BARRED GALAXIES: INTRINSIC OR EXTRINSIC?

Masafumi Noguchi

Astronomical Institute, Tohoku University, Aoba-ku, Sendai 980-77, Japan Received 1995 August 28; accepted 1996 April 16

ABSTRACT

A unified picture is presented of the formation of bar structures in disk galaxies of various morphological types. In order to discuss bar formation in the context of galactic disk formation, a simple analytic model is constructed of the growth of galactic disks by infall of primordial gas from halos and subsequent star formation in the disks. It is monitored during the course of disk growth whether or not the condition for spontaneous bar formation (i.e., bar instability) is fulfilled for the stellar disk component.

It is found that the infall timescale is a key parameter that controls the dynamical property of the resulting stellar disk. Disks that grow fast by rapid infall experience gas-rich phases, in which massive gas clumps arising from gravitational instability in the gas disk heat the stellar disk component dynamically. When the disk has fully grown and becomes mostly stellar, it has already acquired enough random motions to suppress bar instability. On the other hand, when the gas infall from the halo region proceeds slowly, star formation (though less intense than in rapid infall cases) keeps gas mass in the disk low, leading to a dynamically cold stellar component due to lack of strong heating by massive gas clumps. Therefore, the stellar disk becomes unstable and forms a bar once its mass fraction relative to the total galaxy mass reaches a critical value.

Based on this result, we propose that late-type barred galaxies, the disks of which are considered to have formed by slow accretion of the halo gas, have intrinsic origin, whereas the bars in early-type galaxies, whose disks are likely to have grown quickly, have been formed in tidal interactions with other galaxies.

Numerical simulations have been carried out which show that the bars created by tidal perturbations tend to have a relatively flat density profile along the bar major axis with "shoulders" (abrupt steepening of the gradient) at the bar ends, whereas spontaneous bars have a steeper profile. The formation scenario described above, combined with this numerical result, can explain the observed dichotomy that early-type galaxies generally have a flat bar, while late-type galaxies have a bar of exponential type.

Subject headings: galaxies: evolution — galaxies: formation — galaxies: interactions — galaxies: kinematics and dynamics — galaxies: ISM — galaxies: structure

1. INTRODUCTION

The importance of bar structures in the evolution of disk galaxies stems from two major characteristics that bars have: their ubiquity in disk galaxies with a wide range of morphological types and luminosities and their strong dynamical effects arising from nonaxisymmetric gravitational potentials. Barred galaxies occupy roughly one-third of all the disk galaxies. If the intermediate type (designated as SAB) is included, this fraction rises up to about twothirds (e.g., de Vaucouleurs 1963). Once a bar is formed, its strong gravitational torque induces redistribution of mass and angular momentum in the galactic disk (e.g., Hohl 1971). The influence on the interstellar matter is especially large owing to its high responsiveness (e.g., Sorensen, Matsuda, & Fujimoto 1976; Sanders & Tubbs 1980). Strong shocks can occur along the bar and intense star formation will be triggered. The gas infall to nuclear regions driven by bars will fuel the starbursts and active galactic nuclei (e.g., Simkin, Su, & Schwarz 1980; Hawarden et al. 1986). The redistribution of matter will also affect chemical abundance gradients across the disks (e.g., Edmunds & Roy 1993; Martin & Roy 1994). It is possible that bars, once formed, can eventually fade away after inducing these phenomena (e.g., Hasan & Norman 1990). If this is the case, it implies a possibility that more galaxies than observed today as barred galaxies have experienced bar phases in the past, raising the importance of bars even higher. Therefore, a fundamental modification may be required to the types of theoretical models of disk galaxy evolution that assume axisymmetric configurations throughout (e.g., Saio & Yoshii 1990).

Two different possible formation mechanisms of galactic bars have been proposed and studied in the past. Many numerical studies employing N-body simulations have clarified that massive stellar disks with small velocity dispersions develop a strong bar structure spontaneously in a few disk rotation periods (e.g., Hohl 1971; Ostriker & Peebles 1973; Athanassoula & Sellwood 1986). Though the detailed physical mechanism of the bar growth is not yet fully understood theoretically (but see Lynden-Bell 1979), the numerical bars obtained in this way sometimes bear a remarkable resemblance to actual bars in their shape and kinematics (e.g., Sparke & Sellwood 1987). On the other hand, external perturbations such as tidal interactions are found to induce more or less persistent bar structures in a galactic disk which is stable in isolated state (Noguchi 1987). A higher incidence of barred galaxies in paired galaxies than in isolated galaxies (Noguchi 1987; Elmegreen, Elmegreen, & Bellin 1990), and also a preponderance of bars in dense environments such as the central region of the Coma Cluster (Thompson 1981), suggest that this external triggering is actually responsible for the formation of not all but some portion of barred galaxies. To avoid undesirable confusion, we use the term "bar instability" (or "bar-

unstable") in the present paper to imply only the spontaneous bar formation.

Despite these efforts, several fundamental questions remain still unanswered: When and how did actual bars form, and how did they evolve with time? In order to answer such questions with confidence, any theory of bar formation (especially of spontaneous bar formation) should be combined with a theory of galactic disk formation and evolution. Most previous studies on the process of spontaneous bar formation suffer from a serious defect that the numerical models employed in them lack evolutionary continuity to the galaxy (especially disk) formation process. The conventional initial condition adopted widely in bar instability simulations is a completely formed massive disk made solely of stars, which is highly artificial. Although the detailed process of disk formation has not yet been clarified thoroughly (especially in relation to the origin of the Hubble morphological sequence), several lines of argument (e.g., Larson 1976; Lacey & Fall 1985; Sommer-Larsen & Yoshii 1990), based mostly on chemical abundance arguments, suggest gradual disk formation with a timescale of the order of several Gyr. This timescale is not negligible compared with the age of the universe of ~ 12 Gyr and, more importantly, it is considerably larger than the typical growth timescale of unstable bars. Moreover, a significant portion of the disk mass can remain gaseous for a long period, thus invalidating the adoption of a purely stellar disk as initial conditions (This limitation is being relieved in several recent studies, e.g., Friedli & Benz 1993.) In theoretically modeling galactic disk dynamics, the interstellar gas is often treated as a tracer of the underlying gravitational potential presumed to be determined by the stellar component. This is not always justified, however. The possible importance of the interstellar gas of even a small mass fraction has been demonstrated dramatically by a recent theoretical study by Shlosman & Noguchi (1993), which shows that highly dissipative gas effectively stabilizes the stellar disk against bar formation by creating a number of discrete clumps of large individual masses that scatter disk stars. It is naturally conceived that a galactic disk in its early evolution phase contained a much larger fractional mass in the gas component than the disks of present-epoch galaxies. These considerations suggest that a correct account of the gravitational effects from the interstellar matter is crucial in understanding especially the early rapid evolution stages of the galactic disks.

The purpose of the present study is to investigate the formation process of bars in the context of formation and evolution of galactic disks, taking into account the dynamically active nature of the interstellar gas discovered by Shlosman & Noguchi (1993). A simple analytical model of the disk evolution is devised in which bar formation is tied closely with growth of the disk. Emphasis is placed on the disk heating process and the emergence of (stellar) bars. Though a crude treatment, this analysis suggests that the timescale of gas infall to the disk plane (i.e., the timescale of disk formation) is one of the important parameters that govern the fate of the forming disk. Based on this result, it is proposed that the stellar bars in late-type disk galaxies have been formed by bar instability in their disk components, whereas those in early-type disk galaxies resulted from tidal interactions with other galaxies in a manner described by Noguchi (1987). This conclusion is reinforced by supplementary numerical simulations, which reveal a striking difference in the morphological structure of spontaneously formed bars and tidally created ones. Although our primary aim was to understand bar formation processes, the present model seems to provide further implications about disk evolution in general.

In § 2, the analytic models we constructed are described. Results are discussed in § 3. Implications and confrontations with available observational data are given in § 4, and § 5 summarizes conclusions.

2. MODELS

In the present study, we follow the conventional picture of disk galaxy formation, in which the galactic disks form gradually by gas infall in already virialized halos. In this picture, a disk galaxy starts as a purely gaseous cloud embedded in a dark halo component of a similar spatial extent. The halo component is assumed to be already in dynamical equilibrium, while the gas can collapse due to cooling. Initially the gas cloud collapses more or less spherically due to lack of sufficient rotational support arising from ineffectiveness in angular momentum generation in protogalaxies by means of tidal torque from neighbors. This nearly spherical collapse is halted when the ever-increasing centrifugal force finally balances the total gravitational force due to the gas and halo. After this, the gas cloud collapses mainly along its spin axis, giving rise to gas infall perpendicular to the galactic plane. This second stage of collapse seems to be required in any successful theory of disk galaxy formation, regardless of the specific cosmogony on which it is based (e.g., top-down scenarios or bottom-up ones) because the remarkable thinness and rotationally supported nature of the observed galactic disks strongly suggest gentle dissipative formation without large radial mixing of matter. The disk growth due to this second stage of the primordial gas accretion is just what we treat in the present study. We first describe the general features of our models (§§ 2.1 and 2.2) and then discuss the choice of parameter values (§ 2.3).

2.1. Disk Growth Models

We consider a spherical halo with a radius R. The primordial gas initially distributed in this halo region accretes to the disk plane gradually and builds up a galactic disk. As the accretion proceeds, the disk becomes to contain more and more gas (i.e., the interstellar gas), and the stars form from it. We assume that the stars and gas in the disk occupy a flat cylindrical region with a radius R (the same as the halo) and a vertical height of h_s and h_a , respectively (see Fig. 1). The region displayed in Figure 1 should be considered to represent the portion of a disk galaxy within its optical radius. Therefore, the halo in the present paper should be regarded as not representing the entire dark halo (the massive halo), but rather its portion within the optical extent of the galaxy plus any visible spheroidal components. We are not concerned with any structures that might develop in this volume but adopt a "multizone" treatment. For example, we do not consider spiral structures that are likely to form in actual disk evolution. The physical state of each component (i.e., the halo, the stellar disk, and the gas disk) at a given time is specified by several global quantities whose characteristic values are given averaged over the entire region of that component. R, h_s , and h_q are assumed not to depend on time. The total mass for the combined

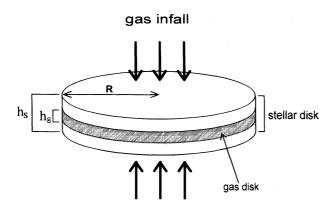


Fig. 1.—Schematic representation of the growing disk models. A gas disk with a thickness of h_g is embedded in a stellar disk with a thickness of h_s . Continuous infall (arrows) of the primordial gas from the halo region increases the total mass of the disk.

system of the stellar and gaseous disks and the halo is assumed to be conserved and denoted by m_t .

Under this simplification, time evolution of masses in the gaseous and stellar disk components is formulated as follows:

$$\frac{dm_g}{dt} = -\frac{m_g}{\alpha} + \frac{m_d}{\beta} \exp\left(-\frac{t}{\beta}\right) - \frac{m_g}{\tau_{fri}},\tag{1}$$

and

$$\frac{dm_s}{dt} = \frac{m_g}{\alpha} \,, \tag{2}$$

where t is the time reckoned from the beginning of gas infall (which is assumed to be 12 Gyr ago), and m_g and m_s are the masses of the gas and the stars in the disk, respectively.

The first term on the right-hand side of both equations represents the effect of star formation and resulting gas consumption. The fundamental star formation process, especially the dependence of star formation efficiency upon physical parameters of the star-forming regions, is poorly known. Here the star formation timescale, α , is set to be proportional to the free-fall time of the gas component in the disk. Namely,

$$\alpha^{-1} = k(\rho_g)^{1/2} , (3)$$

where the mean volume density of the gas within the disk, ρ_a , is given by

$$\rho_g = \frac{m_g}{\pi R^2 h_a} \,, \tag{4}$$

and k is a coefficient that determines the efficiency of star formation. Under the specification adopted here lies a belief that some form of gravitational instability in the disk gas should trigger star formation.

The second term on the right-hand side of equation (1) represents the growth of the disk by gas infall from outside the disk plane, the timescale of which is denoted by β . This means that the mass of the halo decreases correspondingly, because we assume the total mass of the system, m_t , to be conserved (here the halo serves only as a reservoir of the primordial gas, and its evolution is not traced). The final mass of the disk, namely, the total mass of the matter that eventually accretes to the disk, is specified by m_d . The time

evolution of the infall rate will depend on the details of how galaxies separate out from Hubble expansion in the early universe and subsequently collapse to denser states. At present, its functional form is neither predicted definitely by any theory nor constrained by observations. The present form is taken because it is physically reasonable and mathematically convenient. The third term on the right-hand side of equation (1) is included here to make allowance for the depletion of disk gas due to gas flow toward the galactic center. This gas inflow should not be confused with the gas infall from the halo region which builds up the disk. Here the gas flow takes place radially within the disk plane toward the galactic nucleus and deprives the disk of its interstellar gas. This flow is due to the dynamical friction acting on massive gas clumps, as described below.

In the present study, it is assumed that the gas that has reached the disk plane cools quickly by radiative processes. Then the gas disk is expected to be subject to local gravitational instability when the Q-parameter (Toomre 1964) for the gas disk decreases to unity as a result of cooling. This instability is a local one and should be discriminated from the bar instability, which results from coherent response of the entire galactic disk. When Q=1, the gas disk has only one neutrally stable wavelength, which is given by

$$\lambda_c = \frac{2\pi^2 \Sigma_g}{\kappa^2} \,, \tag{5}$$

where the epicyclic frequency, κ , and the surface density of the gas, Σ_g , are given here in the averaged sense by

$$\kappa = (2m_t/R^3)^{0.5},\tag{6}$$

and

$$\Sigma_g = \frac{m_g}{\pi R^2} \,. \tag{7}$$

When Q decreases from a value larger than unity, as is most likely, the disk first becomes unstable to the perturbations having the critical wavelength, λ_c , at the instant when Q becomes 1. The critical wavelength is usually much shorter than the disk radius. Therefore, formation of small clumps corresponding to the critical wavelength is expected. The typical mass of these gas clumps is estimated to be

$$M_{\rm cl} = \pi (0.5\lambda_c)^2 \Sigma_g = \frac{\pi^5 \Sigma_g^3}{\kappa^4} \,. \tag{8}$$

In general, $M_{\rm cl}$ is many order of magnitudes larger than individual stellar masses so that these clumps, while orbiting in the galactic disk, suffer dynamical frictions against the disk stars (and against other less massive gas clumpes as well). A rough estimate of the timescale of spiraling-in due to dynamical frictions is given by Shlosman & Noguchi (1993) as

$$\tau_{\rm fri} \sim \frac{3 \sigma_s^3 \Omega r}{(32\pi)^{1/2} v_a G^2 M_{\rm cl} \rho_s \ln \Lambda}$$
 (9)

Here σ_s is the stellar velocity dispersion in the disk, which is determined by equation (10) as explained later. The angular frequency, Ω , of the disk is set to be a constant, $(m_t/R^3)^{0.5}$, here. The characteristic galactocentric radius, r, is set to be 0.5R. G is the gravitational constant, v_a is the velocity of the gas clump relative to the disk stars (i.e., asym-

metric drift) and is equated to σ_s , and ρ_s is the volume density of the stellar disk and is set to be $m_s/(\pi R^2 h_s)$. The Coulomb logarithm $\ln \Lambda$ is set to be 5 throughout (see Shlosman & Noguchi 1993 for further details). It was found that the use of equation (9) in general leads to too much gas inflow to the galactic center to be compatible with those seen in numerical simulations of the similar two-component gas-plus-star disks. This is not surprising in view of many uncertainties involved in evaluating $\tau_{\rm fri}$. In calculating the models described below, the $\tau_{\rm fri}$ in equation (9) was multiplied by a factor of 15 in order to bring the resultant inflow rates in a rough accord with the results of numerical simulations in Shlosman & Noguchi (1993)

2.2. Disk Heating

The clumping in the interstellar gas mentioned above has a very prominent effect on the dynamical evolution of the disk. Due to their large masses, these clumps effectively deflect orbits of the disk stars. The cumulative effect of individual deflections is a secular increase in the stellar random motions. Namely, the gas clumps heat up the stellar disk dynamically. This heating is a back-reaction of the dynamical friction discussed above. Scattering of the field stars by massive celestial bodies is not a new idea, but it has a record of extensive study in the past, especially in relation to the age dependence of stellar velocity dispersions in the solar neighborhood (e.g., Spitzer & Schwarzschild 1953). What should be noted here is that, as we see later, the heating due to gas clumps can be so great as to suppress bar instability in some cases that are relevant to actual disk evolution (Shlosman & Noguchi 1993).

The degree of heating can be quantified by the increase in the velocity dispersion of the stellar random motions in the disk. Therefore, along with equations (1) and (2), we monitor the change in the stellar velocity dispersion σ_s using the equation

$$\frac{d\sigma_s^2}{dt} = \gamma \sigma_s^{-2} \ . \tag{10}$$

Here the coefficient γ is given as (e.g., Lacey 1984)

$$\gamma = 2N_{\rm cl} M_{\rm cl}^2 \omega F(\xi) \ln \Lambda , \qquad (11)$$

where $N_{\rm cl}$ is the surface number density of the clumps and the vertical epicyclic frequency $\omega \sim (2\pi\Sigma_s)/h_s$, where the stellar surface density $\Sigma_s = m_s/(\pi R^2)$. F takes a value between 0 and 3.1 depending on the ratio $\xi \equiv 2\Omega/\kappa$, which is determined by the rotation law. Thus, the value of F depends in general on the galactocentric radius, so that F is small in nearly rigidly rotating (usually inner) parts and large for the (usually outer) regions with strong differential rotation. Because our treatment of the dynamical heating here does not take radial dependence into account, we take F = 0.47 as its representative value. This value corresponds to a flat rotation, i.e., $\xi = 1.41$. We assume here that $N_{\rm cl} M_{\rm cl} \sim \Sigma_g$, i.e., most of the gas participates in the clumping, as was shown numerically by Shlosman & Noguchi (1993). Among the quantities appearing in the expression of γ , only $M_{\rm cl}$, $N_{\rm cl}$, and ω are time dependent in the present formulation.

Now equations (1), (2), and (10) complete a set of differential equations which determine the temporal evolution of the disk. By integrating these equations numerically starting from an appropriate initial condition, we can obtain, as

a function of time, the disk mass fraction f_d , the gas mass fraction f_g , and the normalized stellar velocity dispersion σ . These are defined by

$$f_d \equiv \frac{m_g + m_s}{m_t}, \quad f_g \equiv \frac{m_g}{m_t}, \quad \sigma \equiv \frac{\sigma_s}{v},$$
 (12)

where $v \equiv (Gm_t/R)^{1/2}$ stands for the characteristic rotational velocity. In the models described below, the integration was carried out over the whole Hubble time from t = 0 to the present epoch t = 12 Gyr, starting from the initial condition $f_d = f_g = \sigma = 0$.

Here a short comment will be worthwhile about the calculation of σ . In a given galactic disk, the velocity dispersion of a stellar population depends in general upon its age such that old populations have a larger dispersion than young ones. We adopt the following procedure to take into account this possible age dependence in calculating σ . The Hubble time has been divided into 100 epochs, with each epoch having a width of 1.2×10^8 yr. When we make new stars at each time step according to equation (2), we note to which generation of these 100 epochs the newly born stars should belong and increase the mass of that generation by an appropriate amount. The mass of each stellar generation is thus calculated. Then equation (10) is applied to each generation respectively to trace the development of its velocity dispersion after birth. Each generation is assumed to have negligible random motions (i.e., $\sigma = 0.05$) at its birth. Finally, the value of σ at a given time is set to be the mass-weighted mean of the velocity dispersions for all the generations present at that time.

As we see in the next section, a set of $[f_d(t), f_g(t), \sigma(t)]$ can be used to judge whether the stellar disk is unstable to spontaneous bar formation at a given time t, by employing the stability criterion derived by Shlosman & Noguchi (1993) for the stellar disks coexistent with dissipative interstellar gas as an extension of the Ostriker-Peebles (1973) criterion for purely stellar disks.

2.3. Choice of Parameter Values

In all the present models, we set R = 10 kpc and $m_t =$ $10^{11} M_{\odot}$. This means that we consider only giant galaxies that are comparable to our Galaxy in mass and size. Any dependence on the galaxy mass (and hence luminosity) is not investigated. The characteristic dynamical time of the system then becomes $\tau_d = (R^3/Gm_t)^{1/2} \sim 5 \times 10^7$ yr, where G is the gravitational constant. We also fix $h_s = 600$ pc, and $h_q = 200$ pc. We note that the adopted value of h_s is comparable to the observed value of the thickness of the stellar disks, $\sim 600-1000$ pc (e.g., van der Kruit & Searle 1982), and h_a is bracketed by the scale heights of the gas layers, ~ 100 pc for molecular gas (e.g., Solomon, Sanders, & Scoville 1979) and ~ 300 pc for H I gas (e.g., van der Kruit & Shostak 1983). The final disk mass, m_d , is set to be 50% of the total mass (i.e., $m_d = 0.5m_t$). This is in accord with the observational evidence (e.g., van der Kruit & Searle 1982; Bahcall & Casertano 1985) that the halo-to-disk mass ratio inside the optical radius is close to unity in the galaxies for which high-quality photometric and kinematic data are available. There is some suggestion that the global disk structure at the current epoch does not vary significantly in different morphological (Hubble) types. The compilation of available observational data by van der Kruit (1987) suggests that the central surface density and scale length of the disk are not systematically different among different Hubble types if dwarf galaxies are excluded. It is nevertheless cautioned that the sizes of the observed samples are not yet sufficiently large, and further observations are required to make it completely clear whether or not the mass ratio m_d changes as a function of morphological type and/or luminosity class (see van den Bergh 1982 for a possible dependence on luminosity).

From the theoretical viewpoint, a disk galaxy with m_d less than ~ 0.3 never develops a bar spontaneously (though the critical mass seems slightly dependent on other characteristics, such as mass concentration toward the center). Only tidal interaction can induce a bar in those relatively light disks (Noguchi 1987). On the other hand, a disk with $m_d \sim$ 0.5 can be bar-unstable or stable depending on the stellar velocity dispersion and the gas mass fraction (Shlosman & Noguchi 1993). As we see later, all the present models that correspond to late Hubble types are shown to develop a bar spontaneously in late phases of their evolution. We, however, observe many unbarred late-type galaxies in the universe. These galaxies may have relatively light (i.e., m_d < 0.5) disks compared with their barred counterparts. Making allowance for the likely distribution of m_d in galaxies of a specific morphological type will be crucial in discussing relative populations of barred and unbarred galaxies in that type. We do not address this point here. Existence of a broad distribution of the disk mass fraction would not alter fundamentally the conclusions of the present study, provided that the form of the distribution is not strongly dependent on the Hubble type.

Concerning star formation, we fix k to be 0.07 in all the models. Namely, the star formation efficiency is not a free parameter, but a universal constant in the present models. Then the star formation rate is determined automatically by the gas density in the disk by equation (3) as the disk evolves. Behind this treatment is our belief that the basic process (whatever it may be) of star formation will not differ largely among different galaxies, so that the law of star formation should not be changed arbitrarily in theoretical modeling. This prescription is in a marked contrast with some galaxy evolution models, which vary the law of star formation as one of the primary factors influencing disk evolution (e.g., Talbot & Arnett 1975; Ferrini & Galli 1988; Galli & Ferrini 1989; Arimoto & Jablonka 1991). In such approaches, it is relatively easy to reconcile the models with the observations by tuning parameters in the adopted star formation law, thus making the conclusions less restrictive. As we will see later, the present choice of k gives the star formation rates in early phases in reasonable agreement with the observational inference, though the simple nature of the present models does not allow finer tuning.

After these choices of parameter values, we are left with only one parameter that we can vary freely and that has physical interest: the infall timescale β . This parameter is considered to be related physically to the global morphology of a galaxy as follows. The collapse of a gaseous protogalaxy is considered to be free fall if the cooling timescale is sufficiently short. In this case, the collapse timescale will vary as the inverse of the square root of the mean density in the protogalaxy. This mean density in turn will depend on the environment in which that galaxy is formed (e.g., in rich clusters or in the fields). One possible inference is that early-type galaxies that are considered to have formed preferentially in high galaxy number density regions have a

relatively short timescale of collapse, whereas the protogalaxy collapse has been prolonged considerably in late-type galaxies, which tend to inhabit more sparse regions (e.g., Giovanelli, Haynes, & Chincarini 1986). This is in qualitative accord with some numerical models of protogalaxy collapse (e.g., Larson 1976). Though the infall timescale β in the present study is relevant not to the whole protogalaxy collapse but only to its late phase (i.e., the epoch at which the gas accretes roughly perpendicularly to the disk plane), we consider hereafter that the Hubble sequence is the sequence of β , with a larger β corresponding to a later morphological type. In agreement with this, Arimoto & Jablonka (1991) found that the infall timescale should increase from early to late morphological types for the observed trends of various photometric properties along the Hubble sequence to be reproduced in their simple evolutionary model. The formation process depicted here may prove to be oversimplified. Even in that case, the disk formation timescale introduced in the present model still has a well-defined meaning and is considered to remain as one of the most important parameters that govern the fate of a growing disk.

3. RESULTS

We have run a series of models in which β is varied from 10 to 80 in units of the dynamical time, which corresponds to (0.5-4) Gyr if scaled to the typical galactic parameters. These values of β should be understood as only rough ones. Real galactic disks are highly inhomogeneous on global scales and are likely to have formed "inside-out" because of increasing density of the matter toward the galactic center. Thus, β will plausibly be an increasing function of the galactocentric distance within a given disk galaxy. The present treatment of the galactic disk neglects this radius-dependent time lag in disk formation completely. Putting this aside, the adopted range of β essentially encompasses the values inferred by other independent studies. The model with $\beta = 0.5$ Gyr represents the limit of fast collapse in the sense that the disk growth timescale is comparable to the rotation period of the disk. On the other hand, the disk grows very slowly in the model with $\beta = 4$ Gyr, with a timescale that is much larger than the rotational period and a sizeable fraction of the Hubble Time.

3.1. Time Evolution

Figure 2 shows the time variation of several quantities of interest for different values of β . In all the panels of Figure 2, the thick portion of each curve indicates that the stellar disk is bar unstable at the corresponding time. The stability has been judged by the method described later. Figure 2a shows just how fast the disk builds up for a given value of β . The mass of the gas disk is plotted in Figure 2b. It is noted that a model with a faster infall develops a higher gas mass at an earlier epoch than that with a slower infall, but it shows a sharper decline of the mass after that, leading to a lower gas mass at the present epoch. The variation of the star formation rate, SFR(t), shows a similar dependence on β to that of the gas mass (Fig. 2c): A faster infall leads to a higher maximum SFR at an earlier epoch but a lower value at present. It is seen that the change in derived SFR(t) from a small β to a larger one corresponds to the observationally inferred change in the star formation history from early to late types along the Hubble sequence. The adopted value of

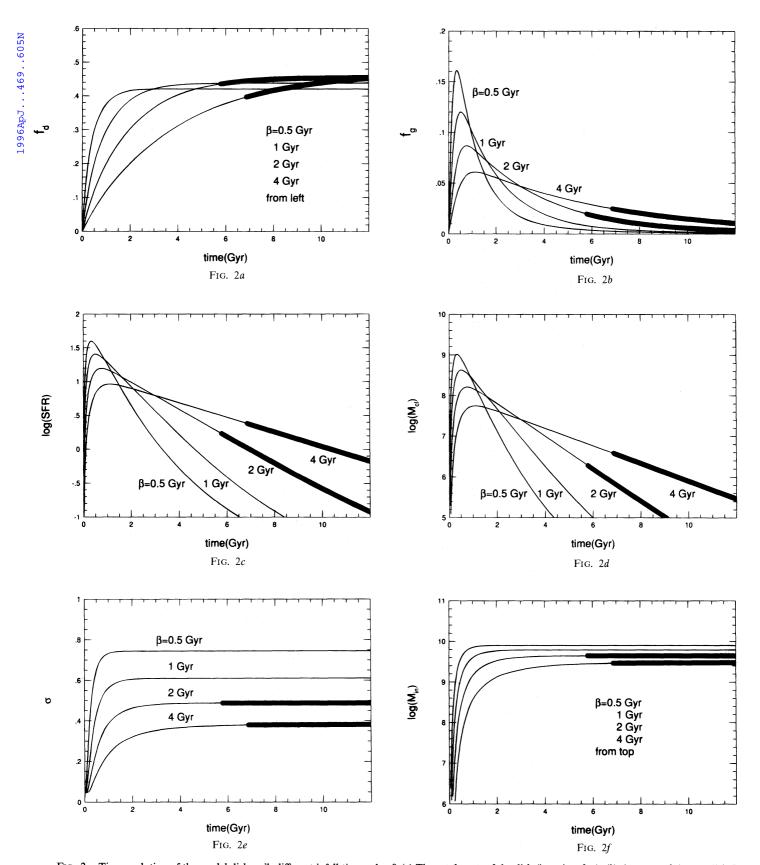


Fig. 2.—Time evolution of the model disks with different infall timescales β . (a) The total mass of the disk (in units of m_t), (b) the mass of the gas disk (in units of m_t), (c) the star formation rate (in units of M_{\odot} yr⁻¹), (d) the typical mass of the gas clumps formed in the disk (in units of M_{\odot}), (e) the velocity dispersion in the stellar disk (in units of the rotational velocity), and (f) the cumulative mass accreted to the galactic center by the dynamical friction—induced gas flow in the disk plane $M_{\rm in}$ (in units of M_{\odot}) are plotted against time for the entire Hubble time. The phase in which the stellar disk is bar unstable is indicated by heavy lines.

the coefficient k (in eq. [3]) = 0.07 gives quantitatively reasonable values for the maximum SFR expected in each type (see Sandage 1986, for example). The gas mass fraction and SFR near the present epoch do not necessarily agree with the observationally inferred values. For example, all the models show a current SFR less than $1 M_{\odot} \text{ yr}^{-1}$, which evidently contradicts observations of late-type spirals. This discrepancy is not serious, however, because the global dynamical property of the stellar disk is determined primarily in early phases, when the gas content in the disk is near its maximum. Therefore, it will be possible to change, for example, the amount of star formation in late phases by introduction of a prolonged gas infall (i.e., nonexponential type) of small amount or by taking account of recycling of gas due to stellar winds and supernova explosions, without affecting the dynamics seriously.

Figure 2d plots the typical mass of the gas clumps, $M_{\rm cl}$, estimated by equation (8) as a function of time. A model with a smaller β shows a larger maximum value of $M_{\rm cl}$ at an earlier epoch but shows a faster decline after that than a model with a larger β . In this case, a larger maximum $M_{\rm cl}$ for a smaller β is caused simply by the larger maximum gas density attained in the disk because κ is identical (see eq. [8]). In early phases (t < a few Gyr), the clump mass exceeds the observed range for giant molecular clouds ($10^5 - 10^6 M_{\odot}$) in our Galaxy and nearby galaxies by several order of magnitude. The fastest growing model disk with $\beta = 0.5$ Gyr develops very massive clumps with $M_{\rm cl} \sim 10^9 M_{\odot}$ in its most gas-rich phase. These clumps have important dynamical effects on disk evolution, namely, the heating of the stellar disk (Fig. 2e) and the gas infall to the nuclear

region (Fig. 2f). As Figure 2e shows, a faster infall brings about more efficient heating due to more massive gas clumps formed in the disk. In the model with $\beta=0.5$ Gyr, the stellar velocity dispersion increases and finally becomes comparable to the rotational velocity, while the heating is reduced greatly in the model with $\beta=4$ Gyr. As we see later, this difference in heating history is the key determinant of whether the stellar disk becomes bar unstable or not. Finally, Figure 2f shows that a model with a faster infall leads to more gas accreted to the galactic center. This is because more massive gas clumps formed in the fast-growing disk experience stronger dynamical frictions from stars, leading to more rapid spiraling-in to the galactic center (see eq. [9]), and because the total amount of available gas itself is larger.

3.2. Bar Instability

The criterion deduced by Shlosman & Noguchi (1993) indicates that the stability of a galactic stellar disk against bar formation is determined essentially by three parameters, namely, the mass fractions of the total disk component, f_d (including both stars and gas), and of the gas disk component, f_g , relative to the total galaxy mass, and the stellar velocity dispersion, σ , in units of the rotational velocity v (see their eq. [24]). This criterion of bar instability is represented as a surface defined on the (f_d, f_g) plane, which gives the critical value of the stellar velocity dispersion, $\sigma_{\rm crit}$, required for bar stability as a function of f_d and f_g . The contours of a constant $\sigma_{\rm crit}$ are drawn in Figure 3 by dashed lines for $\sigma_{\rm crit}/v=0$, 0.2, 0.4, 0.6, and 0.8 (from left to right). Each line represents the neutrally stable states that divide

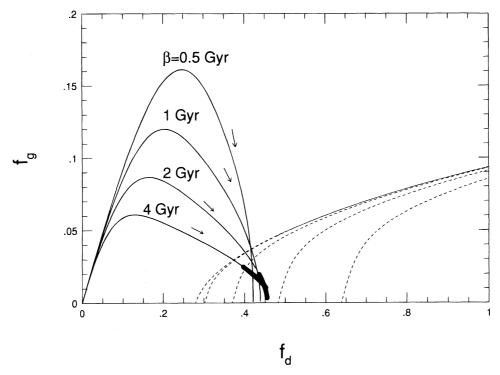


Fig. 3.—Trajectories on the (f_d, f_g) plane of the model disks with different infall timescales β . Here f_d and f_g are the fractions of the masses of the total disk and the gas disk with respect to the total mass of the galaxy (i.e., the disk plus the halo). The phase in which the stellar disk is bar unstable is indicated by heavy curves. Each dashed line represents the neutral stability line for the stellar disk in a two-component star-plus-gas disk model for the given value of the stellar velocity dispersion (normalized by the rotational velocity) from Shlosman & Noguchi (1993). The stellar disk is subject to bar instability if it is located in the domain below the line for the corresponding velocity dispersion. Each model disk starts at the origin, i.e., $f_d = 0$, and $f_g = 0$, and moves with time as indicated by the arrrow.

the whole permissible (f_d, f_g) plane into stable and unstable regions for a given stellar velocity dispersion. The stellar disk of a given model is stable (unstable) if the model is located in the upper (lower) side of the dashed line for the stellar velocity dispersion in that model. In other words, if the actual stellar velocity dispersion, σ , in the model disk is larger than the normalized critical dispersion $\sigma_{\rm crit}/v$ for the values of f_d and f_g of that disk, then the stellar disk remains stable. By tracing the time evolution of the three quantities (f_d, f_g, σ) for a particular model, we can judge whether its stellar disk at any given epoch is subject to bar instability or

The trajectory of each calculated model is drawn in Figure 3. Every model initially starts from the origin ($f_d = 0, f_g = 0$) and then moves to the upper right, with both the gaseous and stellar disks growing in mass. The total disk mass never decreases, whereas the gas disk begins to be depleted at the epoch at which star formation overtakes gas replenishment by infall. This moves the model lower right on the trajectory. Because of the assumed constancy of the final disk mass, m_d , all the trajectories converge to a narrow region around $f_d \sim 0.5$ in Figure 3 as the present epoch is approached. However, dynamical properties of the resulting disk depend on β .

In Figure 3, the bar-unstable states are indicated by thick portions of the evolutionary tracks. It is seen that if β is less than about 1 Gyr, then the disk is kept stable during its whole evolution. In this case, a purely stellar disk with a mass $\sim 50\%$ of the total galaxy mass finally comes out without suffering from bar instability. By the time the disk growth is completed and most of its mass is converted into stars, the stellar disk component has already acquired sufficiently large random motions to assure stability, due to strong dynamical heating caused by very massive clumps in the gas-rich (i.e., $f_q > 0.1$) phase. On the other hand, the heating of the stellar disk is not enough to damp bar instability in a slowly growing disk with $\beta > 2$ Gyr, due to the low maximum gas fraction ($f_g < 0.1$) attained. The stellar disk becomes bar unstable at the instant at which the total disk mass, f_d , exceeds the critical value determined by the gas mass fraction f_q and the stellar velocity dispersion σ . If the effect of dynamical friction (the third term on the righthand side of eq. [1]) is neglected, the results described above are changed slightly, but overall tendency remains the same (of course except Fig. 2f).

The present crude modeling is incapable of clarifying further details of disk formation. For example, the emergence of a bar in a slowly growing disk will not be a "pandisk" event with a well-defined epoch. The buildup of the disk and the subsequent gas depletion due to star formation are likely to proceed faster in the inner regions. Then actual evolution will be such that a small bar first appears in the innermost region and then grows in size as a progressively outer region of the disk is built up and then depleted of the interstellar gas. We discuss this expectation in relation to some observational data in § 4.3.

4. DISCUSSION

In this section, we first discuss the implications of the present results for the formation mechanisms of barred galaxies, with the aid of numerical results from N-body simulations. Although our interest was initially in bar formation, the present study also gives insight into other aspects of disk galaxy evolution. As an important example, a possible

formation mechanism of thick disks is discussed on the basis of the present results. Comparisons with other theoretical works follow. Finally, limitations of the present models are discussed, as well as further possible extensions in the future.

4.1. Bar Formation Processes in Different Hubble Types

Observations show that the fraction of barred galaxies is nearly constant along the Hubble sequence, staying at about one-third (e.g., Elmegreen et al. 1990). There are notable systematic differences in the properties of bars in the early- and late-type galaxies, suggesting different formation mechanisms. Bars in early-type galaxies extend further out relative to the optical radius than bars in late-type galaxies (e.g., Elmegreen & Elmegreen 1985). Surface brightness along the bar major axis decreases slowly or stays almost constant in early types, whereas the bars in late types tend to have exponential brightness profiles declining steeply with the radius (e.g., Elmegreen & Elmegreen 1985; Kent & Glaudell 1989). Another difference is in the shape of isophotos. Rectangular bars are common in early types, while bars in late-type disks have more elliptic isophotos in general (e.g., Ohta, Hamabe, & Wakamatsu 1990). Earlytype bars are sometimes accompanied by lenses and outer rings, whereas inner rings are more common in late-type bars (e.g., Kormendy 1979; Buta & Crocker 1991). There is no convincing theory to explain all of these systematic differences. We propose in the following a dual scenario for the formation of bars in different Hubble types on the basis of the results described in the previous section.

The present model has shown that a key parameter that governs the stability characteristics of evolving disks is the infall timescale, i.e., the growth timescale of the (total) disk. A model with a fast infall develops a stellar disk that has sufficiently large random motions to suppress spontaneous bar formation. On the other hand, a slow infall leads to a stellar disk with a small velocity dispersion, which becomes unstable once its mass reaches a critical value. As one possibility, we have related the infall timescale β to the galaxy morphology, with a smaller β corresponding to an earlier Hubble type. Then one natural consequence from this consideration is that late-type barred galaxies have resulted from bar instability in their disks, whereas bars in early-type galaxies have been formed in tidal interactions.

This scenario can find several independent supports, as described below. One of these comes from environmental consideration. A large body of observational data have established a well-known morphological segregation, i.e., late-type galaxies are located more frequently in the regions in which number density of galaxies is low, whereas earlytype galaxies reside in more crowded regions such as the centers of clusters of galaxies. This trend is true for elliptical versus spiral galaxies (e.g., Dressler 1980) as well as for a finer division, i.e., early-type disks versus late-type ones (Gisler 1980; Giovanelli et al. 1986). Therefore, late-type disk galaxies are considered to have had a smaller probability of encountering or colliding with other galaxies than early-type disks in their life. Then it is quite natural to think that bars in late-type disk galaxies started to form spontaneously at the points at which the disks became unstable without the help of external disturbances. On the other hand, early-type disk galaxies, which cannot form bars by themselves, had enough chance of tidal interactions, which excited bar structures in their disks.

Then one interesting and natural question is: Does the systematic difference in bar structures in different morphological types as described earlier result from the difference in the formation mechanism? The answer to this question from numerical simulations described in the next section constitutes the strongest support to the dual formation scenario proposed here.

4.2. Evidence from N-Body Simulations

We have carried out numerical simulations to see whether spontaneous bars have significantly different properties from tidal bars in their structure. If this is the case, detailed observation of the bar structure in a given galaxy would help to guess its origin. All the numerical simulations described below have been performed by using the tree method (see Barnes & Hut 1986). A detailed description of the numerical treatment is given in the Appendix.

Figure 4 shows a typical example of spontaneous bars. Here an exponential stellar disk having the same mass as the spherical halo and relatively small random motions corresponding to Q=1.5 has been evolved without any external perturbations. The snapshots in Figure 4a indicate that the bar is formed quickly in three disk rotations and reaches a nearly steady state. For a detailed examination of the bar structure, a "slit" is placed along the major axis of the bar for "photometry" at the position angle indicated by a pair of segments in each plot. The resulting surface density profile is shown in Figure 4b. It is noted that the surface density along the bar major axis decreases very rapidly and monotonically with the radius. The density gradient in the inner bar region is steeper than that of the initial exponential disk.

Though we have shown only one example here, the steep density profile is a common feature of spontaneous bars. Ohta et al. (1990) have analyzed a numerical model of a spontaneous bar starting from a Q = 1 exponential disk embedded in a equal mass halo and found a qualitatively similar profile to the model shown in Figure 4. Ohta et al. (1990) have also noted a discrepancy between spontaneous model bars and the bars in early-type galaxies, of which they have made detailed photometry. A wide variety of massive disk models lead to similar final states (e.g., Hohl 1971). Few exceptions to this rule include the simulation by Sparke & Sellwood (1987) of a Kuzmin/Toomre disk having twice as much mass as the Plummer "bulge" accompanying it, which gives rise to a relatively flat profile along the major axis of the resulting bar and lecangular shape of isodensity contours, in fair agreement with the observed early-type bars. The reason for the disagreement of the model by Sparke & Sellwood (1987) with those of the present study and other previous studies is not clear. It should be noted, however, that the initial condition adopted by Sparke & Sellwood (1987) is a Kuzmin/Toomre disk embedded in a spherical Plummer halo with the disk-to-halo mass ratio of 7:3, which is largely different from those taken in the present study. Therefore, it is probable that the difference in initial conditions has led to different morphologies of the final bars.

Let us turn to tidally created bars. Figure 5 shows one example. In this case, the initial velocity dispersion in the disk has been set to Q=3, so that the disk did not develop a bar in isolated state despite its large mass. Except the amount of initial random motions, this model has essen-

tially the same characteristics as the model shown in Figure 4. This model galaxy was then made to have a close encounter with a point-mass perturber of the same mass on a parabolic orbit with a pericenter distance twice the disk radius. The rotation of the disk is in the same direction as the orbital motion of the perturber (i.e., a prograde coplanar encounter). Figure 5a shows that the disk remains nearly axisymmetric until the closest (i.e., pericentric) passage of the perturber, t = 0. A strong bar develops after that and lasts for a long time (at least a few Gyr) with little changes in strength and size. Figure 5b describes the major axis profile of the bar. As seen here, the tidal bar has a much flatter density profile along its major axis than spontaneous bars. Moreover, sudden steepening in density gradient is seen near the ends of the bar, which gives "shoulders" to the density profile. Many simulations for a wide parameter space are still to be done to ascertain if this property is shared generally by tidal bars, although the result of Sundin, Donner, & Sundelius (1993) suggests a similar flattening of the profile for a tidally perturbed Kuzmin/Toomre

Thus, a remarkable qualitative difference is expected to exist in density profiles of spontaneous bars and tidally induced ones. Interestingly enough, this dichotomy in bar structure found in numerical models matches well the observational result for actual barred galaxies. Elmegreen & Elmegreen (1985, hereafter EE) have carried out an extensive blue and near-infrared surface photometry of barred galaxies of various morphological types and found that the observed bars fall into two groups according to the density profile along the bar major axis: "flat" bars and "exponential" bars. The spontaneous bar in our numerical simulations resembles exponential bars in EE that have a steep monotonous decrease of the surface brightness. On the other hand, the numerical tidal bar in Figure 5 bears a qualitative resemblance to flat bars, though some of the flat bars in EE have nearly constant surface brightness up to the bar ends along the bar major axis, which numerical simulations have not successfully reproduced yet. Interestingly, EE have found that flat bars are more frequent in early-type galaxies (i.e., Sa and Sb), whereas late-type galaxies such as Sc and Sd tend to have exponential bars. This observational finding is understood quite naturally in our proposed scenario that early-type barred galaxies have tidal origin while spontaneous bar formation has occurred in late type galaxies, given the structural difference between the two groups of bars revealed by numerical modeling.

Actually the correlation of bar-type with galaxy morphology found by EE is not so tight (see their Fig. 4). Some bars in late-type galaxies do have flat profiles. We have checked the environments of these exceptional galaxies using the atlas by Tully & Fisher (1987) and found that all of these late-type galaxies with flat bars are either members of close galaxy pairs or located in dense galaxy groups. Thus, the tidal origin is suggested also for these late-type barred galaxies. On the other hand, some early-type galaxies with flat bars in EE seem to be isolated, with no discernible nearby companions. Nevertheless, most of these galaxies (except NGC 986) are located in groups or dense parts of large-scale galaxy clustering. Tidal bars are generally considered to have lifetimes longer than several Gyr, as suggested in Figure 5. Within this period, companion galaxies can recede from the perturbed victims by as much as 1 Mpc. Therefore, the isolation at the present epoch does not

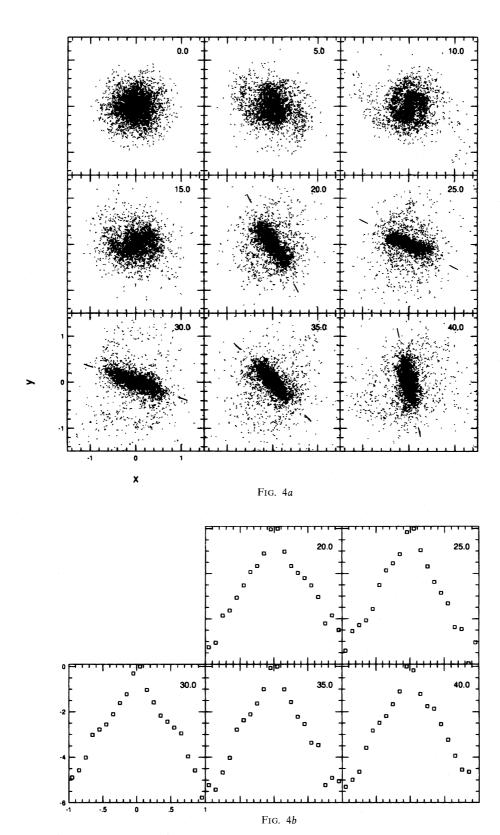


Fig. 4.—An example of spontaneous bars from numerical simulations. An exponential stellar disk embedded in a halo containing the same mass as the disk was evolved in isolation. The disk and the halo have been constructed by 50,000 and 25,000 collisionless particles, respectively. The initial rotational velocity and velocity dispersions in the disk are given in Fig. 6. The Q-value in the disk was set to be 1.5 initially. (a) Snapshots of evolution in face-on view. Only 10,000 disk particles are selected randomly and plotted. Time t in units of the dynamical time is indicated in the upper right corner of each frame. One rotation period at the outer disk edge is 6.28. One time unit is approximately 10^8 yr if scaled to a typical disk galaxy. The coordinates are given in units of the initial disk radius. The disk rotates counterclockwise. (b) The surface density profile along the bar major axis. Each profile was constructed by placing a rectangular slit of a length 2 (i.e., the same as the initial diameter of the disk) and a width of 0.1 on the disk particle distribution at the position angle indicated by a pair of tick marks in the corresponding frame in (a), and counting the number of particles contained in each of 20 squared bins (size of 0.1×0.1) that fill the slit. The surface density is given in units of (negative) magnitude relative to the peak density. The abscissa is the position measured along the slit in units of the initial disk radius, with a positive value corresponding to the position on the slit having a positive x.

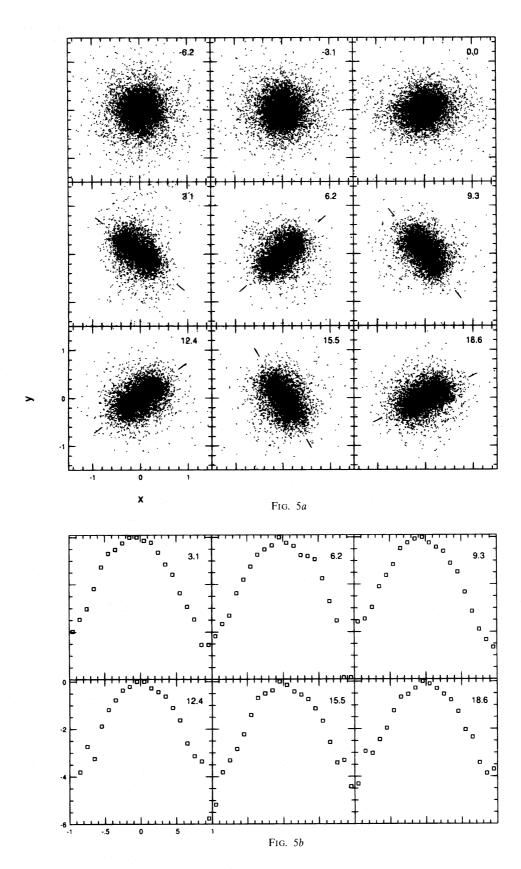


Fig. 5.—Same as Fig. 4, but for an example of tidal bars from numerical simulations. The galaxy model is essentially the same as that described in Fig. 4, but its disk component has a greater velocity dispersion initially (Q = 3) to suppress bar instability. This model keeps a nearly axisymmetric shape for a long time (40 dynamical times) in isolated state. The structure and kinematics of the model at the end of the isolated run are indicated in Fig. 7. A parabolic encounter with a point mass perturber of the same mass gives rise to a strong bar in the disk component, as shown in (a). In this panel, the perturber moves in the disk plane (i.e., the x-y plane) and passes the pericenter (x = 2, y = 0) at t = 0 in the same direction as the disk rotation, i.e., counterclockwise. (b) The surface density distribution along the bar major axis obtained in the same way as in Fig. 4b. Note a much flatter profile than that of the spontaneous bar in Fig. 4b and two shoulders (i.e., abrupt changes in density gradient) at the ends of the bar.

1996ApJ...469..605N

necessarily rule out the possibility of close encounters in a long past. Elmegreen et al. (1990) have found that the fraction of barred galaxies in binary galaxy pairs is $\sim 50\%$ in early types (Sa-Sb) and < 30% in late types (Sbc-Scd), suggesting a more dominant role played by tidal interactions in earlier types. We need not consider that only the relative excess of $\sim 20\%$ seen in early type galaxies is due to tidal interactions. A larger portion may actually be associated with tidally induced bar formation.

Other morphological characteristics (bar sizes relative to the disk radius, isophotal shapes, association with different types of rings) have not been examined here. More realistic numerical treatment will be required to address these points reliably (see § 4.6). It may be necessary to include the interstellar gas in such numerical simulations to give satisfactory explanation to these features, especially the formation of various ring structures.

4.3. Spatial Distribution of Activity

Another interesting observational trend that may have relevance to different formation mechanisms is noted concerning activity in barred galaxies. Devereux (1987) has made ground-based 10 μ m observations of the central regions of 133 nearby galaxies and tried to figure out the spatial concentration of far-infrared emission by comparing 10 μ m data with more poorly resolved IRAS data. He found that some barred galaxies of early types (Sb and earlier) have centrally concentrated active regions (star formation and/or Seyfert activity), whereas early-type unbarred galaxies and late-type galaxies (both barred and unbarred) do not exhibit such concentration. This result suggests different dynamical properties of bars in different morphological types. He points out a possibility that different degrees of bulge dominance in different Hubble types give rise to different resonance characteristics and gas response (see also Athanassoula 1992). According to the present scenario, however, an alternative interpretation is possible as follows. The present model suggests that the spontaneous emergence of a bar in late-type disks is a gradual process that proceeds on the timescale of disk formation or star formation (whichever longer), which is significantly longer than the disk rotation period for late types. The bar is expected to form inside-out as a progressively larger radius of the disk is made up and then depleted of the interstellar gas due to star formation. In this case, we expect active star formation preferentially in the spiral arms just outside the bar, as suggested from some numerical simulations (e.g., Sorensen et al. 1976), but not in the bar itself because depletion of the interstellar gas within a given radius is a necessary condition for the bar to grow radially up to that radius. On the other hand, tidal bar formation, which is considered to be responsible for the bars in early-type galaxies according to the present study, is a quick process completed within roughly one disk rotation period (e.g., Noguchi 1987). If the galactic disk contained enough gas prior to the encounter, then its portion within the length of the bar to be formed will fall to the disk center within nearly one-free fall time and will accumulate there. In this case, the gas does not have enough time to be fully converted into stars before its major portion reaches the galactic center, leading to concentrated activity at the nucleus. This speculation, though still to be confirmed by more elaborate simulations, fits nicely with the observation by Devereux (1987). In this picture, the difference in the spatial distribution of activity in different morphological types is ascribed to the different timescale of bar formation: sudden tidal triggering in earlytype galaxies versus slow spontaneous emergence in latetype ones.

4.4. Thick Disk Formation by Hyper Gas Clouds

We address here the formation mechanism of thick disks insomuch as it is relevant to the present model of galactic disk evolution. There are two fundamentally different ideas about how and when these structures have been formed. The first one interprets thick disks as an intermediate structure between the usual disks (thin disks) and the halo components both chronologically and spatially. According to this idea, thick disks formed after the halo components but before the thin disks. For example, Jones & Wyse (1983) propose a scenario in which the thick disk formed first and the gas shed from evolved stars in the thick disk accreted later on and made a thin disc. The second idea is that the thick disks are a secondary structure formed from the stars which were initially born in the thin disks but afterward were somehow scattered out of the disk plane (e.g., Freeman 1987), though there is no consensus on the responsible scattering mechanism(s).

We propose here a scenario in which the thick disks formed as a result of disk heating by very massive gas clumps, the typical mass of which ranges from $10^8 M_{\odot}$ to higher. Such a very massive gas clump is referred to as a "hypercloud" hereafter by analogy with "superclouds" advocated by Elmegreen & Elmegreen (1983). In contrast to superclouds, which are actually observed as an aggregation with the mass of $\sim 10^7 M_{\odot}$ and the size of ~ 1 kpc, hyperclouds are hypothetical entities at present. No compelling observational evidence exists for hyperclouds, though the giant molecular associations observed in M51 (Rand & Kulkarni 1990), the individual mass of which is estimated to be several times $10^7 M_{\odot}$, may fall in the low-mass end of this category. One fascinating point of these very massive clouds is their effectiveness in rising the stellar random velocities not only in the disk plane, but also in the vertical direction. Though making a precise estimate is difficult, a considerable fattening of the stellar disk is expected in view of efficient heating shown in Figure 2e. It is predicted that early-type galaxies have a thicker stellar disk than late-type ones because hyperclouds are formed preferably in fastgrowing disks corresponding to early-type galaxies according to our analytical model (see Fig. 2d). What is crucial here is the lifetime of hyperclouds. Star formation activity inside a cloud may destroy it through energy injection from stellar winds and supernova explosions. This problem is important and is discussed later. The observational studies in the past have found a marginal evidence that thick disk components are more common in early-type disk galaxies than in late-type ones (Burstein 1979; Tsikoudi 1979; van der Kruit & Searle 1981a, 1981b) in the sense that their existence is closely associated with the presence of an appreciable bulge. The number of the galaxies inspected so far is frustratingly small. If confirmed in a statistically meaningful sample, such an observational trend is in accord with our proposed scenario, though we do not necessarily rule out other possibilities.

The thickness of bars is another matter of debate. If the present scenario is correct, the spontaneous bars should be as thin as the usual thin disks. On the other hand, the tidal bars should be relatively thick because they are formed

mainly from already fattened stellar disks. Noncoplanar (i.e., inclined) galaxy encounters could thicken the resulting bars further. Observational results are sparse and not definitive. Thin bars with an axial ratio larger than 10 are reported by Kormendy (1982) and Wakamatsu & Hamabe (1984). All the thin bars observed to date (Tsikoudi 1980; Wakamatsu & Hamabe 1984; de Carvalho & da Costa 1987; Hamabe & Wakamatsu 1989) are in S0 galaxies, i.e., the earliest disk galaxies. This may be at variance with the prediction above. The problem here is that the B-band luminosity distributions that these authors used for decomposition analysis do not necessarily trace the underlying mass distribution. Young stellar populations that dominate B-band luminosities are considered to have generally a smaller thickness perpendicular to the galactic plane than older ones, which trace the mass distribution more closely (this will be the case even if the disk scale height is not determined during the protogalactic collapse phase but is increased a posteriori by some stirring mechanism, such as scattering by massive gas clouds). Therefore, a quantitative comparison between the theory and observations cannot be done reliably until the scale height of edge-on bars is measured in near-infrared wavelengths, which reflect the underlying mass distribution most faithfully. It seems curious that the claimed edge-on bars in these S0 (lenticular) galaxies are thinner than their disk components. At the moment, there seems to be no idea how such thin bars have been formed. It may be noteworthy in this regard that all three S0 galaxies (NGC 1381, 4452, and 4762) in which a thin bar is suspected have an abnormally flattened global structure. The evolutionary "niche" of such galaxies is itself enigmatic. The extreme flatness of NGC 1381, 4452, and 4762 does not seem to fit with the classification of these galaxies into the So class defined as an intermediate type between elliptical and spiral galaxies. Instead, these galaxies may represent the much evolved state of the ordinary late-type disk galaxies, in which all the interstellar gas has been consumed by star formation and spiral structures are not sustained anymore (leading to S0 classification). Then the observed thin bars could be relics of spontaneous bars that formed in these galaxies when some interstellar gas was still available for the formation of massive stars in the bars. Even putting these issues aside, the theoretical prediction that spontaneous bars are thin while tidal ones are thick may be too simplistic. Combes & Sanders (1981) suggest that a spontaneous bar that is thin at its birth evolves secularly into a thick structure, which we observe as boxy (or peanutshaped) bulges in edge-on galaxies. In order to obtain a correct picture of bar formation, it is important to investigate the correlation of various bar properties and other characteristics such as host galaxy morphology and galaxian environment.

4.5. Comparisons with Other Studies

Recent theoretical studies on the formation and evolution of galactic bars are becoming more and more elaborate, incorporating hitherto neglected gaseous dissipative effects and the star formation process into the numerical scheme. Most of these attempts are, however, not explicitly combined with any specific picture of host galaxy formation.

Combes & Elmegreen (1993) claim to have succeeded in producing a dichotomy in bar structures observed by Elmegreen & Elmegreen (1985), by means of numerical simulation. In their scenario, both the early- and late-type

barred galaxies have resulted from bar instability in fully formed disks. The difference in bar properties is ascribed to the difference in initial mass distribution in the disk, in contrast to the present study, which ascribes different bar properties to the different kinematical history of the galactic disk. They consider that the galaxies of earlier morphological types are characterized by a stronger mass concentration and hence a flatter rotation curve in inner parts than galaxies of later types. Their simulation demonstrates that an initially concentrated galaxy model produces a flat bar like the one observed in early-type galaxies, while a less concentrated disk having a larger extent of rigid rotation produces an exponentially decreasing bar profile. Although their picture may finally turn out to be correct, their numerical models that start from completed disks lack evolutionary perspective, the importance of which was indicated in the present study. Furthermore, Sellwood (1996) has a suspicion that the truncation of the disk at a too small radius in the early-type model of Combes & Elmegreen (1993) has artificially caused the flatness of the major axis density profile of their bar. Friedli & Martinet (1993) have recently investigated numerically the dynamical evolution of galactic disks primarily in relation to bars-in-bars structures (Shlosman, Frank, & Begelman 1989). They advocate a unified scenario for the evolution of barred galaxies of various types. However, this study is also disconnected from the era of disk formation, presuming already existing massive disks as initial conditions.

The present study may be the most relevant to the work by Sellwood & Carlberg (1984), who have done numerical simulations of growing disks. Their main interest was not in bar formation but in the persistence of spiral structures in the presence of cold dissipative gas that falls onto the disk from halo regions. Based on the numerical results, they have remarked that the sequence $SA \rightarrow SAB \rightarrow SB$ is that of increasing disk formation (and gas infall) rate. This is utterly contrary to our conclusion. This discrepancy comes from the different dynamical role of the gas in their models from that in ours. In their models, the gas accreted to the disk plane acts as a coolant that reduces the effective velocity dispersion in the disk. This may seem to be quite natural at the first glance, because the gas produces young stars that inherit small random velocities from their parent gas clouds. In accordance with this idea, they added the gas to the disk as a population of particles with a low velocity dispersion in their numerical models. Therefore, a larger infall rate means more mass contained in the cold stellar disk component at a given time. This makes the fastest growing disk the most unstable to bar formation. However, as the present study suggests, the heating due to very massive gas clumps can overcome and surpass the cooling effect considered by Sellwood & Carlberg (1984).

4.6. Caveats and Future Works

We here discuss limitations of the present study and possible future extensions.

The most serious shortcoming in the present study may be the neglect of (possibly) finite lifetimes of the gas clumps formed in the galactic disk. The treatment in the present models allows the masses of the gas clumps to vary instantaneously as the gas content in the galactic disk changes with time due to the gas infall and star formation (see eq. [8]), but it does not explicitly take into account possible self-destruction of the clumps. Giant molecular clouds

(GMCs) in our Galaxy and nearby galaxies are observed to be sites of active star formation. It is generally believed in general that the internal star formation event will destroy the host cloud eventually by the energy injections from newly born stars via stellar winds and supernova explosions. Solomon et al. (1979) estimate a GMC age of $>3 \times 10^8$ yr with considerable uncertainty. There is some observational evidence for the existence of aggregates of stars and gas much larger and/or heavier than individual GMCs. "Superclouds" advocated by Elmegreen & Elmegreen (1983) have $\sim 10^7~M_{\odot}$ of H I gas and the typical dimension of 1 kpc. They argue that these superclouds are formed by the gravitational instability in a rotating magnetized interstellar gas layer. The "star complexes" found by Efremov (1979) have sizes of 200–1000 pc, age dispersions of $(2-5) \times 10^7$ yr, and masses of $\sim 10^6 M_{\odot}$. A high-resolution millimeter-wave observation has revealed many associations of GMCs in M51, each of which has a mass of (1-6) \times 10⁷ M_{\odot} . The giant molecular associations in the interarm regions in M51 seem to be gravitationally unbound, whereas those in the arm regions appear to be bound (Rand & Kulkarni 1990), suggesting lifetimes of several times 10⁸ yr. These observations and theoretical considerations seem to cast serious doubt on the longevity of massive gas clouds in general. Our treatment, which neglects possible destructions of the gas clumps, may overestimate the dynamical effects (especially the heating of the disk) of these clumps. It should, however, be noted that no detailed study has been carried out on the evolution of clouds with masses exceeding $10^8 M_{\odot}$.

Let us make a crude estimate of easiness with which a gas cloud with a mass M is destroyed by internal star formation. If we assume that the star formation efficiency does not depend on the cloud mass M and that the initial mass function of formed stars is universal, then the total energy input, $E_{\rm in}$, from young massive stars (via stellar winds and supernova explosions) is proportional to M. On the other hand, the gravitational potential energy, E_p , of the cloud will be proportional to M^2 , provided that all clouds are homologous in internal density structure. Easiness of cloud destruction by star formation events may be measured by the ratio $E_{\rm in}/E_p$, which is proportional to M^{-1} under the present assumptions. This argument, though crude, seems to suggest that a more massive cloud is more difficult to dissolve by an internal star formation event. In order to obtain a more quantitative and convincing result, we will have to resort to more direct treatments, such as hydrodynamical simulations of the evolution of star-forming gas clouds.

Another concern is about the heating of stellar disks by spiral arms (e.g., Barbanis & Woltjer 1967; Binney 1981). Neglect of this phenomenon in the present models may not be justified, especially for late-type disk models. If taken into account properly, spiral heating may raise the stellar velocity dispersion to a sufficient level to suppress bar formation by the time the disk growth is completed in a latetype model. However, this seems not to be the case. The aforementioned result by Sellwood & Carlberg (1984) is already suggestive regarding this point. Their simulations, involving spiral forcing as only one heating mechanism, indicate that a spontaneous bar starts to develop whenever the fractional disk mass becomes sufficiently large in later phases of the simulation. This indicates that the spiral heating is not sufficient to damp bar instability, though it is clearly recognizable in the time development of stellar random velocities.

The present study has divided the whole galaxy into the three components, i.e., the halo, the stellar disk, and the gas disk, and it has employed essentially a one-zone approach to describe the dynamical evolution of each component. We had to give up investigating what structures come out within the growing disks and how their internal structure changes with time. It was also impossible to obtain sufficiently quantitative conclusions because of the one-zone nature. Most physical quantities (star formation rate, velocity dispersion, and so on) have been given in averaged values over the entire region of the disk. Undoubtedly these limitations can be relaxed by a fully three-dimensional numerical simulation which includes gasdynamics and the star formation process. Such a work will fill the gap between galaxy formation from density perturbations in the early universe and the subsequent long-term evolution of galactic structures (e.g., Katz & Gunn 1991). However, the state of the art limits such a treatment only to the most global features, such as the emergence of the exponential disks (e.g., Freeman 1970) and flat rotation curves (e.g., Rubin et al. 1985).

Nevertheless, fully three-dimensional simulations seem to be the most attractive approach in the future. One possible compromise that may deserve trying is to carry out threedimensional simulations of gas and stars in already vilialized halos in a similar setting to that postulated in the present study. Unlike the simulations tracing protogalaxy collapse all the way from ~ 100 kpc to ~ 20 kpc scales, such a spatially limited treatment will provide sufficiently fine resolution in time and space to describe convincingly the likely emergence of bars and spiral structures in evolving galactic disks. It is expected that such numerical studies give a clue to several important issues that have not been touched upon in the present study. These include the difference in the length and shape of bars in galaxies of different morphological types, the preponderance of inner rings in late-type barred galaxies and outer rings in early-type barred ones, and the intimate association of lenses to earlytype barred galaxies. An observed overabundance of bars in dwarf galaxies relative to giant galaxies (van den Bergh 1982) should also be addressed. Along with such an evolutional study, we need to clarify why the tidal bars have different structure from the spontaneous bars. This question should be answered from kinematical and dynamical viewpoints. Close examination of orbit populations (e.g., Contopoulos & Papayannopoulos 1980) in both classes of bars seems to be a promising and logical way to obtain insight into this problem.

Let us turn our attention to observations. The current status of our knowledge of fundamental properties of disk galaxies is far from satisfactory when we discuss such a delicate problem as the bar instability. For example, if the estimate of the fractional disk mass is in error by 50%, then the conclusion about bar instability can be turned over. The onset of bar instability is governed largely by three parameters, namely, the mass fractions of the stellar and gaseous disks relative to the total mass (within the optical radius) and the velocity dispersion in disk stars (Shlosman & Noguchi 1993). A more precise determination of these quantities than currently attainable will be vital in making a meaningful comparison between any theoretical model and the observations. In particular, photometric and kinematic observations of nonbarred galaxies (for which we need not

be bothered by nonaxisymmetric configurations and noncircular internal motions) of various morphological types and luminosities will provide a powerful database for theoretical modeling. Although we have assumed a constant value of the disk mass fraction of 0.5 for all the models, this mass ratio may vary significantly among different morphological types and/or luminosity classes (e.g., Athanassoula, Bosma, & Papaioannou 1987). In the present picture, earlytype and late-type galaxies have different bar formation mechanisms because of the different degrees of dynamical heating in their disks. Another version of this dichotomy in bar formation mechanism is still possible. Suppose that early-type galaxies have a less massive disk (relative to the total galaxy mass) than late-type ones on the average, unlike the present assumption that the disk mass fraction is constant along the Hubble sequence. Although the observational data currently available do not seem to point to such a difference, the possibility is not completely ruled out that the sequence from early-type disk galaxies to later ones is an extension of the sequence from elliptical to S0 galaxies in the sense that the mass fraction of the disk component is increasing continuously along this sequence. Then it can happen that early-type disks tend to be more stable to bar formation simply because their disk is not massive enough, and the disks have to find external disturbance to obtain a bar. On the other hand, disks in late-type galaxies might be sufficiently massive for instability to occur. In this case, we expect the same correspondence between morphological types and bar formation mechanisms as proposed in the present study. A quantitative understanding of the global galaxy structure and kinematics for a wide range of morphological types is necessary to discriminate between the two possibilities.

5. CONCLUSIONS

Simple analytical models of growing galactic disks have been devised that take into account infall of the primordial gas from halo regions, as well as gravitational clumping and star formation in the disk gas, in an attempt to understand bar formation processes in the context of disk formation. The dominant heating mechanism of disk stars, which affects bar instability greatly in these models, is the scattering by very massive gas clumps that form in the disk due to gravitational instability in the gas disk component. The rate of gas accretion onto the disk plane (i.e., the gas infall rate from the halo) has been found to be a key parameter that determines global dynamics of the resulting disk and the available bar formation mechanism. Based on the dependence found of disk evolution on the accretion rate, a bimodal formation mechanism has been proposed concerning stellar bars in disk galaxies. The bars in slowly growing disks with a small infall rate are considered to have resulted from the bar instability of the disk component. Slowly growing disks keep their interstellar gas content low at any time. The gas clumps formed in the disk then have small masses and give a negligible kinematical disturbance to the stars formed in the disk. Thus the stellar disk component grows in mass while not greatly increasing its low velocity dispersion that was inherited from the parent interstellar gas. Once its mass reaches a critical value, the stellar disk becomes dynamically unstable and develops a bar. This case is identified with that of late-type disk galaxies, which are considered to have had a small infall rate of the protogalactic gas. On the other hand, tidal interactions are proposed as the mechanism that induces bars in early-type galaxies, the disks of which are considered to have formed relatively rapidly by a fast infall. The gas mass fraction attains a large maximum value in fast-growing disks. During this gas-rich phase, a number of massive gas clumps are formed with individual masses larger than $\sim 10^8 M_{\odot}$. These "hyperclouds," while orbiting in the disk plane, deflect stellar orbits significantly. When the disk growth is completed and most of the disk matter is turned into stars, random motions in the stellar component are already large enough to suppress bar instability. Such a hot stellar disk can generate a bar only through close encounters with other galaxies passing by.

This idea is reinforced by several independent supports, both observational and theoretical. First, the wellconfirmed morphological segregation in disk galaxy distributions implies that early-type galaxies residing predominantly in crowded environments should have had much opportunity for tidal interaction and associated bar formation, while late-type galaxies located mainly in sparse regions should have evolved relatively in isolation. Second, the observed higher incidence of thick disks in early-type galaxies than in late types means more efficient dynamical heating of the stellar disk in early types, for which the heating due to supermassive gas clumps envisaged in the present model seems to provide an attractive mechanism. Third, systematically different spatial distributions of actively characterized by far-infrared emission in the galactic disks of barred galaxies of different Hubble types may be explained by the present picture as a result of different timescales of bar growth in the two mechanisms. Finally and most importantly, numerical simulations have succeeded in producing a dichotomy in bar structure such that spontaneous bars tend to have a steep density profile along the bar major axis, whereas tidally induced bars usually exhibit a flatter profile with shoulders at the ends of the bar. This numerical result, combined with the present dual formation scenario, parallels the observed dichotomy and can explain the tendency that bars of early-type galaxies are systematically flatter than those in late-type galaxies.

APPENDIX

NUMERICAL PROCEDURES

In this Appendix, we give a full account of the numerical method used for the simulations presented in § 4.2.

The galaxy model used in the present N-body simulations is composed of a disk and a halo. The disk and the halo are constructed by 50,000 and 25,000 particles, respectively. The interstellar gaseous component is not included. Both the disk and halo have a mass of 0.5, so that the total mass of the system is unity. The initial radii of the disk and halo are unity, giving the dynamical time of unity for the system. The mutual gravitational interactions between these collisionless particles have

620 NOGUCHI

been calculated by the tree method developed by Barnes & Hut (1986). The gravitational softening radii have been taken to be 0.01 for both the disk and halo particles. The opening angle for the tree code has been set by choosing $\theta = 0.8$.

A1. SPONTANEOUS BAR SIMULATIONS

The initial condition has been set up to embody the models used by Fall & Efstathiou (1980) because this series of models has relatively realistic representation of the observed galactic structures. First, the disk particles have been distributed using a random number generator so as to give an exponential surface density profile with a scale length of 0.25. The disk has been truncated at the galactocentric radius r = 1. Then the gravitational acceleration, $v_d^2(r)/r$, caused by this disk is calculated analytically, where $v_d(r)$ denotes the circular velocity at the radius r balancing the gravity of the disk alone. Before distributing the halo particles, we specify the form of the rotation curve, v(r), following the proposed shape of Fall & Efstathiou (1980). The "turnover" radius, r_m , has been set to 0.7. This radius roughly measures the location at which the disk rotation changes from a nearly rigid rotation of the inner part to a nearly flat one of the outer part. Then the gravitational acceleration that the halo component should exert is calculated by

$$\frac{v_h^2(r)}{r} = \frac{v^2(r)}{r} - \frac{v_d^2(r)}{r} ,$$

where v_h denotes the circular velocity balancing the halo gravitational field. The halo volume density distribution, $\rho_h(r)$, is then calculated so that it produces the correct gravitational field (see Fall & Efstathiou 1980 for further details). The halo is assumed to be spherical. The halo particles are distributed using a random number generator following the calculated profile, ρ_h , and given isotropic Maxwellian random motions with a one-dimensional velocity dispersion $\sigma = [-U(r)/3]^{1/2}$, where U(r) is the gravitational potential per unit mass at the radius r due to the halo component. This specification makes the halo in rough equilibrium under its own gravity.

After setting up the initial configuration, as a preparatory step, we evolve only the halo component with the disk particles held fixed. This step is carried out to bring the halo into dynamical equilibrium in the presence of the disk gravitational field. Using the time step $\Delta t = 0.008$, the halo has been evolved for 16 dynamical times, which was found to be sufficiently long for it to attain nearly equilibrium.

Finally, we proceed to a self-consistent simulation in which both the disk and halo components are evolved under the total gravitational field. In the beginning of this stage, the velocities of the disk particles should be specified. To do this, we first calculate numerically (by the tree code) the gravitational field in the disk plane due to the combination of the disk and halo components and obtain the circular velocity, $v_c(r)$, balancing this gravity. The random motions of the disk particles in the disk plane have been specified by choosing the Q-parameter (Toomre 1964) to be 1.5 and using the epicyclic approximation. The vertical velocity dispersion was taken to be 0.7 times the radial dispersion following the observations of the solar neighborhood. The rotational velocities of the disk particles have then been calculated by applying to the v_c values the stellar hydrodynamical corrections for the random motions. All the disk particles are given the rotational and random velocities calculated like this and then set in motion along with the halo component. The structural and kinematical properties of the galaxy model at this instance are given in Figure 6.

This self-consistent run lasted for 40 dynamical times with $\Delta t = 0.02$, the result of which is depicted in Figure 4. The total energy of the system is conserved within 1.6%.

As a check of the numerical results, we have run another model in which the numbers of the disk and halo particles were reduced to 10,000 and 5000, respectively. Although the global evolution of this system in initial phases showed an appreciable difference from that of the run described in Figure 4, the properties of the bar formed are found to be essentially the same. The referee of the present paper pointed out the possibility that the halo particles experience dynamical frictions against the disk particles due to their being twice as massive as the latter, and the resultant concentration of the halo particles affects the structure of the bar formed. That the dynamical frictions are actually negligible was confirmed by running a model in which the disk and halo components have the same number of particles (i.e., 20,000 for each), and hence individual masses of the disk and the halo particles are the same. Again this model showed essentially the same bar structure as the two models described above, although the initial morphological evolution of the disk was slightly different.

A2. TIDAL BAR SIMULATIONS

The tidal bar simulation depicted in Figure 5 has been run with two stages: the isolated run and the interaction run. The first stage was carried out to produce an equilibrium dynamical model for an isolated galaxy, which is required as the initial condition for the target galaxy in the second interaction stage. The important point in the first stage is to confirm that the galaxy model is really in a steady state within a tolerable range. This stage proceeded in the same way as the spontaneous bar simulations described in \S A1. The only difference is that Q=3 instead of 1.5 in order to make the disk bar stable despite its large mass fraction of 0.5. There is no analytical solution available for the equilibrium velocity distribution functions for the Fall & Efstathiou mass models. Calculation of the velocity dispersions in the disk component and the subsequent correction of the rotational velocities for the contribution made by the random motions have been performed by using the epicyclic approximation and the stellar hydrodynamical treatment. These two treatments are considered to become less and less justified as the random motions dominate the disk dynamics, and the particle system established in this way will become increasingly deviated from the exact dynamical equilibrium. In order to circumvent this difficulty, the system has been fully evolved (i.e., both the disk and halo particles were moved) in isolation for 40 dynamical times after the halo-only evolution was finished and the disk particles were given initial velocities. This was necessary also to confirm that the disk does not develop a bar in absence of external disturbances. The total energy during this stage is conserved within 0.5%. The structural

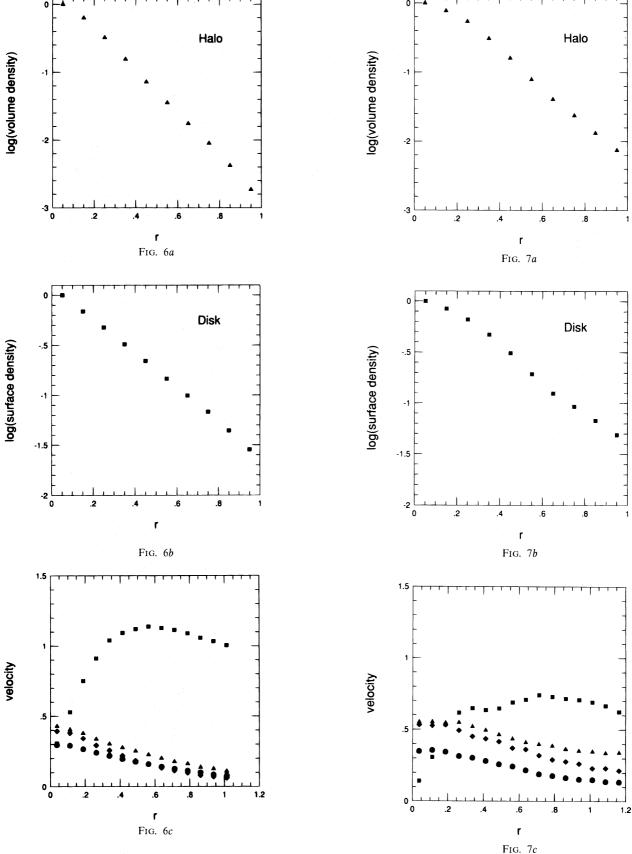


FIG. 6.—Structures and kinematics of the initial galaxy model for the spontaneous bar simulation shown in Fig. 4. (a) The volume density profile of the halo component. (b) The surface density profile of the disk component. (c) The rotational velocities (squares) and the velocity dispersions along the three directions (radial [triangles], azimuthal [diamonds], and vertical [circles]) in the disk component.

Fig. 7.—Same as Fig. 6, but for the initial galaxy model used in the tidal bar simulation shown in Fig. 5.

622 NOGUCHI

and kinematic properties of the galaxy model at the end of this isolated run are given in Figure 7. The second stage makes this isolated galaxy model have a close encounter with a point-mass perturber having the same mass as the galaxy. The galaxy and the perturber were given their initial positions and velocities relative to each other so that they make a closest passage with a perigalactic distance of 2 (i.e., twice the disk radius) at 12.4 dynamical times after the simulation starts. This condition gives the initial separation of the galaxy and the perturber of 9.6, which is thought to be large enough to avoid the undesirable effects of the sudden appearance of the perturber. The orbit was chosen to be parabolic. The interaction run has been performed for 31 dynamical times with the time step $\Delta t = 0.031$. The reliability of the numerical results has been checked by running another model employing smaller numbers of the disk and halo particles, 10,000 and 5000, respectively. The properties of the resulting bar (the length, the density profile, etc.) are found not to depend so strongly on the numbers of the used particles as to alter the conclusions stated in this paper.

REFERENCES