THE ASTROPHYSICAL JOURNAL, 499:93–111, 1998 May 20

SECULAR EVOLUTION OF SPIRAL GALAXIES. II. FORMATION OF QUASI-STATIONARY SPIRAL MODES

XIAOLEI ZHANG
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Mail Stop 78, Cambridge, MA 02138
Received 1995 November 1; accepted 1997 September 23

ABSTRACT

In this second of a series of three papers, we demonstrate that the kind of spiral patterns in disk galaxies which can remain quasi-stationary on the timescale of a Hubble time are spontaneously formed spiral modes. The amplitude of the spiral mode is limited to a finite value, determined by the basic state characteristics, through a collective dissipation process, introduced in Paper I of this series.

We show that the radial distribution of the total torque coupling integral of a spontaneously formed spiral mode must be (and is found to be) of a characteristic bell shape, with the peak of the bell located at the corotation radius of the mode. This distribution of the total torque coupling is concordant with the radial distribution of the phase shift between the potential and density spirals (Paper I). Whereas the existence and the sign of the total torque coupling imply that there is angular momentum being carried outward by a trailing spiral density wave mode, the bell shape further indicates that there is angular momentum being deposited on each annular ring of the galactic disk en route of this outward angular momentum transport. The amount of the deposited angular momentum is negative inside corotation and is positive outside corotation. Since a spiral mode has negative angular momentum density inside corotation and positive angular momentum density outside corotation, it follows that the bell-shaped torque coupling of a trailing spiral model leads to its own spontaneous growth in the linear regime. In going from the linear to the progressively nonlinear regimes, an increasingly larger fraction of the deposited angular momentum is channeled onto the basic state of the disk through a collisionless shock, rather than being used entirely for the wave mode growth. Finally, in the fully nonlinear regime (the quasi-steady state of the wave mode), all of the angular momentum deposited by the trailing spiral mode is given to the basic state. The spiral mode can then remain quasi-stationary on the order of a Hubble time, at the expense of a continuous dissipative basic state evolution.

It is argued that the spontaneous formation and stabilization of a large-scale spiral mode in a disk galaxy is an example of a nonequilibrium phase transition.

Subject headings: galaxies: evolution — galaxies: kinematics and dynamics — galaxies: spiral — galaxies: structures

1. INTRODUCTION

It has long been realized that the type of global structures formed spontaneously in nonequilibrium systems are not neutral structures. They are, rather, large-scale instabilities of the underlying systems, which are sustained by a continuous flux of energy and entropy, as well as by a local dissipation process which offsets their spontaneous growth tendency. The belief that long-lasting coherent structures in nonequilibrium systems are in a state of dynamical equilibrium is reinforced by the well-known "fluctuation-dissipation" theorem for thermodynamic systems, which implies that no finite amplitude fluctuation can be sustained without invoking a corresponding dissipation process.

The spiral structure of galaxies appears to be another example of a coherent nonequilibrium structure, so we expect that the maintenance of spiral patterns involves a similar competition and cooperation between a spontaneous growth tendency and local dissipation. The capability of a trailing spiral mode to transport angular momentum outward (Lynden-Bell & Kalnajs 1972; hereafter LBK) and to deposit angular momentum in the course of this outward transport (this paper) allows the spiral mode to grow spontaneously. On the other hand, the local instability condition at the spiral arms and the presence of a phase shift between the potential and density spirals allow the deposited angular momentum to be channeled onto the basic state through a collisionless shock, at a finite amplitude of the wave (Zhang 1996, hereafter Paper I). A quasi steady state is obtained when the above two processes reach dynamical equilibrium.

We show in the following sections, through analytical arguments and numerical simulations, that the mechanism outlined above appears to be the only possible explanation for the quasi-stationary spiral modes obtained in the N-body simulations. In particular, a separate dissipative gas component is not required, nor is excessive heating of the disk stars needed in order to achieve wave amplitude stabilization. We further demonstrate that it is not sufficient for the mechanism responsible for wave stabilization to be merely nonlinear (while reversible). The relevant mechanism must also involve a dissipative energy and angular momentum exchange between the basic state of the disk and the spiral density wave.

The organization of the paper is as follows. In § 2 we describe a spontaneously formed N-body spiral mode, which gives us some hints on the general characteristics of the mechanism underlying the wave stabilization process. In § 3, we present analytical arguments to show that a coherent picture of the behavior of a spontaneous spiral mode in the linear regime as well as in the quasi steady state can be obtained from analyzing the radial distribution of its total torque coupling integral. Furthermore, we show that consistent results can be obtained for the modal growth and stabilization process from either the two-wave super-

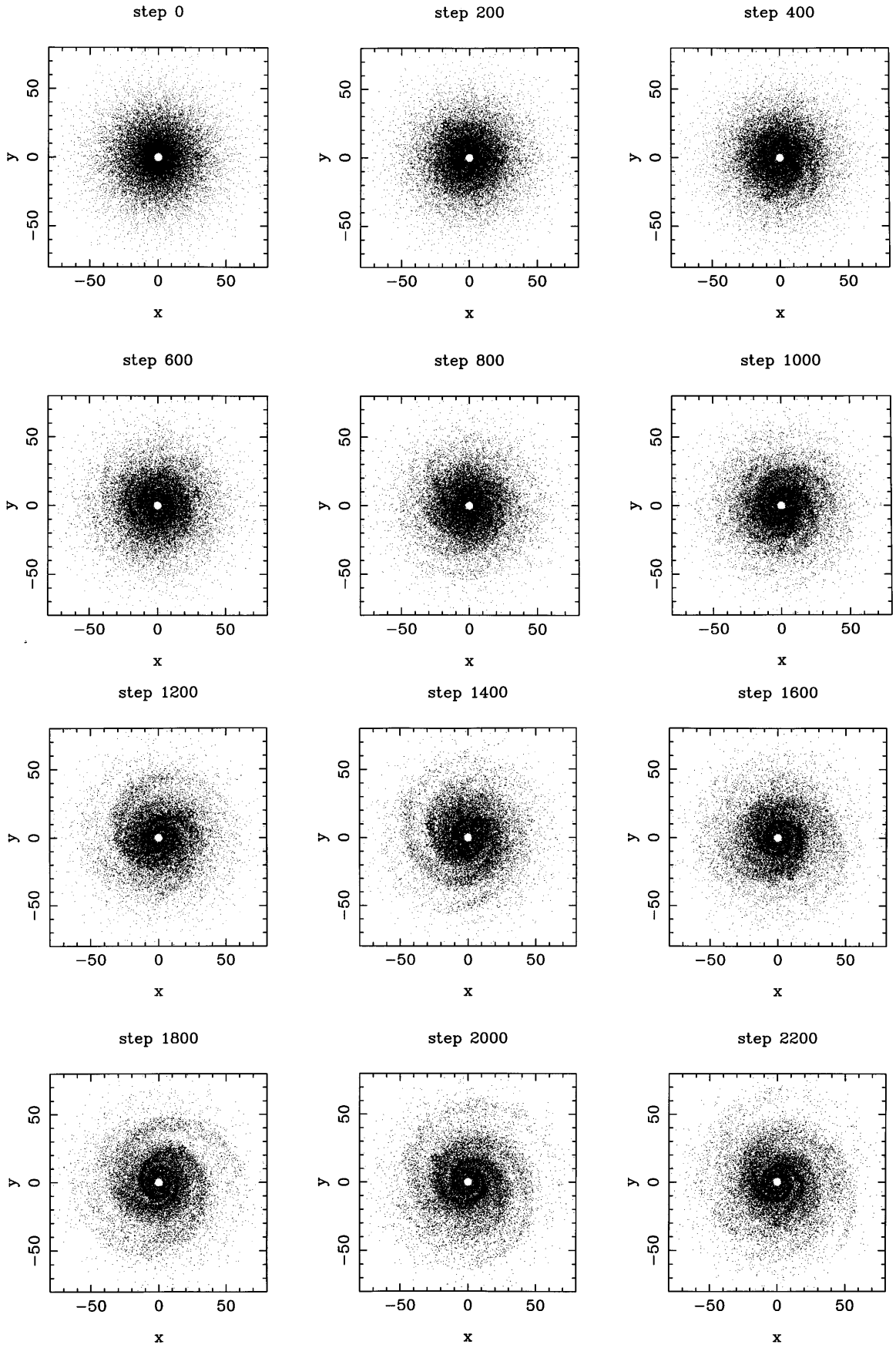

FIG. 1.—Morphological evolution of an N-body spiral mode in a purely stellar disk. The simulation and the basic state parameters used are described in § 2 of the text. About 20,000 randomly chosen particles are plotted in each frame. The star symbol denotes a particle at the corotation radius which revolves around the galactic center with the pattern speed of the spiral.

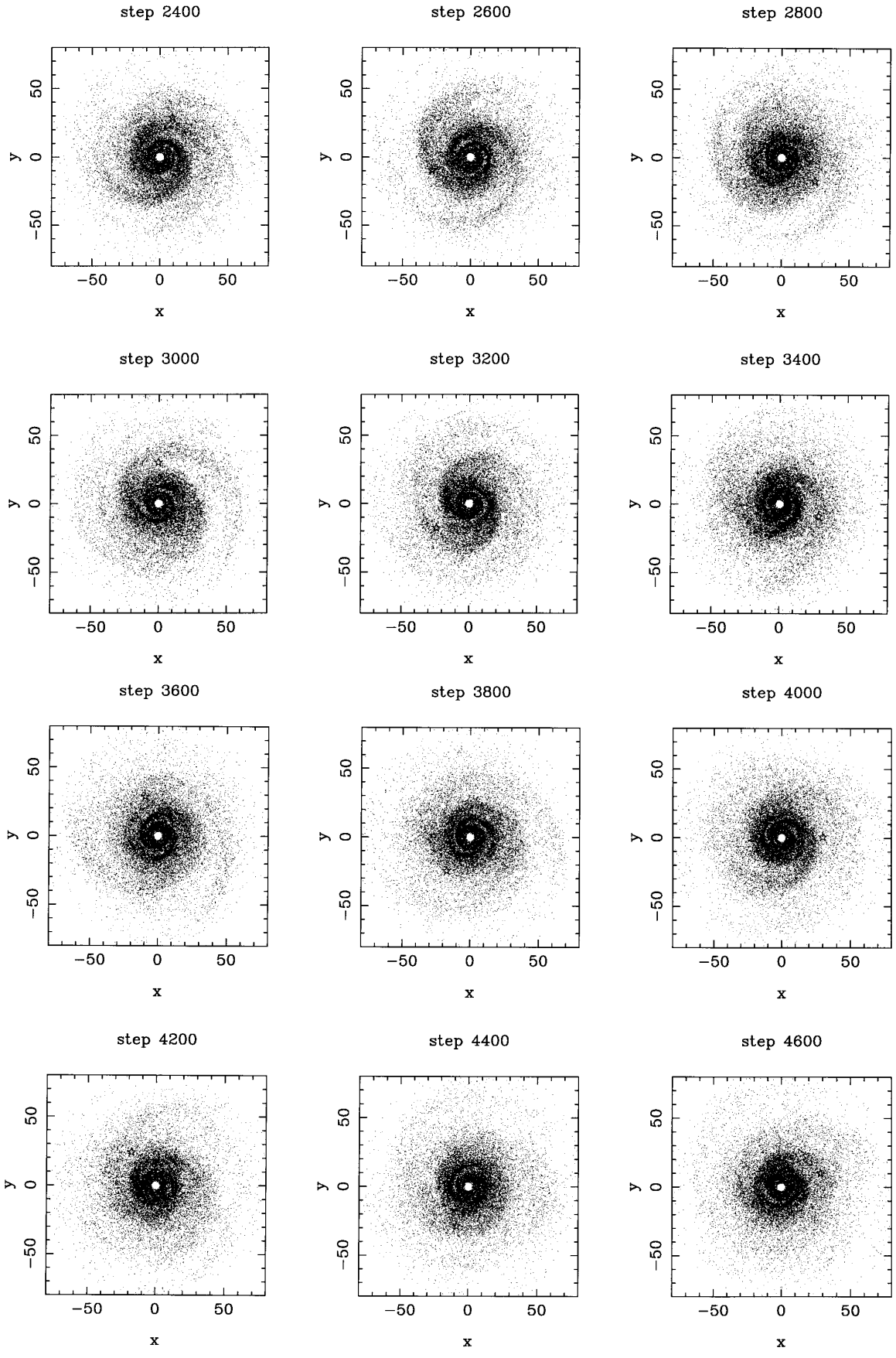

FIG. 1—*Continued*

position point of view or from the modal point of view. This reconciles the results of the current work with some of the previous published ones. § 4 analyzes the role of gas in the spiral wave stabilization process. In § 5, we show that by a proper reinterpretation and rescaling of the N-body results, we could arrive at the conclusion that certain members of the physical spiral galaxies are able to remain quasi-stationary on the order of a Hubble time. In § 6, we discuss the relation of the spontaneous formation and stabilization of spiral modes to nonequilibrium phase transitions in general, as well as the sense and the role of "dissipation" in an essentially Hamiltonian N-body system. Finally in § 7, we summarize the conclusions of this paper.

2. AN EXAMPLE OF A SPONTANEOUSLY FORMED, N-BODY QUASI-STATIONARY SPIRAL MODE

The N-body code used for performing the following simulation is a two-dimensional polar grid code. More detailed description of this code, including the units and the normalization scheme used, can be found in Thomasson (1989), Donner & Thomasson (1994), as well as in Paper I.

In the simulation described in this section, we have used a polar grid of 55 radial divisions and 64 azimuthal divisions. The time step used corresponds to 314 time steps per revolution period at radius 20. About 100,000 equal-massed particles are used to represent the active disk component. The above spatial and temporal resolutions are both twice as coarse as that used in the Paper I simulation. A gravity softening length of 1.5 times the grid length unit is used.

The basic state specification is similar to that of Paper I, with a modified-exponential-disk component and a rigid bulge and rigid halo components. The initial value for the stability parameter Q is 1.1. However, in the current case, the mass ratio of disk:halo:bulge is 0.4:0.5:0.1 instead of 0.5:0.4:0.1 as in Paper I. This reduced active disk mass avoided the early central bar formation observed in the Paper I simulation, and this allowed the spiral mode to survive for a much longer time.

In the different frames of Figure 1, we plot the evolution of the disk morphology at various time steps. For ease of keeping track of the pattern rotation, a tracer star at the corotation radius ($r = 30$), whch revolves around the galactic center at the local circular speed, is also drawn onto each frame. The spiral pattern is seen to have kept the global coherence throughout the period from time step 1200 to 4600, which is about six pattern revolutions (the pattern speed is about 0.011 radian per time step). The inner spiral pattern inside the corotation radius is seen to have maintained the longest period of pattern coherence: it in fact survived for a couple more revolutions after step 4600. This is the longest period that a coherent spiral shock pattern (not just the underlying $m = 2$ mode) has been maintained in N-body simulations of spontaneous spirals. Careful treatment of the initial velocity assignment scheme to preserve the self-consistency of the basic state, as well as the choice of basic state itself both contributed to the longevity of this mode.

At larger time steps, the nonlinear spiral pattern gradually fades away due to the heating of the disk stars. However, Fourier analysis shows that the underlying $m = 2$ mode retains nearly constant amplitude, after the initial exponential growth, until the very end of the simulation run (time step 8000), which is about 14 pattern revolutions (Fig. 2). The pattern speed of the $m = 2$ component is also found to be nearly constant throughout the simulation run. In § 5, we will show that the heating of the disk stars in the N-body simulation is much exaggerated due to the small number of particles used, compared to that which is present in a physical galaxy. The amount of this excess heating can be calculated, and true heating due to the spiral pattern alone can be obtained after a simple rescaling of the N-body results. We will then see that for many disk galaxies the nonlinear spiral shock patterns can survive on the order of a Hubble time.

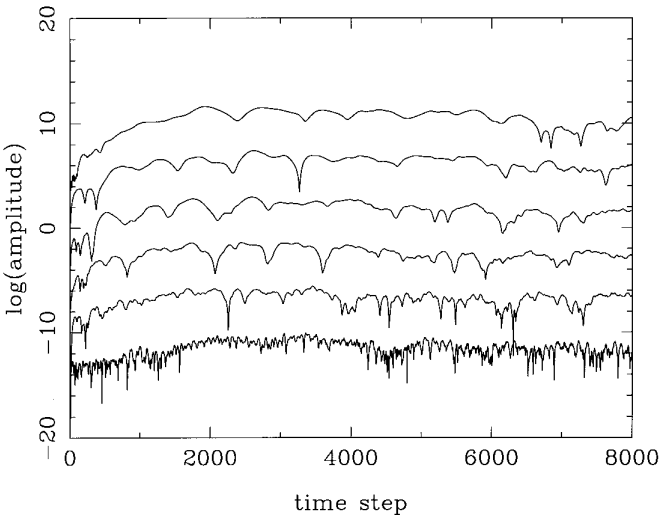

FIG. 2.—Amplitude evolution of the $m = 2$ component of the spiral mode in Fig. 1. From bottom to top, the different curves are for radius 10, 20, 30, 40, 50, and 60, respectively. The curves are shifted in the vertical scale by multiples of 5 with respect to the $r = 10$ curve.

In Figure 3, we plot the evolution of Toomre's Q parameter (Toomre 1964). It can be seen that the Q parameter is monotonically increasing during the entire simulation run and that the stabilization of the wave amplitude after time step 1200, observed in Figures 1 and 2, has no observable correlation with the increase of the Q parameter, which measures the degree of disk heating in this case. In fact, at step 1200 when the wave amplitude saturates, the disk is still rather cool.

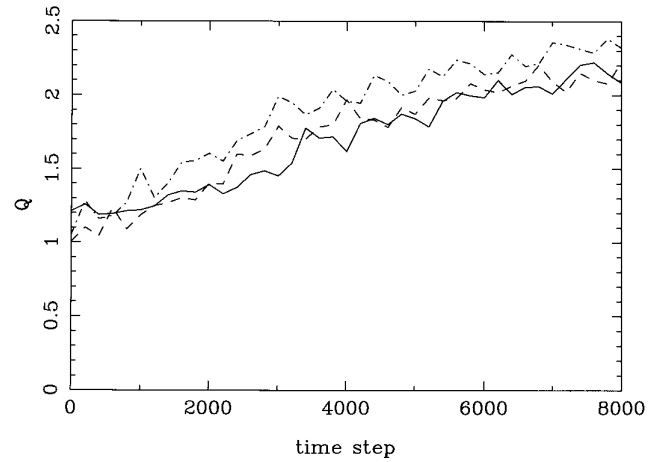

FIG. 3.—Evolution of Q-parameter at radius $r = 20$ (*solid*), $r = 30$ (*dash*), and $r = 40$ (*dash-dot*), respectively, with time for the N-body mode of Fig. 2.

We can thus make the following observations about the amplitude saturation process of this N-body spiral mode:

1. The amplitude stabilization is not due to the dissipation in the gas component since no such component is included in the current simulation.
2. The amplitude stabilization is not due to the excess heating of the disk stars either since the equilibrium amplitude is reached when the disk is still rather cool and the subsequent continued heating of the disk does not significantly change the equilibrium amplitude.
3. The amplitude stabilization is not due to the channeling of wave energy into multi-armed spiral modes since the spiral pattern remains predominantly two-armed throughout the simulation run.
4. The presence of a rigid halo component did not stop the initial exponential growth of the wave mode, and the subsequent wave amplitude saturation could not be due solely to the presence of the halo component.

Some nonlinear and dissipative mechanisms appears to underlie the wave-amplitude stabilization process. In the next section, we will analyze the nature of this process.

3. THE PROCESS OF REACHING THE QUASI STEADY STATE

The quasi-stationary spiral pattern we obtained in the last section was clearly not a single traveling wave since the group velocity of such a wave (Toomre 1969) would cause the wave to travel to beyond the outer extent of the disk, or to reach the inner boundary, in just a few orbital times. The well-defined and nearly constant pattern speed of the spiral pattern indicates that it is an unstable mode of the underlying disk, amplified through the overreflection mechanism at the corotation region (Goldreich & Lynden-Bell 1965; Julian & Toomre 1966; Mark 1976; Toomre 1981), as well as a feedback mechanism near the center of the disk (Lin & Lau 1979).

An unstable mode is capable of growing spontaneously and homogeneously. Since a spiral mode has negative energy and angular momentum densities inside corotation and positive densities outside corotation, the spontaneous growth of the spiral mode requires that the radial derivative of its total torque coupling integral $C(r)$, which determines how much angular momentum is left onto each annular ring as a result of the torquing by the disk material on both sides of the annular ring, must allow negative angular momentum to be deposited onto each annular ring inside corotation and positive angular momentum to be deposited outside corotation. In another words, a spiral perturbation will grow spontaneously if the shape of its $C(r)$ in the linear regime allows it to be self-enhancing.

We show in this section that a characteristic bell shape of the torque coupling integral, which is consistent with the presence and the sign of the potential/density phase shift at both sides of corotation, is responsible both for the spontaneous growth of the spiral mode in the linear regime and for the dissipative basic state evolution at the quasi steady state of the wave mode. This appears to be an extension of the well-known "fluctuation-dissipation theorem" into the far-from-equilibrium regime, i.e., the kind of spiral modes capable of spontaneously emerging as large-scale fluctuations have built-in characteristics such that they can only be sustained as quasi-stationary patterns through the action of local dissipation.

3.1. *Radial Distribution of the Torque Coupling Integral for an Unstable Mode in the Linear Regime*

The total torque coupling integral due to a spiral structure is the sum of the gravitational torque coupling integral C_g and the advective torque coupling integral C_a (LBK). The requirement of local angular momentum balance in an annular ring of unit width at radius r, during the outward angular momentum transport process by a trailing spiral wave, leads to (see § A1 of the Appendix of this paper)

$$\frac{dC}{dr} = \frac{d(C_a + C_g)}{dr} = -2\pi r \frac{d\bar{L}}{dt} , \tag{1}$$

where $\bar{L}$ is the azimuthally averaged angular momentum density at radius r and where $C > 0$ for a trailing spiral mode (Paper I).

Now $\bar{L}$ can be decomposed into $\bar{L}_{\text{basic state}} + \bar{L}_{\text{wave}}$. During the linear modal growth process, $d\bar{L}_{\text{basic state}}/dt = 0$ away from the resonances,[1] and the angular momentum deposited by C is used entirely for the increase of the wave amplitude. So we have

$$\frac{dC}{dr} = \frac{d(C_a + C_g)}{dr} = -2\pi r \frac{d\bar{L}_1}{dt} , \tag{2}$$

where $\bar{L}_1$ is the azimuthally averaged angular momentum density of the wave mode at radius r.

Since inside corotation, $\bar{L}_1 < 0$, modal growth means that $d\bar{L}_1/dt < 0$ inside corotation. This demands further, through equation (2), that $dC/dr > 0$ inside corotation. Similarly, modal growth demands $dC/dr < 0$ outside corotation. Therefore we arrive at the conclusion that *for an unstable mode in the linear regime, the radial profile of its total torque coupling integral must be that of a characteristic bell shape, with the top of the bell at the corotation radius.*

For *homogeneous* modal growth throughout the disk, we must further have

$$\frac{d\bar{L}_1}{dt} = 2\gamma_{g0} \bar{L}_1 , \tag{3}$$

where γ_{g0} is the (constant) amplitude growth rate of the mode in the linear regime (thus $2\gamma_{g0}$ for the angular momentum growth rate).

We thus have, from equations (2) and (3),

$$\gamma_{g0} = -\frac{dC/dr}{4\pi r \bar{L}_1} = \text{constant} ; \tag{4}$$

in other words, the magnitude of the linear growth rate γ_{g0} of an unstable mode is *determined* by its radial gradient dC/dr of the total torque coupling $C(r)$, as well as by its angular momentum density $\bar{L}_1$, with the ratio of the two being a constant independent of radius.

In the Appendix, § A2, we show that in the linear regime, the gravitational torque coupling C_g is related to the torque integral $\bar{\mathcal{T}}$ (Paper I) through

$$\frac{dC_g}{dr} = -2\pi r \bar{\mathcal{T}}(r) ,$$

$$= r \int_0^{2\pi} \Sigma_1(r, \phi) \frac{\partial \mathcal{V}_1(r, \phi)}{\partial \phi} d\phi , \tag{5}$$

[1] This result comes about simply from the definition of the basic state, which is the steady azimuthally symmetric part of the disk on top of which linear perturbations of sinusoidal form are superimposed in the linear modal calculation.

and in the Appendix § A2, we have also defined a new variable $T_1(r) \equiv -2\pi r\bar{\mathcal{F}}(r)$.

Since in the linear regime, the distribution of $C(r) = C_a(r) + C_g(r)$ for an unstable spiral mode is dominated by the distribution of $C_g(r)$[2], we see that the bell-shaped $C(r)$ distribution is consistent with the existence and the sign of phase shift for a trailing spiral, which gives a positive phase shift and a positive $T_1(r)$ inside corotation and negative ones outside corotation. In other words, *the existence and the sign of the phase shift between the potential and density spirals of a trailing spiral mode is what leads to the spontaneous growth of such a mode in the linear regime.*

3.2. *From the Linear Regime to the Quasi Steady State*

In our N-body simulations, as the wave amplitude increases away from the linear regime, it is found that the disk stars begin to show radial migration: the stars inside corotation spiral inward, and the stars outside corotation spiral outward. The exponential growth of the wave amplitude gradually saturates, and eventually the wave growth halts completely.

When the wave amplitude stops growing, the spiral pattern still has a very coherent appearance, and calculations show that the total torque coupling due to the spiral pattern retains a bell shape throughout the modal growth process, as well as at the quasi-steady state. The typical distribution of the different torque coupling integrals at the quasi steady state is shown in Figure 4, for the spiral mode calculated in Paper I (we used this spiral mode instead of the one calculated in § 2 because the one in Paper I was calculated with a much higher resolution and is also the one with more coherent and smooth distributions of torque couplings. The modal characteristics, including the locations of the various resonances, however, are very similar between the two cases). The total torque coupling integral in Figure 4 clearly appears to have a bell shape, which is peaked at the corotation radius ($r = 30$). Since the wave amplitude at that time step is quite steady, there is little doubt that such a bell shape for the torque couplings is intrinsic to the quasi-stationary spiral mode under concern and is not due to some spurious effect of the N-body simulation, such as the broadening of the resonances due to fluctuating wave amplitude. In fact, in Figure 15 of Donner & Thomasson (1994), which calculates the evolution of Toomre's Q parameter as a result of heating by the very same spiral mode as presented here, we can clearly see a local peak in the heating curves near $r = 10$, the location of the inner Lindblad resonance for this mode. Therefore the broadening of that resonance could not be more than the width of that narrow peak and certainly could not have covered the entire disk as our current bell shape covers.

Note that in the various torque coupling curves shown in Figure 4, there is also a second weak hump at the outer disk. This second hump is likely to be due to the reflection of the wave energy from the outer disk edge, which results in the formation of a secondary mode outside the outer Lindblad resonance of the dominant mode, which is located at $r = 42$ (Donner & Thomasson 1994).

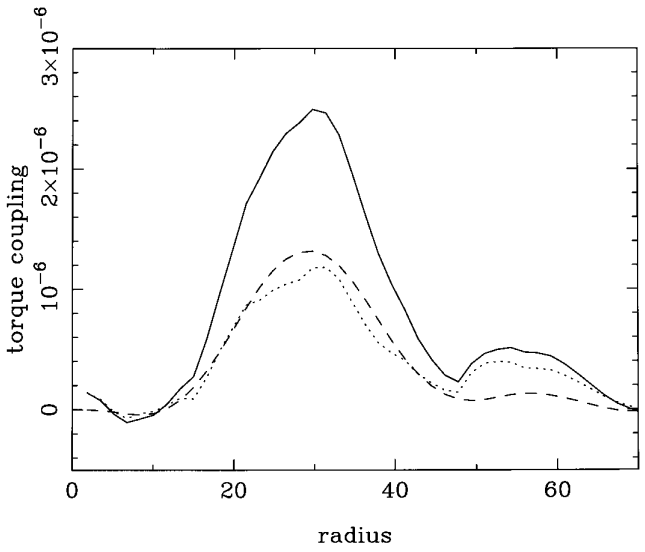

FIG. 4.—Gravitational torque coupling (*dotted line*), advective torque coupling (*dashed line*), and total torque coupling integral at time steps 1600, for the N-body spiral mode of Paper I. For this mode, the corotation radius is $r = 30$, the inner Lindblad resonance is at $r = 10$, and outer Lindblad resonance is at $r = 42$.

The bell-shaped total torque-coupling integral such as shown in Figure 4 indicates that the spiral pattern continues depositing energy and angular momentum onto the disk throughout the modal growth process, as well as at the quasi steady state of the wave mode. Since at the quasi steady state the deposited energy and angular momentum no longer increase the amplitude of the wave, they can only be left onto the basic state. This is consistent with the observed radial migrations of the basic state stars beyond the linear regime.

This channeling of the deposited wave energy and angular momentum onto the basic state is achieved by the spiral gravitational shock at the spiral arms (Paper I). When stars are crossing the spiral arms, the near-collision condition produced by the shock destroys the phase coherence of the orbits, so stars can no longer align themselves perfectly to that which corresponds to a kinematic spiral (Kalnajs 1973) in order to reinforce the wave amplitude. The sudden velocity deceleration at the shock, for a star inside corotation, also means that the star loses part of its orbital angular momentum to the wave (or the wave deposits negative angular momentum onto the basic state star), and therefore spirals inward. As the wave amplitude increases and therefore the shock gets narrower and stronger, an increasingly larger fraction of the deposited negative energy and angular momentum by the wave is used for the braking (deceleration) of the star's orbital motion, rather than for the realignment of particles to contribute to the growth of the wave. This characteristic of the nonlinear regime is also partially reflected in the sign of C_a shown in Figure 4, which is the same as the sign of C_g, unlike what we should expect for such an open wave in the linear regime.

At the quasi steady state of the wave mode, instead of

[2] This is because in the linear regime, in the hydrodynamic limit of $ka < 1$, where k is the wave number and a is the sound speed of the medium, C_a always opposes C_g in terms of the direction of angular momentum flux it carries (LBK; Goldreich & Tremaine 1979), in between the inner and outer Lindblad resonances (which is the area generally covering most of the galactic disk). Since a trailing spiral mode can grow only by transporting angular momentum outward, and since C_g for a trailing spiral pattern is always positive (which leads to outward angular momentum transport), it follows that the kind of trailing spiral modes capable of spontaneous growth *must* have its C_g larger in magnitude then C_a in the linear regime. This conclusion is also borne out from analyzing the detailed expressions of C_a and C_g for the various wave branches of WKBJ waves (Goldreich & Tremaine 1979).

having $dC_g/dr = T_1(r)$ as is the case for the linear regime, the proper closure relation now appears to be $dC/dr = dC_a/dr + dC_g/dr = T_1(r)$ (Appendix, § A2). All the angular momentum deposit indicated by dC/dr [or equivalently $T(r)$] now goes to the basic state, and none is used for wave growth.

3.3. *Characteristics of the Quasi Steady State*

As we have already mentioned, the quasi steady state for a spontaneously formed spiral mode is likely to be in a dynamical equilibrium state with the spontaneous growth tendency of the mode balancing the dissipative damping tendency. In some of the published work on the subject of wave-damping mechanisms, it has often been speculated that the relevant damping mechanism must be nonlinear. In this section, however, we will show that any reversible nonlinear mechanism (which serves to deform the azimuthal profile of the wave) by and of itself is not capable of limiting the wave amplitude to a finite value in a finite time. The relevant wave stabilization mechanism must also involve a dissipative energy and angular momentum exchange process between the spiral density wave and the basic state of the disk.

From past studies, we know that the characteristics of the waves in the nonlinear regime include the modifications of the dispersion relation (Norman 1978), as well as the change of the azimuthal profile of the wave (Lubow, Balbus, & Cowie 1986), from their corresponding forms in the linear regime. However, it was found that the amplitude of the wave usually remains as an undetermined free parameter in these nonlinear wave solutions (Yueh 1981), which indicates that, in general, a reversible nonlinear deformation of the wave profile does not lead to the prediction of a finite equilibrium wave amplitude.[3]

The generic form of the amplitude evolution equation of a spontaneously formed spiral mode is

$$\frac{dA}{dt} = \gamma_g(A)A \ . \tag{6}$$

The influence of nonlinearity on the growth of a spontaneous spiral mode is expected to be chiefly a reduction of the growth rate $\gamma_g(A)$. This is because as the wave amplitude increases as a result of overreflection at the corotation and feedback, the azimuthal profile of the wave gradually becomes nonsinusoidal. The growth rate $\gamma_g(A)$ of the mode thus decreases since part of the wave energy now is channeled into the higher order harmonics in azimuth, which cannot be directly amplified by the galactic resonant cavity. However, the growth rate will never actually reach zero through the deformation of the azimuthal wave profile alone, i.e., we have $\gamma_g(A) \to 0$ as A increases, but always $\gamma_g(A) > 0$ for an unstable mode in the nonlinear regime. This is especially true for the spiral patterns in physical galaxies, where the departure of the azimuthal wave profiles from sinusoidal is never serious, and the $m = 2$ sinusoidal modal component is always dominating. Thus the amplitude of the mode will grow without bound in the presence of a nonlinear but dissipationless mechanism alone.

However, the dissipative channeling of the wave energy and angular momentum onto the basic state through the action of the spiral shock gives rise to an effective amplitude damping rate γ_d, with the net growth rate of the mode being $\gamma_g' = \gamma_g - \gamma_d$, which *replaces* the γ_g in equation (6).

It is expected that γ_d also increases with increasing wave amplitude. This is because γ_d, being the manifestation of a *local* dissipation process, depends mainly on the *steepness* of the shock, which is itself determined by local values of the wave amplitude and potential/density phase shift (Paper I). Through a direct numerical calculation it can be shown that the phase shift itself is mainly determined by the pitch angle of the spiral and by its radial density envelope function and is not sensitively dependent on the azimuthal (nonlinear) profiles of the spiral pattern. In fact, for azimuthal density profiles as diverse as that shown in Figure 5, the calculated phase shift distribution,[4] as shown in Figure 6, are amazingly similar. Therefore, $\gamma_d[A, \phi_0(A)] \approx \gamma_d(A, \phi_0)$, and it tends to increase with increasing wave amplitude.

The trend of variation of $\gamma_g(A)$ and $\gamma_d(A)$ stated above means that the effective growth rate $\gamma' = \gamma_g(A) - \gamma_d(A)$ will reach zero at a finite wave amplitude: this amplitude is the equilibrium amplitude of the wave mode, at which $dA/dt = 0$ according to equation (6). Since the manner in which the growth rate decreases with the wave amplitude, as well as the manner in which the dissipation rate increases with the wave amplitude, are both determined by the basic state, as well as by the modal properties, with the latter determined ultimately also by the basic state, we see that the equilibrium amplitude of a spiral mode depends only on the basic state characteristics.

3.4. *Reconciliation of the Two-Wave Superposition and the Unstable-Mode Points of View*

It is a well-known fact from general wave theory that two oppositely propagating traveling waves of the same frequency and the same amplitude sum up to a standing wave. However, the issue of what kind of characteristics the two traveling waves must have in order that their superposition leads to a spontaneously growing unstable mode is hardly ever addressed. We show below that in a simplified one-dimensional case, an exponentially decaying profile for the wave action (we use the phrases "wave action" and "total torque coupling integral" interchangeably in the following) is needed along each wave's propagation path, which, together with the overreflection at corotation and the feedback near the galactic center, leads to a spontaneously and homogeneously growing wave mode. This understanding also helps to bridge a connection between the current work and some of the previously published work, notably that of LBK and Goldreich & Tremaine (1979).

[3] An exception is the WKBJ soliton found in Norman (1978), which has an equilibrium amplitude determined by the pitch angle of the wave. This special case of the soliton does not bear directly on the nature of an open and unstable spiral mode since a soliton does not have a spontaneous growth tendency and thus it is possible for nonlinearity to offset the dispersiveness of the wave to achieve quasi-steady state.

[4] A reasonable choice for the definition of an equivalent phase shift ϕ_0 between two nonlinear waveforms is

$$\phi_0 = \frac{1}{m}\sin^{-1}\left[\frac{1}{m}\frac{\int_0^{2\pi}\Sigma_1(\partial\mathscr{V}_1/\partial\phi)d\phi}{(\int_0^{2\pi}\mathscr{V}_1^2\,d\phi)^{1/2}(\int_0^{2\pi}\Sigma_1^2\,d\phi)^{1/2}}\right],$$

where Σ_1 represents the perturbation density wave form and $\mathscr{V}_1$ the perturbation potential waveform, and m is the number of spiral arms. The equivalent phase shift is the amount of phase shift which would be present between two sinusoidal waveforms if each one is endowed with the same energy as the corresponding nonlinear waveform, and which would lead to the same value for the torque integral as would the nonlinear waveforms. Note that in the above expression the perturbation waveforms must have their azimuthal mean values subtracted.

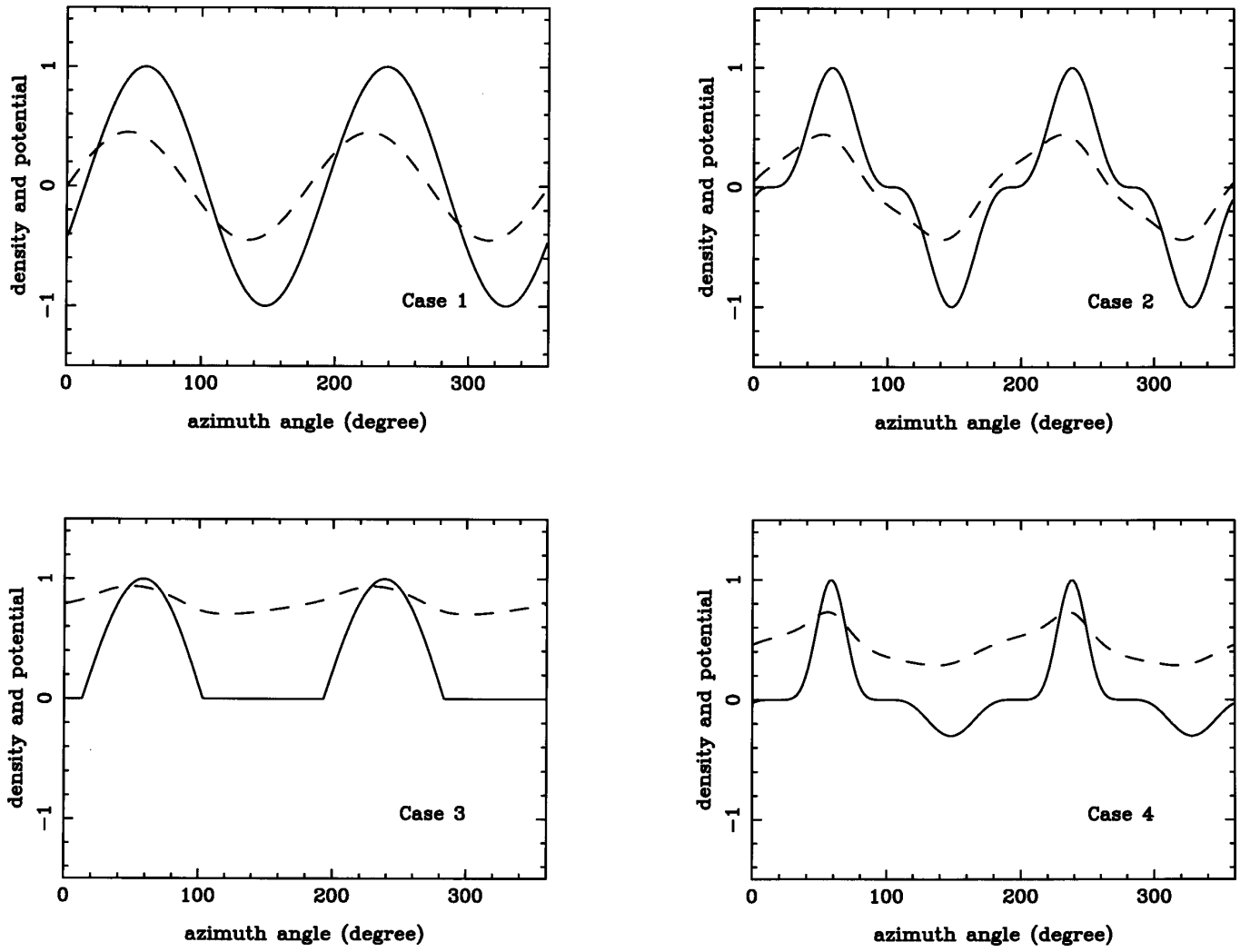

FIG. 5.—Azimuthal profiles of the spiral density (*solid lines*) and spiral potential (*dashed lines*) calculated through the Poisson integral of the corresponding density, for several truncated spiral waves.

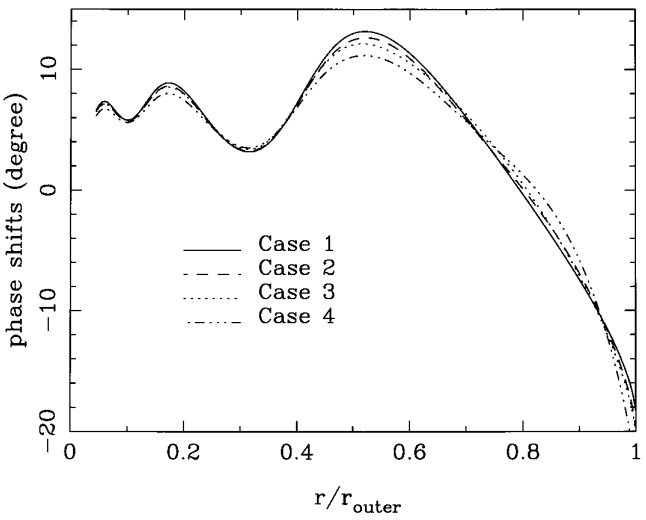

FIG. 6.—Radial dependence of the equivalent phase shifts for the truncated spiral waves of Fig. 5. The ripples observed are due to the truncation of the computation domain at r_{outer}. The phase shifts should be a constant independent of radius for an infinitely long constant amplitude spiral wave.

Since in this section we are mainly interested in the general trend of the distribution of the magnitude of $C(r)$, we assume, without loss of generality, a one-dimensional wave propagation and reflection scenario. The two dimensional picture will be needed when considering the detailed phase relation of the different wave branches, the superposition of which gives rise to the normal modes of a given basic state. The direction of wave propagation in our one-dimensional case would correspond to the radial direction in a physical galactic disk. We assume further that the wave is amplified only at corotation radius r_{co}, with an overreflection factor Γ for the wave action. The overreflection mechanism at corotation, besides connecting the two oppositely propagating waves inside r_{co}, also produces an outgoing wave outside r_{co} (Mark 1976). We also assume a constant group velocity c_g for all the wave trains, as well as an inner-turning point of the wave located close to the center of the galaxy.

The exponential profiles we assumed for the three waves are illustrated in Figure 7. For $r \leq r_{co}$, the spatial distribution of the inwardly propagating wave C_1 is

$$C_1(r) = C_0 \exp\left(\frac{r - r_{co}}{r_0}\right), \tag{7}$$

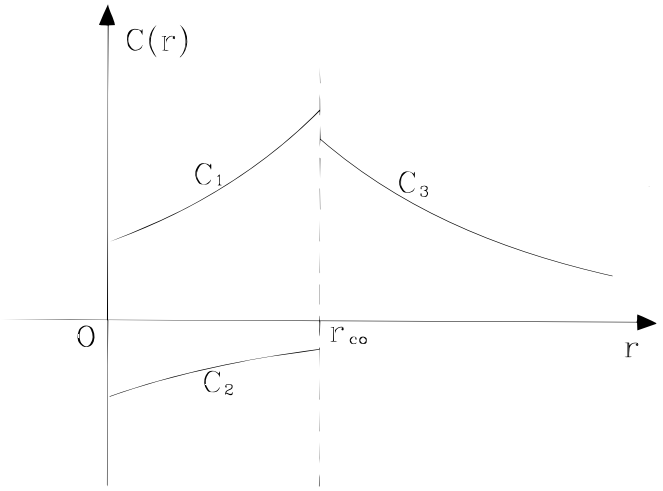

FIG. 7.—Superposition of wave trains to form an unstable mode

and the spatial distribution for the outwardly propagating wave C_2 is

$$C_2(r) = -C_0 \exp\left(\frac{-r - r_{co}}{r_0}\right), \tag{8}$$

where r_0 is a constant scale length. The overall intensity of the wave inside corotation is the linear superposition of C_1 and C_2, i.e.,

$$C_{12}(r) = C_1(r) + C_2(r),$$
$$= C_0 \exp\left(-\frac{r_{co}}{r_0}\right)\left[\exp\left(\frac{r}{r_0}\right) - \exp\left(-\frac{r}{r_0}\right)\right], \tag{9}$$

The radial gradient of C_{12} is now (for $r < r_{co}$)

$$\frac{dC_{12}(r)}{dr} = \frac{C_0}{r_0} \exp\left(-\frac{r_{co}}{r_0}\right)\left[\exp\left(\frac{r}{r_0}\right) + \exp\left(-\frac{r}{r_0}\right)\right], \tag{10}$$

which is positive.

To write also the expression for the wave branch outside corotation, we note that $C_1(r_{co}) = C_0$, and $C_2(r_{co}) = -C_0 \exp[-2(r_{co}/r_0)]$, therefore the conservation of wave action across corotation requires that

$$C_3(r) = [C_1(r_{co}) + C_2(r_{co})] \exp\left(-\frac{r - r_{co}}{r_0}\right),$$
$$= C_0\left[1 - \exp\left(-2\frac{r_{co}}{r_0}\right)\right] \exp\left(-\frac{r - r_{co}}{r_0}\right), \tag{11}$$

for $r > r_{co}$. The shape of C_{12} for $r < r_{co}$ plus C_3 for $r > r_{co}$ can be easily verified to be of a bell shape peaked at corotation.[5]

If we put in the appropriate time dependence in the above three profiles to account for the fact that these are really traveling waves, i.e., we change variable from r to $(r + c_g t)$ in C_1, and from r to $(r - c_g t)$ in C_2 and C_3, where c_g is the constant group velocity of the wave, we immediately obtain

$$C_1(r, t) = C_0 \exp\left[\frac{(r + c_g t) - r_{co}}{r_0}\right], \tag{12}$$

$$C_2(r, t) = -C_0 \exp\left[\frac{-(r - c_g t) - r_{co}}{r_0}\right], \tag{13}$$

and

$$C_3(r, t) = C_0\left[1 - \exp\left(-2\frac{r_{co}}{r_0}\right)\right] \exp\left[-\frac{(r - c_g t) - r_{co}}{r_0}\right], \tag{14}$$

which are temporally exponentially growing waves (mode) *at a given radius*, with the constant growth rate being

$$2\gamma_g \equiv \frac{c_g}{r_0}, \tag{15}$$

where the factor of 2 is there because again we have chosen to use γ_g to represent the amplitude growth rate, hence $2\gamma_g$ for the wave action growth rate.

Furthermore, since the amplification of the wave is restricted to the corotation region, the amplitude growth rate of a wave mode γ_g should be related to the over-reflection factor Γ at corotation, where $\Gamma = |\{[C_1(r_{co})]/[C_2(r_{co})]\}| = e^{2(r_{co}/r_0)}$, as well as the time it takes for the wave to travel from the corotation radius to the center of the galaxy τ through

$$\exp(2\gamma_g \times 2\tau) = \Gamma, \tag{16}$$

where

$$\tau = \frac{r_{co}}{c_g}. \tag{17}$$

It can be easily checked that with the relation of γ_g and r_0 as in equation (15), the total change of wave action C along the wave propagation path, after a round trip on the C_1 and C_2 branches, is indeed $\exp(2\gamma_g \times 2\tau) = e^{2(r_{co}/r_0)} = \Gamma$, so the entire feedback loop is closed.

The past work of LBK and Goldreich & Tremaine (1979) have both shown that the total torque coupling is equal to the product of wave angular momentum density and the group velocity of the wave. In the linear regime, this product should be conserved along the path of wave propagation. We now see that the requirement of conservation of wave action along the path of wave propagation in the linear regime does not contradict our conclusion of a bell-shaped $C(r)$ function for a *spontaneously growing* open spiral wave mode, since this conservation is in the comoving frame of each wave train and is *not* across the galactic radius r at a given time.

At the quasi steady state of the wave mode, the distribution of $C(r)$ is expected to still have the same bell-shaped general profile, as is confirmed in the N-body simulations, so are the distributions of each wave branch. The difference is that the deposited wave angular momentum is now no longer used for wave growth but rather is left onto the basic state. In the nonlinear regime, we do not simply have two propagating waves inside corotation, but rather propagating/decaying waves in the comoving frame of each

[5] In fact, the peak has a cusp shape at corotation. However, the N-body spirals (as well as the physical spiral galaxies!) always tend to smooth the cusp into a bell. The differing distributions between the two-dimensional and one-dimensional cases for the wave actions will also contribute to the fact that the actual observed action distribution in N-body simulations is a bell rather than a cusp.

wave. Note that without decay, the propagation of the C_1 wave from $r + \Delta r$ to r would lead to the time increase of wave action at r, but the dissipative channeling of wave action onto the basic state makes the wave at the quasi steady state have the same action at all times when it arrives at r.

Finally, we mention that in the above, although we have shown that the spontaneous and homogeneous modal growth can be achieved if the traveling wave components of the mode have exponential spatial distribution, and that such a distribution is self-sustaining once it is established, we have not addressed the question of how such exponentially distributed wave action initially got established. The answer to this question is apparent when we realize that a spiral mode is an instability of the underlying disk. The spontaneous growth tendency of such an instability, in a particle disk, implies that, of all the possible *initial positions and velocities* of the N particles in the axisymmetric disk, there is a far greater portion of the $6N$-dimensional initial-condition phase space which, when evolved with time, results in a spontaneously growing spiral mode (which can then be decomposed into two exponentially distributed waves inside corotation), than the portion of this phase space which does not result in such a spontaneous growth. So by randomly choosing the initial positions and velocities of the disk particles (while satisfying the large-scale equilibrium constraints) we are almost certain to land on an initial-condition phase space which leads to the spontaneous growth of spiral instability. The subsequent seemingly more ordered state of the disk with a spontaneously formed spiral mode is, in fact, of higher entropy than the state of the original smooth (but unstable) disk. We will address the issue of entropy evolution in nonequilibrium systems further in § 6.

The fact of the *homogeneous* growth we almost always observe in a spontaneous N-body spiral is likely to be due to the fact that for randomly distributed initial seed noise, there should *not* be a region of the disk which is favorably selected for an initial faster growth. The bumps and irregularities in the modal distribution could also be filtered out in time by the galactic resonant cavity between corotation and inner disk, once the two-dimensional nature of the problem and the phases of the constituent waves are taken into account.

4. THE ROLE OF GAS

In § 2, we have seen that an N-body quasi-stationary spiral mode can form without a separate dissipative gas component. However, one still wonders what role the interstellar medium, if present together with the stars, plays in a physical spiral galaxy. We show in the following that in such a case stars and gas play more or less parallel roles in limiting the wave amplitude growth. In particular, we argue that even the gas clouds contribute to the wave damping process mainly through participating in the collisionless shock that the stars are supporting.

We first present the result of a two-component (star plus gas) N-body simulation, with the gas component going through inelastic collisions when approaching each other at a short distance. The details of the gas collisional cooling treatment can be found in Thomasson (1989). We have chosen to use a gas cloud collisional cross section of half of the mesh length unit and a coefficient of restitution of 0.5. We use 25,000 particles to represent stars and another 25,000 particles to represent gas clouds; the two components together make up the active disk mass. The star and gas particles are chosen to have the same mass. The computation mesh as well as the basic state used are identical to that in § 2.

In Figure 8, we plot the stellar and gaseous morphologies at three different time steps. We observe that the gas clouds are condensed into much narrower spiral arms than stars and show filamentary structures. A "twin-peaks-and-bar" morphology near the inner Lindblad resonances, similar to that observed for the galaxy NGC 3627 (Zhang, Wright, & Alexander 1993) is seen to develop at time step 2400.

In Figure 9, we plot the calculated phase shifts of the stellar and gaseous spiral densities with respect to a common spiral potential, at time step 1600. We observe that the two phase shifts have similar profiles, although the amplitude of the phase shift for gas is somewhat larger than for stars. However, the relative phase shift between the stellar and gaseous densities is obviously much smaller than the phase shift of each component with respect to their common potential. Since at the quasi steady state, the potential and density phase shift of each component indicates the contribution of each component to the dissipative damping of the potential wave, the result of Figure 9 shows that stars and gas play parallel roles in limiting the wave amplitude growth, and the stellar wave is not damped solely at the expense of gas. A plot of surface density evolution of the two components (not shown here) reinforces this conclusion.

At this point, a curious reader may wonder how the idea of using gaseous waves to damp stellar waves ever took existence and persisted as a major wave damping mechanism for more than two decades. A brief account here of the evolution of ideas related to wave damping shows that this is nothing unexpected.

One of the earliest versions of the spiral density wave theory treats the tightly wrapped waves in stellar disks using the lowest order WKBJ approximation (Lin & Shu 1964, 1966). In this approximation, the potential and density spiral patterns are everywhere proportional to each other, so in this sense it is a local theory, and no self-consistent potential/density phase shift is allowed in this theory. Roberts (1969) calculated the nonlinear gaseous response under the *enforced* (non–self-consistent) WKBJ spiral potential and found that shocks usually form. Kalnajs (1972) correctly pointed out that such an induced gaseous shock is phase shifted from the forcing spiral potential as a result of the action of the shock dissipation and therefore should tend to damp the forcing spiral potential of the stars. Kalnajs' work was supplemented by that of Roberts & Shu (1972), and this mechanism has since become one of the major candidates for the damping of growing spiral wave modes.

However, what was not realized at the time was the fact that these WKBJ shock waves can *never* be made self-consistent within the framework of WKBJ analysis. The presence of the phase shift between the potential and density spirals of an open spiral shock solution contradicts the WKBJ form of the Poisson equation, which requires the potential and density to be exactly in phase. Unfortunately, this subtle point of non–self-consistency in the WKBJ shock solutions was not appreciated by the majority of the workers in the field, and the impression that a gaseous

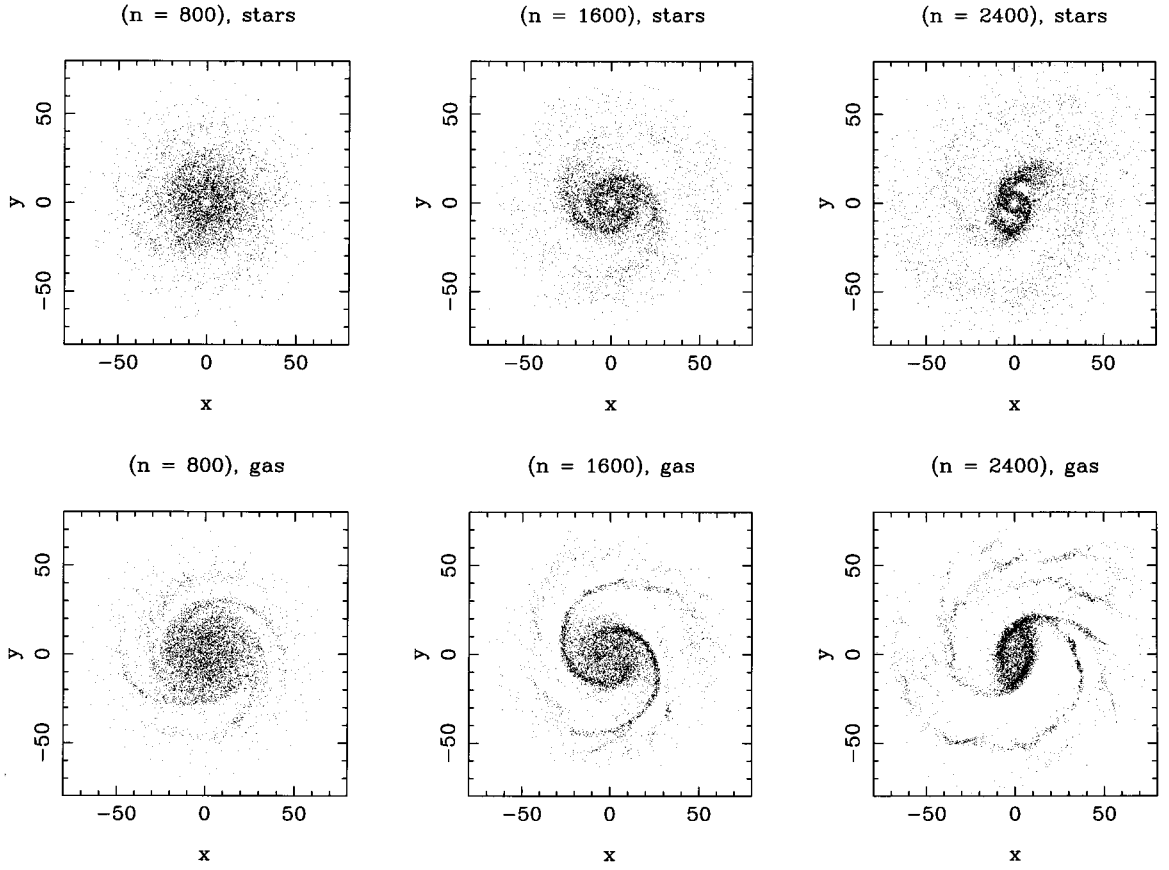

FIG. 8.—Star and gas morphologies at three different time steps. The simulation and the basic state parameters are described in § 4 of the text. About 5000 randomly chosen particles are plotted in each frame.

spiral wave serves to damp the stellar wave has become widespread, even though all the self-consistent two-component N-body simulations conducted in the interim had shown no evidence of a significant star-gas phase shift (see, e.g., Carlberg & Freedman 1985).

Our current work shows that only *open* spiral waves can self-consistently admit a spiral shock since only open waves can self-consistently admit a potential/density phase shift, which is an unavoidable element in a viscous spiral shock solution. Furthermore, the amount of shock dissipation at the quasi steady state of the wave is always given by the global phase shift closure relation whether a separate dissipative gas component is present or not. This characteristic of global-determines-local, or hierarchical, control is typical of the coherent structures formed in nonequilibrium systems (§ 6).

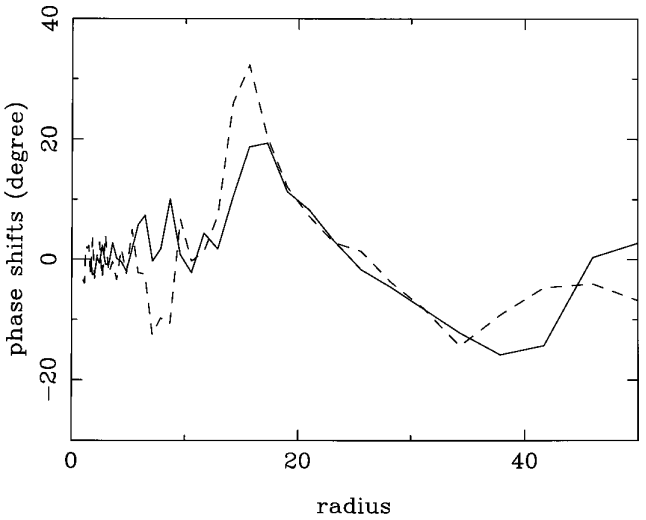

FIG. 9.—Potential and density phase shift for the stellar component (*solid*) and the gas component (*dashed*), respectively, for the spiral mode of § 4 at time step 1600.

In physical spiral galaxies, most of the mass of the interstellar medium is in the form of molecular clouds, especially at the spiral arm region. A good fraction of these clouds are self-gravitating, with the mean free path of cloud-cloud collisions on the order of a few hundred parsec. In Paper I, we showed that the size of star-gas two-fluid instability at the spiral arms, which is the effective particle mean free path in the spiral collisionless shock, is on the order of 1 kpc, which is not much larger than the gas-cloud mean free path in the absence of the shock. Therefore, it appears that a large fraction of the gas clouds' contribution to the dissipative damping of the wave is through participating in the spiral gravitational shock, and the scattering of these clouds in the collisionless shock is at least as important as the direct collisions of clouds due to their random velocities. In fact, a good fraction of the cloud-cloud collisions themselves happen as a result of the gas' response to the spiral-arm local gravitational instability.

Within the confines of the 1 kpc sized spiral gravitational instability, a series of secondary shocks is expected to develop in the gas component, as a result of the supersonic turbulent cascade. This may underlie the observed cloud-size/velocity dispersion relation for the molecular clouds of

our Galaxy (Larson 1981), with the source of the largest scale energy injection due to the fact that the pattern speed of the wave and the circular speed at a given galactic radius is generally different (§ 5). Radiative cooling is only initiated when the spiral shock energy input is cascaded down to hydrodynamic shock scales.

5. HOW LONG COULD THE SPIRAL PATTERNS IN PHYSICAL GALAXIES REMAIN QUASI-STATIONARY?

During the past three decades' work on the density wave theory of spiral structure, one of the central issues of interest (and of controversy) has been the longevity of spiral patterns in disk galaxies (see, e.g., Binney & Tremaine 1987, chap. 6). Although the statistical studies of grand design spiral galaxies in groups show evidence that spiral patterns are long lived (Elmegreen & Elmegreen 1983, 1989), the mechanism through which this longevity is achieved has not been firmly established. In recent years, it has gradually become apparent that the long-lived spiral patterns in disk galaxies are normal modes of the underlying disks (see, e.g., Bertin et al. 1989a, 1989b, and the references therein). However, there remains the question of how to offset the violent growth tendency of a large fraction of these spiral modes to obtain modes with amplitudes that are approximately steady.

We have shown in the last few sections that a quasi-stationary spiral mode can be sustained through the dynamical balance of its spontaneous growth tendency and a local dissipation process, as well as by a flux of matter, energy, and angular momentum through the system. Therefore, in principle, a spiral pattern can remain quasi-stationary as long as this flux of matter and energy can be maintained. However, we have also shown that the source and sink of this flux lie in the basic state of the galactic disk, so the price to pay for maintaining a quasi-stationary spiral structure is the continuous evolution of the basic state. The question of "how long the spiral patterns in disk galaxies can remain quasi-stationary" thus becomes how long it takes the basic state to change significantly so that it will no longer be able to support a spiral pattern.

In the following, we derive analytical expressions for the rates of the spiral-induced basic state evolution. We then estimate these rates for the N-body spiral mode of § 2 and compare them against the actually measured basic state evolution rates in the same N-body calculations. A confirmation of the evolution rates serves to validate both our analytical expressions and the N-body approach itself. Once such confirmation is made, we could use the N-body results to infer the lifetime of spiral patterns in physical disk galaxies. Further discussions of the astrophysical consequences of the spiral-induced basic state evolution is given in Zhang (1997, hereafter Paper III).

5.1. *Secular Changes Produced by a Quasi-stationary Spiral Structure*

From the discussion in § 3, we know that in the quasi-steady state, the rate of angular momentum exchange between an open spiral pattern and the basic state of the disk, per unit area, is given by

$$\overline{\frac{dL}{dt}}(r) = -\frac{1}{2\pi}\int_0^{2\pi} \Sigma_1(r,\phi)\frac{\partial \mathscr{V}_1(r,\phi)}{\partial\phi}\,d\phi\,, \tag{18}$$

which, for two sinusoidal waveforms is given by

$$\overline{\frac{dL}{dt}}(r) = \left(\frac{m}{2}\right)A_\Sigma A_v \sin(m\phi_0)\,, \tag{19}$$

where A_Σ and A_v are the amplitudes of the density and potential waves, respectively, and ϕ_0 is the phase shift between these two waveforms. We have assumed that ϕ_0 is greater than zero when the density spiral leads the potential spiral.

Furthermore, the fractional amplitude F_v of the potential wave is related to the potential wave amplitude A_v itself through

$$F_v = \frac{|k|A_v}{r\Omega^2}\,, \tag{20}$$

where k is the wave number. This relation can be derived as follows: assuming a perturbation potential wave of the form

$$\mathscr{V}_1(r,\phi) = A_v \cos\left(m\phi + m\frac{\ln r}{\tan i}\right), \tag{21}$$

where i is the pitch angle, and m is the number of arms of the spiral pattern. If the potential wave amplitude A_v does not change with radius rapidly, we have that

$$\frac{\partial \mathscr{V}_1}{\partial r} = -A_v\frac{1}{r}\frac{m}{\tan i}\sin\left(m\phi + m\frac{\ln r}{\tan i}\right). \tag{22}$$

Since the radial derivative for the axisymmetric component of the potential is

$$\frac{\partial \mathscr{V}_0}{\partial r} = \Omega^2 r\,, \tag{23}$$

it follows that

$$F_v \equiv \frac{|\partial\mathscr{V}_1/\partial r|_{\max}}{|\partial\mathscr{V}_0/\partial r|} = \frac{|kA_v|}{\Omega^2 r}\,, \tag{24}$$

where we have used $\tan i = m/|k|r$.

We therefore have

$$\begin{aligned}\overline{\frac{dL}{dt}}(r) &= \left(\frac{m}{2}\right)A_\Sigma A_v \sin(m\phi_0)\,,\\ &= \frac{1}{2}F^2 v_c^2 \tan i \sin(m\phi_0)\Sigma_0\,,\end{aligned} \tag{25}$$

where we have used $A_\Sigma = F_\Sigma\Sigma_0$, with Σ_0 being the surface density of the disk; $F^2 \equiv F_\Sigma F_v$; $v_c = \Omega r$; as well as $m/|k|r = \tan i$.

For a single average star of mass M, the rate of its angular momentum loss is therefore

$$\overline{\frac{dL^*}{dt}}(r) = \frac{1}{2}F^2 v_c^2 \tan i \sin(m\phi_0)M\,. \tag{26}$$

Since $\overline{dL^*/dt}$ is also equal to (for a nearly flat-rotation–curved galaxy)

$$\frac{dL^*}{dt}(r) = -v_c M\frac{dr_*}{dt}\,, \tag{27}$$

where r_* is the mean radius of the star under consideration, we have

$$\frac{dr_*}{dt} = -\frac{1}{2}F^2 v_c \tan i \sin(m\phi_0)\,. \tag{28}$$

We have thus obtained the rate of secular change of the mean radius of a star due to the spiral-induced wave/basic state interaction. Since the value of ϕ_0 is positive inside corotation and negative outside, the resulting orbital change leads to the change of disk surface density toward configurations of ever-increasing central concentration, together with an extended outer envelope.

Besides inducing the stellar mean radius change, the wave/basic state interaction also leads to a secular increase in the stellar epicycle radius. This is due to the difference in the ratio of stellar energy and angular momentum loss (gain), compared to the ratio of the wave energy and angular momentum gain (loss), at a location other than corotation. Specifically, we have that the rates of loss of orbital energy and angular momentum for basic state stars are related through

$$\frac{dE_{\text{basic state}}}{dt} = \Omega \frac{dL_{\text{basic state}}}{dt} , \tag{29}$$

and the rates of gain of energy and angular momentum by the wave are related through

$$\frac{dE_{\text{wave}}}{dt} = \Omega_p \frac{dL_{\text{wave}}}{dt} . \tag{30}$$

Since $dL_{\text{basic state}}/dt$ is equal in magnitude to dL_{wave}/dt, we have that the rate of random energy gain (per unit area) of the disk stars is related to the angular momentum exchange rate (per unit area) through

$$\frac{d\,\Delta E}{dt} \equiv \frac{d(E_{\text{basic state}} - E_{\text{wave}})}{dt} = (\Omega - \Omega_p) \frac{dL_{\text{wave}}}{dt} , \tag{31}$$

where L_{wave} is the angular momentum density of the wave. Note that this expression is true (and has a positive sign) both inside and outside corotation, since both $\Omega - \Omega_p$ and dL_{wave}/dt change sign across corotation.

Now dL_{wave}/dt is equal in magnitude to $\overline{dL/dt}$ of equation (25), we can thus write

$$\frac{d\,\Delta E}{dt} = \frac{1}{2} (\Omega - \Omega_p) F^2 v_c^2 \tan i \sin (m\phi_0) \Sigma_0 . \tag{32}$$

Equation (32) gives the rate of random energy increase of stars per unit time and per unit area. The rate of random energy increase for a single star of mass M is then given by

$$\frac{d\,\Delta E^*}{dt} = \frac{1}{2} (\Omega - \Omega_p) F^2 v_c^2 \tan i \sin (m\phi_0) M , \tag{33}$$

where M is the mass of the star.

Since we also have

$$\frac{d\Delta E^*}{dt} = \frac{d[(1/2)M\sigma^2]}{dt} , \tag{34}$$

where σ is the (three-dimensional) velocity dispersion of stars, we finally obtain

$$\frac{d\sigma^2}{dt} \equiv D^{(3d)} = (\Omega - \Omega_p) F^2 v_c^2 \tan i \sin (m\phi_0) , \tag{35}$$

where D^{3d} is the diffusion constant of the space velocity of a star.

We will show in Paper III that the rates of secular basic state changes we derived in this section could account for the formation of the Galactic bulge in a Hubble time and could also quantitatively explain the observed age-velocity dispersion relation of the solar neighborhood stars.

5.2. *Verification of the Rates of Secular Basic State Evolution Using N-Body Simulations*

5.2.1. *Faithfulness of the N-Body Simulation Approach in Modeling the Long-Term Evolution of Spiral Galaxies*

In this section we will use the results of N-body simulations to confirm the analytical results we obtained in the previous section. For this purpose, we first need to address the issue of how faithfully the N-body technique can model the secular evolution of the disks of spiral galaxies.

The N-body technique has been employed to model disk galaxies for several decades (see, e.g., Miller & Prendergast 1968; Hohl & Hockney 1969; Sellwood 1987, and the references therein). Since the number of particles used in such simulations is usually 10^6–10^7 times less than what is actually present in physical galaxies, much attention has been paid to the issue of whether there is exaggerated rate of relaxation due to the small number-of-particle effect. In a two-dimensional disk, it is found that even though the two-body relaxation timescale is comparable to the local dynamical timescale (Rybicki 1971), when properly chosen softened gravity and mesh-grid are used, the two-body relaxation becomes negligible over the timescale of interests, at least in cases where the disk is stable against large-scale gravitational instabilities (Hohl 1973; Romeo 1994).

However, as has been cautioned by various authors (e.g., Sellwood 1987; Weinberg 1993), the emergence of large-scale gravitational instabilities could significantly alter the picture. In the case of the past two-dimensional modeling of spiral galaxies, which generally employed 50,000–100,000 particles, significant heating associated with the emergence of spiral structure can always be observed, and the spiral patterns formed quickly decay as a result of this self-induced heating. Since the calculated two-body relaxation timescale in such cases is always longer than the time span of the simulation run, this heating effect in the N-body simulations has been taken by many as indicating that spiral patterns in physical galaxies cannot be long-lived since their self-induced heating always leads to their rapid self-destruction.

In the following, we will demonstrate that the rate of heating observed in the N-body spiral disks is much larger than in physical galaxies, and this rapid heating is produced partly by the particle scattering process at the spiral arms (Paper I). Since this scattering process is an integral part of the spontaneous formation of spiral pattern, the N-body simulations of unstable spiral modes (as opposed to tidal spirals in originally stable disks) can *never* be made truly collisionless. This seems to suggest that we are out of luck in trying to use a small number of particles to simulate the evolution of a physical spiral disk since in the presence of particle scatterings the number of particles used in the simulations *will* matter. We show in the following that the situation is in fact a lot more optimistic. It turns out that the N-body simulations using the presently employed number of particles can accurately model the angular momentum evolution (or the mass distribution evolution) of the basic state. It is also shown that, indeed, the small number of particles used exaggerates the heating of the disk. But through a proper re-scaling of the excess-heating rate, using

the dependence of the Poisson noise on the number of particles, we can also obtain an estimate of the actal heating rate in the physical disk galaxies.

Another caution often raised by the previous practitioners of the N-body technique is the large errors often found in the individual particle's trajectories. A calculated orbit generally diverges exponentially from the true orbit in the presence of small random error (Miller 1967). Furthermore, the kind of N-body system which admits a spontaneous global instability structure, such as a spiral pattern or a bar, is known to show an even stronger level of such dynamical instability, which, as we have stated before, is due to the collective dissipation process associated with these global structures. Collisions or close encounters are known to increase the dynamical instability of an N-body calculation (Miller 1967).

In the face of such dynamical instability, one wonders how and in what sense our N-body calculation could simulate a realistic physical system. We point out here that if our aim is to model the global behavior of an instability structure, the sensitivity of individual orbits to random errors need *not* worry us at all. This is because the very same collective dissipation process associated with such a global instability makes the system *effectively* a dissipative system, whether it is nominally a Hamiltonian system or not. As is well known, the large-scale structures formed in dissipative systems possess asymptotic stability, and are *insensitive* to the details of small random errors. This asymptotic stability is responsible for the faithfulness of the N-body simulations in modeling the macroscopic properties of large-scale instability structures in nonequilibrium systems.

5.2.2. *Three Test Runs Using Different Numbers of Particles: Mutual Verification of the Theoretical Prediction and the N-Body Simulation Approach*

In this section, we present the result of the N-body simulations of a spontaneously formed spiral mode. This simulation uses the same basic state as that in the simulation of § 2, but it employs 100,000, 200,000, and 300,000 particles, respectively, to represent the active disk component, for the three simulation runs. It is found that the global morphologies of the spiral mode formed are very similar in all three cases even though the spiral pattern in the smaller number-of-particle case emerges a little earlier and fades a little sooner. Fourier analysis shows that the pattern speeds and equilibrium amplitudes of the $m = 2$ spiral mode formed are also similar for the three cases.

When plotting out the evolution of the basic state parameters in the three cases, we found that the circular velocity v_c and the surface density Σ have very similar evolution behavior. The values of radial velocity dispersion v_r, however, show marked difference. In particular the stars in the smaller number-of-particle run are found to heat up significantly more than those in the larger number-of-particle run. Another observation is that the heating effect is the smallest near the corotation radius, for all the cases.

In order to quantify the above observations, we now calculate the expected radial mass accretion rates and heating rates at a particular radius inside corotation for the spiral patterns obtained in these simulations and compare them to that actually observed in each case.

We begun by choosing an arbitrary radius of $r \approx 13.5$ (which is inside corotation) to verify the radial mass accretion equation (28). We found that at this radius $v_c = 0.4$,

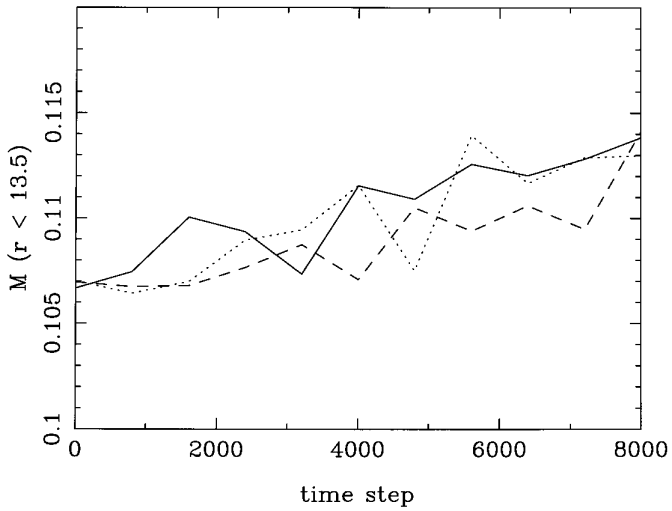

FIG. 10.—Evolution of the total mass within $r = 13.5$. *Solid:* $N = 100{,}000$ run; *dashed:* $N = 200{,}000$ run; *dotted:* $N = 300{,}000$ run.

$F = 0.1$,[6] $i = 17°$, and $\phi_0 = 5°$. Therefore, we obtain dr_*/dt (theoretical) $= -0.0001$ (per time step), or a $\Delta r = -0.3$ in 3000 time steps, from equation (28). This is obviously a rather small rate of radial drift, due to the small equilibrium amplitude for this particular spiral mode (the spiral mode calculated in Paper I, for example, has a much larger equilibrium amplitude, and thus much larger basic state evolution rate). This makes the direct comparison to the actually observed orbital decay rate in the simulations rather difficult. So instead, we calculate the expected rate of increase of total mass inside the radius $r = 13.5$, as a result of the stellar orbital decay,

$$\frac{dM}{dt}(r \le 13.5)_{\text{(theoretical)}} = \left(-\frac{dr_*}{dt}\right)_{\text{(theoretical)}} \times 2\pi r \Sigma(r)\,|_{r=13.5} . \quad (36)$$

Now the local surface density at $r = 13.5$ is about 1.5×10^{-4}. Therefore we obtain that the expected mass increase rate is $(dM/dt)(r \le 13.5)_{\text{(theoretical)}} \approx 1.2 \times 10^{-6}$.

We then integrate the actual surface density obtained in the N-body simulations from $r = 0$ to $r = 13.5$ and obtain the $M(t)_{\text{(measured)}}$ for the three runs with different numbers of particles, which we plot in Figure 10. Immediately we see that the long term trends of variation of the surface density in these three runs are very similar. From the slope of these curves, we obtain $dM/dt(r \le 13.5)_{\text{(measured)}} \approx 10^{-6}$ between time step 3000 and 6000, which is quite close to the theoretical prediction (eq. [36]). This gives us confidence that the theoretical radial orbital decay rate eq. [28] gives a good description of the actual mass accretion observed in the N-body simulations. The close agreement of the three curves in Figure 10 also shows that N-body simulations using small number of particles can quite accurately model the angular momentum evolution of the basic state induced by a galactic spiral structure.

Next, we look into the issue of disk heating in the N-body experiment. Again we consider the disk region near radius $r = 13.5$. We have that $\Omega \approx 0.029$, $\Omega_p \approx 0.01$, as well as

[6] The value of F is calculated by first calculating the fractional potential wave amplitude F_v and density wave amplitude F_Σ and demanding that $F^2 = F_v F_\Sigma$. In the current case, at $r = 13.5$, we found that $F_v = (kA_v/\Omega^2 r) = 0.04$, where $A_v \equiv \{[\int_0^{2\pi} (v - \bar{v})^2\, d\phi]/(2\pi)^{1/2}\}^{1/2}$; and $F_\Sigma = (A_\Sigma/\Sigma_0) = 0.25$, where $A_\Sigma \equiv \{[\int_0^{2\pi} (\Sigma - \bar{\Sigma})^2\, d\phi]^{1/2}/(2\pi)^{1/2}\}^{1/2}$.

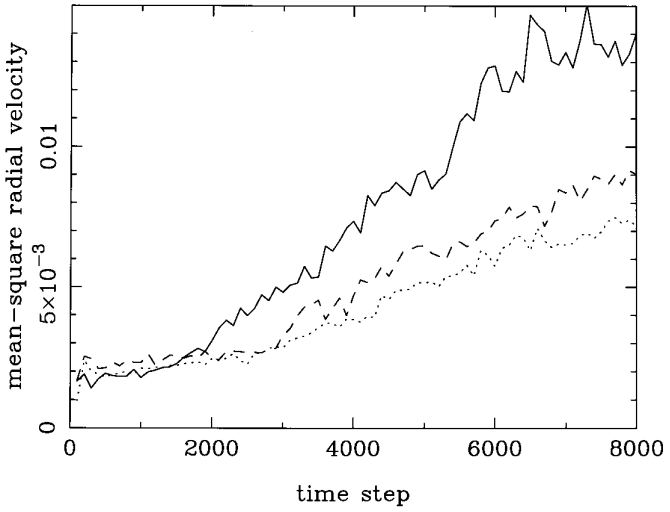

FIG. 11.—Evolution of mean-square radial velocity for $N = 100{,}000$ run (*solid*), $N = 200{,}000$ (*dashed*), and $N = 300{,}000$ runs. About 500 particles are used in deriving the mean-square radial velocity in the first case, 1000 particles in the second case, and 1500 particles in the third case, all have the same mean radius of $r = 13.5$.

$F \approx 0.1$, $i \approx 17°$, $\phi_0(r = 13.5) \approx 5°$, $v_c(r = 13.5) = 0.4$ as before. This gives us $D^{2d}(\text{theoretical}) \approx 1.6 \times 10^{-6}$ according to equation (35).[7] Therefore $D^{1d}_{(\text{theoretical})} = 0.8 \times 10^{-6}$.

In Figure 11, we plot the actual evolution of the mean-square radial velocity for a group of stars near $r = 13.5$.[8] The first thing we notice is that the three curves are no longer close to one another as was the case for the total mass evolution, but rather diverge from one another as the time step increases. Furthermore, as we increase the number of particles the trend of variation of the heating curves does not tend toward zero (as would be expected if the heating were entirely due to the particle Poisson noise) but rather toward a finite heating rate. We expect that this converged heating rate is equal to that given by the theoretical heating rate $D^{1d}_{(\text{theoretical})}$, which is the systematic heating due to the spiral pattern.

We now try to verify this speculation. During the time period between step 3000 and step 6000, the spiral pattern in all the runs are found to have similar amplitudes. The measured heating rates, however, give $D^{1d}(N = 100{,}000) = 2.3 \times 10^{-6}$, $D^{1d}(N = 200{,}000) = 1.5 \times 10^{-6}$, and $D^{1d}(N = 300{,}000) = 1.3 \times 10^{-6}$. Assuming that the theoretical value of $D^{1d}_{(\text{theoretical})} = 0.8 \times 10^{-6}$ gives the amount of heating which is produced by the spiral pattern itself, we arrived at that the amount of excess-heating due the particle Poisson noise is $\Delta D^{1d}(N = 100{,}000) = 1.5 \times 10^{-6}$, $\Delta D^{1d}(N = 200{,}000) = 0.7 \times 10^{-6}$, and $\Delta D^{1d}(N = 300{,}000) = 0.5 \times 10^{-6}$, respectively. We see right away that the amount of excess-heating in the three cases satisfies

$$\frac{\Delta D^{1d}(N_1 = 100{,}000)}{\Delta D^{1d}(N_2 = 200{,}000)} \approx 2 \approx \frac{N_2}{N_1}, \tag{37}$$

and

$$\frac{\Delta D^{1d}(N_2 = 200{,}000)}{\Delta D^{1d}(N_3 = 300{,}000)} \approx 1.5 \approx \frac{N_3}{N_2}, \tag{38}$$

which is just the expected behavior of the Poisson random noise![9]

Therefore, we arrive at the conclusion that besides incorporating the intrinsic heating due to the large-scale spiral pattern itself, the N-body simulation introduces excess heating due to the small number of particles used.

5.3. *Lifetime of a Spiral Pattern Inferred from N-Body Simulations*

For physical spiral galaxies, the number of stars present are on the order of 10^{11}, therefore the scaling relation we had in the last section would give $\Delta D^{1d}(N = 10^{11}) = 7 \times 10^{-13}$ in the normalized units of the simulation. Therefore, $\Delta D^{1d}(N = 10^{11})/D^{1d}_{(\text{theoretical})} = 8 \times 10^{-7}$, which shows that in physical galaxies, heating due to the random fluctuations of individual stars is negligible compared to that due to a quasi-stationary spiral structure.

Recall that the $m = 2$ mode in all of our N-body simulations has survived throughout the simulation runs. This indicates that the (nonlinear) spiral mode would be prominent if not because of the (exaggerated) large epicycle radius of the particles towards the end of the simulation runs. We can therefore estimate the lifetime of the nonlinear spiral patterns in physical galaxies from the lifetime of our N-body spiral patterns through a proper rescaling of the excess heating effect due to the small number of particles. For example, the spiral shock pattern (especially the part inside corotation) for the run using 200,000 particles was found to last about 10 revolutions (6000 time steps) with the exaggerated heating rate of $D^{1d} = 1.5 \times 10^6$. Therefore with the realistic heating rate $D^{1d}(\text{theoretical}) = 0.8 \times 10^6$, the spiral pattern in physical galaxies could last (but will gradually tighten up in pitch angle) for at least 20 revolution periods.[10] This is usually comparable to two-thirds of the Hubble time, which is also close to the timescale whereby the dominant disk galaxy type changes from Sc to S0 in groups and clusters. The secular evolution induced by the galactic spiral structure causes both the central concentration of stellar mass as well as the increase in velocity dispersion of stars. In a quasi-equilibrium state the two should satisfy the virial relation.

The result of our N-body simulations therefore indicates that the spiral patterns in physical galaxies could survive on the order of a Hubble time, with the exact lifetime of an individual spiral galaxy depending upon its total mass, its environment, as well as the detailed distribution of its basic state. In particular, at the advanced stage of a disk galaxy's secular evolution, the type of spiral modes present are usually tightly wound and have small equilibrium amplitude (Paper III), so the speed for further evolution is reduced, which means that the spiral pattern could survive

[7] Note that since we are here performing a two-dimensional simulation, the overall energy input from the spiral is distributed in two dimensions, rather than in three dimensions as is the case for physical galaxies.

[8] In the actual N-body calculation, since the particle numbers are assigned sequentially starting from the inner disk, the particles used for calculating the mean-square radial velocity are those with serial numbers between $N/4$ and $N/4 + N/200$.

[9] Note that since D is the rate of increase of mean-square velocity, it has a $1/N$ dependence on the number of particles instead of a $1/(N)^{1/2}$ dependence.

[10] The actual heating rate in physical galaxies of comparable basic state and spiral wave amplitude is, in fact, even smaller, when we realize that in a realistic disk the amount of random energy input due to the spiral pattern is distributed in all three spatial dimensions, instead of in the two dimensions as in our simulations. The net heating rate in physical galaxies can also be lowered through gas cooling.

even longer than if the pattern is a vigorous one throughout its lifetime.

We thus see that the result of our current work finally vindicates the quasi-stationary spiral structure hypothesis of Lindblad (1963) and Lin & Shu (1964, 1966) made more than three decades ago.

6. DISCUSSION

6.1. *The Formation of Quasi-stationary Spiral Modes as an Example of Nonequilibrium Phase Transitions*

It is well known that for an isolated system, the direction of entropy evolution is toward an increasing degree of macroscopic uniformity. For open systems at far-from-equilibrium conditions, however, it often happens that the usual thermodynamic branch of the solution becomes unstable, and new types of highly organized spatial-temporal structures emerge spontaneously. Owing to their similarity to equilibrium phase transitions, these kinds of spontaneous structure formations in nonequilibrium systems have been termed "nonequilibrium phase transitions," and the structures formed "dissipative structures" (Glansdorff & Prigogine 1977; Nicolis & Prigogine 1977) to emphasize the constructive role of dissipation in the maintenance of these nonequilibrium structures.

The mechanisms for the spontaneous formation of organized structures in nonequilibrium systems are the amplification of fluctuations and feedback. In a realistic physical system, there is always a small level of fluctuation around the thermodynamic branch of the solution (in the case of the disk galaxy, the thermodynamic branch corresponds to the axisymmetric basic state). However, below a certain critical value of the gradient of macroscopic variables (measured in our case chiefly by Toomre's Q parameter), these fluctuations are damped. On the other hand, above this critical value of the gradient, certain components of the fluctuations can be amplified and give rise to a macroscopic current. The macroscopic fluctuation is finally stabilized by the opposing effects of global amplification/feedback and local dissipation, as well as the exchange of energy and matter with the outside world.

The large-scale coherent structures in open and nonequilibrium systems are functional as well as architectural. One of the important functions of these "dissipative structures" is to greatly accelerate the speed of entropy evolution of these systems towards reaching thermodynamic equilibrium.[11] In essence, the local highly ordered structure (which has low entropy) maintains its constant entropy in the meta-stable state by continuously exporting the entropy it produces to its environment. As a result, the entropy of the structure plus the environment increases at a much faster rate than when the system was still on the thermodynamic branch of the solution.

The spiral structure of galaxies has many characteristics of a typical "dissipative structure." First, as we have shown, a quasi-stationary spiral mode is maintained by the opposing effect of the spontaneous growth tendency and local dissipation, with a continuous flux of energy, angular momentum, and entropy through the system carried by the trailing spiral wave itself. Second, it can be shown that the formation of spiral structures greatly accelerates the speed of entropy evolution of a disk galaxy compared to that of a uniform disk (Zhang 1992). Third, since a spiral mode is a global instability in the underlying basic state of the disk, the spontaneous emergence of the spiral pattern (which is obviously a global symmetry-breaking process) happens as long as the disk satisfies certain far-from-equilibrium constraints (i.e., the basic state characteristics must allow the linear growth rate of a spiral mode to be greater than zero). Last, the characteristics of the quasi-stationary spiral pattern formed are determined solely by the property of the basic state and not by the accidentals of the external perturbations. This last point was reinforced by the N-body simulations of tidal spiral patterns in slightly unstable disks, where it was found that after the initial transient state, the characteristics of the tidally induced patterns correlate strongly with the properties of the basic state, rather than with the nature of the encounter (Donner & Thomasson 1994). These characteristics of the spiral structure clearly identify it as an example of a "dissipative structure" defined by Glansdorff & Prigogine (1971), and the spontaneous formation and stabilization of a large-scale spiral mode as an example of a nonequilibrium phase transition.

6.2. *Dissipation-like and Attractor-like Behavior in Many Degrees-of-Freedom, Unbounded Hamiltonian Systems*

The spiral structures in disk galaxies belong to the class of "dissipative structures" in nonequilibrium systems. One interesting feature about the spiral structures in the kind of disks which contain only stars is that such systems do not contain "dissipation" in the usual sense of the word, i.e., energy loss. An N-body stellar disk is clearly a Hamiltonian system. However, it appears that Hamiltonian systems are not all alike. There is a type of Hamiltonian system that is (1) unbounded, i.e., has infinite spatial extent; and (2) has many degrees of freedom so that it can be treated as a thermodynamical system. When such a system is far from equilibrium, it can, for all practical purposes, resemble a dissipative system.

A spiral galaxy is only an extreme case of the kind of Hamiltonian systems which can display dissipation-like behavior. In milder forms, the collisionless relaxation and dynamical friction in many self-gravitating stellar systems are also examples of the dissipation-like behavior in these nominally Hamiltonian systems. It can be shown that the (coarse-grained) entropy is always increased by such dissipation-like processes. Therefore, such unbounded, self-gravitating systems possess an arrow of time.

The distinctive feature of spiral galaxies among all the self-gravitating Hamiltonian systems which display dissipation-like behavior is that the macroscopic evolution of spiral galaxies is greatly speeded up by the emergence of large-scale spiral structures. A spiral structure therefore resembles the "attractor" concept in truly dissipative systems: it behaves like a meta-stable strange attractor, in fact. On such an attractor, neighboring orbits diverge exponentially but remain bounded (at least for a finite period of time, in the case of a spiral galaxy). The global organization of the macroscopic pattern in such a system persists despite the microscopic chaotic behavior of each individual orbit.

[11] For the case of a self-gravitating system, a true equilibrium configuration does not exist due to the presence of gravity. A self-gravitating system always evolves towards configuration of increasing central concentration. The current state of the universe, where self-gravitating subsystems of matter are either located very far apart, or are still moving even further apart, appears to be the result of the big bang initial condition, which resulted in an extremely low-entropy initial configuration so that further evolution on all scales could happen at the expense of entropy increase.

7. CONCLUSIONS

In this paper we have demonstrated that the spiral patterns in disk galaxies are capable of remaining quasi-stationary on the order of a Hubble time. The quasi steady state is a dynamical equilibrium state achieved through the balance of the spontaneous growth tendency of an unstable spiral mode and the dissipative channeling of the wave energy and angular momentum onto the basic state of the galactic disk. The spiral pattern is maintained in the quasi steady state at the expense of a continuous basic state evolution, as well as a continuous increase of the entropy of the whole system. Stars and gas are found to play parallel roles in limiting the wave amplitude growth. It is further demonstrated that the spontaneous formation and stabilization of the large-scale spiral instability is an example of the general phenomenon of nonequilibrium phase transitions.

The author was indebted to Steven Balbus, Francoise Combes, Bruce Elmegreen, Paul Jaminet, Agris Kalnajs, C. C. Lin, Colin Masson, Angel Plastino, Alessandro Romeo, Frank Shu, Magnus Thomasson, Alar Toomre, Scott Tremaine, and Hongren Zhang, as well as an anonymous referee for helpful discussion. This research was conducted while the author was working with the U.C. Berkeley radio astronomy lab and with the Smithsonian Astrophysical Observatory Submillimeter Array project.

APPENDIX

TORQUE COUPLING AND ANGULAR MOMENTUM TRANSPORT IN A SPIRAL DISK

The total torque coupling produced by a self-sustained galactic spiral density wave, which gives rise to the rate of angular momentum transport across a particular galactic radius, generally consists of two parts (see, e.g., LBK; Goldreich & Tremaine 1979). The first part is the gravitational torque coupling C_g owing to the mass distribution of the spiral, with its rate of the outward angular momentum transport across galactic radius r given by

$$C_g(r) = \frac{1}{4\pi G} r \int_{-\infty}^{\infty} \int_0^{2\pi} \frac{\partial \mathscr{V}}{\partial \phi} \frac{\partial \mathscr{V}}{\partial r} d\phi \, dz \, , \tag{39}$$

where $\mathscr{V}$ is the gravitational potential and G is the constant of gravitation.

The second part is the advective torque coupling, or coupling through Reynolds stress, owing to the large-scale velocity distribution produced by the spiral structure. In the Eulerian fluid formulation, the rate of advective angular momentum transport across the galactic radius r is given by

$$C_a(r) = r^2 \int_0^{2\pi} \Sigma V_r V_\phi \, d\phi \, , \tag{40}$$

where Σ is the surface density, and where V_r and V_ϕ are the radial and azimuthal velocity perturbation relative to the circular velocity, respectively.

A1. THE RELATION BETWEEN THE TORQUE COUPLINGS AND THE RATE OF ANGULAR MOMENTUM CHANGE

That the rate of change of angular momentum within a galactic radius r is given by the sum of $C_g(r)$ and $C_a(r)$ was assumed obvious in LBK and in some of the following works. Here we present a proof of this result, which was kindly provided by the referee of the current paper.

The rate of decrease of total angular momentum $\boldsymbol{J}$ contained within galactic radius r is given by

$$-\frac{d\boldsymbol{J}}{dt}(r) = -\frac{d}{dt} \int_{|\boldsymbol{r}'| \le |\boldsymbol{r}|} \int_v (\boldsymbol{r}' \times \boldsymbol{v}) f(\boldsymbol{r}', \boldsymbol{v}, t) d^3v \, d^3r' = -\int_{|\boldsymbol{r}'| \le |\boldsymbol{r}|} \int_v (\boldsymbol{r}' \times \boldsymbol{v}) \frac{\partial f}{\partial t} d^3v \, d^3r' \, .$$

By the continuity equation,

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \boldsymbol{r}} (f\boldsymbol{v}) - \frac{\partial}{\partial \boldsymbol{v}} \left(f \frac{\partial \mathscr{V}}{\partial \boldsymbol{r}} \right) = 0 \, , \tag{42}$$

hence we have

$$-\frac{d\boldsymbol{J}}{dt}(r) = \int_{|\boldsymbol{r}'| \le |\boldsymbol{r}|} \int_v (\boldsymbol{r}' \times \boldsymbol{v}) \left[\frac{\partial}{\partial \boldsymbol{r}'} (f\boldsymbol{v}) - \frac{\partial}{\partial \boldsymbol{v}} \left(f \frac{\partial \mathscr{V}}{\partial \boldsymbol{r}'} \right) \right] d^3v \, d^3r' \, . \tag{43}$$

The final term in the integrand is

$$-\frac{\partial}{\partial \boldsymbol{v}} \left[(\boldsymbol{r}' \times \boldsymbol{v}) f \frac{\partial \mathscr{V}}{\partial \boldsymbol{r}'} \right] + \boldsymbol{r}' \times \frac{\partial \mathscr{V}}{\partial \boldsymbol{r}'} f \, , \tag{44}$$

so integrating in v and remembering that as $|\boldsymbol{v}| \to \infty, f \to 0$ and that $\int f d^3v = \rho$, we have

$$-\frac{d\boldsymbol{J}}{dt}(r) = \int_{|\boldsymbol{r}'| \le |\boldsymbol{r}|} \int_v (\boldsymbol{r}' \times \boldsymbol{v}) \frac{\partial}{\partial \boldsymbol{r}'} (f\boldsymbol{v}) d^3v \, d^3r' + \int_{|\boldsymbol{r}'| \le |\boldsymbol{r}|} \boldsymbol{r}' \times \frac{\partial \mathscr{V}}{\partial \boldsymbol{r}'} \rho \, d^3r' \, , \tag{45}$$

with the final term reduces to the gravitational torque coupling $C_g(r)$ in (39) under the disk geometry; and the integrand of the first term can be written as

$$\frac{\partial}{\partial \boldsymbol{r}} [f\boldsymbol{v}(\boldsymbol{r}' \times \boldsymbol{v})] - f\left(\boldsymbol{v} \cdot \frac{\partial}{\partial \boldsymbol{r}'}\right)(\boldsymbol{r}' \times \boldsymbol{v}) ; \tag{46}$$

here the second term is equal to zero because $\boldsymbol{v} \cdot I \times \boldsymbol{v} = 0$, where I is the identity matrix.

The first term when integrated gives

$$\int_{|\boldsymbol{r}'|=|\boldsymbol{r}|} \boldsymbol{r}' \times \int_v f\boldsymbol{v}\boldsymbol{v}\, d^3v\, \boldsymbol{ds} = \int_{|\boldsymbol{r}'|=|\boldsymbol{r}|} \boldsymbol{r}' \times S^* \cdot \boldsymbol{ds} , \tag{47}$$

where $\int_v f\boldsymbol{v}\boldsymbol{v}\, d^3v$ is the total stress tensor S^* of LBK.

Under the disk geometry, the integral of (47) further reduces to

$$\int_\phi \int_r \int_v f v_r v_\phi\, d^3v\, d\phi\, dz . \tag{48}$$

In the fluid limit, we can define the averaged velocity dispersions V_r and V_ϕ such that

$$V_r V_\phi \equiv \frac{\int_v f v_r v_\phi\, d^3v}{\int_v f d^3v} = \frac{\int_v f v_r v_\phi\, d^3v}{\rho} . \tag{49}$$

Therefore the first term in $-dJ/dt$ becomes (note that in disk geometry $\boldsymbol{J} = J\hat{z}$)

$$\int_z \int_\phi r^2 V_r V_\phi \rho\, d\phi\, dz , \tag{50}$$

which reduces to the expression of $C_a(r)$ in (40) after the z integration.

Thus we have proved

$$-\frac{dJ}{dt}(r) = C_a(r) + C_g(r) , \tag{51}$$

i.e., the decrease of total angular momentum within a galactic radius is equal to the sum of gravitational and advective torque couplings through the boundary at the same radius. Taking the derivative of the above equation, we then obtained the results used in the main text,

$$-2\pi r \frac{\overline{dL}}{dt}(r) = \frac{d[C_a(r) + C_g(r)]}{dr} , \tag{52}$$

where $\bar{L}$ is the azimuth-averaged angular momentum density at a particular radius.

In § 3 of the main text, we found that $C(r)$ is a bell-shaped curve in the N-body spiral mode, during *both* the linear modal growth stage and at the quasi-steady state of the spiral structure. Since $(\overline{dL}/dt)(r)$ is thus nonzero, equation (52) shows that we can never have a completely steady state. However, nonlinear dissipative structures have mean states that secularly change as a result of dissipation. Thus equation (52) is to be interpreted as the rate of deposition of angular momentum into the mean background state at the quasi steady state of the wave mode.

A2. THE RELATION BETWEEN THE TORQUE INTEGRAL $\bar{\mathscr{T}}$ AND THE TORQUE COUPLING INTEGRALS C_g AND C_a

The torque integral/phase shift relation we included in Paper I,

$$\bar{\mathscr{T}}(r) = -\frac{1}{2\pi} \int \Sigma_1 \frac{\partial \mathscr{V}_1}{\partial \phi}\, d\phi \equiv -\frac{1}{2\pi r} T_1(r) , \tag{53}$$

describes the averaged torque density of the spiral potential field on the disk material for an annular ring of unit width at disk radius r, therefore it also indicates the rate of angular momentum exchange between the density wave and disk matter. In the following, we first show that in the linear regime $T_1(r)$ is related to the gravitational torque coupling integral C_g as $dC_g/dr = T_1(r)$ (S. Tremaine 1995, private communication), as follows.

Differentiating the expression of C_g in (39) with respect to r, we have

$$\frac{dC_g}{dr} = \frac{1}{4\pi G} \int dz \int d\phi \left(\frac{\partial \mathscr{V}}{\partial \phi} \frac{\partial}{\partial r} r \frac{\partial \mathscr{V}}{\partial r} + r \frac{\partial \mathscr{V}}{\partial r} \frac{\partial^2 \mathscr{V}}{\partial \phi\, \partial r} \right) . \tag{54}$$

But the second term, which can be written as $(r/2)(\partial/\partial\phi)(\partial \mathscr{V}/\partial r)^2$, integrates to zero over ϕ.

Now the differential form of the Poisson equation is

$$\frac{1}{r} \frac{\partial}{\partial r} r \frac{\partial \mathscr{V}}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \mathscr{V}}{\partial \phi^2} + \frac{\partial^2 \mathscr{V}}{\partial z^2} = 4\pi G\rho , \tag{55}$$

where ρ is the volume density of disk matter. Therefore,

$$\frac{dC_g}{dr} = \frac{1}{4\pi G} \int dz \int d\phi \, \frac{\partial \mathscr{V}}{\partial \phi} \left(4\pi G\rho r - \frac{1}{r} \frac{\partial^2 \mathscr{V}}{\partial \phi^2} - r \frac{\partial^2 \mathscr{V}}{\partial z^2} \right). \tag{56}$$

The second term gives

$$\int dz \int d\phi \, \frac{\partial \mathscr{V}}{\partial \phi} \frac{\partial^2 \mathscr{V}}{\partial \phi^2} = \frac{1}{2} \int dz \int d\phi \, \frac{\partial}{\partial \phi} \left(\frac{\partial \mathscr{V}}{\partial \phi} \right)^2 = 0 \, . \tag{57}$$

The third term gives

$$\int dz \int d\phi \, \frac{\partial \mathscr{V}}{\partial \phi} \frac{\partial^2 \mathscr{V}}{\partial z^2} = - \int dz \int d\phi \, \frac{\partial \mathscr{V}}{\partial z} \frac{\partial^2 \mathscr{V}}{\partial \phi \, \partial z} = - \frac{1}{2} \int dz \int d\phi \, \frac{\partial}{\partial \phi} \left(\frac{\partial \mathscr{V}}{\partial z} \right)^2 = 0 \, . \tag{58}$$

Therefore,

$$\frac{dC_g}{dr} = r \int dz \int d\phi \rho \, \frac{\partial \mathscr{V}}{\partial \phi} = r \int d\phi \Sigma \, \frac{\partial \mathscr{V}}{\partial \phi} = r \int d\phi \Sigma_1 \, \frac{\partial \mathscr{V}_1}{\partial \phi} = -2\pi r \bar{\mathscr{T}}(r) = T_1(r) \tag{59}$$

in the linear regime. Q.E.D.

The relation we demonstrated above, i.e., $dC_g/dr = T_1(r)$, is not expected to hold away from the linear regime, after the spiral shock forms. This is because the local gravitational instability condition at the spiral arms, which is associated with the formation of spiral shock, makes the differential form of the Poisson equation (55) no longer valid between the gravitational potential at an arbitrary position, and the averaged volume density of the disk at the corresponding location. This can be seen as follows. If the differential form of the Poisson equation actually is valid inside a spiral instability, with the density ρ obtained through averaging the local grainy particle distribution, we can then integrate this differential form around an annular ring and obtain a corresponding spiral potential distribution which, in the case of steady wave amplitude, leads to no secular orbital change, a result obviously contradicting what we have found out in the N-body simulations of both Paper I and in the current paper. Another way of saying that the differential form of the Poisson equation fails inside a collective spiral instability is that, *using the continuum formulation we can never uncover all the physics, including the effect of secular matter migration in a spiral disk*. In a Eulerian formulation, we would instead observe the steepening of the spiral shock until the singularity forms at the shock front, which invalidates the Eulerian formulation and which further requires the introduction of the viscous effect in the spiral shock region. Such viscous effect, however, cannot be properly modeled by introducing a uniform viscosity parameter into the Eulerian equation set to make it into a Navier-Stokes equation set since the operation of the spiral shock viscosity is not homogeneous (it is present only at the spiral arms). So the particle disk formulation is the only one which incorporates naturally all the physics in a realistic spiral disk, some of which obviously cannot be properly described by the differential form of the Poisson equation.

The situation is essentially the same for a fluid disk as well, since once the shock forms, the effective viscosity it requires needs also to be generated by the local inhomogeneities within the spiral arm local gravitational instability, through a constrained scattering process.

At the quasi steady state of the wave mode, the result of the N-body simulations indicates that the torque integral $T_1(r)$ is close in both magnitude and shape to the summation of $dC_a/dr(r)$ and $dC_g/dr(r)$. The rigorous demonstration of the relation between these different torque integrals will be given in a later paper.

REFERENCES

Bertin, G., Lin, C. C., Lowe, S. A., & Thurstans, R. P. 1989a, ApJ, 338, 78
———. 1989b, ApJ, 338, 104
Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
Carlberg, R. G., & Freedman, W. L. 1985, ApJ, 298, 486
Donner, K. J., & Thomasson, M. 1994, A&A, 290, 785
Elmegreen, B. G., & Elmegreen, D. M. 1983, ApJ, 267, 31
———. 1989, in Evolutionary Phenomena in Galaxies, ed. J. E. Beckman & B. E. J. Pagel (Cambridge: Cambridge Univ. Press), 83
Glansdorff, P., & Prigogine, I. 1971, Thermodynamic Theory of Structure, Stability and Fluctuations (New York: Wiley)
Goldreich, P., & Lynden-Bell, D. 1965, MNRAS, 130, 125
Goldreich, P., & Tremaine, S. 1979, ApJ, 233, 857
Hohl, F. 1973, ApJ, 184, 353
Hohl, F., & Hockney, R. W. 1969, J. Comput. Phys., 4, 306
Julian, W. H., & Toomre, A. 1966, ApJ, 146, 810
Kalnajs, A. J. 1972, ApJ, 11, L41
———. 1973, Proc. Astron. Soc. Australia, 2(4), 174
Larson, R. B. 1981, MNRAS, 194, 809
Lin, C. C., & Lau, Y. Y. 1979, Stud. Appl. Math., 60, 97
Lin, C. C., & Shu, F. H. 1964, ApJ, 140, 646
———. 1966, Proc. Natl. Acad. Sci., 55, 229
Lindblad, B. 1963, Stockholm Obs. Ann., 22, No. 5
Lubow, S. H., Balbus, S. A., & Cowie, L. L. 1986, ApJ, 309, 496
Lynden-Bell, D., & Kalnajs, A. J. 1972, MNRAS, 157, 1 (LBK)
Mark, J. W.-K. 1976, ApJ, 206, 418
Miller, R. H. 1967, J. Comput. Phys., 2, 1
Miller, R., & Prendergast, K. H. 1968, ApJ, 151, 699
Nicolis, G., & Prigogine, I. 1977, Self-Organization in Nonequilibrium Systems (New York: Wiley)
Norman, C. A. 1978, MNRAS, 182, 457
Rybicki, G. B. 1971, Ap&SS, 14, 15
Roberts, W. W. 1969, ApJ, 158, 123
Roberts, W. W., & Shu, F. H. 1972, Astrophys. Lett., 12, 49
Romeo, A. 1994, A&A, 286, 799
Sellwood, J. A. 1987, ARA&A Suppl., 25, 151
Thomasson, M. 1989, Research Rep. No. 162, Department of Radio and Space Science with Onsala Space Observatory, Chalmers Univ. of Techn. Goteborg
Toomre, A. 1964, ApJ, 139, 1217
———. 1969, ApJ, 158, 899
———. 1981, in Structure and Dynamics of Normal Galaxies, ed. S. M. Fall & D. Lynden-Bell (Cambridge: Cambridge Univ. Press), 111
Weinberg, M. D. 1993, ApJ, 410, 543
Yueh, T. Y. 1981, SIAM, 65, 1
Zhang, X. 1992, Ph.D. thesis, Univ. California, Berkeley
———. 1996, ApJ, 457, 125 (Paper I)
———. 1997, ApJ, submitted (Paper III)
Zhang, X., Wright, M., & Alexander, P. 1993, ApJ, 418, 100