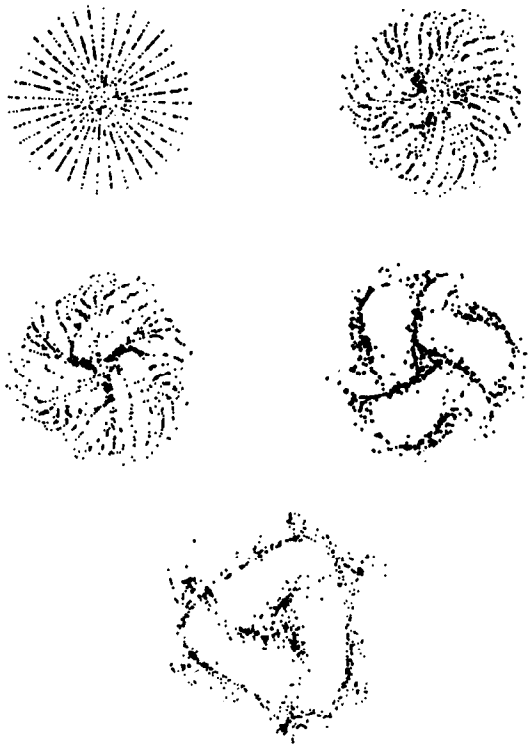

FIG. 1. "Cold" disk with threefold symmetric perturbation. A sample of 1000 particles plotted in configuration space at various times (in units of the rotation period of the initial disk). These configurations retain exact threefold symmetry; the pictures are not exactly threefold symmetric because of the sampling method used to select particles for plotting. Upper left frame at $t = 0$; upper right, $t = 0.72$; middle left, $t = 0.83$; middle right, $t = 1.00$; bottom, $t = 1.24$.

strong condensations (Fig. 1), but the only nonzero amplitudes were in Fourier angular components $m = 0, 3, 6, 9,\ldots$. Similarly, a "cold" disk, perturbed by the "cartesian" potential, had only $m = 0, 4, 8, 12$, and 16 terms after a full rotation, even though the system had developed very strong features.

The standard checks of total energy and total angular momentum conservation are routinely applied.

4. RESULTS AND DISCUSSION

Several sequences of experiments have been run with the polar program. We first describe the phenomena observed in these experiments, then turn to the discussion of questions concerning the validity of conclusions inferred from numerical experiments.

A. *Disks*

An equilibrium "cold" disk was constructed, but it is unstable, confirming a well-known result. One extra particle, placed somewhere among the 60 000, is sufficient to upset the equilibrium. Growth rates are around 10 e-foldings per rotation period, but no one angular Fourier component grows substantially faster than the others.

A second disk, consisting of two counter-rotating "cold disks" was also an unstable equilibrium model with growth rates nearly the same as those for the "cold" disk, again with no outstanding Fourier component. Hohl [23] reported earlier that the counter-rotating disk was "violently unstable," although an equilibrium counter-rotating configuration was not demonstrated. The counter-rotating disk has zero mean velocity, but it has a large (anisotropic) velocity dispersion. It clearly shows that velocity dispersion is not sufficient for stability. However, its velocity at any point in the configuration space is a two-stream distribution, which might be expected to be unstable.

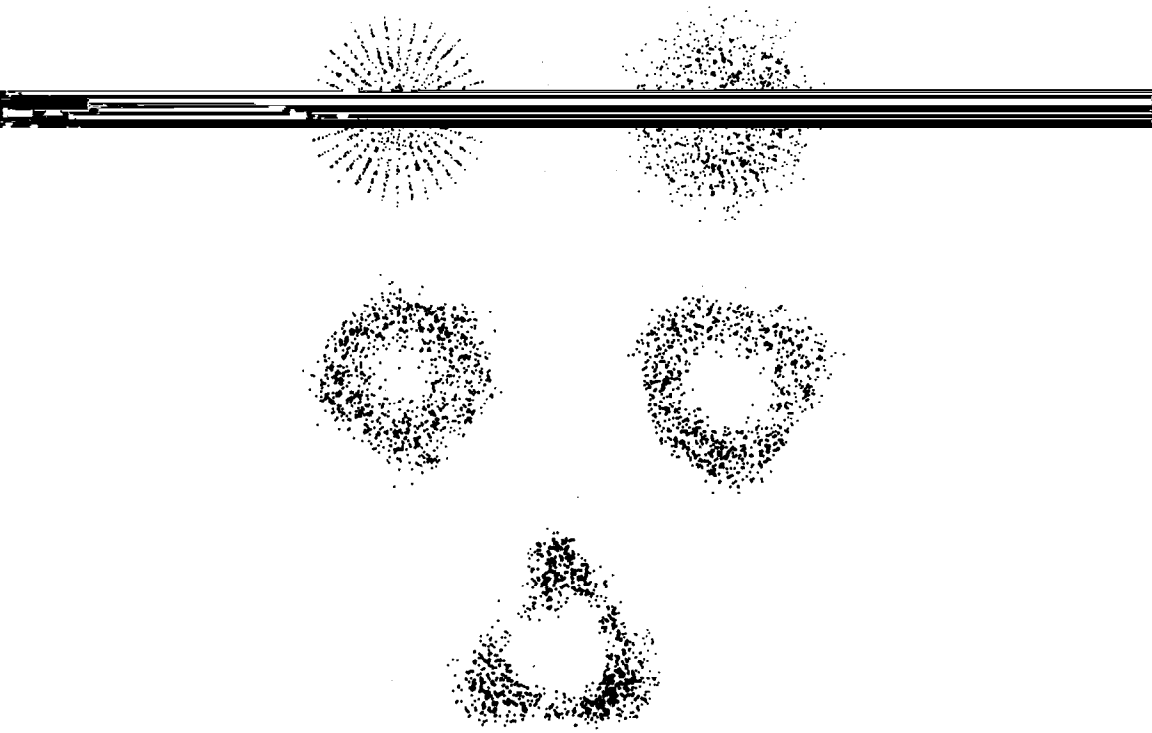

FIG. 2. Kalnajs disk with $\Omega = 0.9$; a sample of 1000 particles plotted at $t = 0, 0.5, 1.0, 1.5,$ and 2.0 rotation periods (top left–top right–middle left–middle right–bottom).

The example previously cited with threefold symmetrical perturbations to an equilibrium "cold" disk (Fig. 1) provided confirmation that $m = 3, 6, 9,...$ angular Fourier components also display exponential growths with comparable growth rates, even in the absence of excitations with different m-values. We have not isolated single unstable modes experimentally. This complicates the discussion somewhat. Normal modes are discussed later (Sect. 4.E).

Kalnajs disks with $\Omega = 0.9$ (Fig. 2) formed a ringlike structure, which later broke up in a nonaxisymmetric way. Although starting conditions with $\Omega \neq 1$ were designed to be equilibrium models, equilibrium could not be confirmed experimentally, so stability cannot be demonstrated directly. Their behavior can, however, be compared with the stability properties of the Kalnajs omega-models. These disks with $\Omega = 0.9$ show an initial axisymmetric development such as might be expected from the known instability of the Kalnajs omega models to axisymmetric disturbances with $\Omega > 0.811$. But those Kalnajs models also have many nonaxisymmetric instabilities with comparable growth rates as well. It is interesting that this model selected the axisymmetric route. Several distinct models with $\Omega = 0.9$ have been run; all went into ring-structures which later took on nonaxisymmetrical forms. The distinct models used different runs of pseudorandom numbers to generate the initial load, but were otherwise identical. One of the other models went to a fourfold nonaxisymmetric form, indicating that rather small changes can alter the form of the nonaxisymmetric development.

The next check was to see whether the disk started to develop structure with $m \neq 0$ which somehow triggered the $m = 0$ (axisymmetric) development. This was tested in an experiment with only $m = 0$ potentials allowed. A ring-structure formed on the same time-scale (Fig. 3). As expected, this ring did not develop any nonaxisymmetric form. Kalnajs disks with $\Omega = 0.9$ apparently developed rings directly from the initial state.

The appearance of ring structures with all of these examples is interesting. In linear stability theory, there is no reason why axisymmetrical disturbances should have negative amplitude at the origin. Systems that develop condensations at the center should occur as often as ring systems. The explanation is that a finite amplitude effect is observed. Particles that can reach the center have small angular momenta; there are too few of them to build up a condensation at the center. This provides further evidence that the effects observed are beyond the range of linear stability theory.

The breakup of the ring into nonaxisymmetric condensations on the time-scale of these simulations has important implications for the lifetime and frequency of occurrence of ring-systems in astronomy. This provided the motivation to undertake the experiments with a ring circulating around a central mass.

Next, Kalnajs disks with $\Omega = 0.8$ were run (Fig. 4). These did not form rings; rather, they developed large density variations in angular Fourier components

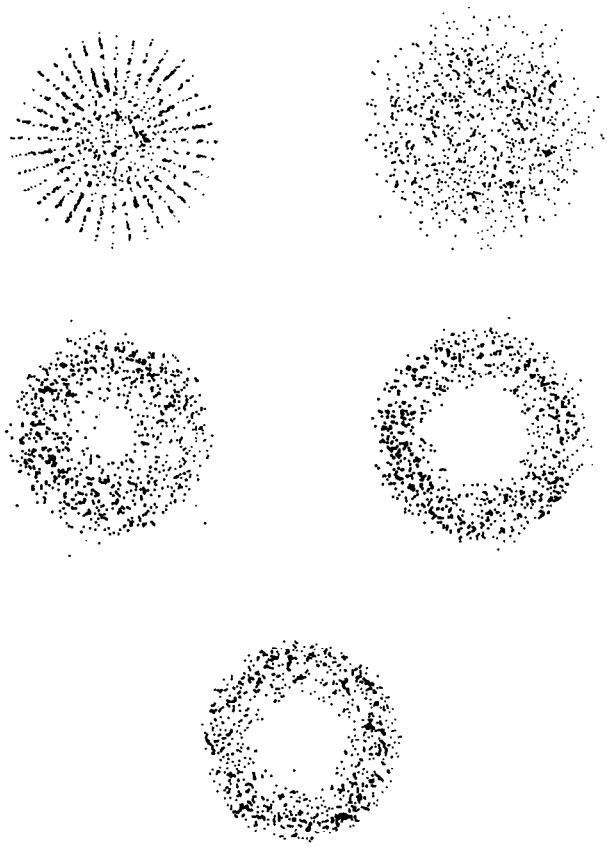

FIG. 3. Kalnajs disk with $\Omega = 0.9$; purely axisymmetric, but otherwise self-consistent potentials. Sampling and times as in Fig. 2.

with $m \geqslant 2$. The Kalnajs "omega model" analogous to this disk is stable against axisymmetric disturbances, but $\Omega = 0.8$ is very close to the limit at 0.811. Like the $\Omega = 0.9$ cases, nonaxisymmetrical forms become apparent in the Figures beyond $t \sim 1.5$ rotations. Once more $m = 3$ appeared. The predominance of $m = 3$ is apparent in the numerical summaries long before it is noticeable in the pictures. Because this model used the same sequence of pseudorandom numbers as that of Fig. 2, this may result from the particular sequence of pseudorandom numbers.

The important feature of Fig. 4 is that two of the prominences present at $t = 2.5$ have moved toward each other by $t = 3$, and a barlike structure is present at $t = 3.5$. This simulation was terminated after $t = 3.5$. This provides the essential clue that leads to an understanding of the preference for barlike structures. We shall return to this point later (Sect. 4.C).

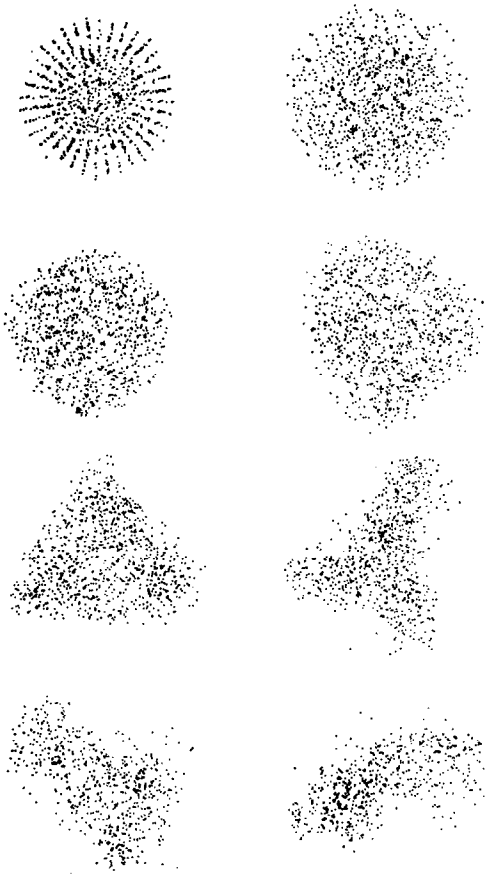

FIG. 4. Kalnajs disk with $\Omega = 0.8$; a sample of 1000 particles at $t = 0$, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 rotations of the equivalent cold disk (top left–bottom right), showing development of threefold state, and its changeover to a barlike structure.

B. *Rings*

The ring model contains two parameters, q and Q. One of these (Q) describes the mass in the center, and the other describes the velocity dispersion in the ring. Stable models should be attainable with either of these parameters sufficiently large. The experimental problem is to determine the curve in the (Q, q) plane that separates stable from unstable models. With the Saturn Rings problem as a guide, it was expected that very large values of Q would be required for stability at small values of q [24, 25]; values on the order of 100 to 1000. Experimentally, stability was found at $Q = 4$, while models with $Q = 3$ were unstable with $q = 0.1$ and 0.03.

The ring with $q = 0$ is an equilibrium configuration experimentally, and since models with $q \leqslant 0.1$ were stable at $Q \geqslant 4$, it is reasonable to treat these ring configurations for all Q, with $q \leqslant 0.1$ as equilibrium models. We can discuss the stability of these models.

The experimental distinction between stable and unstable models is clear-cut and impressive. Examples with various values of Q at $q = 0.1$ provide a good illustration. The model at $Q = 1$ was unstable, but that at $Q = 10$ was stable. Next, $Q = 5$ was found to be stable, $Q = 3$ unstable, and $Q = 4$ stable. The model at $Q = 3$ showed an exponential growth in Fourier components of the density that were detectable above the "noise" at $t = 0.6$ rotations, with growth rates on the order of a factor 30 per rotation-time. The "noise" results from fluctuations in Fourier amplitudes of the density due to motions following the pseudo-random initial conditions. By contrast, the model with $Q = 4$ still had the appearance of noise at 3.5 rotation-times. The noise amplitudes appear to have a very slow growth at $Q = 4$. A typical noise amplitude was around 60 or 70 at the beginning, around 150 at $t = 3.5$; $N^{1/2}$ noise is around 230. The $Q = 3$ calculation started with similar noise values, and had Fourier amplitudes around 2000 to 4000 by the end of the run at $t = 2.3$ rotations. It is possible that the $Q = 4$ case could take off with rapid growth of a disturbance at some later time, if the calculation were run long enough.

The experimental analog of a stability threshold is that, as the parameter that controls stability is varied from the unstable regime toward stability, models show unstable behavior over a wide range of values of the parameter. In this unstable range, unstable behavior normally develops very quickly. Even unstable models may "linger" in a relatively quiet state for a while before the explosive changes that accompany unstable behavior manifest themselves. But a value of the parameter can then be found such that small additional changes in the parameter substantially lengthen the time that the models "linger." There is, of course, no assurance that models with the parameter well beyond the "threshold" would not eventually go "unstable." The region of rapid lengthening of the "lingering" time is the experimental analog of the stability threshold. Systems that display this general pattern in which there is little change as a parameter is varied, followed by a narrow region in which there is rapid change in qualitative behavior, are the experimental analog of systems that can be described by a stability analysis. This same general pattern obtained with the earlier experiments in which Toomre stability was studied in a disk galaxy with pseudoparticles in the form of rings [16].

From the standpoint of the interpretation of the numerical experiments, the existence of a threshold behavior is important; the numerical value of the para-

as the individual particles oscillate about the equilibrium orbit, now leave all

particles in the same radial bin for the potential calculation. With larger values of q some particles cross the boundary and are tabulated in different radial bins; this increases the effective "noise" level by a very large factor, and is equivalent to a much stronger excitation. The instabilities reported here result from particle oscillations in the angular coordinate. Maxwell [24] argued that tangential oscillations should be the most important for determining stability. The tangential oscillations affect the radial form of the potential through the different radial dependences of potential components that arise from the different angular Fourier components in the density.

All Fourier components of the density participate in the unstable developments, but components with $m \sim 6$ to 8 tend to have somewhat larger amplitudes when the exponential growth has attained values well beyond the initial "noise." Because of the noise, it is not possible to identity any one component as dominant at small amplitude, or as having triggered the instability. In particular, there is no evidence that $m = 18$ was responsible [24]. The "softening" of the potential is insufficient to overcome the expected dominance of the $m = 18$ component; it produces at most a 20% reduction in the value of the potential.

C. *Preference for Barlike Structures*

Barlike structures developed with these disk galaxy simulations just as they had in earlier calculations. However, barlike structures never developed directly. In every example, higher angular components showed early strength, but did not predominate very long. By the time certain angular Fourier components could be clearly identified above the "noise," $m = 4$ or $m = 3$ predominated. This is shown in Figs. 5 and 6, where a few Fourier components of the density and of the potential are plotted as functions of time for two different radii. The $m = 3$ density and potential terms dominate from the time those terms emerge from the noise around $t = 1$ or 1.5 up to $t = 3$. Thereafter $m = 2$ dominates.

In the general pattern whereby these disks reach a barlike form, the dominant angular Fourier component first noticed is of higher order than $m = 2$. But the order of the dominant component then decreases, one by one, until $m = 2$ is reached. For example, a system dominated by $m = 4$ soon absorbs a corner and becomes triangular, dominated by $m = 3$. The triangular shape typically develops like that in Fig. 4, in which the three prominences become more distinct, then two move together to form a barlike ($m = 2$) pattern. In the process, $m = 3$ dominates quite a bit longer than did $m = 4$. The $m = 2$ state dominates much longer yet, but other experiments indicate that it, too, finally "washes out." But $m = 2$ dominates for quite a long time.

This same early dominance by higher-order angular terms is also present in picture sequences published by others. There are many examples, but [4, Fig. 4],

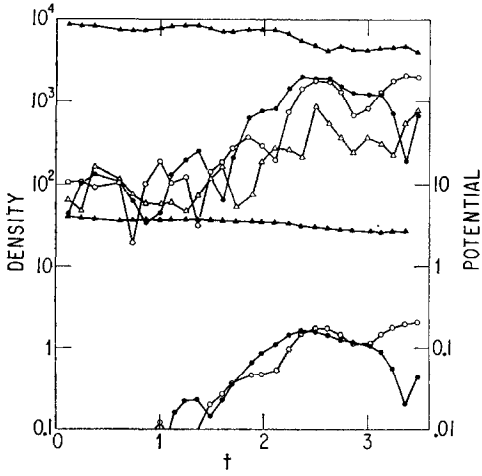

FIG. 5. Amplitudes of angular Fourier components of density (top plots) and potential (bottom plots) as a function of time, at $u = 20$ ($r = 27.5L$). The plot with filled triangles shows the $m = 0$ component; the open circles are $m = 2$; filled circles represent $m = 3$, and open triangles are $m = 5$. The $m = 5$ potential values were too small to represent on this plot. Both the density and potential were dominated by $m = 3$ from the time the patterns overpowered the noise around $t = 1.5$ until about $t = 3$; thereafter both were dominated by $m = 2$. The dominance in both the density and the potential switched from $m = 3$ to $m = 2$ at about the same time. These plots refer to a Kalnajs disk with $\Omega = 0.8$, the same calculation as shown in Fig. 4.

[10, Figs. 2a, 2b], [18, Figs. 1b, 2a, 3b, 5a, and 6a] illustrate the point. These all formed barlike structures later.

Barlike patterns do not form directly. More complicated patterns change form until the barlike forms remain. This is an important result for the astronomical applications. The early stages of astronomical objects can hardly be expected to have been smooth and axisymmetrical. To account for the observed predominance of two-armed spirals (approximate $m = 2$ symmetry), there must be a mechanism whereby other forms can reach a barlike or two-armed symmetry. We have seen such a process at work in these simulations.

The rearrangements observed here are permissible for n-body systems which conserve energy, angular momentum, and centroid positions. Indeed, a threefold symmetrical system can reduce its potential energy by bringing two of its projections together.

Checks of elements of the matrix, $\hat{D}_n(r_i, r_j)$ of Eq. (6) bring out one of the reasons why $m = 2$, or barlike, condensations live longer than other forms. The potentials, $\hat{\mathcal{V}}_m(r)$, are by far strongest at $m = 0$ (the monopole term), as would be expected. Then $\hat{\mathcal{V}}_1(r)$ is much smaller, $\hat{\mathcal{V}}_2(r)$ much smaller still, and so on. The

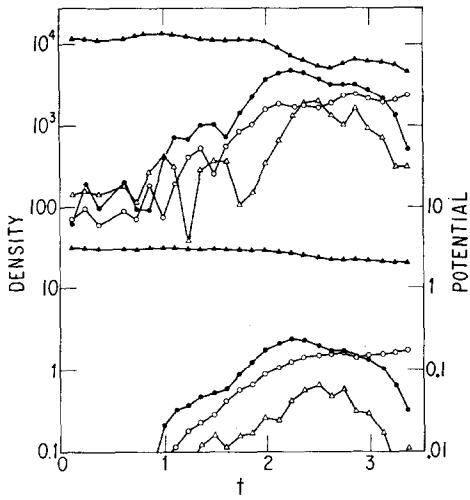

FIG. 6. Amplitudes of angular Fourier components of density and potential as a function of time, at $u = 22$ ($r = 39.1\ L$), near the outer boundary of the initial configuration (which was at $r = 46.5\ L$), for the same calculation as Fig. 5. Plots and labeling same as Fig. 5. Patterns overpowered the noise at an earlier time than in Fig. 5, and the potential switched from being dominated by $m = 3$ to domination by $m = 2$ somewhat earlier than the density at this radius.

different potential terms decrease exponentially with increasing m. The $m = 1$ terms are not expected to be important on symmetry grounds; they refer to displacements of the center of mass. The strongest nonaxisymmetric term to be taken into account is $m = 2$, which drives barlike condensations. All higher terms are exponentially weaker.

This tendency can be seen from Eq. (7) for $\hat{\mathscr{D}}_m(r_1, r_2)$ in terms of a hypergeometric function. Long-range effects arise between rings of quite different radii, so z is small ($\frac{3}{4}$ or less); in these cases $F(\frac{1}{2}, m + \frac{1}{2}; m + 1; z^2)$ has a value near unity and changes little with m at fixed z. The dominant m-dependence in the potential is in the $z^{m+1/2}$ term, which decreases exponentially with increasing m.

These properties are illustrated in the plots of elements of $\hat{D}_m$ in Figs. 7 and 8. The ordinates in these figures represent the potential energy of interaction between a ring of particles and one particle with fixed radius.

The radius of the ring is plotted on the abscissa. There are 36 point particles in the ring. The curves in each figure are labelled according to the value of m which they represent. In Fig. 7, the particle is at $u = 10$, while in Fig. 8 it is at $u = 20$. These plots were made with $a = 3L$; all the potentials would become infinite when the two radii were equal if the "softening parameter," a, were set to zero. The extremely rapid decrease in the potential with increasing m is apparent.

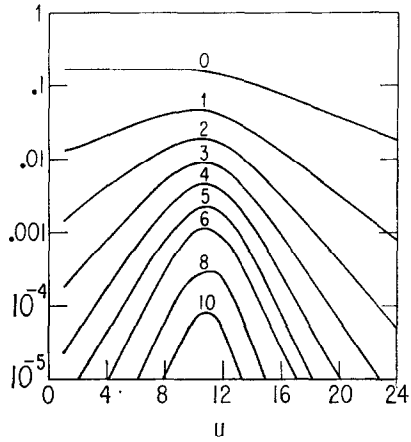

FIG. 7. Potentials in various Fourier angular components at various radii, generated by one particle at $u = 10$. The abscissa represents u, essentially the logarithm of the radius; the ordinate is the value of the potential. The various curves represent different angular Fourier components; each is labeled by its m-value. At sensible distances from the source, the potentials in different harmonics decrease nearly exponentially with increasing m-value.

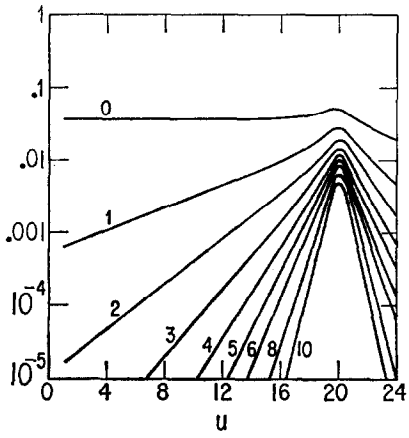

FIG. 8. Potentials in various Fourier angular components at various radii, generated by one particle at $u = 20$. Axes and labeling some as in Fig. 7. These figures are based on entries in the matrix $\hat{D}_m(r_1, r_2)$ (Eq. (6)).

Particles can move over the small potential hills and valleys of components with large m; they tend to remain trapped for longer times by the larger valleys of the $m = 2$ component. This explains why systems started with larger velocity dispersions do not develop the "small condensations" observed with "cold" systems [23]: the mean particle kinetic energy is sufficient to overcome the potential energy variations with the larger m-values.

The dominance of $m = 2$ is a characteristic long-range effect. The occurrence in nature of $m = 2$ symmetries is evidence that long-range effects, extending over the entire system, are responsible for the appearance of these objects. An alternative way of regarding this effect is in terms of multipoles; the low-order multipoles produce effects that extend to long ranges.

Within condensations, the higher-order harmonics are more important (z is then near unity for the contributing terms). Then still more detail can develop. But that detail need not be twofold symmetric.

The $m = 1$ terms cannot be dismissed entirely. A nonzero $\hat{\sigma}_1(r)$ implies that the centroid of the ring at r is shifted off the origin; but different rings can shift differently. The physical condition that the centroid must remain at the origin is an integral condition that corresponds to the sum of all $\hat{\sigma}_1(r)$ contributions.

The systems are dominated by the same angular Fourier component over most of the disk. There is no evidence that dominance by $m = 2$ starts near the center and works outward. It is sometimes stated that computer models display an "instability to barlike disturbances in the inner regions." We have been unable to confirm this statement. There is no evidence that a nearly axisymmetrical configuration tends to organize itself into a pattern dominated by $m = 2$ in preference to larger m-values. A threefold symmetrical system does go over to $m = 2$, however.

A physical process has been identified that leads to the observed preference for barlike, or $m = 2$, structures. This has been possible because the polar programs permit these systems to be studied with fewer distracting computational features than was possible with the cartesian formulation. The identification of an intuitively appealing physical mechanism is far more convincing than arguments to eliminate possible objections.

D. *Cartesian Coordinates*

The possibility that the preference for barlike structures might result from performing the calculations in cartesian coordinates (point 4 of Sect. 2) has been eliminated by the observations and discussions of the previous subsection. However, experiments run using the method of Section 3.E help to assess other effects of cartesian coordinates. These experiments were started from Kalnajs disks with an outer radius at $u = 22$ ($r/L = 46.5$).

A "cold" disk run with the cartesian perturbing potential at unit strength

immediately developed an exponentially growing disturbance with fourfold symmetry. But a Kalnajs disk with $\Omega = 0.8$, identical to the $\Omega = 0.8$ model described earlier except for the cartesian perturbing potential at unit strength, showed no noticeable difference. The strongest density variations occurred at $m = 3$, just as they had with the unperturbed model. Sensitivity to the cartesian perturbing potential clearly depends on the model being studied: "cold" disks feel the potential very strongly, models with some initial velocity distribution feel it much less strongly.

Kalnajs disks with $\Omega = 0.9$ were studied in more detail. Disks, otherwise identical, were tested with 1, 1.5, 2, and 4 times the normal strength of cartesian perturbing potentials. The model with twice normal strength showed an early short-lived predominance of $m = 4$ on a ring-like structure, but a corner was quickly absorbed, the model developed dominant $m = 3$ angular component, and then went into a T-shape. The Kalnajs models perturbed by 1 and 1.5 normal strength cartesian potentials were very similar to the unperturbed model.

The model with $\Omega = 0.9$ and 4 times normal strength did not develop the central hole of a ring structure; rather, it went through a sequence something like Fig. 4. It showed an early $m = 4$ dominance, but by 1.3 rotations, it had a triangular shape and an $m = 3$ dominance. By $t = 2$, $m = 2$ was dominant, and continued to dominate until the calculation was terminated around $t = 3$. In this model of four times the normal strength, the transition from $m = 4$ to $m = 3$ dominance took place simultaneously over a fairly wide radial range, but the outermost parts of the disk changed over a little bit later. The subsequent change from $m = 3$ to $m = 2$ behaved similarly. There was no place within the disk where $m = 4$ gave way to $m = 2$, and the $m = 4$ dominance was completely absent from all parts of the disk before $m = 2$ began to dominate anywhere. This was the usual manner of switching from domination by one Fourier component to another in all of the experiments.

The failure to form a ring indicates that the cartesian perturbation is strong enough that individual particles do not conserve angular momentum. Enough particles have lost angular momentum to keep the center filled. The cartesian perturbation of four times normal strength with which this occurred is equivalent to a cartesian calculation carried out on a 64×64 active grid. The failure to conserve angular momentum was not noticeable at 1 and 2 times normal strength; apparently 128×128 active grids are satisfactory with velocity dispersions like those present in disks with $\Omega = 0.9$. And even the four times normal strength appears satisfactory at $\Omega = 0.8$.

Periodic boundary conditions. The effect of using periodic boundary conditions in the potential calculation can be compared to the perturbation due to the use of cartesian coordinates. The periodic boundary condition corresponds to infinite

periodic replication of the galaxies; the effects are usually described as being due to neighbors.

Like the cartesian potential, neighbors produce only fourth, eighth, etc.,angular Fourier components in the potential; the fourth is by far the strongest. Amplitudes can be estimated by summing terms of Eq. (7) over the nearest few sets of neighbors, with each neighbor treated as if the entire mass were concentrated at its center. The argument, z, is so small that all the F's are well approximated by 1; the z^n term dominates. This leads to a (distance from the center)^4 dependence of the potential in the fourth angular Fourier component. With a 128×128 active grid (unit strength of the cartesian potential) the perturbation due to the cartesian potential is stronger than that due to the neighbors for $u < 18$ ($r/L = 23$); it is much stronger over most of the configuration, and over the region that contains most of the particles. With a 64×64 active grid (four times normal strength; equivalent to a grid spacing of $2L$), the cartesian perturbation is stronger than neighbor perturbations for $u < 20$ ($r/L = 33$). Only those particles at the very edge of the configuration feel the neighbors as strongly as they feel the cartesian grid. Over most of the configuration, the effect of the neighbors is much weaker than that of the cartesian grid. There is a large effect, of course, near the boundary between neighbors.

Periodic boundary conditions were used with the earliest simulations to speed up the potential calculation by the use of Fourier transforms [1]; the neighbors have been eliminated in more recent calculations through the use of empty regions. It was expected, in the early runs, that neighbors were less important than some of the other artifacts of the calculation. While this expectation has been questioned recently [18], we have here quantitative confirmation that it is indeed negligible in comparison to perturbations produced by a 64×64 active cartesian grid. The use of a cartesian grid is more harmful to the simulations than the neighbors present with periodic boundary conditions. The importance of the cartesian grid and of the neighbors depends on the actual mass distribution. Configurations with stronger central concentrations are more affected by cartesian grid effects and less affected by neighbors.

Cartesian potential calculations can be designed with some number, N_s, of grid points along an edge. This defines the length of the Fourier transformations. Two different designs have been used: (1) Particles may be assigned to the full grid, leading to periodic boundary conditions and to neighbors, and (2) particles can be confined to half the linear dimension of the grid with the other half-empty to eliminate the effects of the neighbors. Particles feel a spurious fourth angular Fourier component due to the cartesian grid in both cases, but that component in the potential is stronger with a coarser active grid. There is an additional fourth angular component due to the neighbors in case (1). The comparisons reported here show that most particles in a configuration experience larger perturbations in

the fourth angular component of the potential in case (2) than in case (1), for fixed N_s. Since computation time and memory requirements are largely determined by N_s, a calculation which gains in detail at the expense of allowing perturbations due to neighbors actually has smaller spurious effects due to the grid and generally makes more effective use of the computer. However, some workers object to neighbors, largely on aesthetic grounds; the choice is largely a matter of taste.

E. *Stability and Normal Modes*

The ring-models provide an example of a stable system in the computer. The different behavior at different values of Q illustrates the distinction between stable and unstable systems. This example strengthens the earlier conclusions that we can recognize stability, and that numerical effects in the integration do not prevent the computed system from showing stable behavior. These were discussed in Section 2 as points 1, 2, and 8.

The examples in which stability is clear-cut all start from an equilibrium model. Some unstable equilibrium models are recognizable because features of the computation create a small island of stability in an otherwise unstable environment. This is usually caused by roundoff. A perturbation large enough to remove the system from that island shows the unstable behavior. The equilibrium "cold" disk, the counterrotating disk of Section 4.A, and the rings with $Q \leqslant 3$ are unstable equilibrium models; it would be difficult to argue that they are true equilibria if they did not show the island of stability.

Questions of stability present no problem with the ring systems, where equilibrium models and a stability threshold could be demonstrated. Similarly, they present no problem with respect to the question about dominance of $m = 2$; structures dominated by $m = 2$ developed out of more complicated structures by an orderly physical process without need to invoke the notion of stability.

The critical question that remains in the discussion of computer models of disk galaxies is whether models may exist with low velocity dispersions which are stable equilibria in the sense that any changes of form, velocity dispersion, etc., proceed slowly compared to a dynamical time-scale. From the experiments reported here, and from all of their predecessors, the only models found that change slowly have had large velocity dispersions. Models started with small velocity dispersions have undergone violent changes (large changes on dynamical time-scales), and finally settled down to relatively slowly changing states that have large velocity dispersions. If slowly changing models with low velocity dispersions exist, they have not yet been found by the computer experiments. Such states cannot have escaped detection at any stage of the computer experiments. We are therefore left to conclude that initial conditions that might lead to such models must be very improbable. This conclusion is strengthened by the extreme sensitivity of initial conditions to the least failure to meet the self-consistency conditions; the intol-

erance of any failure of exact self-consistency implies that any islands of stability must be very small indeed.

Linear stability theory could be helpful if it pointed out some stable equilibria with low velocity dispersions that might be good candidates to test by numerical experiments. This has not been successful. Failing that, an orderly procedure to study the properties of disk galaxies would be through responses of the disk to selected disturbances. As normal modes of the theoretical disk, determined by linear stability theory, should be ideally suited, this goal motivates the investigation of normal modes. This effort has not been successful either. First, it turns out that the normal modes of this computer model are different from the normal modes of the Kalnajs disks, and that the Kalnajs normal modes cannot even be synthesized from linear combinations of degenerate normal modes of the computer model. Next, we find that this must be true of any N-body simulation. These investigations follow.

Normal modes: *Computer systems.* Comparison of computer disk galaxy models with the stability properties of the theoretical models after which they were patterned was a useful guide in the study of the disks. The normal modes for the quiet disk of particles are not the normal modes of the Kalnajs omega models [20]. The distinction brings out one of the complications in attempts to compare the results of numerical experiments with theory. The normal modes of the Kalnajs models are continuous disturbances with a certain spatial form, a certain shape in velocity space, and a certain time-dependence (represented by a complex frequency).

The independent motion of an individual particle is a normal mode of the particle system in the computer. To demonstrate this, consider oscillations of one particle about its equilibrium orbit, with all other particles undergoing undisturbed motions on their respective equilibrium orbits. The other particles enter and leave the cells into which the configuration space is divided for purposes of the potential calculation in a manner that maintains constant numbers of particles in each cell all the time. The field of computed potentials does not change until oscillations of the one particle carry it across the boundary of one of the cells. However, in the limit of vanishingly small oscillation amplitude, the particle almost never crosses the boundary. The potentials do not change for a single particle oscillating at infinitesimal amplitude; each particle then oscillates independently. This argument may not apply with other methods of assigning particles to grid points to obtain densities for use in the Poisson equation.

The stability of the computer model is determined by the orbital stability of each of the individual particles in its equilibrium potential field. Every orbit must be stable. The number of independent normal modes is equal to the number of particles; the Kalnajs models have an infinite number of normal modes.

Once an oscillation has carried a particle across a cell boundary the character of

the problem changes, and stability conditions are different (and more difficult to determine). Since oscillations in the neighborhood of an equilibrium orbit involve departures in both radial and angular directions, any orbit with a detectably nonzero amplitude will eventually lead the particle to cross a cell boundary. This limits the amplitude of the oscillations to the roundoff level in the computer. But within this restricted domain, stable orbits can exist even if orbits with larger oscillation amplitudes are unstable. This is the island of stability referred to earlier.

There are degenerate normal modes in the computer as well: linear combinations of these are also normal modes. If the computer disk had been perfectly designed, all these frequencies would be equal; in fact they are not all precisely equal. However, no linear combinations can be formed that approximate even the lowest order Kalnajs modes. The frequencies cannot be matched. Even if the frequencies could be matched for one or two of the Kalnajs normal modes, the spatial structure would be completely different. A single Kalnajs normal mode cannot be isolated for study in the computer experiment.

Normal modes: Particle systems. The ideas of Monte Carlo sampling can be used to show that the normal modes for the Kalnajs models are not normal modes for any particle system which they may represent. Kalnajs [20] considered disks which are the particle analogs of the completely flattened MacLaurin spheroid, and admitted a certain amount of internal motion. A disk of radius R_0 has surface density

$$S(r) = S_0\xi, \tag{10}$$

where $\xi = (1 - (r^2/R_0^2))^{1/2}$, and S_0 is the surface density at the center. The gravitational potential in this disk is the potential of an isotropic harmonic oscillator. Kalnajs showed how to build a disk galaxy out of an assembly of particles moving on harmonic oscillator orbits. The mean circular velocity of the particles is a fraction, $0 \leqslant \Omega \leqslant 1$ of the angular velocity of a particle on a circular orbit in this potential. These disks are equilibrium solutions of the self-consistency problem defined by the combined Poisson and Vlasov equations.

Kalnajs [20] discussed the stability of these "omega models" by means of a normal modes analysis; stability was determined by an investigation of the frequencies of normal modes of a linearized Vlasov equation. A remarkable feature is that the potentials of the normal modes of all the omega models have the same spatial form, irrespective of the value of Ω. On the plane of the disk, the potential of one of these modes is given by

$$\Psi_l^m = p(l, m)\, e^{i(m\varphi - \omega t)} P_l^m(\xi), \tag{11}$$

which implies that the projection of the normal mode onto the configuration space produces a surface-density perturbation

$$\sigma_l^m = e^{i(m\varphi-\omega t)}\, P_l^m(\xi)/\xi. \tag{12}$$

In these equations, m is the azimuthal Fourier component, ω is the frequency of oscillation of the normal mode, P_l^m is an associated Legendre function, $p(l, m)$ is a numerical coefficient that comes from the solution of the Poisson equation, and $(l - m)$ must be even. The details are given by Kalnajs [20].

Both the normal modes and any disturbances are representable in a 4-dimensional phase space; $\sigma_l^m(\xi, \varphi; t)$ is the projection of one of these normal modes onto the 2-dimensional configuration space at time t. These projections onto the configuration space at some instant of time are sufficient for our purposes. An arbitrary disturbance can be projected onto the configuration space as well, to yield a surface density $\sigma(\xi, \varphi)$; and the projection can be analyzed into associated Legendre functions [26]:

$$\xi\sigma(\xi, \varphi) = \sum_{l,m} c_l^m P_l^m(\xi)\, e^{im\varphi} \tag{13}$$

to yield a set of expansion coefficients, c_l^m.

Consider a Monte Carlo simulation of a Kalnajs model with N particles; the instantaneous location of the ith particle is described by ξ_i , φ_i . The projection onto the configuration space yields the instantaneous surface density

$$\sigma(\xi, \varphi) = \sum_{i=1}^{N} \delta(\xi - \xi_i)\, \delta(\varphi - \varphi_i)/R_0^2\xi_i\,, \tag{14}$$

where the $R_0^2\xi_i$ in the denominator comes from the transformation of δ-functions into ξ, φ coordinates. The Monte Carlo simulation can be described as a sampling of ξ_i , φ_i from a probability distribution, $\Pi(\xi, \varphi) = S(\xi)/M$, where $M = N\mu$ is the total mass of the configuration. We propose to expand the surface density (14) into Legendre functions, and then to evaluate the expectation and covariances of the expansion coefficients, c_l^m by means of an ensemble average. These are all standard arguments. There results:

$$c_l^m = \sum_{i=1}^{N} c_l^m(i), \tag{15}$$

$$c_l^m(i) = \frac{(2l+1)}{2\pi}\frac{(l-m)!}{(l+m)!}\exp(-im\varphi_i)\, P_l^m(\xi_i), \tag{16}$$

$$E(c_l^m) = N\int_0^{2\pi} d\varphi_i \int_0^1 \xi_i\, d\xi_i\, \Pi(\xi_i\,, \varphi_i)\, c_l^m(i), \tag{17}$$

and

$$E(c_l^m c_{l'}^{m'*}) = N \int_0^{2\pi} d\varphi_i \int_0^1 \xi_i \, d\xi_i \, \Pi(\xi_i, \varphi_i) \, c_l^m(i) \, c_{l'}^{m'}(i)^*. \tag{18}$$

The evaluation of the integrals is facilitated by expressing $\xi\Pi(\xi, \varphi)$ in terms of Legendre functions, and by treating $\xi\Pi(\xi)$ as an even function to extend the integration on ξ from -1 to $+1$, as described by Hunter [26];

$$\xi\Pi(\xi, \varphi) = (1/2\pi)[P_0^0(\xi) + 2P_2^0(\xi)]. \tag{19}$$

In many Monte Carlo simulations, the basic probability density can be expressed in terms of only the lowest-order eigenfunction; here, the second term gives rise to some interesting features. The expectation values give

$$E(c_0^0) = N/2\pi, \tag{20}$$

$$E(c_2^0) = N/\pi, \tag{21}$$

and

$$E(c_l^m c_{l'}^{m'*}) = \frac{(2l+1)}{(2\pi)^2} N \frac{(l-m)!}{(l+m)!} \delta_{mm'} \delta_{ll'} + \frac{(2l+1)(2l'+1)}{(2\pi)^2} N \frac{(l-m)!\,(l'-m)!}{(l+m)!\,(l'+m)!} \delta_{mm'} \int_{-1}^{1} d\xi \, P_2^0 P_l^m P_{l'}^m. \tag{22}$$

The covariances can be obtained from the expectation for the product, (22). The first term on the righthand side of (22) is the usual diagonal part of the covariance matrix. The second term is more interesting: the integral can be expressed in terms of 3-j symbols [27],

$$\begin{pmatrix} l+2 & l & 2 \\ m & -m & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} l & l & 2 \\ m & -m & 0 \end{pmatrix}.$$

The symbol with $l+1$ does not enter because of the $(l-m)$-even rule, but there is another at $l-2$. It suffices to note the nonzero terms; explicit evaluations of these integrals are not required.

The covariance matrix splits into block-diagonal form, with a separete (infinite) block for each value of m. Each block, labeled by successive even or odd values of l, has symmetrical tridiagonal structure, with the $l' = l-2$, l, and $l+2$ elements nonzero for $l, l' \geqslant 0$.

The diagonal elements of the covariance matrix represent the "noise" mentioned repeatedly.

The important point is that the covariance matrix is *not* diagonal. Thus the Legendre functions cannot be normal modes for the ensemble of N-particle systems used to sample the Kalnajs disk. Furthermore, all elements are of order

N; the covariance matrix does not become diagonally dominant as $N \to \infty$. Even in that limit, the Legendre functions cannot be normal modes of a particle system. This conclusion applies to the linear combinations of omega-models described by Kalnajs [20] as well.

A diagonal covariance matrix may not suffice to assure a set of normal modes. It does not seem worthwhile to undertake the diagonalization of the covariance matrix.

The normal modes of the Kalnajs disks cannot be normal modes for an ensemble of particle systems, even though the same distribution-function was used to describe both ensembles. Particle positions were selected to avoid positional correlations for the Monte Carlo simulation, but that is consistent with the Vlasov equation. This is a disturbing result because the Kalnajs disk is intended to represent just such particle systems.

The omission of time-dependence is not important. As the particle system evolves, it passes through a sequence of spatial configurations, and the argument applies to each of them.

The decomposition of the density and potential into angular Fourier components that has been used throughout this paper remains valid, since the covariance matrix is diagonal in m. The troubles lie in the set of radial functions.

The nondiagonal covariance matrix implies that amplitudes of the "normal modes" cannot be varied independently. This means that a single normal mode cannot be isolated for study in any N-body calculation. There are other difficulties in attempting to isolate independent normal modes of the Kalnajs omega models in an initial value calculation, but there is no need to discuss them. The potentially powerful tool of studying the response of the disk to selected disturbances is not available because there is no means to excite the disk with any disturbance that should produce a response readily amenable to analysis. The best that can be done is to conduct experiments in which a whole class of disturbances is stable or unstable, as was done with the studies of the Kalnajs disks with Ω chosen such that the configuration should be stable or unstable to all axisymmetric disturbances. Individual axisymmetrical modes cannot be studied.

A similar result obtains with deterministic, rather than randomly selected, particle positions. An infinite number of particles is required to construct a configuration in which only one Kalnajs normal mode is excited.

5. Conclusions

The principal motivation for undertaking the polar calculation was to demonstrate that the preference for $m = 2$ features in disk galaxy simulations did not result from the special symmetry of a cartesian grid (point 4 of Sect. 2). Several m-values have equal status in the polar program, so features that depend on the

symmetry of the grid could take a variety of forms. This test could be carried to an extreme by choosing a prime number of angular grid points, N_a, but that has not been necessary.

The resulting program provided sufficiently the fine control over the numerical experiments that we were able to identify physical processes that cause self-gravitating systems to prefer $m = 2$ symmetries (Sect. 4.C). The fine control also permitted useful study of other questions. In particular, several equilibrium models were constructed. One of these (the ring model) showed a stability threshold, which provided additional experimental evidence that stability can be studied in these numerical experiments. This strengthens the arguments concerning points 1, 2, and 8 of Section 2.

The problem of large velocity dispersions in "equilibrium" models remains. The disagreement with observation cannot be a feature of the numerical experiments (Sect. 2), but must result from an incomplete physical model. There must be more to a galaxy than a set of mass-points that interact through forces of self-gravitation.

Natural processes, starting from a variety of uncontrolled initial conditions, could not systematically seek out very special states so critical that they cannot be found in numerical experiments. All the arguments of Section 2 confirmed by the additional evidence reported here, indicate that any large regions of "stability" such as we seek should have been apparent in the computer experiments if they actually existed. Quite a variety of initial conditions has been tried, including some (those with "gentle cooling" [4]) that were close to an equilibrium—and even these would not remain with low velocity dispersions. The failure to find such a model in the computer is not simply an inability to construct an equilibrium model (point 7 of Sect. 2).

A long-lived condition with low velocity dispersion would have to be attained from a wide variety of initial conditions to represent the kind of solution we seek. It should arise as easily as the $m = 2$ features. No such state is found, forcing the conclusion that the physical model is deficient. We shall not speculate on possible modifications to the physical model.

This raises one more serious question: given that the physical model is deficient in that it does not yield "equilibria" with low velocity dispersions, why should our conclusions concerning $m = 2$ features be valid? The $m = 2$ features now stand on a physical process, which can operate with a variety of physical models. The physical process was discovered with the help of the numerical experiments. This underlines the importance of identifying a physical process.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge a National Research Council Senior Postdoctoral Resident Research Associateship at the Ames Research Center, supported by the National Aeronautics

and Space Administration. Mr. David G. Monet of the University of Chicago helped considerably in writing, testing, and initial operation of the polar programs, and especially with pictures of the particle configurations.

REFERENCES

1. R. H. MILLER AND K. H. PRENDERGAST, *Astrophys. J.* **151** (1968), 699.
2. R. H. MILLER, K. H. PRENDERGAST, AND W. J. QUIRK, *Astrophys. J.* **161** (1970), 903.
3. F. HOHL AND R. W. HOCKNEY, *J. Computational Physics* **4** (1969), 306.
4. F. HOHL, *Astrophys. J.* **168** (1971), 343.
5. R. L. MORSE, *in* "Methods of Computational Physics," Vol. 9, p. 213, Academic Press, New York, 1970; C. K. BIRDSALL, A. B. LANGDON, AND H. OKUDA, *in* "Methods of Computational Physics," Vol. 9, p. 241.
6. R. H. MILLER, *Astrophys. Space Sci.* **14** (1971), 73.
7. A. B. LANGDON, "Third Quarterly Progress Report of 1969 on Plasma Research," Electronics Research Laboratory, University of California, Berkeley, pp. 70–71. J. A. BYERS AND M. GREWAL, *Phys. Fluids* **13** (1970), 1819; J. A. BYERS, *in* "Proceedings of the Fourth Conference on Numerical Simulation of Plasmas" (J. P. Boris and R. A. Shanny, Ed.) pp. 496–510, Office of Naval Research, Arlington, Va., 1970.
8. R. H. MILLER, *J. Computational Physics* **6** (1970), 449.
9. R. W. HOCKNEY, *in* "Methods of Computational Physics," Vol. 9, p. 136, Academic Press, New York, 1970.
10. F. HOHL, *J. Computational Physics* **9** (1972), 10.
11. A. TOOMRE, *Astrophys. J.* **139** (1964), 1217.
12. F. HOHL, "Collisionless Relaxation in Spiral Galaxy Models," NASA TN D-7561, 1974.
13. A. B. LANGDON, *J. Computational Physics* **6** (1970), 247.
14. F. HOHL, *Astrophys. J.* **184** (1973), 353.
15. R. W. HOCKNEY, *J. Computational Physics* **8** (1971), 19.
16. R. H. MILLER, *Astrophys. J.* **190** (1974), 539.
17. P. O. VANDERVOORT, *Astrophys. J.* **162** (1970), 453.
18. R. W. HOCKNEY AND D. R. K. BROWNRIGG, *Mon. Not. Roy. Astron. Soc.* **167** (1974), 351.
19. S. CHANDRASEKHAR, "Principles of Stellar Dynamics," Sect. 4.8, Dover, New York, 1960.
20. A. J. KALNAJS, *Astrophys. J.* **175** (1972), 63.
21. O. BUNEMAN, *J. Computational Physics* **1** (1967), 517.
22. A. TOOMRE, *Astrophys. J.* **138** (1963), 385; M. CLUTTON-BROCK, *Astrophys. Space Sci.* **16** (1972), 101; B. BARBANIS AND K. H. PRENDERGAST, *Astron. J.* **72** (1967), 215.
23. F. HOHL, "Dynamics of Disk Galaxies," NASA TR R-343, 1970.
24. J. C. MAXWELL, "On the Stability of the Motions of Saturn's Rings," Macmillan, Cambridge/London, 1859. Also reprinted in "Scientific Papers of James Clerk Maxwell," Vol. I, Cambridge Univ. Press, 1890.
25. A. F. COOK AND F. A. FRANKLIN, *Astron. J.* **69** (1964), 173.
26. C. HUNTER, *Mon. Not. Roy. Astron. Soc.* **126** (1963), 299.
27. A. R. EDMONDS, "Angular Momentum in Quantum Mechanics," Princeton Univ. Press, Princeton, N. J., 1957.