THE ASTROPHYSICAL JOURNAL, 514:77–95, 1999 March 20

EARLY EVOLUTION OF DISK GALAXIES: FORMATION OF BULGES IN CLUMPY YOUNG GALACTIC DISKS

MASAFUMI NOGUCHI
Astronomical Institute, Tohoku University, Aoba, Sendai 980-8578, Japan; noguchi@astroa.astr.tohoku.ac.jp
Received 1998 June 22; accepted 1998 October 26

ABSTRACT

A new idea is proposed for the origin of bulges in spiral galaxies. Numerical simulations of protogalactic collapse suggest strongly that galactic bulges have been assembled from massive clumps formed in galactic disks in their early evolutionary phase. These clumps result from the gravitational instability of the gas-rich disks of young galaxies. Owing to dynamical frictions, those massive clumps, individual masses of which can be as large as $\sim 10^9\ M_\odot$, are able to spiral toward the galactic center within a few Gyr. Inward transport of disk matter by this process leads to the formation of a galactic bulge. A simple analytical model has been constructed in which the clumpy evolution of a disk galaxy is controlled by two parameters: the timescale with which the primordial gas in the halo accretes onto the disk plane (i.e., the collapse timescale) and the initial mass fraction of the gas relative to the galaxy total mass. Under plausible assumptions for the variation of these parameters among spiral galaxies, the clumpy evolution model can explain an observed trend in which the bulge-to-disk ratio increases as the total mass or the internal density of the galaxy increases. This success suggests that the clumpy evolution of the galactic disk constitutes an important ingredient of disk galaxy evolution. Star formation in primeval disk galaxies takes place mostly in the clumps. The resulting knotty appearance of these systems may explain the peculiar morphology observed in a number of high-redshift galaxies.

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

1. INTRODUCTION

One of the most notable structural features of disk galaxies is that they are composed of two distinct components: disks and bulges. The origin of this two-component structure remains unclear, although it constitutes an important backbone of disk galaxies. Several theoretical studies have addressed this problem (e.g., Larson 1976; Gott &Thuan 1976). Conventional understanding is that galactic bulges were formed as a single body within a relatively short period as a result of the collapse of a gaseous protogalaxy and that the disk components were built up by the later accretion of residual primordial gas. The pioneering models by Larson (1976) for the formation of disk galaxies are among the most extensive numerical models that have embodied this idea, but they have the potential problem that two different regimes of star formation must be assumed in order to realize distinct separation between a bulge and a disk. Recent observations of the Milky Way bulge (e.g., Rich 1996; McWilliam & Rich 1994) and the bulges of other galaxies have stimulated attempts to elucidate when and how galactic bulges form and the nature of the bulges themselves (e.g., Matteucci & Brocato 1990; Peletier & Balcells 1996; Kormendy 1993).

On the other hand, recent observations by the *Hubble Space Telescope* (*HST*) and other new-generation ground-based telescopes are revealing properties of distant galaxies at epochs when the universe was younger than half its present age, and are thus providing direct access to the initial evolutionary phase of galaxies. Those galaxies having a redshift $z > \sim 1$ generally exhibit morphologically peculiar structures that defy application of the traditional Hubble classification scheme (e.g., van den Bergh et al. 1996; Abraham et al. 1996; Glazebrook et al. 1994; Glazebrook et al. 1995; Griffiths et al. 1994; Driver, Windhorst, & Griffiths 1995; Cowie, Hu, & Songaila 1995). The most remarkable class of these high-redshift galaxies are systems that consist of a few blobs (clumps) embedded in a common diffuse envelope. For example, Griffiths et al. (1994) note several galaxies with prominent bright knots in their sample from the *HST* Medium Deep Survey. Galaxies having similar morphology are also found in images examined by other groups (e.g., Giavalisco, Steidel, & Macceto 1996b; Koo et al. 1996; van den Bergh et al. 1996).

Some of the peculiarities observed in these high-redshift galaxies may originate in the copious interstellar gas existing in the early evolutionary phase of disk galaxies. Even with a small mass fraction, owing to its dissipativeness, the interstellar gas is sometimes manifested as a dynamically important component. A good example is provided by the spiral structures, ubiquitous in present-day disk galaxies, in which the gas mass fraction is typically no more than several percent. It is not difficult to imagine that the younger and more heavily dominated by interstellar gas the galaxy is, the more important the role the gas plays will be. In a gas-rich regime in which the gas mass fraction relative to the total galaxy mass exceeds $\sim 10\%$, the interstellar gas is very active *dynamically*. The gas, being dissipative and self-gravitating, tends to form numerous massive clumps, each of which is gravitationally bound. These massive clumps are involved with various dynamical processes (Noguchi 1998). For example, they deflect stellar motions effectively and heat up the stellar disk component (i.e., increase the stellar velocity dispersion), thus affecting the stability of the galactic disk (Shlosman & Noguchi 1993; Noguchi 1996). Noguchi (1998) has proposed a gravitational instability origin for the subgalactic clumps frequently observed in high-redshift galaxies and has argued that they may serve as building blocks of galactic bulges. The present paper expands upon this idea and tries to discuss the evolution of young disk galaxies in more detail, with much emphasis placed on the formation of galactic bulges and dependence of its efficiency on galaxy parameters.

Sections 2 and 3 describe the numerical simulation that has inspired this new idea of bulge formation. In § 4 analytical modeling is devised for the growth of galactic disks and is applied to bulge formation. Observational data about global properties of spiral galaxies are summarized in § 5 and compared with the theoretical results in § 6. Discussion and conclusions are given in §§ 7 and 8, respectively.

2. NUMERICAL SIMULATION

Collapse of a protogalaxy composed of the dark matter and primordial gas has been simulated by an N-body method including a star formation algorithm. Similar approaches have been taken by many researchers (e.g., Baron & White 1987; Katz & Gunn 1991; Katz 1992; Steinmetz & Müller 1995; Navarro & White 1994). Unlike most of these works, which employed initial conditions based on a specific type of cosmology such as the cold dark matter model, the present work does not assume a particular cosmology. I start from an idealized initial condition and attempt to predict properties of young disk galaxies that will be robust to the nature of the assumed cosmology. The primary interest here is in investigating the detailed morphological and dynamical evolution of forming disk galaxies. Therefore, the simulation is restricted to a three-dimensional volume that encloses a single galaxy. The factor by which a galaxy has collapsed after turn-around (i.e., separation from Hubble expansion) is considered to be large, a factor of 10 (e.g., Fall & Efstathiou 1980). The present study does not deal with the whole process of this collapse but only with the evolution of the galaxy after it has shrunk to the present size and has become balanced by increased centrifugal force in the direction perpendicular to the rotation axis. Therefore, the halo (the dark matter) mentioned hereafter means only the portion contained in the optical radius of the galaxy. This limitation in space makes it possible to investigate more detailed structure and kinematics of the forming galaxy than is possible with other cosmologically implemented simulations.

The primordial gas that was destined to make the visible parts of the present-day galaxies is considered to have been more or less clumpy (e.g., Fall & Rees 1985). The size and mass of these clumps is a matter of debate. Though recent progress in observational technique is enabling access to these pregalactic entities in the form of quasar absorption-line systems, for example, the physical properties deduced from those observations are diverse. For example, Steidel (1990) estimates the typical size of 1–15 kpc and the mass of 10^6–10^9 $M_\odot$ for the Lyman limit systems, which are considered to constitute halos of galaxies located on the lines of sight to more distant quasars. Lack of exact knowledge about pregalactic material hampers a reliable treatment of the formation process of galaxies in a numerical study.

The formation process of elliptical galaxies could be much different, depending on the assumed properties of individual pregalactic units. Stellar dynamical processes are considered to have nonnegligible effects on the formation and evolution of elliptical galaxies, and because of their collisionless nature, purely stellar dynamical processes retain memory of initial conditions to some extent. On the other hand, evolution of disk galaxies is less likely to be affected by initial conditions. The observed thinness of the galactic disks suggests that the assembling of pregalactic units into galactic disks has been a highly dissipative process. Merger of nondissipative stellar systems would puff up the disk (e.g., Tóth & Ostriker 1992). In other words, pregalactic bodies, whatever their sizes and masses may have been, should have been primarily gaseous when they coagulated to make a galactic disk. The dissipative nature of these bodies must have erased the identity of the individual bodies quickly. Thus, all the disk galaxies, in their early evolutionary phase, should have had a relatively smooth and mostly gaseous galactic disk with roughly the same radius and thickness as the present-day stellar disk.

2.1. *Initial Conditions*

As a device to realize the dissipative growth of a galactic disk, we consider a spherical protogalaxy consisting of dark matter and primordial gas. These two components have uniform density distributions with the same cut-off radius, $R = 15$ kpc. The total mass of the system is $M = 1.5 \times 10^{11}$ $M_\odot$, with the masses of the dark matter and gas components being $M_h = 0.75 \times 10^{11}$ $M_\odot$ and $M_g = 0.75 \times 10^{11}$ $M_\odot$, respectively. With this choice, the dynamical timescale of the system, $(R^3/GM)^{1/2} = 7.06 \times 10^7$ yr, where G is the gravitational constant. The unit velocity becomes $(GM/R)^{1/2} = 208$ km s^{-1}. The dark matter and gas components are modeled as systems comprising $N_h = 20{,}000$ and $N_g = 30{,}000$ particles, respectively. The adopted gas fraction, $M_g/M = 0.5$, is based on the observations that the mass fraction of the dark halo component *within the optical radius* clusters around 0.5 for most disk galaxies for which the rotation curve and the distribution of the luminous component are observed with high accuracy (van der Kruit & Searle 1982; Bahcall & Casertano 1985; but see also the discussion in § 4.6). There is no stellar particle initially ($N_s = 0$).

The dark matter particles are given isotropic random motions with a one-dimensional velocity dispersion of $\sigma = 122$ km s^{-1}, bringing the dark halo roughly into virial equilibrium. The gas particles are given a uniform rotation with an angular frequency of $\Omega = 13.9$ km s^{-1} kpc^{-1}. Therefore, the gas system is already in rough centrifugal equilibrium, owing to enough rotational support, and is expected to not collapse much in the radial direction but mostly along the spin axis. The direction of the angular momentum vector for this rotational motion is taken as the z-axis hereafter, with the x- and y-axes lying in the plane perpendicular to the z-axis. The evolution of the system under its self-gravity is simulated by an N-body method.

The dark matter particles are treated as collisionless. Namely, their motions are determined solely by the global gravitational field they feel. On the other hand, gas particles are modeled as particles that collide with each other inelastically, thus dissipating their kinetic energy. Star formation process from the gas is also included, as described shortly. The gravitational force exerted on each particle is calculated by the GRAPE, which is a dedicated device for N-body calculations (Ebisuzaki et al. 1993). The softening radii, $\varepsilon_h = 300$ pc and $\varepsilon_g = 150$ pc, are introduced for dark matter and gas particles, respectively, to suppress undesirable two-body effects.

2.2. *Gas Dynamics and Star Formation*

Gas dynamics is treated by the so-called sticky particle method. Inelastic collisions between the gas particles are introduced to model the dissipative nature of the interstellar medium (Levinson & Roberts 1981; Roberts & Hausman 1984; Hausman & Roberts 1984). All the cloud particles are assumed to have the same finite radius $r_c = 38$ pc. In the simulation, two overlapping clouds are made to collide inelastically, provided that they are approaching each other. After collision, the radial component of the mutual velocity is multiplied by $f_{\rm col} = 0.01$, and its sign is reversed, while the tangential component is unchanged, in order to mimic energy dissipation. The evolution of the system does not change appreciably as far as $f_{\rm col} < \sim 0.5$.

The star formation process is simulated by changing a gas cloud particle to a stellar particle. This conversion is performed with a probability that is related to the local gas density around the cloud particle as follows. The local gas density, ρ, for a given cloud particle is determined by counting the number of the gas clouds residing in the sphere of a radius $r_{\rm dens} = 750$ pc centered on the cloud in question. Then the probability, p, for star formation for this cloud in the current time step is calculated by $p = K_{\rm star}\, dt\, \rho^{1/2}$. Here $K_{\rm star}$ is a coefficient that controls the star formation efficiency, and dt is the size of the time step, which is 7.06×10^5 yr (i.e., one hundredth of the dynamical time). The value of $K_{\rm star}$ is determined empirically so that the time variation of star formation rate is in a rough agreement with the one inferred for disk galaxies from various observational data (see § 3). The square root dependence on the local gas density expressed in the above expression for p is specified by the consideration that the star formation process, though its detailed nature is not well understood, originates in some form of local gravitational instability in the interstellar medium and hence its characteristic timescale must be related to the free-fall timescale of the local unstable region.

After getting p by the above equation, one number ξ is created in the range (0, 1) using the uniform random number generator. If $\xi < p$, that cloud particle is changed into a stellar one. If $\xi > p$, no such transformation is made. This recipe for star formation, namely, converting a whole cloud into a stellar particle with some probability instead of converting some fraction (p in this case) of one cloud at every time step, was taken to avoid intolerable increase of the number of particles in the model as new stars form. This procedure keeps the sum of the number of gas particles, N_g, and that of stellar particles, N_s, constant during the course of simulation. One important addition to this recipe is specification of a threshold density for star formation. Star formation is inhibited if the local gas density is less than a critical value $\rho_{\rm th}$ (see below). Stellar particles are treated as collisionless, with a gravitational softening parameter, $\varepsilon_s =$ 150 pc.

The threshold density was set to be $\rho_{\rm th} = 0.1$ $(M_\odot)$ pc^{-3}. Observations of nearby disk galaxies suggest that the threshold *surface* gas density, $\mu_{\rm th}$, for star formation is $\sim$(1–10) $M_\odot$ pc^{-2} (Kennicutt 1989). Because the gas in the present model occupies a three-dimensional volume in general and is not necessarily confined in a planar configuration, the application of a surface density appears questionable and a threshold *volume* density will be more relevant. Using the observed thickness, $h_g \sim 100$ pc, of the distribution of the interstellar molecular gas in the disk of the Milky Way galaxy (e.g., Solomon, Sanders, & Scoville 1979), I converted the threshold surface density to the threshold volume density with $\rho_{\rm th} \sim \mu_{\rm th}/h_g \sim (0.01 - 0.1) M_\odot$ pc^{-3}. The value adopted for $\rho_{\rm th}$ in the present simulation is the upper limit of this range, though such a translation from the two-dimensional density to a three-dimensional one may not be fully justified because the configuration of the interstellar matter could affect its stability and ability to form stars.

3. NUMERICAL RESULTS

Morphological evolution of this model is shown in Figure 1. As a galactic disk builds up by infall of the gas from the halo region, several massive clumps are formed in the disk plane ($\sim 0.6 < t < \sim 0.7$, especially Fig. 1*c*). These clumps follow the global rotation of the galactic disk. Some of these clumps are observed to merge and make larger clumps while orbiting in the disk, as clearly seen in Figure 1*c*. Because these clumps suffer from dynamical friction against surrounding disk material, they gradually sink to the central region of the disk. This transfer of matter leads to the accumulation of a large mass in the central region. The side view of the system is quite intriguing (Fig. 1*d*). It is noted that this inward motion of clumps leads to the formation of a spheroidal component in the galactic center. It is natural to regard this component as a galactic bulge. Almost all the star formation takes place in the disk plane, so stars younger than $\sim 10^7$ yr are distributed in a very thin disk. The bulge component is thus not made by early star formation occurring in a spheroidal space at the galactic center but is assembled gradually as individual clumps formed in the disk plane spiral into the galactic center.

The clumpiness of the system makes definition of the galactic center not a trivial problem. In the analyses described below, the galactic center is defined to be the center of the most massive clump at each epoch, which is called the bulge. The galactic center thus defined can be displaced from the coordinate origin, $x = y = z = 0$. The mean disk plane, defined as the peak position in the z-distribution of the gas clouds, also drifts slightly in the negative z-direction during the simulation. These deviations from the coordinate system are taken into account in the analysis when appropriate.

Specification of a threshold gas density for star formation is definitely responsible for the lack of star formation before the epoch of disk formation. Volume density of the gas is very low when the gas occupies a three-dimensional spherical volume of the halo region, but it increases by a large amount when the gas is gathered into a two-dimensional disk configuration. This change of the dimensionality initiates vigorous star formation in the galactic disk. On the other hand, the column density of the gas hardly changes, because the collapse occurs mostly in the vertical direction.

Clumpiness observed in the numerical model is a direct consequence of the gas-richness of the disk, which forms by a relatively rapid infall. Time evolution of the masses of the total disk and the gas disk is shown in Figure 2. Here the gas particles whose distance from the mean disk plane is less than 0.1 R are considered to constitute the gas disk, and all the stellar particles are regarded as members of the stellar disk. The regions with the cylindrical radius $r < 0.3R$ or $r > 1.5\,R$ are excluded. Initial evolution is the growth of a mostly gaseous galactic disk along the line $f_g = f_d$. After the disk formation has been almost completed, star formation starts to deplete the gas component, with the model point going down nearly along the line, $f_d =$ constant, in the left panel. A slight decrease in f_d is due to inflow of the clumps

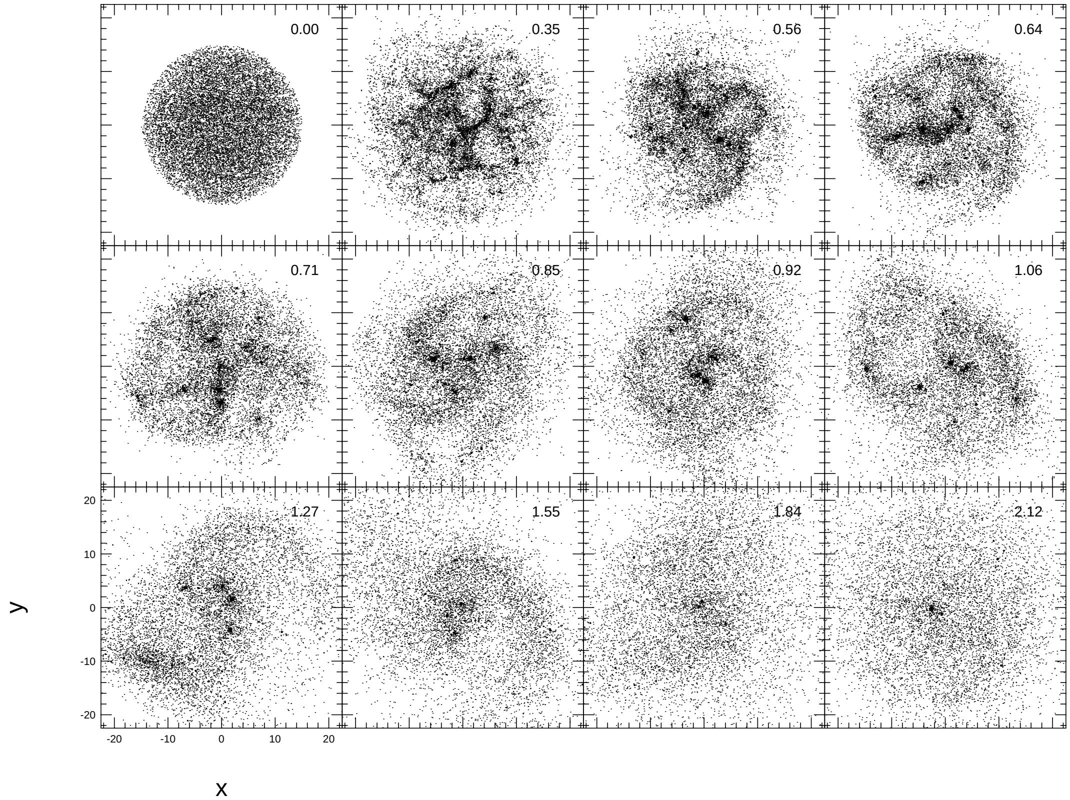

FIG. 1*a*

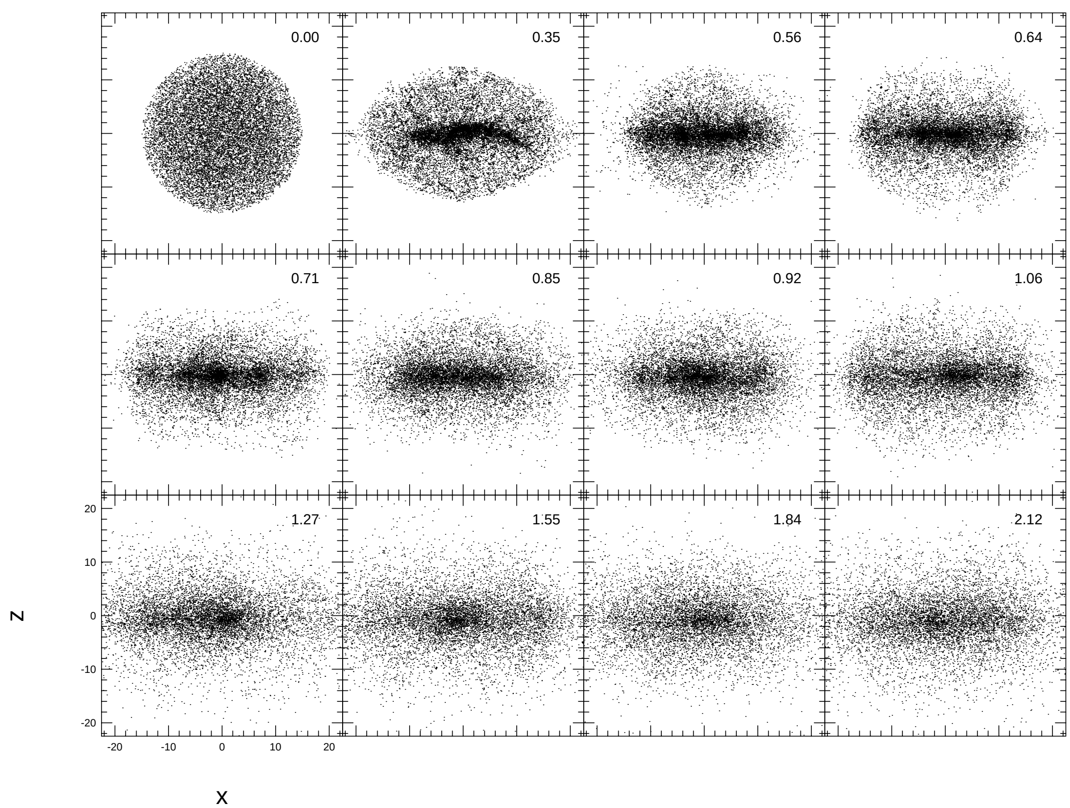

FIG. 1*b*

FIG. 1.—Morphological evolution of the numerical model. (*a*) Projection of the gas cloud particles onto the *x*-*y* plane (i.e., the disk plane). Only half the particles selected at random are displayed. Time, *t*, in units of Gyr is indicated in the upper right corner of each frame. Coordinates are given in units of kpc. (*b*) The same as (*a*), but the projection is onto the *x*-*z* plane. The *x*-*y* and *x*-*z* projections of all stellar particles older than 10^7 yr are given in (*c*) and (*d*), respectively. In panel (*c*), the five massive clumps used for subsequent analysis are indicated at $t = 0.64$ Gyr, enclosed by circles.

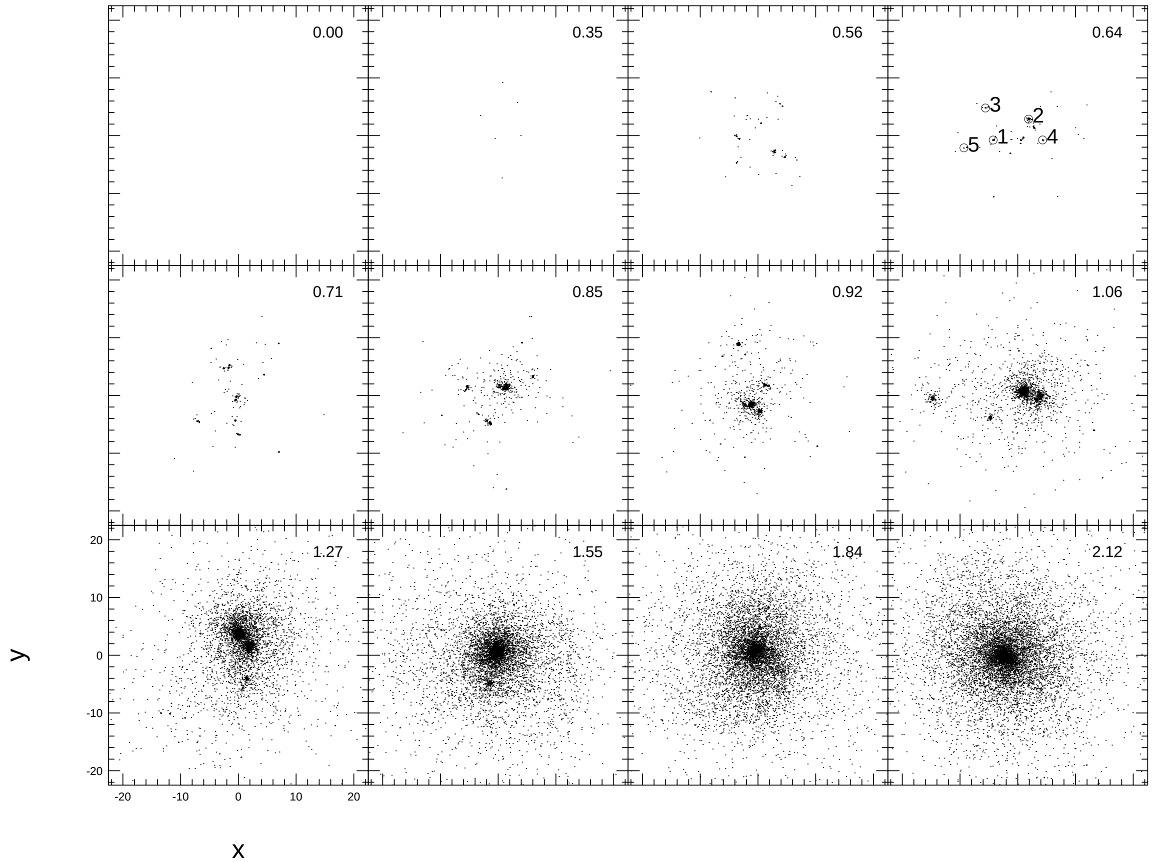
0.00
0.35
0.56
0.64
3
2
5
1
4
0.71
0.85
0.92
1.06
1.27
1.55
1.84
2.12
20
10
0
-10
-20
y
-20
-10
0
10
20
x

FIG. 1*c*

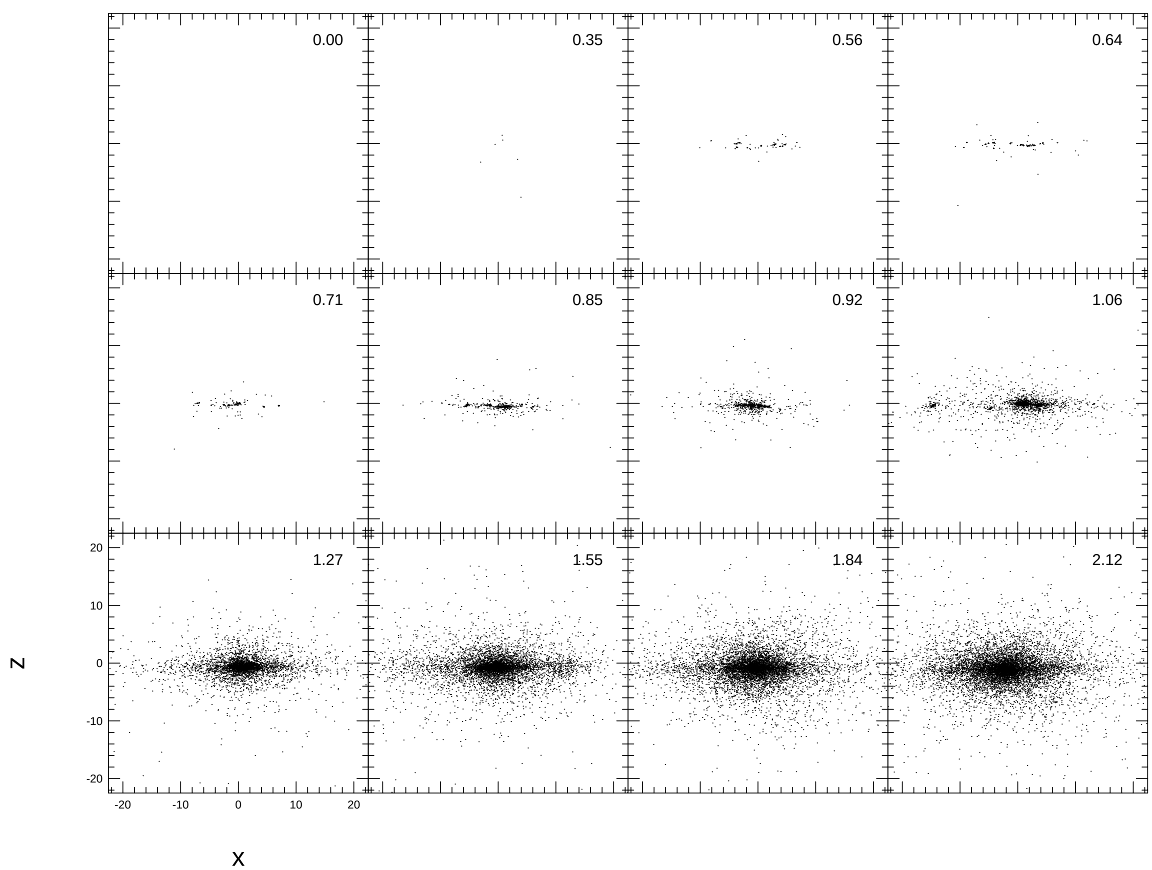
0.00
0.35
0.56
0.64
0.71
0.85
0.92
1.06
1.27
1.55
1.84
2.12
20
10
0
-10
-20
z
-20
-10
0
10
20
x

FIG. 1*d*

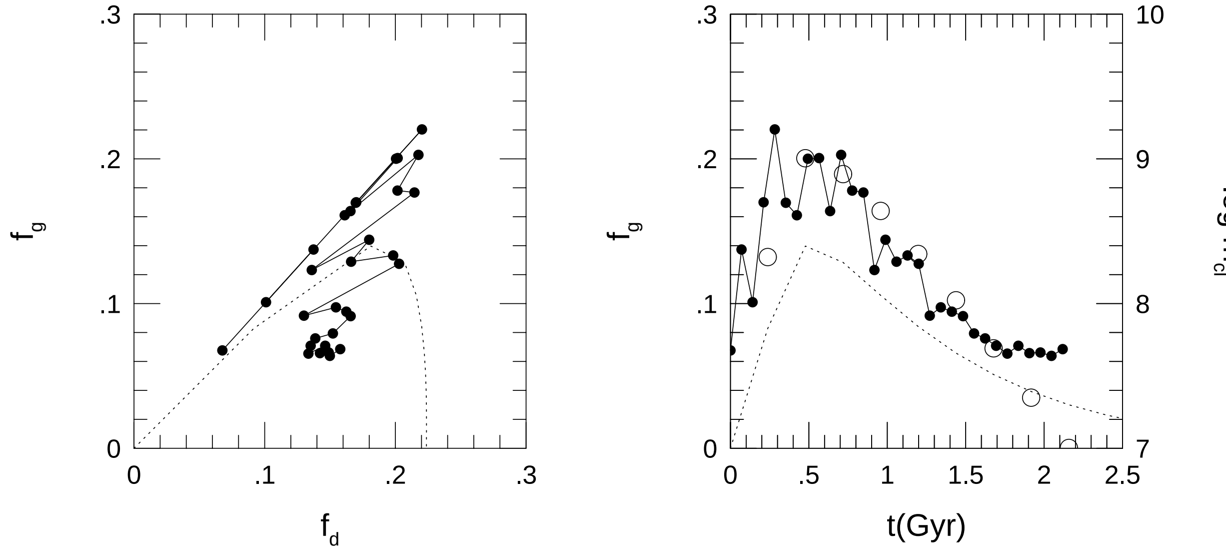

FIG. 2.—Mass fractions of the gaseous and total disks with respect to the total mass in the numerical model (*solid lines*). The fraction of the gaseous disk, f_g, is a summation over the gas particles with $|z - z_c| < 0.1R$ and $0.3R < r < 1.5R$, where z_c denotes the z-coordinate of the mean disk plane at each time and r is the cylindrical radius measured from the z-axis. The mass fraction of the total disk, f_d, is the sum of f_g and the mass fraction of the stellar disk. The latter is summed over all the stellar particles with $0.3R < r < 1.5R$. The left panel indicates the evolutionary track on the f_d-f_g diagram, whereas the time evolution of f_g is shown in the right panel. Time, t, is given in units of Gyr. The dashed lines indicate the evolution of the analytical multicomponent model explained in § 4. The analytical model plotted here has the same mass and radius as the numerical model. The initial fraction of the gas is also the same as in the numerical model (i.e., $\Gamma = 0.5$), and the collapse timescale was set to $\beta = 0.3$ Gyr. The dynamical friction timescale given in eq. (5) of the text was multiplied by 0.3 in this model. Open circles in the right panel indicate the expected mass, m_{cl}, of the individual clumps (in units of $M_\odot$) in the analytical model, which is evaluated as explained in § 4.3.

toward the galactic center, which builds up the bulge component.

The clumps in this numerical model start to appear distinctly at $t \sim 0.6$ Gyr, when the value of f_g is around its maximum of 0.19. I selected a few typical clumps at $t = 0.64$ Gyr, as shown in Figure 1*c* and investigated their internal structure. Examination of the surface density profile made it possible to draw a boundary between the main body of the clump and its envelope and to calculate the mass of the main body as the clump mass. Masses of individual clumps are found to be of the order of 0.01 M ($\sim 10^9$ $M_\odot$). This value far exceeds those of any molecular complexes observed in nearby galaxies (i.e., the most massive known entities that constitute the galactic disks in the present epoch) and can be well compared to the masses of dwarf galaxies.

The dashed line in Figure 2 indicates the evolutionary track of an analytical model that is described later. This analytical model, with a collapse timescale of 0.3 Gyr and an initial fraction of the gas mass of 0.5, represents the behavior of the numerical model fairly well. I calculated the expected clump mass as the instantaneous Jeans mass in the growing model disk of this analytical model (*open circles* in Fig. 2, *right panel*). The clump mass thus obtained reaches the maximum value of $\sim 10^9$ $M_\odot$ at $t \sim 0.5$ Gyr when f_g (*dashed line*) reaches the maximum of ~ 0.14, which is roughly equal to the maximum of f_g in the numerical model. The agreement between the numerical and analytical clump masses suggests that the seeds of those clumps in the numerical model are created by local gravitational instability in the gaseous disk component. We should, however, be careful in relating the clump mass to the Jeans mass in the disk at current epoch, because masses of individual clumps in the numerical model increase continuously in general because of mergers among themselves and later accretion of the surrounding gas, whereas the analytically calculated clump mass starts to decrease as f_g decreases.

Figure 3 shows the time development of the star formation rate (SFR). The SFR starts to increase abruptly ~ 0.5 Gyr after the start of the simulation and attains a peak rate of ~ 40 $M_\odot$ yr^{-1} at $t \sim 0.7$ Gyr. This is just the epoch when most of the gas has settled to the disk (see Fig. 1*b*). The SFR declines monotonically after the maximum except for a few weak but distinct burstlike increases. By $t = 2.1$ Gyr, 37%

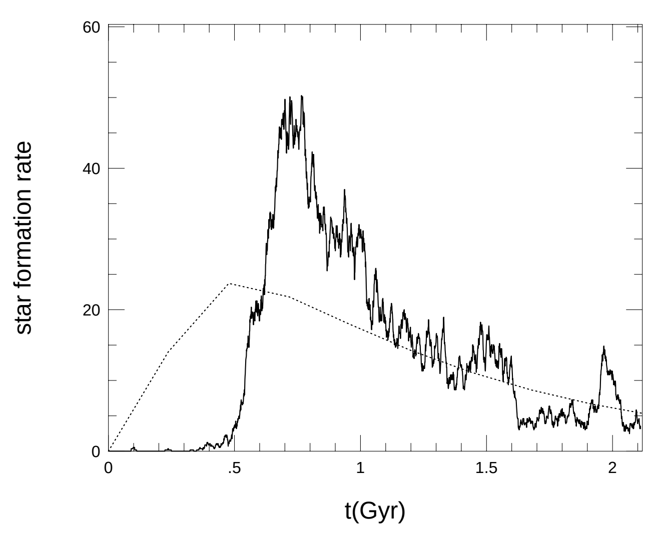

FIG. 3.—Time evolution of the star formation rate in the numerical model (*solid line*). Time, t, and the star formation rate are given in units of Gyr and $M_\odot$ yr^{-1}, respectively. The dashed line indicates the SFR in the analytical model plotted in Figure 2.

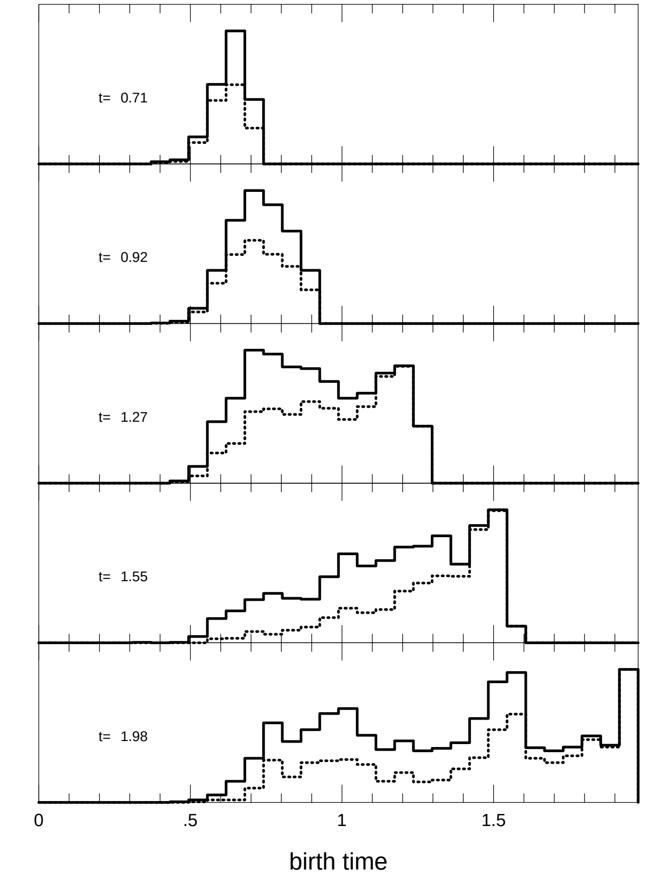

FIG. 4.—Distribution of the age for the stars contained in the bulge region in the numerical model. The center of the bugle is defined as that of the most massive clump. Each histogram shows, for the given epoch, the relative number of the stars that were born at the time indicated in the abscissa. Solid lines are distributions for stars located within $0.1R$ of the bulge center, whereas the dotted lines are for stars located within $0.025R$. Time is in units of Gyr.

of the gas has been converted into stars. The peak SFR of $\sim 40\ M_{\odot}\ \mathrm{yr}^{-1}$ is within the range of 4–75 $M_{\odot}\ \mathrm{yr}^{-1}$ deduced by Steidel et al. (1996b) for a population of $z > 3$ galaxies. The temporal change of the SFR in the numerical model agrees also with the history of star formation in early-type disk galaxies inferred from the observations (Sandage 1986). Note also that the overall behavior of the SFR is similar to that of the SFR obtained in the analytical model.

It is noted that the SFR reaches the maximum roughly at the time when the clumpy structure is most prominent (Fig. 3). This means that during the course of evolution, the model galaxy will be observed as a clumpy object at the highest probability, provided the observation is made in the rest-frame *UV* band, which is sensitive to young massive stars. Also note the global asymmetry of the stellar component in the present model, which is most noticeable at $t = 0.9$ Gyr to 1.3 Gyr (see Fig. 1*c*).

The asymmetric and clumpy disk seen in this model is strongly reminiscent of the images of many galaxies at large redshift obtained by the *HST* (e.g., van den Bergh et al. 1996; Griffiths et al. 1994; Koo et al. 1996). Although the irregular and clumpy nature is most noticeably found in the objects with a medium redshift (i.e., $\sim 0.5 < z < \sim 1$), it may be traced into a larger redshift. Steidel et al. (1996a) investigated the morphology of Lyman break galaxies with $2.3 < z < 3.4$ found in the Hubble deep field and noticed some objects with multicomponent structure. They also mention diffuse asymmetric "halos" around these high-redshift galaxies. In view of the present numerical result, some parts of these morphological peculiarities may be caused by gravitational instability of the galactic disks in their early evolution phases.

The clumps discussed above experience strong dynamical friction owing to their large masses. Resulting accumulation of clumps to the galactic center makes a bulge, as already stated. The growth of the bulge is stepwise, because accretion of clumps takes place in a discrete manner. Figure 4 shows the distribution of the age for the bulge stars at several epochs. The discrete nature of the bulge growth is manifested as a few local maxima in the age distribution (see especially the bottom panel), which correspond respectively to starbursts triggered by the falling-in of a massive clump into the bulge region. At the final epoch indicated ($t = 1.98$ Gyr), the age spread is as large as ~ 1 Gyr.

4. ANALYTICAL MULTICOMPONENT MODELS

The numerical model discussed above is enlightening, but it has a difficulty. As shown below, one of the most important parameters that governs the evolution of disk galaxies is the accretion timescale, namely, the timescale of gas infall from the halo. It is difficult to vary this timescale freely by employing the sticky particle method. Accretion can be accelerated or slowed down in the sticky particle model by varying the size of the clouds and/or the restitution coefficient f_{col} in the cloud collisions. However, it is not fully clear how these parameters are related to the observed properties of the interstellar gas. For example, the giant molecular clouds of the Milky Way have a very complicated internal structure comprising many hierarchical levels in spatial scale, and no single spatial scale seems to dominate (e.g., Henriksen 1991). Also, the accretion of gas to the disk plane is likely to be affected by other effects than dissipative cloud-cloud collisions as discussed below. These considerations have motivated development of analytical models that introduce the accretion timescale as a free parameter. In this section, I discuss the analytical model of disk growth to remedy the limitation possessed by the numerical simulation. The analytical model developed here has essentially the same ingredients as the previous version described in Noguchi (1996), but it is improved in several respects.

4.1. *Basic Equations*

I consider a spherical halo with a radius R. The primordial gas initially distributed in this halo region gradually accretes to the disk plane and builds up a galactic disk. As the accretion proceeds, the disk comes to contain more and more gas (i.e., the interstellar gas) and stars form from it. I assume that the stars and the gas in the disk occupy a flat cylindrical region with the same radius as the halo. The whole system considered here is meant to represent the portion of a disk galaxy within its optical radius, so the halo in the present model should not be regarded as representing the entire dark halo (the massive halo) but only the portion within the optical extent of the galaxy, in accordance with the numerical model presented above. The system is divided into four components: the halo, the stellar disk, the gas disk, and the bulge. The mass of each component is denoted by m_h, m_s, m_g, and m_b, respectively. Total mass is denoted by M ($= m_h + m_s + m_g + m_b$). I am not concerned with any structures that might develop in each component and treat each component as a single zone. Namely, the physical state of each component at a given time is specified by several global quantities whose value should be understood as a characteristic one, namely the average over the entire region of that component.

Under this simplification, time evolution of masses in respective components is formulated as follows:

$$\frac{dm_g}{dt} = -\mathrm{SFR} - \frac{m_g}{\tau_{fri}} + \frac{M\Gamma t}{\beta^2} \exp\left(-\frac{t}{\beta}\right), \qquad (1)$$

$$\frac{dm_s}{dt} = \mathrm{SFR}, \qquad (2)$$

and

$$\frac{dm_b}{dt} = \frac{m_g}{\tau_{fri}}, \qquad (3)$$

where t is the time reckoned from the beginning of gas infall (which is here assumed to be 12 Gyr ago). The first term in the right-hand side of equations (1) and (2) denotes the effect of star formation, and SFR stands for the star formation rate for the whole galaxy; τ_{fri} is the timescale of the inward motion of clumps formed in the disk. This inflow is caused by dynamical frictions acting upon the clumps, and its timescale determines the growth rate of the bulge component. The bulge growth is determined by equation (3) in this analytical formulation. The third term in the right-hand side of equation (1) represents addition of gas to the disk by the accretion (gas infall) from outside the disk plane, the timescale of which is denoted by β. Hereafter, β is referred to interchangeably as the accretion timescale, the collapse timescale, or the infall timescale. As the infall proceeds, the mass of the halo decreases correspondingly, because the total mass of the system, M, is assumed to be conserved (here the halo serves only as a reservoir of the primordial gas and its evolution is not traced). The total mass of the matter that eventually accretes to the disk is specified by its fraction, Γ, with respect to the total mass of the galaxy. Namely, Γ is the fraction of the gas mass relative to the total galaxy mass at the initial instant. The functional form of the time variation of the infall rate in equation (1) was adopted because it is physically reasonable and mathematically convenient. The time dependence here is slightly modified from the previous one adopted in Noguchi (1996).

4.2. *Star Formation Rate*

The fundamental process of star formation in the interstellar medium is not yet well known. Phenomenologically, the star formation rate is often represented as a power of the amount of interstellar gas (Schmidt law). The present study follows the result by Kennicutt (1989), who has found from a compilation of data for a number of nearby disk galaxies that the star formation rate per unit area of the galactic disk is proportional to Nth power of the surface density of the gas, which includes both molecular and atomic hydrogen. In the present model, this relation is represented as

$$\mathrm{SFR} = \alpha \Sigma_g^N R^2, \qquad (4)$$

where Σ_g denotes the gas surface density in the disk and is approximated by $m_g/(\pi R^2)$ in the present multicomponent treatment, and α is a coefficient that determines the absolute value of the star formation rate. The range of the power in the above Schmidt law is $N = 1.3 \pm 0.3$ according to Kennicutt (1989). It was found that setting $\alpha = 355\ M_\odot\ \mathrm{yr}^{-1}$ produces a range of star formation rate compatible with the observational inference when the mass, m_g, and the radius, R, are expressed in units of $10^{11}\ M_\odot$ and 10 kpc, respectively. Therefore α is hereafter fixed to $355\ M_\odot\ \mathrm{yr}^{-1}$ and $N = 1$. Adoption of an universal value of α for all the models is based on the inference that the fundamental star formation process is the same in all the disk galaxies.

The work by Kennicutt (1989) suggests that a certain threshold exists for gas density, below which star formation is effectively inhibited. One plausible interpretation is that the star formation activity is associated closely with the gravitational instability of the gas disk and the threshold corresponds to neutral stability. How is the threshold density determined? One possibility is that galactic disks have an intrinsic lower limit of velocity dispersion in the gas cloud motion and that a disk with too small a surface density is stabilized by the random motion of the interstellar gas clouds. Indeed, H I and CO observations of nearby galaxies and the Milky Way suggest near constancy among different galaxies of the velocity dispersion in the atomic and molecular gas clouds (e.g., van der Kruit & Shostak 1984; Stark & Brand 1989). Therefore, I introduced a threshold for star formation by specifying a minimum velocity dispersion that the gas component of the galactic disk can take.

The star formation threshold introduced in the present model works as follows. As explained shortly, the gas disk is assumed to stay at the marginal stability defined by $Q = 1$, where Q denotes the stability parameter introduced by Toomre (1964). At this state, the surface density of the gas and the velocity dispersion (or the sound velocity), σ, in the gas are related by $\pi G \Sigma_g = \sigma\kappa$, where κ is the epicyclic frequency. As Σ_g decreases in late phase of evolution, σ also decreases correspondingly. However, σ cannot become less than σ_{min}, which is the specified lower limit of the velocity dispersion. Therefore, star formation stops at the instant when σ reaches σ_{min}. The threshold gas surface density is then determined by $\Sigma_{\mathrm{min}} = \sigma_{\mathrm{min}}\kappa/\pi G$. After this instant, the amount of the gas consumed by star formation is balanced with the amount of the gas added to the disk by accretion, so that the gas surface density is always kept nearly at the threshold value (i.e., $\Sigma_g \sim \Sigma_{\mathrm{min}}$). As a special case, setting $\sigma_{\mathrm{min}} = 0$ allows star formation to proceed continuously, depending on the current gas density.

4.3. *Dynamical Friction Timescale*

Noguchi (1996) applied the classical Chandrasekhar formula in evaluating τ_{fri}. It is, however, doubtful that this formula, which assumes a homogeneous and infinite background of randomly moving light particles, can be safely applied to a heavy body orbiting in a highly flattened and systematically rotating disk component. Quinn & Goodman (1986) have extensively discussed the sinking process of satellite galaxies through galactic disks and suggest several analytical evaluations of the orbital decay timescale.

Instead of applying a certain analytical formula, I have resorted to a more empirical approach in the evaluation of τ_{fri}. I have run a number of numerical simulations for a massive rigid body (hereafter the "clump") orbiting in the disk plane of a galaxy model, varying the clump mass, the disk surface density, the velocity dispersion of disk stars, and the shape of the rotation curve (see Appendix for numerical details). Because all the models show a remarkably similar time variation of the galactocentric distance of the clump when the time axis is suitably adjusted, I decided to define $\tau_{\rm fri}$ as the time it takes the clump to move from the initial radius, $r = 0.8$, to the final radius, $r = 0.2$.

Measurement of $\tau_{\rm fri}$ in these experiments is summarized in Table 1. It is evident that the mass of the clump and the surface density of the disk are the key parameters that determine the dynamical friction timescale. Dependence of $\tau_{\rm fri}$ on other parameters is weak and can be neglected for the practical range of interest. This is fortunate, because little is known in this analytical treatment about the time variation of the disk velocity dispersion or the shape of the rotation curve. Examination of the parameter dependence has led to the following empirical formula for $\tau_{\rm fri}$.

$$\frac{\tau_{\rm fri}}{\tau_{\rm dyn}} = 0.25\left(\frac{m_{\rm cl}}{M}\right)^{-0.5}\left(\frac{\Sigma}{M/R^2}\right)^{-0.67} . \quad (5)$$

Here $\tau_{\rm dyn}[\equiv(GM/R^3)^{-1/2}]$ is the dynamical time of the galaxy, $m_{\rm cl}$ is the mass of the clump, and Σ is the surface density of the total disk, which includes both gas and stars, i.e., $\Sigma = (m_g + m_s)/(\pi R^2)$.

In calculating the clump mass, it is assumed that the gas disk is always maintained in a marginally unstable state with $Q = 1$. This assumption finds justification as follows. If $Q > 1$ at a certain moment, no star formation is expected to occur, because the gas disk is stable gravitationally to small-scale perturbations. Then Q decreases, because heat input from massive stars is lacking while the gas radiates its energy. When the decreasing Q cuts the value of unity and becomes slightly smaller than unity, the instability sets in and stars begin to form. These stars provide the gas with energy through supernova explosions and stellar winds. Then Q is elevated above unity again. Thus the value of Q will oscillate around $Q = 1$, and $Q = 1$ is a good approximation of the dynamical state of the gas disk.

In the state of marginal instability (or stability), the mass of the clump formed is given by

$$m_{\rm cl} = \pi(0.5\lambda_c)^2\Sigma_g = \frac{\pi^5\Sigma_g^3}{\kappa^4} , \quad (6)$$

where the critical wavelength $\lambda_c = 2\pi^2\Sigma_g/\kappa^2$ and the epicyclic frequency is approximated as $\kappa = (2M/R^3)^{0.5}$. Clump formation is related closely to star formation in the present model. To be consistent with introduction of a star formation threshold, clump formation is inhibited when $\Sigma_g < \Sigma_{\rm min}$. The instantaneous value of $\tau_{\rm fri}$ is determined by combining equations (5) and (6). Thus the bulge mass, m_b, calculated through equation (3), may be uncertain by a considerable amount because of many simplifications made in the present formulation. However, relative comparison of m_b between different models will be meaningful, and I focus on the qualitative behavior of the bulge growth in what follows.

Now, equations (1), (2), and (3), combined with auxiliary equations (4), (5), and (6), complete a set of equations that determine the temporal evolution of the system. By integrating these equations numerically starting from the initial condition, $m_g = m_s = m_b = 0$, we know time evolution of any physical quantities of interest.

4.4. *Important Parameters: β and Γ*

Evolution of a galactic disk in the present formulation is determined mostly by two parameters, the accretion timescale, β, and the initial mass fraction of the gas, Γ. Before dealing with more complicated cases, I here discuss what effect these parameters have on the disk evolution. Two model series are considered in order to isolate the effect of varying each parameter. In series A, only the value of β is changed with all other model parameters being equal, while all the models in series B have the same parameters but Γ. The mass and the radius of the galaxy are fixed to $M = 10^{11}$ $M_\odot$ and $R = 10$ kpc in both series.

Figure 5 shows the time evolution of the models in series A. It is seen that as β decreases, the peak values of the gas mass fraction (m_g/M), the star formation rate, and the clump mass become larger and are attained at a progressively earlier epoch. On the other hand, the present-day values of the gas mass fraction, the star formation rate, and the clump mass are larger for a model with slower accretion. The bulge formation starts and finishes at an earlier epoch as β decreases. A smaller β also leads to a larger bulge. The mass of the stellar disk at the present epoch is smaller for a smaller β, because a larger fraction of the accreted material

TABLE 1

TIMESCALE OF DYNAMICAL FRICTION-INDUCED SINKING

Model	m_d	$m_{\rm cl}$	r_m	Q	$\tau_{\rm fri}$ [a]
1	0.3	0.01	0.3	1.5	5.41
2	0.1	0.01	0.3	1.5	11.4
3	0.2	0.01	0.3	1.5	7.32
4	0.3	0.03	0.3	1.5	3.78
5	0.3	0.003	0.3	1.5	11.1
6[b]	0.3	0.001	0.3	1.5	21.7
7	0.3	0.01	0.1	1.5	5.31
8	0.3	0.01	0.7	1.5	6.32
9	0.3	0.01	0.3	1.0	5.29
10	0.3	0.01	0.3	2.0	5.96
11[b]	0.3	0.01	0.3	1.5	4.75
12[c]	0.3	0.01	0.3	1.5	4.92

NOTE.—The effective radius of the clump, r_s, is 0.07 unless specified.

[a] The timescale, $\tau_{\rm fri}$, is the time it takes the clump to move from $r = 0.8$ to $r = 0.2$, and its unit is the dynamical time of the galaxy.

[b] Effective radius $r_s = 0.2$.

[c] Effective radius $r_s = 0.4$.

goes to the bulge component in this case, owing to more efficient inflow of matter caused by more massive gas clumps. Thus the models in series A show qualitatively different time evolution depending upon the value of β. It should be noted that the bulge growth caused by the inflow is heavily reduced or stopped when the disk becomes mostly stellar.

What about the effect of varying Γ? All the models in series B, in which Γ is varied from 0.1 to 0.5 and β is fixed to 2 Gyr, have turned out to exhibit a qualitatively similar time evolution. Each quantity changes with time in a similar way, and no remarkable effect of varying Γ is observed. Effect of varying Γ manifests only as the difference in the absolute value. It was found that the values of m_s, m_g, and SFR are roughly proportional to the mass fraction, Γ. On the other hand, the normalized clump mass, m_{cl}/Γ, and the normalized bulge mass, m_b/Γ, show a difference of ~ 10 among the calculated models. Thus these quantities behave in a highly nonlinear way with respect to Γ. This strong nonlinearity leads to a large difference in the ratio of the bulge to the disk, or the B/T.

Real galaxies have a large range in both mass and size, so that the two model series discussed above (which assume fixed M and R) are highly idealized. It is conceivable that two galaxies differing in their mass or size have largely different collapse time β or mass fraction Γ. Likely dependence of β and Γ on the galaxy property should be taken into account in order to construct more realistic models.

4.5. *Collapse Timescale: β*

Infall of the gas from outside the disk plane is a natural consequence of galaxy formation from extended halos. It has been also introduced into chemical evolution models to reproduce the metallicity distribution in the solar neighborhood and the age-metallicity relation (e.g., Lacey & Fall 1985).

It is very difficult to deduce quantitatively the timescale of the protogalaxy collapse from observational data. However, several correlations observed among spiral galaxy properties, especially colors and gas contents, provide circumstantial evidence that this timescale, β, varies

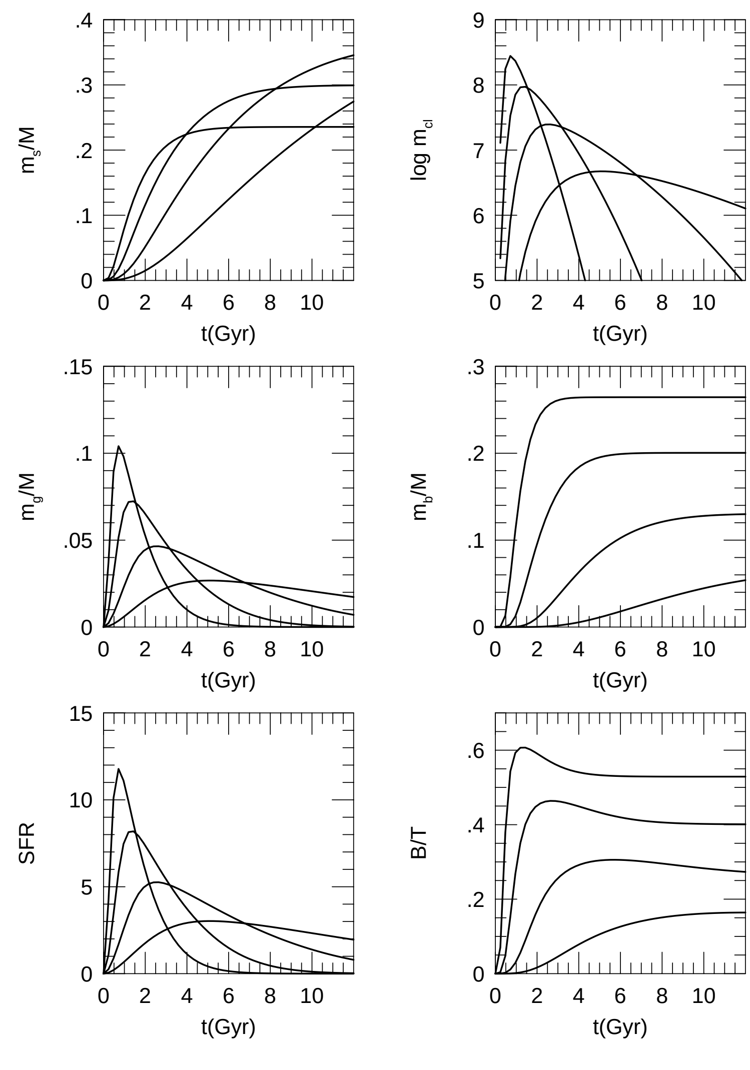

FIG. 5.—Time evolution of the models in series A. All the models have the same mass, $M = 10^{11}\, M_\odot$, and the same radius, $R = 10$ kpc. Only the accretion timescale, β, has been varied. From the most rapidly increasing curve, $\beta = 0.5$, 1.08, 2.32, and 5.0 Gyr in all the panels. All the models have $\Gamma = 0.5$. The dynamical friction timescale given in eq. (5) of the text was multiplied by 0.3 in all the models. The star formation rate, SFR, and the mass of the clump, m_{cl}, are in units of $M_\odot\ \mathrm{yr}^{-1}$ and $M_\odot$, respectively. The mass ratio of the bulge to the total luminous matter, B/T, is defined by $m_b/(m_b + m_s)$.

from galaxy to galaxy in a systematic way. Late-type disk galaxies have a larger mass fraction of the interstellar medium relative to the galaxy's total mass (e.g., Young 1990; Casoli et al. 1998) and bluer total colors (e.g., de Vaucouleurs 1974; Gavazzi 1993) than early-type spirals. In view of the results for the model series A discussed above, these characteristics indicate that the stellar population in late-type galaxies is relatively younger, suggesting a slower buildup of their disks. Although variation along the Hubble sequence appears substantial, dependence on the galaxy luminosity seems to be much larger. The well-known color-magnitude relation (e.g., Tully, Mould, & Aaronson 1882) states that the galaxy becomes bluer as it becomes fainter. Recent analysis by Gavazzi (1993) has found that about two-thirds of the total variation in spiral galaxy color is due to luminosity difference and only one-third is contributed by the dependence on the Hubble morphological type. The galaxy luminosity also has a large influence on the gas content. Gavazzi's (1993) plots for separate morphological classes show that at a fixed morphological type the relative gas content increases by ~ 10 as the galaxy luminosity decreases by 4–5 magnitudes. Actually, the dependence on the luminosity seems to be stronger than the morphological-type dependence, being consistent with large scatter of the gas mass fraction at a fixed morphological type seen in Young (1990).

Systematic change of the collapse timescale along the Hubble type and the galaxy luminosity is also plausible from a theoretical point of view. A spiral galaxy of an earlier morphology tends to have a higher density of matter within its optical extent at the same luminosity, as suggested by its larger rotational velocity (e.g., Rubin et al. 1985). It is likely that the collapse timescale is significantly governed by the radiative cooling of the primordial gas in the protogalaxy, and if this is the case, a higher density should have led to a more rapid collapse owing to more efficient cooling. On the other hand, a smaller galaxy may have been more strongly affected by the energy injection from internal star formation process than a more massive galaxy. Then the collapse of a small galaxy must have been prolonged considerably by the feedback from the initial star formation (Such feedback mechanisms would have caused mass loss from the system, or even to disintegration of the system in the case of sufficiently intense star formation, as expected in dwarf ellipticals; see, e.g., Yoshii & Arimoto 1987.)

These considerations have led to a specification of the collapse timescale as a decreasing function of both the galaxy mass and the internal density. The present model assumes a generalized parameterization of this dependence as follows:

$$\beta = c\left(\frac{M}{10^{11}\ M_\odot}\right)^a \left(\frac{\rho}{0.1\ M_\odot\ \mathrm{pc}^{-3}}\right)^b \ (\mathrm{Gyr}) \ . \qquad (7)$$

Here the density of the galaxy is defined by $\rho = M/R^3$, and c is the coefficient to determine the absolute value of the collapse timescale. The power indices, a and b, determine the steepness of the dependence, and a larger value adopted for each results in a larger range in the collapse time. From the discussion above, both indices are likely to be negative.

A simple consideration suggests that the range in β should be smaller than a factor of 10 for the whole population of disk galaxies. First of all, a collapse time smaller than a few times 10^8 yr will lead to formation of an elliptical galaxy rather than a spiral, because such a rapid collapse within a few dynamical times of the system will initiate fast star formation that consumes most of the gas before the system reaches a centrifugal equilibrium. A plausible upper limit to the collapse timescale comes from the fact that most disk galaxies appear to have already finished collapse to the disk plane by the present epoch. If the collapse is still continuing, we should see a considerable amount of gas at a large distance from the galactic plane, which should be emitting X-ray radiation corresponding to the virial temperature of the galaxy. *ROSAT* observations of nearby spiral galaxies (Read, Ponman, & Strickland 1997) indicate that the amount of such hot gas outside the galactic plane is less than $\sim 10^9\ M_\odot$, which is less than 10% of the galaxy total mass for S0-Sc galaxies in their sample. This observation, though yet to be extended to a larger sample, seems to preclude a collapse time that is a large fraction of the age of the universe.

To summarize, I considered the two cases: (1) $a = -\frac{1}{2}$, $b = 0$, and $c = 2.0$; and (2) $a = -\frac{1}{3}$, $b = -\frac{1}{3}$, and $c = 2.5$. A steeper power of a and a slightly smaller value of c in case (1) have been taken to render the range of β in the considered (ρ, M) domain nearly the same in both cases. As an additional constraint, an upper limit of 5 Gyr is imposed upon β. Actually, the value of β is likely to depend also on the position in the galaxy (e.g., Matteucci & François 1989; Lacey & Fall 1985; Larson 1976). The value that is used in the present multicomponent modeling should be regarded as an averaged value characteristic of the whole galaxy.

4.6. *Initial Gas Fraction:* Γ

Another important parameter, Γ, i.e., the fraction of mass that eventually accretes to the galactic plane, is not well constrained from observations either. It seems reasonable to assume that Γ is equal to the combined mass fraction (relative to the total galaxy mass) of the bulge and the disk, including the interstellar medium, because there is no firm evidence that at the present epoch a significant amount of matter resides in the halo component except dark matter, as stated above. Under this assumption, reliable determination of Γ is hampered. One ambiguity arises from our poor knowledge about the mass-to-luminosity ratio for the luminous components. This ratio depends strongly on the formation history (i.e., the time variation of the star formation rate) and the initial mass function of the stellar population considered, which are generally difficult to deduce. On the other hand, the total mass, M, seems to be well determined from the rotation curve, at least for nonbarred galaxies, for which the circular rotation provides a fairly exact description of the disk kinematics.

Comparing the empirical relation between the mass-to-luminosity ratio, M/L_B, and the $B-V$ color of spirals with the theoretical one predicted by Larson & Tinsley's (1978) photometric evolution model, Tinsley (1981) found that the observed increase of M/L_B with the $B-V$ is much shallower than that theoretically predicted, suggesting that late-type spirals have relatively more dark matter than early-type ones. Thus, Tinsley (1981) has claimed that the halo mass fraction is the dominant parameter controlling the morphological type. Athanassoula et al. (1987) also find that the dominance of the dark matter is larger in bluer galaxies. However, Jablonka & Arimoto (1992) have recently concluded that the halo mass ratio is universal among spiral galaxies (at least from Sa to Sc), based on a detailed population synthesis analysis treating the bulge and the disk separately. Also, decomposition models for a selected sample of spirals having well-measured rotation curves seem to show no systematic variation of this ratio along the Hubble sequence (e.g., van der Kruit & Searle 1982; Bahcall & Casertano 1985).

Persic & Salucci (1988), on the other hand, conclude that the ratio of the dark mass to the luminous mass within the optical radius increases as the galaxy becomes less luminous. According to the result by Ashman (1990), which is based on Persic & Salucci (1988), a galaxy with a mass of 10^{10} $M_\odot$, for example, has a mass fraction of dark matter twice as large on average as that of the most massive galaxies, although the dispersion at a fixed mass is considerably large (see Fig. 2 of Ashman 1990).

In view of these arguments, I considered the following three cases. (1) $\Gamma = 0.12 \log M - 1.00$; (2) $\Gamma = 0.64 \log M + 0.64 \log \rho - 0.297$; and (3) $\Gamma = 0.50$. It should be noted here that the mass fraction, Γ, does not have a direct relationship with the cosmological baryon fraction. It is hard to imagine that the latter quantity is spatially variable. However, the ratio, Γ, is the mass fraction relative to the total mass *within the present optical size of the galaxy* and can be changeable during galaxy formation processes. For example, the difference in the spin parameter, λ, of the protogalactic halo may have led to different Γ, so that the gas in a halo with a smaller λ has collapsed radially by a larger factor relative to the halo and realized a larger Γ (e.g., Dalcanton, Spergel, & Summers 1997). Segregation of the gas and dark matter in the protogalaxy and resulting transfer of the angular momentum from the gas to the dark matter in protogalactic mergers are another potential mechanism to cause a variety in Γ. Note that the average matter density within the optical radius is also affected in general by the processes changing Γ.

4.7. *Model Families*

I have calculated two families of models that differ in the star formation process. In one family, star formation is allowed to take place in dependence on gas density by taking $\sigma_{\min} = 0$. These models are called continuous models. A finite threshold, $\sigma_{\min} = 3$ km s^{-1}, was specified for the other family (hereafter called threshold models). Each calculated model is specified by three parameters as (c or t)-(β type)-(Γ type), where "c" and "t" denote continuous and threshold specification for star formation, respectively, and types for β and Γ denote the corresponding specifications given in §§ 4.5 and 4.6, respectively. For example, t-1–3 means a model in which β is a function only of the mass, Γ is constant, and the star formation threshold is introduced. All possible combinations of the three parameters have been computed, leading to 12 models in total. In each model, the galaxy evolution has been calculated at 8×8 grid points on the (ρ, M) plane, which are equally spaced both in $\log \rho$ and in $\log M$. The ranges of mass and density are 10^{10} $M_\odot \le M \le 3 \times 10^{12}$ $M_\odot$ and 6×10^{-3} $M_\odot$ pc$^{-3} \le \rho \le 1$ $M_\odot$ pc^{-3}, respectively.

5. OBSERVATIONAL MATERIAL

In the present study, I confine comparison between models and observations to the "classical" spirals, i.e., to the morphological types from Sa to Sc. The observational data, especially those for the bulge-to-disk luminosity ratio, are well accumulated only for this range of morphology.

5.1. *Bulge-to-Disk Ratio*

Several studies have tried to measure the luminosity ratio of the bulge and disk components by the decomposition technique (e.g., Yoshizawa & Wakamatsu 1975; Kent 1985; Simien & de Vaucouleurs 1986). Figure 6 shows the observed bulge-to-total luminosity ratio as a function of galaxy mass and density for the sample used by Whitmore (1984), which is itself based on the observations of Rubin, Ford, & Thonnard (1980) and Rubin et al. (1982). I calculated the mass and density for the sample galaxies from the optical radius (i.e., the radius at which the surface brightness in B-band is 25 mag arcsec^{-2}), R_{25}, and the rotational velocity, V_{25}, at the optical radius as

$$M = G R_{25} V_{25}^2$$

and

$$\rho = M R_{25}^{-3} ,$$

where G is the gravitational constant. Both R_{25} and V_{25} have been taken from Rubin et al. (1980, 1982). The mass thus calculated will give an approximately correct value for the optical part of the galaxy, but the density calculated here should be regarded as a rough characteristic value because of the steep density gradient inside the galaxy.

It is recognized in Figure 6 that, on the average, the B/T increases as the galaxy mass increases at a fixed galaxy density, and at a fixed galaxy mass it increases with the density. A statistical analysis carried out for this sample also confirms this correlation. However, it is difficult to represent the bulge-to-total luminosity ratio as a function of one single parameter. The principal component analysis carried out by Whitmore (1984) indicates that the surface brightness of the galaxy is only one physical quantity (among the parameters he investigated) that shows significant correlation with the B/T. Indeed, the lines of constant surface mass density plotted in Figure 8 run roughly in parallel with loci of constant B/T, although the irregularity in the distribution of the B/T is large. Another compilation by Dale et al. (1997) also supports the dependence of the B/T on the galaxy mass and the galaxy density displayed in Figure 6. Their data show that a more massive or denser galaxy tends to have an earlier morphological type and that the regions occupied by galaxies with the same Hubble morphological type are elongated in directions similar to the constant surface density lines.

The observed values of the B/T should be taken with caution. The measurement of the bulge-to-disk luminosity ratio is a very delicate task. Usually a light distribution model composed of a de Vaucouleurs' bulge (i.e., the $r^{1/4}$ law) and an exponential disk is fitted to the observed surface brightness profile. However, this method has been criticized recently (e.g., Andredakis & Sanders 1994; de Jong 1996) on the ground that the de Vaucouleurs' density distribution does not decrease sufficiently fast with the radius and thus the fitting is significantly influenced by the irregular light distribution sometimes observed at larger radii. Andredakis & Sanders (1994) propose an exponential form for both the bulge and the disk for better decomposition. The result by de Jong (1996) using this method shows a clear trend in which the B/T decreases as the morphological type becomes later both in the B and K bands, as expected. The

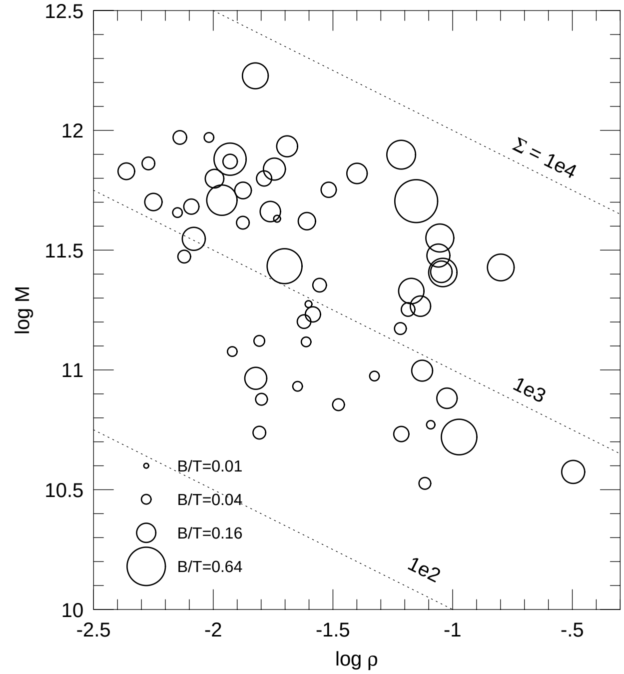

FIG. 6.—Ratios of bulge to total luminosity for a sample of spiral galaxies with morphological type of Sa to Sc, taken from Whitmore (1984). The ordinate indicates the mass, M, of the galaxy inside the optical radius in units of $M_\odot$, whereas the abscissa indicates the density, ρ, within the optical radius in units of $M_\odot$ pc^{-3}. The area of each circle is proportional to the B/T ratio for the corresponding galaxy. The dashed lines specify the constancy of the surface density defined by $\Sigma \equiv M/R^2$, where the mass, M, is in units of $M_\odot$, and the optical radius, R, is in units of parsec.

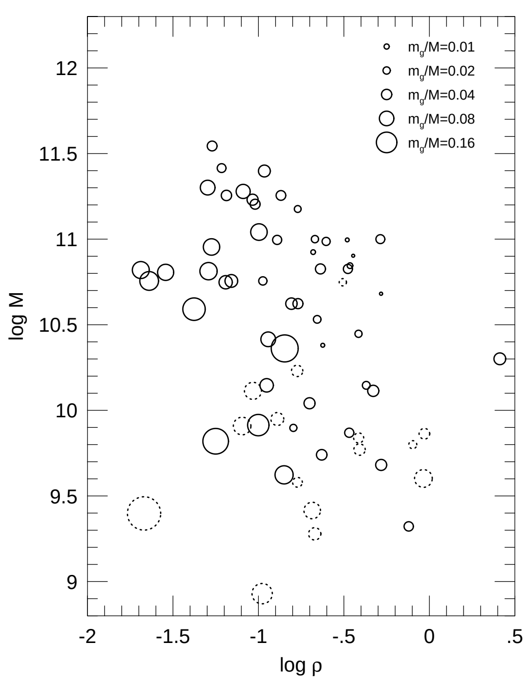

FIG. 7.—Mass fraction of the interstellar gas relative to the total matter inside the optical radius of the galaxy. The galaxy sample is taken from Sage (1993). The ordinate indicates the mass, M, of the galaxy inside the optical radius in units of $M_\odot$, whereas the abscissa indicates the density, ρ, within the optical radius in units of $M_\odot$ pc^{-3}. Here the interstellar gas means the sum of the molecular gas and the atomic gas. The area of the circle is proportional to the gas mass fraction. The dotted circles indicate the upper limits in molecular mass.

scatter around the mean relation is large, however. Also the measured B/T with an exponential bulge is systematically smaller than the one obtained by the de Vaucouleurs' model by a factor of 3–5. It should also be noted that any decomposition technique works best when the contributions from the bulge and the disk are comparable. On the other hand, detection of a faint disk in the presence of a luminous bulge or a small bulge embedded in a bright disk is difficult.

5.2. *Epoch of Bulge Formation*

When and how rapidly the galactic bulges formed is one of the most important but unsolved problems regarding disk galaxies. For example, the age of the Milky Way bulge and the age-spread in the bulge stars have not yet been tied down with sufficient accuracy (e.g., Matteucci & Brocato 1990; Holtzman et al. 1993; Rich 1996; Norris 1996). Although quantitative assessment is difficult, it is possible that the age of the bulge is different from galaxy to galaxy. Metallicity observations by Jablonka, Martin, & Aritomo (1996) show that the value [Mg/Fe] decreases as the bulge becomes fainter. This may indicate that the smaller bulges possessed in general by galaxies of the later morphological types (see Fig. 1 of Jablonka et al. 1996) have been formed over a more extended period, though a definite answer must await accurate models of chemical evolution.

Peletier & Balcells (1996) have recently found that the optical and near-infrared colors of the bulge are very similar to those of the disk in a number of spiral galaxies of type S0 to Sbc; A blue bulge is likely to be associated with a blue disk. One possible interpretation is that the age of the bulge is correlated with that of the disk and the age difference is a small portion of the age of the universe. At the same time, their data seem to suggest a large difference in the bulge age among different galaxies. Fitting of single age, single-metallicity stellar population models by Vazdekis et al. (1996) to the $U-R$ and $R-K$ colors of the observed bulges suggests that a number of bulges can have ages as young as 4 Gyr and a few bulges may have even younger ages of ~ 1 Gyr. Nevertheless, reliable age determination for old stellar populations is difficult because of degeneracy of their colors.

5.3. *Gas Content and Star Formation Rate*

Although the primary aim of the present study is to understand the formation of galactic bulges, any successful model of galaxy evolution should be able to explain the observed trend in the gas content and the star formation activity among galaxies of different type and luminosity.

Young (1990) shows that the mass ratio of gas including both neutral and molecular hydrogens relative to the dynamical mass calculated from the rotation curve decreases systematically as the morphological type becomes earlier. The gas fraction given by Young (1990) ranges from 0.03 for Sa-Sab galaxies to 0.3 for Scd galaxies. However, the scatter within the same morphological type is as large as ~ 10. The ratio of molecular to neutral hydrogens decreases as the Hubble type becomes later. Gavazzi (1993) provides extensive data for H I content in disk galaxies. His data show that H I flux per unit H flux increases as the morphological type becomes later, as expected. At a fixed type, H I/H is an increasing function of the H-band luminosity of the galaxy, indicating that less massive galaxies are more gas-rich on the average. The sample of Sage (1993) has been used to produce Figure 7. One sees in this figure a global trend in which the gas mass ratio increases as the galaxy mass and/or density decreases, though the scatter in the ratio is considerably large.

The largest set of star formation rate measurements is given by Kennicutt (1983), who calculated the SFR for a number of spiral galaxies from H_α luminosity. Analysis of his sample has revealed that the normalized star formation rate, i.e., SFR/M, increases as the galaxy mass decreases. There appears to be no clear trend from Kennicutt's (1983)

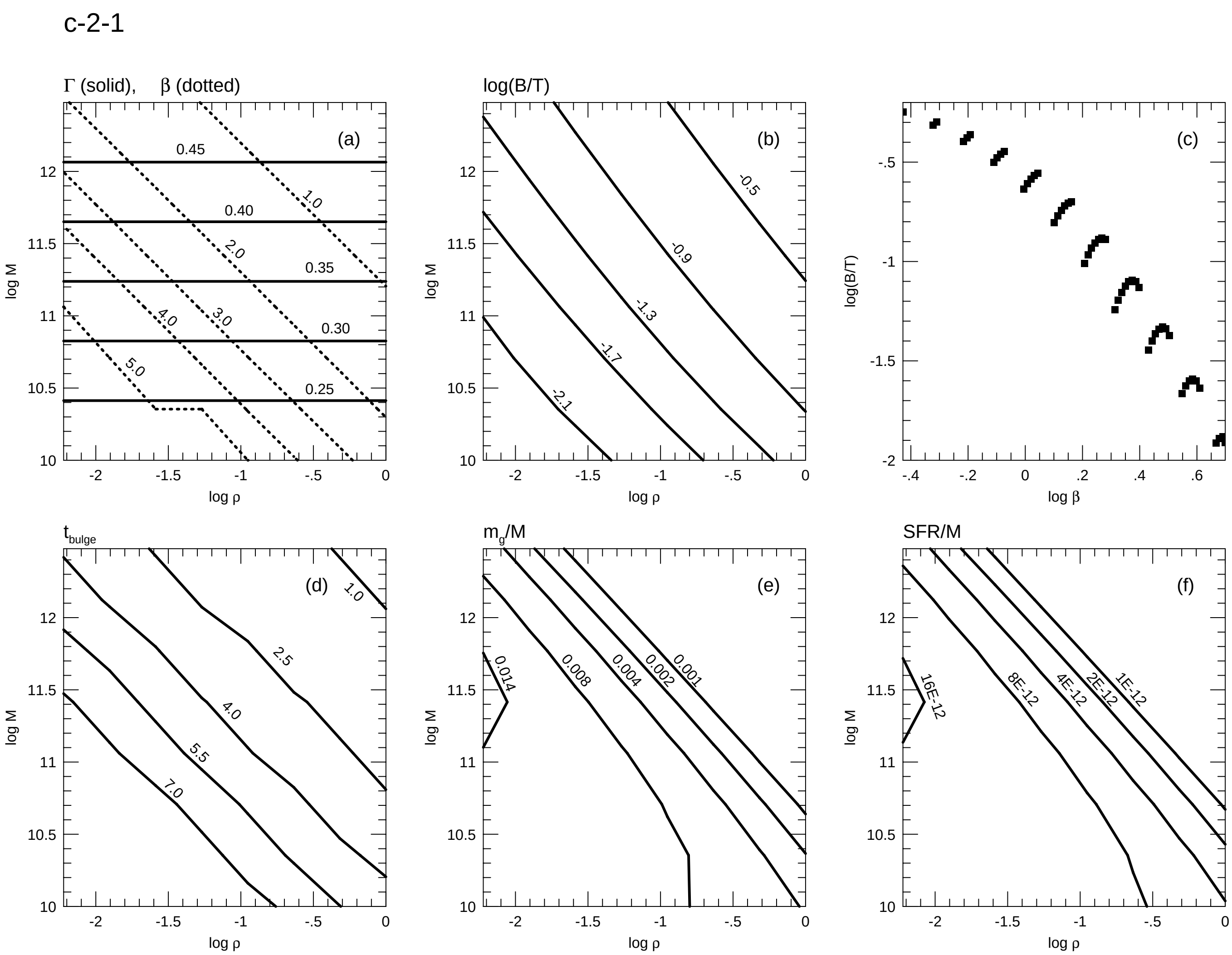

FIG. 8.—Results for the analytical model c-2–1. (*a*) The accretion timescale, β, and the initial mass fraction of the gas, Γ. The ordinate indicates the mass, M, of the galaxy inside the optical radius in units of $M_\odot$, whereas the abscissa indicates the density, ρ, within the optical radius in units of $M_\odot$ pc^{-3}. (*b*) The ratio of the bulge mass to the total luminous mass at the present epoch, defined by $B/T = m_b/(m_b + m_s)$, where m_b and m_s are masses of the bulge and the stellar disk, respectively. (*c*) The ratio of the bulge mass to the luminous mass, B/T, plotted against the accretion timescale, β (in Gyr). (*d*) The epoch of bulge formation, $t_{\rm bulge}$, which is defined to be the epoch where the bulge mass reached half of the present value. (*e*) The ratio of the gas mass to the total galaxy mass, m_g/M, at the present epoch. (*f*) The ratio of the star formation rate to the total galaxy mass, SFR/M, at the present epoch.

data that the normalized SFR increases for less dense (hence correspondingly later on average) galaxies. This seems to be at odds with the behavior of the gas fraction plotted in Figure 7, which suggests a stronger dependence of the gas content on the galaxy density than on the galaxy mass. This may be due partly to heavy obscuration of H_α emission by dust, which should be more severe in late-type spirals. These somewhat puzzling results indicate a necessity for improved and more extensive observational data.

6. MODEL RESULTS

As stated in § 4.7, twelve model families have been calculated. Results for only one representative family, c-2-1, are presented here, because the other families all show qualitatively similar behaviors in many respects. This family has a collapse timescale, β, depending on both the mass and the density of the galaxy, while the initial gas fraction, Γ, is dependent only on the mass (Fig. 8*a*). Comparison is carried out using the (ρ, M) plane. In fitting the models to the observations the ρ-axis and/or the M-axis were allowed to slide by a small amount. Because the present analytical model treats only *characteristic* values of the galaxy, there is some ambiguity about galaxy mass and density. In particular, the density is a poorly defined quantity because of its steep variation with the galactocentric radius. Furthermore, even if the galaxy density is defined as an average density within the optical radius (typically R_{25}), a discrepancy of this radius between different authors (because of different Hubble constants adopted, for example) would be much exaggerated in the evaluation of the density, through the inverse cubic dependence of the density on the radius. These ambiguities should be kept in mind whenever the models are compared with the observations.

6.1. *Bulge-to-Disk Ratio*

Figure 8*b* plots the mass ratio of the bulge relative to the total luminous matter, $B/T \equiv m_b/(m_b + m_s)$, for the family, c-2–1. Comparison of these values with the observed B/T luminosity ratios should be done carefully. The difference in the mass-to-luminosity ratio in bulges and disks will not totally justify equating the mass ratio to the luminosity ratio. Also, the observed luminosity ratio depends on the decomposition technique used (§ 5.1). Thus I pay attention primarily to qualitative behavior of the B/T ratio. The calculated B/T increases as the density and/or the mass of the galaxy increases, showing behavior qualitatively the same as the observation suggests. It is encouraging that the range in the B/T plotted in Figure 8*b* is roughly equal to the observed range given by de Jong's (1996) K-band photometry for galaxies from Sa to Sc. The other 11 models also exhibit variation of the bulge-to-disk ratio as a function of galaxy mass and density, which is qualitatively consistent with the observation.

Figure 8*c* plots the B/T against β. Although the c-2–1 family shows a tight correlation between the two quantities, this is not always the case. Modulation by the galaxy density or the mass fraction, Γ, is not neglected in general. Therefore, the conclusion by Noguchi (1998) that the bulge-to-disk ratio is determined primarily by the collapse timescale (i.e., the disk formation timescale) is oversimplified. The previous conclusion was derived from more limited calculations in which both the galaxy mass and the galaxy density are fixed and only the collapse timescale was varied (such as series A in § 4.4), and it may be erroneous. Nevertheless, it is true that the accretion timescale has a strong influence on the resulting bulge-to-disk ratio.

6.2. *Formation Epoch of Bulges*

Figure 8*d* indicates the epoch of bulge formation, $t_{\rm bulge}$. Here $t_{\rm bulge}$ is defined as the epoch at which half the final bulge mass has accumulated. The bulge formation epoch in all the calculated models becomes earlier as the mass and/or the density of the galaxy increases. The B/T ratio is roughly anticorrelated with the formation epoch, which may agree with the observed decrease of the [Mg/Fe] ratio as the bulge becomes fainter (Jablonka et al. 1996) and the preponderance of irregular bulges in late-type spirals (Carollo et al. 1997). A good positive correlation is found between $t_{\rm bulge}$ and β. Because β also determines the major epoch of disk formation, this correlation may explain the similarity in colors between the bulge and the disk in many spiral galaxies observed by Peletier & Balcells (1996). The other 11 models exhibit behaviors similar to those stated here.

It may be argued that the age of the bulge indicated in Figure 8*d* is too large, especially for less massive and/or less dense galaxies. The present multizone model cannot make allowance for the likely variation of the collapse timescale at different radii in the galaxy. The value of β should be regarded as a kind of average over the entire disk, and it tends to overestimate the actual accretion timescale in the inner disk, which contributes much to the bulge formation. This limitation presumably leads to an overestimate of $t_{\rm bulge}$. In any case, because of the large differences of the accretion timescales at different radii in the disk component, measured age difference between the bulge and the disk in a galaxy will depend strongly on how extended a part of the disk is considered.

6.3. *Present Gas Content and Star Formation Rate*

Figure 8*e* displays the mass fraction of the gas component, m_g/M, at the present epoch. The mass fraction is seen to decrease as the mass and/or density of the galaxy increases in most domains of the (ρ, M) plane, in agreement with the observation. This behavior is shared by all the continuous models having the β type of 2. The three continuous models with β type 1 all show gas mass fractions not significantly dependent on galaxy density and may contradict the observation shown in Figure 7. One possible problem with all the continuous models is that the absolute value of the gas mass fraction tends to be considerably smaller than observed, by as much as a factor of ~ 10. This discrepancy may not rule out these models convincingly, however. There is a substantial discrepancy between the interstellar gas masses determined by different authors. For example, Casoli et al. (1998) give gas fractions ranging from $\sim 5 \times 10^{-3}$ for Sa galaxies to $\sim 2 \times 10^{-2}$ for Sc galaxies, which are considerably smaller than the values given by Young (1990) and are in better agreement with the models.

Introduction of a threshold gas density controlled by a constant minimum velocity dispersion of the gas disk improves the result remarkably. In this case, the gas content at the present epoch is determined solely by the galaxy parameters, M and ρ (through the epicyclic frequency), and does not depend on the star formation history of the galaxy. This is because by the present epoch the galaxy, in the considered (ρ, M) domain, has already reached the infall-limited regime of star formation. The gas mass fraction as a function of M and ρ is the same in all the threshold models, and it increases as M and/or ρ decreases, in qualitative agreement with the observation. Inhibition of star formation by the threshold boosts the present-day gas content and brings the models into a better agreement with the observational data.

The specific star formation rate, SFR/M, follows the same behavior as the gas fraction on the (ρ, M) plane in all the continuous models (Fig. 8*f*). Introduction of the star formation threshold leads to sporadic star formation in late phases of evolution and makes the dependence of SFR/M on M and ρ quite irregular. Adoption or rejection of any model in terms of the present star formation activity seems impossible because of the already mentioned uncertainty in the observational data (§ 5.3).

7. DISCUSSION

It has been demonstrated that the clumpy evolution model can reproduce the observed variation of the bulge-to-disk ratio among spiral galaxies for a range of possibility wide enough to encompass the real situation. Nevertheless, the quality of available observational data and limitations in the theoretical modeling seem to hamper narrowing further the parameter range. Here I discuss a few implications of the present study and the limitations of the models.

7.1. *Appearance of Primeval Disk Galaxies*

Only few studies have explored what primeval galaxies might really look like (e.g., Meier 1976; Baron & White 1987; Katz 1992). Theoretical predictions about the appearance of young galaxies are becoming increasingly important because the direct observation of high-redshift galaxies enabled by *HST* and other new instruments is starting to provide powerful constraints on theoretical evolutionary models from morphological and dynamical viewpoints.

CDM cosmogony predicts that the clumpy appearance of primeval galaxies is a direct consequence of dominant small-scale density perturbations imposed on the matter distribution in the early universe. Indeed, Baron & White (1987) demonstrate by numerical simulation that a young *elliptical* galaxy should not be observed as a bright single body but as a conglomeration of several discrete blobs connected by a common faint envelope. Katz's (1992) dissipational formation model for a spiral galaxy, which includes star formation processes, also produces clumpy appearance at high redshift, because of imposed initial density perturbations. In this case, the clumps form during the collapse of the galaxy.

In contrast to these CDM-based simulations, the clumps advocated in the present study have no causal relationship with initial density perturbations in the universe. The clumpy nature of primeval disk galaxies in the present model originates in the gravitational instability of gas-rich galactic disks formed in the early phase of disk galaxy evolution. It is encouraging that the *HST* and large ground-based telescopes have recently found clumpy structures in a number of high-redshift galaxies, although they may be manifestation of CDM clumps. Usually few clumps are contained in one object, but the number may be severely affected by the spatial resolution and the surface-brightness limit of the instrument used. Even chain galaxies (e.g., Cowie et al. 1995), which are elongated objects containing several bright blobs, may be primeval disk galaxies viewed edge-on in which the clumps scattered in the disk component are viewed within the projected disk plane. (Another possibility, suggested by Dalcanton & Shectman 1996, is that they represent edge-on low surface-brightness galaxies.) Head-tail systems called "tadpole" galaxies by van den Bergh et al. (1996) may not be rectilinear objects but edge-on manifestations of clumpy disks in which one of several clumps is particularly large.

Possible evidence for clumpy structures other than direct imaging comes from the analysis of correlation functions. Infante, de Mello, & Menanteau (1996) found, for galaxies with the average redshift of $\langle z \rangle \sim 0.35$, a discontinuity of the correlation function at the separation of $\sim 6''$. This suggests strong clustering of faint galaxies within ~ 20 kpc of individual galaxies, which may be caused by clumping in a single galactic disk. Because the mean redshift of their sample is relatively small, it would be desirable to extend a similar analysis into a larger redshift.

One caveat in interpreting images of distant galaxies is that morphology of objects at large redshift is strongly influenced by *k*-correction and the steep dependence of surface brightness on redshift [of the form $(1 + z)^{-5}$]. Bohlin et al. (1991) and Giavalisco et al. (1996a) have cautioned that those clumpy structures observed in medium- to high-redshift galaxies may not be features characteristic of early evolutionary phases but simply exaggerated manifestations of the irregular distribution of star-forming regions observed in some galaxies (usually of late types) in the local universe.

Interactions or mergers with smaller satellite galaxies are sometimes invoked in the interpretation of the peculiar appearance of high-redshift galaxies (e.g., Griffiths et al. 1994; van den Bergh et al. 1996). The most straightforward and powerful test for discriminating between the instability hypothesis proposed here and the merger/interaction scenario is to examine the kinematics of "satellites." In the merger/interaction scenario, we expect random orientation of clump orbits relative to the primary because there is no reason to think that the bombardment of other galaxies from outside has any preferred geometry with respect to the primary. On the other hand, it is inevitable in the present scenario that the motions of "satellites" are coplanar and that all satellites rotate in the same direction around a common center. Spectroscopic observations will provide a direct check.

7.2. *Creation of the Bulge-Disk Structure*

Dynamical study of the formation and evolution process of galaxies has a long history (e.g., Eggen, Lynden-Bell, & Sandage 1962). One of the most important goals in this area is to understand the origins of the observed variety of galaxies. A series of numerical works carried out by Larson (1969, 1974, 1975, 1976) stands out as a landmark in the theoretical galactic astronomy. In his models, the formation of a disk galaxy proceeds in two stages characterized by star formation processes that operate at very different rates. First there is a rapid star formation process that forms a spheroidal component, and later a much slower star formation permits most of the residual gas to condense into a disk before it is consumed by star formation. Larson (1976) envisages two possible causes of the reduction of star formation efficiency after the bulge has formed, which is required for the formation of a significant disk component. The first is inhibition of star formation by the tidal force exerted on the gas clouds by the already existing bulge. The second is that the protogalactic gas has a two-phase structure, with dense clouds that rapidly form stars in a spheroidal component and less dense intercloud gas that does not form stars until it has settled to a disk. In order to get distinct separation between the disk and the bulge, he has also prescribed turbulent viscosity that is large in the early phase of evolution, when the bulge grows, but is reduced in later phases.

In marked contrast with Larson's (1976) model, the present study proposes a more disordered and chaotic formation of disk galaxies. The dominance of massive subgalactic clumps and resulting dynamical processes constitute the main ingredients of the present model. The bulge is assembled from those clumps formed by the local gravitational instability in the disk component. In the present model, the clumpy nature of the young gas-rich disk provides the required viscosity to form a bulge (e.g., Lin & Pringle 1987). Larson (1976) had to introduce viscosity a priori because the axisymmetry of the configuration imposed on his models could not allow radial transport of angular momentum by nonaxisymmetric perturbations such as clumps, bars, and spiral arms. Although the present study stresses the secondary nature of the galactic bulges, this does not mean that the *whole* disk has formed before the bulge, as has been stated repeatedly. Only inner parts of the disk contribute to bulge formation; the outer parts are considered to form much later than the bulge by a slow accretion of primordial gas from the outer halo. In this respect, the present scenario is not so drastically different from Larson's (1976) model as it might seem.

Any theory for the formation and evolution of galaxies should ultimately be incorporated into a cosmological model that specifies the initial and boundary conditions for galaxy evolution. In the CDM cosmology, which seems to be the most promising at present, past numerical studies indicate that the galactic bulges are formed from smaller clumps that form in the intergalactic space before assemblage of individual galaxies (e.g., Katz 1992). However, the limited resolution in numerical simulations makes it impossible to examine detailed morphological and kinematical structures of the formed bulges. It also remains unclear how the bulge-to-disk ratio depends on various physical parameters.

Which is the more dominant of these mechanisms, merger of CDM clumps or the clumping and collapse of inner galactic disks? It may turn out that several different processes contribute to bulge formation. Recent imaging observations of the nearby galaxy bulges carried out by Carollo et al. (1997, 1998) and Carollo & Stiavelli (1998) with the *HST*/WFPC2 are revealing that there are (at least) two kinds of galactic bulges (see also Phillips et al. 1996). The first includes the "classical" bulges that obey the de Vaucouleurs' density profile (the $R^{1/4}$ law), and the second includes those bulges that are best fitted by exponential density profiles. These two classes seem to constitute separate sequences on the ($M_{\rm bulge}$, γ) plane, where $M_{\rm bulge}$ is the absolute magnitude of the bulge and γ is the nuclear slope in the luminosity profile. The classical bulges trace the same relation between $M_{\rm bulge}$ and γ as elliptical galaxies. The existence of galactic bulges with morphological or kinematical properties different from those of elliptical galaxies is also pointed out by Kormendy (1993). This observed dichotomy may suggest operation of two different mechanisms in bulge formation. One possibility is that the $R^{1/4}$ bulges were made by mergers of CDM clumps when the galaxies were assembled, whereas exponential bulges originate in the clumping of the inner galactic disks, as proposed in the present study.

7.3. *Longevity of Clumps*

The present study stresses the importance of heavy clumps formed in the early galactic disks in driving long-term disk galaxy evolution. Therefore, the longevity of these clumps is a key factor of the present model. Although the present study takes into account the contribution from formed clumps of all mass scales, it may be that clumps with smaller masses are in general prone to destruction. Possible destruction processes include energy injection from internal star formation and tidal destruction due to the galactic gravitational field. The strong concentration of gas clouds into narrow spiral arms observed in nearby galaxies points to relatively short lifetimes for those clouds. For example, giant molecular clouds, with masses of $\sim 10^6\ M_\odot$, are believed to have lifetimes shorter than a few times 10^8 yr, perhaps owing to energy deposit from young massive stars born in them. Even giant molecular associations (GMAs), with estimated masses of several times $10^7\ M_\odot$, seem to be transient. Rand & Kulkarni (1990) found that GMAs in the interarm regions of M51 are gravitationally unbound and argued that they may be disintegrating owing to tidal shearing by the background gravitational field of the galaxy. If this is correct, the estimated lifetime of these GMAs is several times 10^8 yr. The present model sometimes develops clumps with masses of $10^{8-9}\ M_\odot$. The fate of these extremely massive clumps is not clear, because they are absent from nearby (i.e., present-day) galaxies. Such clumps may also suffer from disintegration.

Is the treatment in the present study then totally unrealistic? The answer is probably "No." Assuming that a clump of any mass has a finite lifetime, probably depending on its internal structure and the strength of the external tidal field, the interstellar gas in real galaxies will circulate through different phases as follows. First a group of clumps will be formed owing to the gravitational instability in the gas disk. Until some mechanism destroys these clumps, they will experience dynamical frictions and move inward to the galactic center. When the clumps are destroyed, clump material is dispersed into the interstellar space and the inflow driven by dynamical friction stops. However, after a certain period, a new generation of clumps is formed from this diffuse material by gravitational instability and they resume inward motion. Actually, this recycling will not be a coherent process over the entire galactic disk, and every local region in the disk will cycle through these phases independently of others.

The numerical simulation described in §§ 2 and 3 differs from the situation considered here, despite the fact that it includes energy feedback from star formation events in a simple form. The massive clumps formed in the numerical model maintain their identity until they merge with other clumps or are swallowed by the bulge. On the other hand, the analytical multicomponent models presented in § 4 seem to fit in better with the circulation picture of the interstellar medium. They calculate the typical mass of clumps from the instantaneous gas surface density. Therefore, the clump mass is changing as a function of time. In other words, the models do not assume the identity of each clump for the whole period of galaxy evolution. Thus those analytic models provide a fairly good description of actual situation, provided that the average lifetime of clumps is a significant fraction of the whole period of one circulation cycle.

8. CONCLUSIONS

The collapse of a protogalaxy composed of dark matter and primordial gas has been investigated by numerical simulations and analytical multizone modeling in an attempt to examine the early evolution of disk galaxies. The importance of the ample interstellar matter existing in young galactic disks has been highlighted. Confrontation of the theoretical results with the available observational data has led to a new picture of disk galaxy evolution in which the bulge is the secondary object formed from disk matter.

As the protogalaxy collapses, a gaseous disk starts to form. Because the gas density is low, very few stars form before the disk formation. Formation of the disk causes a drastic increase in gas density and initiates the star formation process. In early evolution phases, the galactic disk is rich in interstellar gas and efficient energy dissipation keeps the disk dynamically cold. Then gravitational instability sets in, leading to the formation of massive clumps rotating in the disk plane. The individual mass of these clumps can be as large as $\sim 10^9\ M_\odot$. Intense star formation occurs in these clumps. Therefore, the galaxy at this epoch exhibits a clumpy and irregular appearance in optical wavelengths, which may explain the morphologically peculiar galaxies observed at high redshift. While orbiting in the disk plane, the clumps tend to merge with each other and make successively larger clumps. The clumps lose their orbital kinetic energy through dynamical friction against surrounding stars and gas clouds, and thus accumulate in the central region, forming a spheroidal bulge. The collapse timescale of the protogalaxy will be larger in its outer parts. Slower accretion of primordial gas establishes the outer parts of the disk after the bulge formation is mostly finished.

An analytical modeling of the clumpy galaxy evolution picture has been carried out in which a disk galaxy is described as a multicomponent system comprising a dark halo, gaseous and stellar disks, and a bulge. Evolution of these components is controlled by two parameters: the accretion timescale, i.e., the rapidness with which the primordial gas contained in the halo region accretes to the disk plane; and the mass fraction of the gas at the beginning. Based on observational evidence, the accretion timescale in real spiral galaxies has been assumed to be a decreasing function of both the galaxy mass and the galaxy density, whereas the initial fraction of the gas has been assumed to increase as the galaxy mass and/or the galaxy density increases. Under this specification, the models have succeeded in reproducing the observational result that a galaxy having a large total mass and/or a large internal density tends to have a large bulge-to-disk ratio in the morphology range of Sa–Sc.

The author thanks the anonymous referee for many constructive comments that contributed to improvement of the paper.

APPENDIX A

NUMERICAL EVALUATION OF SINKING TIMESCALE OF A CLUMP

A massive body orbiting in the disk component of a galaxy spirals into the galactic center by the action of dynamical friction. The timescale of this inward motion of the clump has been evaluated by numerical simulations as follows. Parameters of each simulation are listed in Table 1.

N-body models for a disk galaxy, which consists of a halo and a disk, have been constructed first. The models contain no interstellar gas. Here the halo is not meant to represent the entire (dark) halo that might surround the visible galaxy but only the portion of the dark halo *inside* the optical radius plus any luminous spheroidal components such as a bulge. The total mass of the galaxy is unity and the mass of the disk component is m_d. Both components are truncated at the galactocentric radius of unity. The halo and the disk are constructed by 5000 and 50,000 collisionless particles, respectively. The gravitational softening radius is 0.04 and 0.02 for the halo and disk particles, respectively.

The model is fundamentally based on Fall & Efstathiou (1980). The stellar disk has an exponential surface density distribution with a scale length of 0.25, in agreement with observations. The disk rotates nearly rigidly in the inner parts and at a nearly constant velocity in the outer parts. The turnover radius that divides these two parts is denoted by r_m. Thus the global shape of the rotation curve is determined by specifying r_m. The halo is assumed to be spherically symmetric, and its volume density distribution is determined so that the rotational velocity in the disk plane due to the combined gravitational field of the halo and the disk matches the specified rotation curve.

The velocity dispersion of the halo is chosen as follows. First, the isotropic velocity dispersion at each radius is calculated so that the condition for "local virial equilibrium" is satisfied everywhere (see Noguchi 1991 for details). A trial simulation showed that such a condition does not lead to virial equilibrium for the entire system. In order to alleviate this, the velocity dispersion was multiplied by a factor of 0.65.

This disk galaxy model is evolved in isolation before the sinking simulations. First only the halo component is evolved for 15 dynamical times with the disk component fixed. After the halo is relaxed, the disk is activated. At this time, the gravitational force acting on each disk particle is calculated and the circular velocity is given to that particle so that the centrifugal force is exactly balanced with the gravity. Next, small random velocities are given to disk particles, which correspond to a specified Q parameter of Toomre (1964). The rotational velocity of each star is then corrected for the contribution from this random motion. This state just after the activation of the disk is adopted as the initial condition for the disk galaxy in the sinking simulations.

The clump is treated as a particle that has a mass, $m_{\rm cl}$, and a gravitational softening radius of r_s, which is regarded as the effective radius of the clump. The mass and radius of the clump given in Table 1 are also in units of those of the disk galaxy.

The clump starts at the galactocentric radius $r = 0.8$ in the disk plane. The initial velocity is that of the circular motion, and the orbital motion is in the same direction as the disk rotation. This initial condition is appropriate for the clumps formed from the disk material. A number of simulations have been carried out by varying m_d, m_{cl}, r_m, and Q. Each simulation is performed until the galactocentric radius of the clump decreases to 0.2. The epoch of this moment is taken to be the dynamical friction timescale, τ_{fri}, and listed in Table 1. Gravitational interactions between all the particles have been calculated by the tree code (e.g., Barnes & Hut 1986), and the orbit integration has been done by the leap-frog scheme with a time step of 0.01, i.e., one hundredth of the dynamical timescale (a time step of 0.015 was used in the halo relaxation phase).

REFERENCES

Abraham, R. G., Tanvir, N. R., Santiago, B. X., Ellis, R. S., Glazebrook, K., & van den Bergh, S. 1996, MNRAS, 279, L47
Andredakis, Y. C., & Sanders, R. H. 1994, MNRAS, 267, 283
Ashman, K. M. 1990, ApJ, 359, 15
Athanassoula, E., Bosma, A., & Papaioannou, S. 1987, A&A, 179, 23
Bahcall, J. N., & Casertano, S. 1985, ApJ, 293, L7
Barnes, J. E., & Hut, P. 1986, Nature, 324, 446
Baron, E., & White, S. D. M. 1987, ApJ, 322, 585
Bohlin, R. C., et al. 1991, ApJ, 368, 12
Carollo, C. M., & Stiavelli, M. 1998, AJ, 115, 2306
Carollo, C. M., Stiavelli, M., & Mack, J. 1998, AJ, 116, 68
Carollo, C. M., Stiavelli, M., de Zeeuw, P. T., & Mack, J. 1997, AJ, 114, 2366
Casoli, F., et al. 1998, A&A, 331, 451
Cowie, L. L., Hu, E. M., & Songaila, A. 1995, AJ, 110, 1576
Dalcanton, J. J., & Shectman, S. A. 1996, ApJ, 465, L9
Dalcanton, J. J., Spergel, D. N., & Summers, F. J. 1997, ApJ, 482, 659
Dale, D. A., Giovanelli, R., Haynes, M. P., Scodeggio, M., Hardy, E., & Campusano, L. E. 1997, AJ, 114, 455
de Jong, R. S. 1996, A&A, 313, 45
de Vaucouleurs, G. 1974, in IAU Symp. 58, The Formation and Dynamics of Galaxies, ed. J. R. Shakeshaft (Dordrecht: Reidel), 1
Driver, S. P., Windhorst, R. A., & Griffiths, R. E. 1995, ApJ, 453, 48
Ebisuzaki, T., Makino, J., Fukushige, T., Sugimoto, D., Ito, T., & Okumura, S. K. 1993, PASJ, 45, 269
Eggen, O. J., Lynden-Bell, D., & Sandage, A. R. 1962, ApJ, 136, 748
Fall, S. M., & Efstathiou, G. 1980, MNRAS, 193, 189
Fall, S. M., & Rees, M. J. 1985, ApJ, 298, 18
Gavazzi, G. 1993, ApJ, 419, 469
Giavalisco, M., Livio, M., Bohlin, R. C., Macchetto, F. D., & Stecher, T. P. 1996a, AJ, 112, 369
Giavalisco, M., Steidel, C. C., & Macchetto, F. D. 1996b, ApJ, 470, 189
Glazebrook, K., Ellis, R., Santiago, B., & Griffiths, R. 1995, MNRAS, 275, L19
Glazebrook, K., Lehár, J., Ellis, R., Aragón-Salamanca, A., & Griffiths, R. 1994, MNRAS, 270, L63
Gott, J. R., & Thuan, T. X. 1976, ApJ, 204, 649
Griffiths, R. E., et al. 1994, ApJ, 435, L19
Hausman., M. A., & Roberts,W. W. 1984, ApJ, 282,106
Henriksen, R. N. 1991, ApJ, 377, 500
Holtzman, J. A., et al. 1993, AJ, 106, 1826
Infante, L., de Mello, D. F., & Menanteau, F. 1996, ApJ, 469, L85
Jablonka, P., & Arimoto, N. 1992, A&A, 255, 63
Jablonka, P., Martin, P., & Arimoto, N. 1996, AJ, 112, 1415
Katz, N. 1992, ApJ, 391, 502
Katz, N., & Gunn, J. E. 1991, ApJ, 377, 365
Kennicutt, R. C. 1983, ApJ, 272, 54
———. 1989, ApJ, 344, 685
Kent, S. M. 1985, ApJS, 59, 115
Koo, D. C., et al. 1996, ApJ, 469, 535
Kormendy, J. 1993, in IAU Symp. 153, Galactic Bulges, ed. H. Dejonghe & H. J. Habing (Dordrecht: Kluwer), 209
Lacey, C. G., & Fall, S. M. 1985, ApJ, 290, 154
Larson, R. B. 1969, MNRAS, 145, 405
———. 1974, MNRAS, 166, 585
———. 1975, MNRAS, 173, 671
———. 1976, MNRAS, 176, 31
Larson, R. B., & Tinsley, B. M. 1978, ApJ, 219, 46
Levinson, F. H., & Roberts,W. W. 1981, ApJ, 245, 465
Lin, D. N. C., & Pringle, J. E. 1987, MNRAS, 225, 607
Matteucci, F., & Brocato, E. 1990, ApJ, 365, 539
Matteucci, F., & François, P. 1989, MNRAS, 239, 885
McWilliam, A., & Rich, R. M. 1994, ApJS, 91, 749
Meier, D. L. 1976, ApJ, 207, 343
Navarro, J. F., & White, S. D. M. 1994, MNRAS, 267, 401
Noguchi, M. 1991, MNRAS, 251, 360
———. 1996, ApJ, 469, 605
———. 1998, Nature, 392, 253
Norris, J. E. 1996, in IAU Symp. 169, Unsolved Problems of the Milky Way, ed. L. Blitz & P. Teuben (Dordrecht: Kluwer), 353
Peletier, R. F., & Balcells, M. 1996, AJ, 111, 2238
Persic, M., & Salucci, P. 1988, MNRAS, 234, 131
Phillips, A. C., Illingworth, G. D., MacKenty, J. W., & Franx, M. 1996, AJ, 111, 1566
Quinn, P. J., & Goodman, J. 1986, ApJ, 309, 472
Rand, R. J., & Kulkarni, S. 1990, ApJ, 349, L43
Read, A. M., Ponman, T. J., & Strickland, D. K. 1997, MNRAS, 286, 626
Rich, R. M. 1996, in IAU Symp. 169, Unsolved Problems of the Milky Way, ed. L. Blitz & P. Teuben (Dordrecht: Kluwer), 403
Roberts, W. W., & Hausman, M. A. 1984, ApJ, 277, 744
Rubin, V. C., Burstein, D., Ford, W. K., Thonnard, N. 1985, ApJ, 289, 81
Rubin, V. C., Ford, W. K., Jr., & Thonnard, N. 1980, ApJ, 238, 471
Rubin, V. C., Ford, W. K., Jr., Thonnard, N., & Burstein, D. 1982, ApJ, 261, 439
Sage, L. J. 1993, A&A, 272, 123
Sandage, A. 1986, A&A, 161, 89
Shlosman, I., & Noguchi, M. 1993, ApJ, 414, 474
Simien, F., & de Vaucouleurs, G. 1986, ApJ, 302, 564
Solomon, P. M., Sanders, D. B., & Scoville, N. Z. 1979, in IAU Symp. 84, The Large-Scale Characteristics of the Galaxy, ed. W. B. Burton (Dordrecht: Reidel), 35
Stark, A. A., & Brand, J. 1989, ApJ, 339, 763
Steidel, C. C. 1990, ApJS, 74, 37
Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996a, AJ, 112, 352
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996b, ApJ, 462, L17
Steinmetz, M., & Müller, E. 1995, MNRAS, 276, 549
Tinsley, B. M. 1981, MNRAS, 194, 63
Tóth, G., & Ostriker, J. P. 1992, ApJ, 389, 5
Toomre, A. 1964, ApJ, 139, 1217
Tully, R. B., Mould, J. R., & Aaronson, M. 1982, ApJ, 257, 527
van den Bergh, S., Abraham, R. G., Ellis, R. S., Tanvir, N. R., Santiago, B. X., & Glazebrook, K. G. 1996, AJ, 112, 359
van der Kruit, P. C., & Searle, L. 1982, A&A, 110, 61
van der Kruit, P. C., & Shostak, G. S. 1984, A&A, 134, 258
Vazdekis, A., Casuso, E., Peletier, R. F., & Beckman, J. E. 1996, ApJS, 106, 307
Whitmore, B. C. 1984, ApJ, 278, 61
Yoshii, Y., & Arimoto, N. 1987, A&A, 188, 13
Yoshizawa, M., Wakamatsu, K. 1975, A&A, 44, 363
Young, J. S. 1990, in The Interstellar Medium in Galaxies, eds. H. A. Thronson, & J. M. Shull (Dordrecht: Kluwer), 67