The Astrophysical Journal, 634:1067–1084, 2005 December 1

ON STAR FORMATION AND THE NONEXISTENCE OF DARK GALAXIES

Edward N. Taylor[1] and Rachel L. Webster
School of Physics, University of Melbourne, Parkville VIC 3010, Australia;
rwebster@physics.unimelb.edu.au
Received 2004 August 13; accepted 2005 August 16

ABSTRACT

We investigate whether a baryonic dark galaxy or "galaxy without stars" could persist indefinitely in the local universe, while remaining stable against star formation. To this end, a simple model has been constructed to determine the equilibrium distribution and composition of a gaseous protogalactic disk. Specifically, we determine the amount of gas that will transit to a Toomre unstable cold phase via the H_2 cooling channel in the presence of a UV–X-ray cosmic background radiation field. All but one of the models are predicted to become unstable to star formation: we find that, in the absence of an internal radiation field, the majority of gas will become Toomre unstable in all putative dark galaxies with baryonic masses greater than $10^9\,M_\odot$, and in at least half of those greater than $10^6\,M_\odot$. Moreover, we find that all our model objects would be detectable via H I line emission, even in the case that star formation is potentially avoided. These results are consistent with the nondetection of isolated extragalactic H I clouds with no optical counterpart (galaxies without stars) by the H I Parkes All-Sky Survey. Additionally, where star formation is predicted to occur, we determine the minimum interstellar radiation field required to restore gravothermal stability, which we then relate to a minimum global star formation rate. This leads to the prediction of a previously undocumented relation between H I mass and star formation rate that is observed for a wide variety of dwarf galaxies in the H I mass range 10^8–$10^{10}\,M_\odot$. The existence of such a relation strongly supports the notion that the well-observed population of dwarf galaxies represents the minimum rates of self-regulating star formation in the universe.

Subject headings: atomic processes — galaxies: formation — galaxies: ISM — galaxies: kinematics and dynamics — molecular processes — stars: formation

Online material: color figures

1. INTRODUCTION

There has been long-standing speculation that optically selected galaxy surveys might be missing a population of very low surface brightness (LSB) objects (Disney 1976; Disney & Phillipps 1987). It has recently become possible to address this question observationally: the H I Parkes All-Sky Survey (HIPASS) has now mapped the entire southern sky at ~21 cm. Since the HIPASS catalog (Meyer et al. 2004) represents an optically unbiased sample of neutral, atomic hydrogen (H I) in the local ($z \lesssim 0.04$) universe, it is ideal for the detection of H I–rich, baryonic dark galaxies, i.e., "galaxies without stars."

The result of the HIPASS search for dark galaxies is emphatic and surprising. With few possible exceptions (e.g., Kilborn et al. 2000; Ryder et al. 2001; see also Minchin et al. 2005), no isolated, extragalactic H I clouds were discovered for which an optical counterpart could not be identified (Doyle et al. 2005); it seems there are no dark galaxies.

Prior to HIPASS, the discovery of a number of "dim" galaxies—"galaxies where luminous matter is only a minor component of the total galaxy mass" (Carignan & Freeman 1988)—had led to the expectation that star formation (SF) might be suppressed completely within some galaxies. The first such examples were dwarf irregular galaxies with extended H I envelopes; DDO 154 serves as the archetype of this class of gas-rich LSB dwarfs (Krumm & Burstein 1984). At the other end of the mass scale, the serendipitous discovery of Malin 1 (Bothun et al. 1987; Impey & Bothun 1989), a very massive, H I–dominated LSB disk galaxy, provided an example of the class of "crouching giants" predicted by Disney (1976).

A common feature of Malin 1 and DDO 154 is their high H I mass-to-light ratios: $M_{\rm H\,I}/L_B \sim 5$ (compared to $M_{\rm H\,I}/L_B \sim 0.1$-1 for "normal" H I–detected LSB galaxies; Fisher & Tully 1975; Waugh et al. 2002). In addition to the observational problem of determining their abundance, which HIPASS has begun to address, galaxies like DDO 154 and Malin 1 pose the following theoretical problem: "Why was there so little gas processed into stars" in these systems (Carignan & Beaulieu 1989)?

Are these objects exceptional in the way in which they are presently forming stars? How far might this dim population extend toward complete darkness? And if dark galaxies can exist, is there some mechanism to prevent their detection by HIPASS?

In this paper we consider the viability of dark galaxies in the local universe, that is, whether a dark galaxy, having formed, can remain dark, or whether it will inevitably "light up." A simple model has been developed for the distribution and composition of primordial gas in a protogalactic disk that is in chemodynamical equilibrium, in the presence of a photodestructive and photoheating UV–X-ray cosmic background radiation (CBR) field. The specific goal is to determine where gravothermal instability (and hence SF) is initiated through the H_2 cooling channel. Additionally, where H_2 cooling cannot be checked by the CBR alone, we place a lower bound on the global star formation rate (SFR) in these objects by determining the additional interstellar radiation field (ISRF) required to restore thermal balance.

The discussion proceeds as follows. First, an overview of the argument is provided in § 2, as well as physical motivation for

[1] Sterrewacht Leiden, Leiden Observatory, P.O. Box 9513, Leiden 2300 RA, Netherlands; ent@strw.leidenuniv.nl

our model for a protogalaxy. In § 3, the link between H_2 cooling and Toomre instability is discussed, and we advance our specific criterion for SF. We provide a technical discussion of the processes by which we determine the gas distribution and composition in §§ 4 and 5, respectively. The second calculation, in which we estimate the global SFR, is then described in § 6. After a discussion of fiducial parameter choices in § 7, the results of our modeling procedure are given in § 8 and discussed in § 9. A summary of the salient and novel aspects of this work can be found in § 10.

2. OVERVIEW AND SUPPORTING CALCULATIONS

2.1. *Overview of the Argument*

Our primary goal is to determine whether a putative dark galaxy is gravothermally stable in its long-time/steady state configuration. Where this is the case, it is conceivable that such an object (assuming that it can form) might remain dark indefinitely; otherwise, from this point, if not well before, the galaxy will inevitably light up.

Whereas rotational support stabilizes against the collapse of larger scale density fluctuations, small-scale fluctuations can only be stabilized by thermal pressure (see § 3). Once the galaxy formation process is complete, the level of rotational support becomes fixed; stability is then principally determined by temperature, T. In particular, as shown in § 3, most low-mass protogalactic disks are stable at $T = 10^4$ K, while essentially all will become unstable if they are able to cool to $T \sim 300$ K.

The final temperature of the gas depends on its composition. Internal cooling of the gas proceeds by radiative de-excitation of collisionally excited atoms/molecules. Whereas H I cooling processes rapidly become inefficient (in the sense that the timescale for cooling becomes greater than the age of the universe) as collisional excitation ceases below $\sim 10^4$ K, H_2 cooling remains efficient above $\sim$300 K. The key to initiating gravothermal instability and SF in low-mass disks is therefore efficient H_2 cooling; conversely, the prevention of SF requires the prevention of H_2 cooling.

Within this paradigm, SF is thus a process that must be prevented, rather than initiated: a source of dissociating radiation is required to prevent the partial conversion of H I to H_2. Within the model, the CBR is the agent for dissociation, as well as the source of heating that opposes H_2 cooling. SF is deemed to be preventable wherever the rate of photoheating due to the CBR exceeds the rate of H_2 cooling within the gas; otherwise, SF is inevitable.

We thus reduce the problem of determining the in/stability of the protogalactic disk to determining the in/efficiency of H_2 cooling: we assume that gravothermal stability implies and is implied by thermal balance at $T \sim 10^4$ K. (We are able to explicitly check the converse assumption, that thermal instability at $T = 10^4$ K implies cooling to Toomre unstable temperatures, in § 9.2.) Because we are interested primarily in the situation of permanent gravothermal stability, we first assume that the gaseous disk is in dynamical and chemical equilibrium at $T = 10^4$ K, i.e., a long-time configuration at the end of H I–dominated thermal evolution.

This configuration represents the long-time state of a galaxy that has managed, for whatever reason, to avoid SF, since the gas is Toomre stable so long as $T \gtrsim 10^4$ K; if, in this configuration, the gas cannot initiate H_2 cooling, this model therefore represents the final state of a galaxy without stars. Conversely, where H_2 cooling is inescapable, the result will be thermal, chemical, and dynamical instabilities, leading necessarily to gravothermal collapse and SF.

In other words, following authors like Corbelli & Salpeter (1995), Haiman et al. (1996b), Oh & Haiman (2002), and Schaye (2004), we link SF to the existence of a cold phase within the gas and examine under what conditions the transition to the cold phase is initiated (see, too, Elmegreen & Parravano 1994; but also Young et al. 2003). Note that turbulent effects, which can both initiate and prevent SF, are neglected; all discussion of this point is deferred until § 9.1. We have thus constructed a static problem: we make no attempt to track the evolution of the gas at any stage.

Wherever possible, our assumptions have been tailored to minimize the rates of both H_2 production and H_2 cooling, so as to make SF as difficult as possible. We thus place a firm lower limit on the mass of gas that can exist in a steady state without undergoing any SF.

Where we find that SF cannot be prevented by the action of the CBR, we perform a second calculation to place a lower limit on the expected SFR. This is done by introducing a diffuse ISRF as a second source of photoheating and photodissociation. Specifically, we determine the minimum ISRF intensity required to maintain thermal balance at $T = 10^4$ K and so restore gravothermal stability. With some simple assumptions (see § 6), this ISRF can then be related to a minimum global SFR within the model galaxy.

2.2. *The Back of the Envelope Calculation*

The CBR can act to check H_2 cooling in two ways: by direct photodestruction of H_2 and by overwhelming any H_2 cooling with photoheating. The cloud's only defense is self-shielding, i.e., absorption of enough of the CBR by the outermost gas to shield inner regions from its influence.

The action of the CBR is thus to prevent SF in a shielding "skin" that covers both faces of a protogalactic disk, in a manner analogous to photodissociation regions around UV sources in our own Galaxy (Hollenbach & Tielens 1999); SF is only possible in the self-shielded center. The thickness of this skin can be characterized by the critical surface density of gas, $\Sigma_{\rm crit}$, below which the CBR is able to keep the gas completely photodissociated.

We obtain a first estimate for $\Sigma_{\rm crit}$ by noting that for the CBR to prevent the production of H_2, the total rate of photodestruction in a column (which is just the flux of photodestructive photons) must exceed the total recombination rate, viz.,

$$\frac{J_\nu}{h\nu}\,\Delta\nu \gtrsim \dot{n} Z_d. \tag{1}$$

Here the flux of destructive photons is approximately the product of the number of photons at the threshold frequency, $J_\nu/h\nu$ (where J_ν is the CBR flux at the threshold frequency ν), and the frequency interval over which photodestructive absorption can occur, $\Delta\nu$. Similarly, the recombination rate in the column is approximately the product of the recombination rate per unit volume, $\dot{n}$, and some characteristic column height, Z_d.

With the approximations $\Sigma \approx \rho_0 Z_d$ (where ρ_0 is, for now, a characteristic density) and $\dot{n} \approx \alpha x_i x_j \rho_0^2/\mu^2$ (α is the rate coefficient of the reaction between species i and j that replenishes the cloud's H_2, x_i is the relative abundance of species i, and $\mu \approx 1.22 m_p$ is

the mean molecular mass of particles in the gas), it is possible to rewrite the inequality given by equation (1) as

$$\Sigma \lesssim \left(\frac{J_\nu \Delta\nu}{h\nu}\frac{\mu^2}{\alpha x_i x_j}\frac{\sigma^2}{2\pi G}\right)^{1/3} = \Sigma_{\rm crit}. \quad (2)$$

In writing this expression, we have also made use of equation (17) to relate ρ_0 and Z_d, which introduces the gas velocity dispersion, σ.

Using equation (2), we can now obtain a quantitative estimate for $\Sigma_{\rm crit}$, assuming $T = 10^4$ K (which fixes σ and the value of α), chemical equilibrium (which similarly fixes x_i and x_j), and a putative CBR spectrum $J(\nu)$. To estimate $\Delta\nu$, we assume that H_2 shielding occurs between the Lyman-Werner and Lyα threshold energies, 11.2–13.6 eV. With these numbers, we calculate $\Sigma_{\rm crit} \approx$ 3–4 $M_\odot$ pc^{-2}, depending on the character of the CBR spectrum (see § 7.3).

In reality, the CBR will produce at least two layers of skin: one in which it prevents H I production (i.e., an ionized layer) and, internal to that, one in which H_2 production is prevented (i.e., a neutral, atomic layer). Using the same argument as above, we find that the surface density of the ionized layer is $\approx 1\ M_\odot$ pc^{-2}.

We therefore expect that the CBR will completely prevent SF where the surface density of the protogalactic disk is less than several $M_\odot$ pc^{-2}. (For comparison, observations are consistent with a fixed SF threshold of $\sim 10\ M_\odot$ pc^{-2}; Taylor et al. 1994; see also Martin & Kennicutt 2001 and discussion in Schaye 2004.)

2.3. *Characteristic Timescales*

We assume chemodynamical equilibrium in our modeling. Thus, while our results do not rely on any particular galaxy formation scenario, they do depend on these equilibrium states being accessible to each system within its present lifetime. Our model does not apply to *forming* galaxies, nor does it attempt to follow the behavior of the gas as it transits from warm to cold.

In particular, our criterion for SF implicitly assumes that enough time is available for the gas to cool from $T \sim 10^4$ K to $T \sim 300$ K after the onset of H_2 cooling. In the remainder of this section we present some simple calculations in support of this assumption.

First, we can evaluate the timescale for cooling $t_c = \varepsilon/\dot{\varepsilon}$, where $\varepsilon \sim nkT$ is the thermal energy density (n is the number density of the gas; k is Boltzmann's constant) and $\dot{\varepsilon} = \Lambda$ is the net cooling rate per unit volume. Using the rates of cooling listed in Appendix A and assuming gas composition in chemical equilibrium at $T = 10^4$ K (see Fig. 3), we find $t_c \approx 1.5 \times 10^{-3} n^{-1} t_{\rm H} = 20 n^{-1}$ Myr, where $t_{\rm H} \approx 13.7$ Gyr is the Hubble time and n has units of cm^{-3}. (Close to the galactic plane, n is typically in the range 10^{-2} to 10^1 cm^{-3}; see Fig. 4.)

Under these conditions, the H I and H_2 cooling rates are approximately equal. In contrast, the same composition at 8000 K leads to cooling times for H I and H_2 cooling of $2.3 n^{-1} t_{\rm H}$ and $15 n^{-1}$ Myr, respectively: H_2 cooling remains efficient well below the point at which H I–driven thermal evolution effectively ceases.

As we argue in § 5.3, in gas near thermal equilibrium, the H_2 abundance will always approach equilibrium from above; in this case the equilibrium value therefore provides a lower limit to the actual H_2 abundance. For an H_2 abundance 10 times greater than the equilibrium value, $x_{H_2} \sim 10^{-3}$, the cooling time at $T = 10^4$ K is reduced by a factor of 30 to $\sim 0.67 n^{-1}$ Myr. Among tidal dwarf galaxies, the only observed population of presently forming galaxies, the H_2 abundance is typically as high as $x_{H_2} \sim 0.20$ (Braine et al. 2001); the cooling time in this case is reduced to just $\sim 10^{-3} n^{-1}$ Myr.

Next, we can make a rough estimate for the time needed to produce such an overabundance, $t_{H_2} = x_{H_2}/\dot{x}_{H_2}$. Again assuming equilibrium abundances at 10^4 K and using the rate for the dominant H_2 production process (reaction 12 in Appendix A), $\dot{x}_{H_2} \approx 2 \times 10^{-17} n^2$ s^{-1}, we find $t_{H_2} \approx 1.6 n^{-1}$ Myr to reach $x_{H_2} = 10^{-3}$. This value is of the same order of magnitude as if we were to take a semiempirical rate for H_2 production on dust grains at $T \lesssim 10^3$ K (two-body H_2 production becomes inefficient at these temperatures), $\dot{x}_{H_2} \sim 10^{-17} n^2$ s^{-1} (Hollenbach & Tielens 1999).

Thus, as equilibrium at 10^4 K is disturbed, we have $t_{H_2} < t_c$, and x_{H_2} increases above the equilibrium value; t_c is accordingly diminished. At the same time, t_{H_2} increases as the temperature falls, until $t_c \ll t_{H_2}$. At this point two-body chemical evolution effectively ceases, leaving a "freezeout" abundance of H_2: time-integrated studies of shock-heated (i.e., $T > 10^4$ K, $x_{e^-} \approx 1$) gas cooling catastrophically below 10^4 K (Shapiro & Kang 1987; Oh & Haiman 2002) find that x_{H_2} asymptotes to a "universal" freezeout abundance of order 10^{-3} at $T \sim 4000$ K. With the provision that this freezeout H_2 abundance can survive photodissociation (see § 9.2) from this point on, the gas continues to cool until $t_c > t_{\rm H}$ at $T \lesssim 300$ K.

Once the warm ($T \sim 10^4$ K) gas initiates H_2 cooling, it will therefore transit to a cold ($T \sim 300$ K) state in $\lesssim 20 n^{-1}$ Myr. This should be compared to the characteristic timescale for the first stage of SF, the free-fall time for a collapsing cloud, $t_f = (3\pi/16 G\bar{\rho})^{1/2} \sim 52 n^{-1/2}$ Myr. Where H_2 cooling occurs, the cloud will therefore cool on a timescale *at least* a few times faster than the collapse timescale; depending on the actual H_2 abundance, the factor is more likely $\gtrsim 30$.

3. THE MODEL. I. THE CRITERION FOR STAR FORMATION

The standard criterion for the gravothermal stability of a rotating disk, the Toomre criterion (Toomre 1964), can be obtained from the dispersion relation for an axisymmetric density perturbation (i.e., a sound wave) with wavelength l and frequency ω (§ 6.2.3 of Binney & Tremaine 1987):

$$\omega^2 = \kappa^2 - 2\pi G\Sigma\frac{2\pi}{l} + \left(\frac{2\pi}{l}\right)^2 c_s^2. \quad (3)$$

Here $\kappa(R)$ is the epicyclic frequency of the orbit and $c_s(T)$ is the sound speed within the gas. Since the general waveform's time evolution is like $e^{-i\omega t}$, stability requires that $\omega^2 > 0$, and hence

$$\left(\frac{c_s\kappa}{\pi G\Sigma}\right)^2 \equiv Q^2 \geq 1. \quad (4)$$

This is the Toomre stability criterion, written in terms of the Toomre parameter, Q.

In order to determine the range of stable l-values, we combine equation (3) with the condition $\omega^2 > 0$ to obtain

$$\left(\frac{c_s\kappa}{\pi G\Sigma}\right)^2 > 4\left(\frac{l}{l_{\rm crit}} - \frac{l^2}{l_{\rm crit}^2}\right), \quad (5)$$

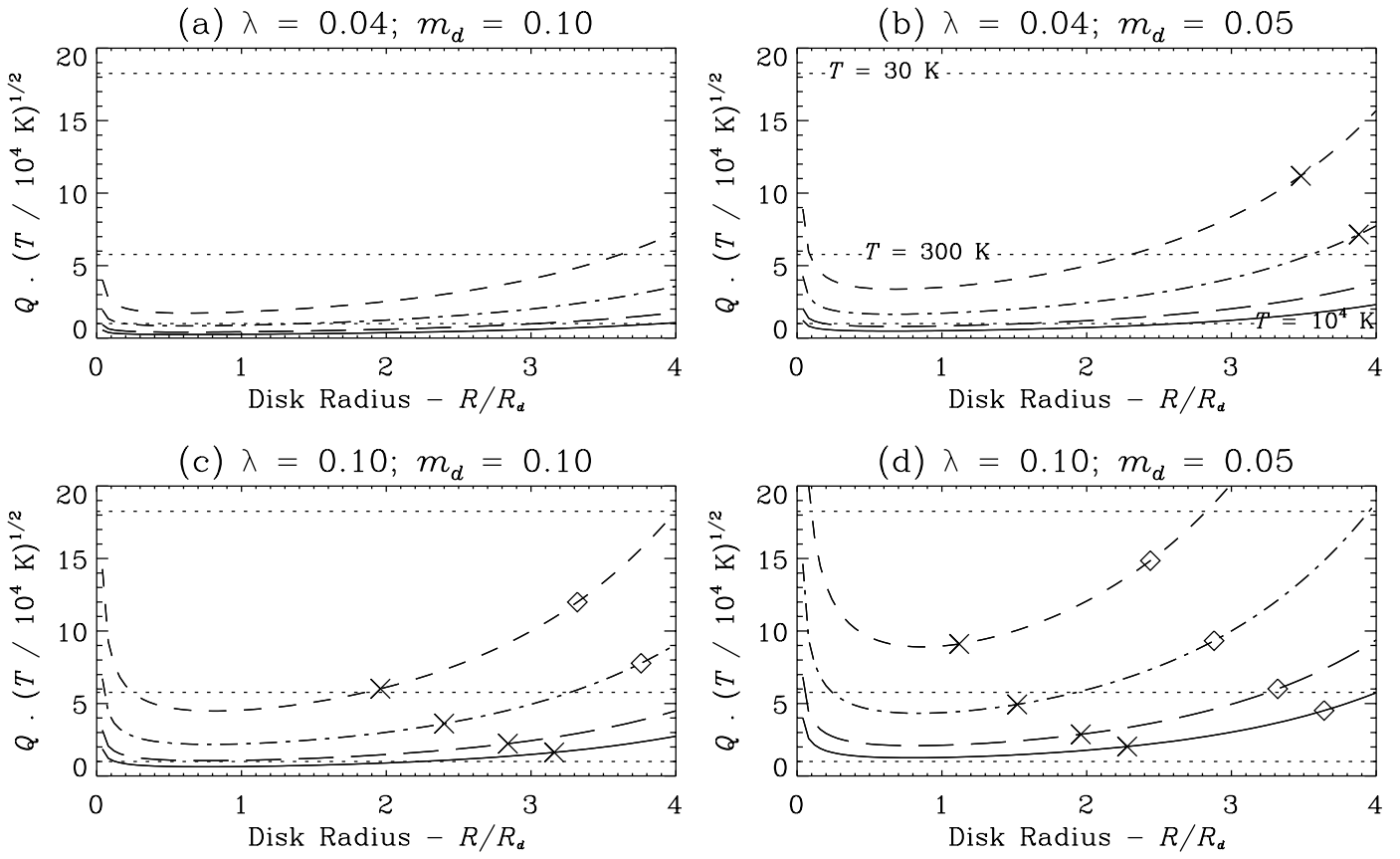

FIG. 1.—Toomre profiles $Q(R)$ for various parameter sets. Panels (*a*) through (*d*) show the effect of different choices for the spin parameter, λ, and the disk mass fraction, m_d (see § 7 for a detailed description); each panel differs in only one parameter from those adjacent to it. Within each panel, profiles are shown for virial masses of 1×10^8 (*short-dashed line*), 1×10^9 (*dot-dashed line*), 1×10^{10} (*long-dashed line*), and $5\times10^{10}\ M_\odot$ (*solid line*). The Toomre parameter is temperature dependent: within each panel, dotted lines show where the Toomre threshold ($Q = 1$) would lie if the gas were at temperatures of 10^4 (*bottom*), 300 (*middle*), and 30 K (*top*). Disks are unstable below the Toomre threshold. To illustrate the relationship between Toomre stability and self-shielding, we also show the radius within which the gas is self-shielding (see § 8) for two different CBR spectral indices, $\eta = 0.7$ (*diamonds*) and 2.0 (*crosses*). [*See the electronic edition of the Journal for a color version of this figure.*]

where the left-hand side is now recognizable as Q^2. Since $Q^2 \geq 0$, stability is assured for

$$l > l_{\rm crit} \equiv \frac{4\pi^2 G\Sigma}{\kappa^2}. \tag{6}$$

Also, for $l^2 \ll l_{\rm crit}^2$, the inequality given by equation (5) reduces to

$$l \lesssim \frac{c_s^2}{G\Sigma}, \tag{7}$$

which is just the Jeans criterion for the gravothermal stability of a thin disk (see also Schaye 2004). These two cases can be interpreted as the zero pressure and zero rotation limits, respectively. Large perturbations are stabilized by the effects of rotation (inequality [6]), while small perturbations are stabilized by internal pressure (inequality [7]).

Once the disk and halo have settled into dynamical equilibrium, all gross dynamical quantities (e.g., ϱ_h, Σ, V, κ) become fixed. The only variable quantity in the inequality given by equation (4) that can influence Q is then the sound speed, $c_s = (kT/\mu)^{1/2}$. The amount of the disk that is unstable therefore depends strongly on T. This is illustrated in Figure 1, which shows profiles of $Q(R)$ for the various parameter sets that we consider and the regions in which these disks would be unstable at temperatures of 10^4, 300, and 30 K.

As mentioned in § 2, for low-mass disks, thermal stability at 10^4 K implies and is implied by gravothermal stability; conversely, low-mass disks only become Toomre unstable at temperatures that are only accessible via the H_2 cooling channel. Once H_2 cooling is initiated, the gas makes a rapid transition to a cold phase with $T \lesssim 300$ K (Norman & Spaans 1997), in which case it is virtually guaranteed (at least for $M_{200} \gtrsim 10^9\ M_\odot$) to be Toomre unstable. For these reasons, we use the H_2 cooling rate of the gas, in chemical equilibrium and at $T = 10^4$ K, in conjunction with the Toomre criterion for gravothermal instability to define our criterion for SF.

Specifically, at a given point, star formation is deemed to occur wherever the H_2 cooling rate exceeds the total rate of photoheating, provided that the gas is Toomre unstable at $T \geq 300$ K. Figure 1 illustrates the relation between these two criteria: gravothermally unstable regions lie below the Toomre threshold; we also show the radius within which the gas is self-shielding.

4. THE MODEL. II. THE DISTRIBUTION OF MATTER

In the next three sections we provide a detailed, technical discussion of our three-step modeling procedure. The means by which we determine the distribution of matter in the protogalactic disk is described in this section. We discuss the solution for the

equilibrium chemical composition of the gas in § 5, and the calculation of the global minimum SFR is discussed in § 6. A reader wishing to avoid such a technical discussion may choose to skip to § 7.

The solution for the volume distribution of the gas involves two separate calculations. First, the analytic model of Mo et al. (1998, hereafter MMW98) is used to determine the radial distribution of matter in the model protogalaxies. The MMW98 model has been shown to predict some global characteristics (e.g., size distribution, velocity profiles, Tully-Fisher relation) of nearby disk galaxies and damped Lyα systems, based on more or less generic predictions of N-body cosmological simulations (MMW98). Secondly, once the radial mass distribution is known, the vertical distribution of mass in the disk is determined using an extension of the formalism of Bahcall (1984) and Dove & Thronson (1993). These two solution processes are described in §§ 4.1 and 4.2, respectively.

In this section, and throughout the rest of this work, the symbols R and Z are used to represent the (cylindrical) galactocentric radius and the vertical distance from the midplane, respectively; the lowercase r is reserved for spherical coordinates.

4.1. *The MMW98 Solution for the Radial Mass Distribution of a Protogalactic Disk*

The MMW98 model operates in the "adiabatic limit," in which the galaxy formation process conserves an "adiabatic invariant" (Barnes & White 1984). The final configuration is thus made independent of the assembly process, making it possible, for a known initial mass distribution, to determine the final configuration, provided that the form of the final disk mass distribution is specified.

Following MMW98, the dark matter and baryonic matter are assumed to follow the same initial mass distribution, specifically the fitting formula proposed by Navarro et al. (1997):

$$\rho_h(r) = \frac{\rho_{\rm crit}\delta_c}{r/r_h(1 + r/r_h)^2}, \tag{8}$$

which has been shown to describe equilibrated halos in ΛCDM cosmological simulations.

The halo scale length, r_h, is defined in relation to the quantity r_{200}, which is chosen such that the mean density within a sphere centered on the halo and with radius r_{200} is 200 times the critical density for closure, $\rho_{\rm crit}(z) = 3H^2(z)/8\pi G$; i.e.,

$$M_{200} \equiv M(r_{200}) = 200\rho_{\rm crit}\frac{4}{3}\pi r_{200}^3. \tag{9}$$

Numerical experiments show r_{200} to approximately demarcate between regions in which material is still infalling and those in which material tends toward equilibrated (radial) orbits (Cole & Lacey 1996); r_{200} and M_{200} are therefore referred to as the virial radius and mass, respectively. (See Bullock et al. [2001b] for a more detailed discussion of the relation between r_{200}/M_{200} and the virial radius/mass.) The precise relation between r_{200} and r_h is set by the concentration parameter, $c = r_{200}/r_h$. The normalization factor δ_c in equation (8) is defined as

$$\delta_c \equiv \frac{200}{3}c^3\left[\ln(1 + c) - \frac{c}{1 + c}\right]^{-1}. \tag{10}$$

Secondly, motivated by the observed distribution of *light* in nearby galaxies (Freeman 1970), the assembled *gaseous* disk is assumed to have an exponential surface density profile, $\Sigma(R)$:

$$\Sigma(R) = \frac{m_d M_{200}}{2\pi R_d^2}e^{-R/R_d}, \tag{11}$$

where the normalization factor is chosen so that the total disk mass is a parametric fraction, m_d, of the total dark plus baryonic mass. Thus, for a given M_{200} and m_d, $\Sigma(R)$ is completely specified by a scale length, R_d.

The iterative MMW98 solution for R_d relates the total energy of the preassembly halo, E, to the angular momentum of the assembled disk, J_d, using the definition of the spin parameter, λ:

$$\lambda^2 \equiv \frac{J^2|E|}{G^2M_{200}^5}, \tag{12}$$

where $J_d = j_d J$ is a parametric fraction, j_d, of the total angular momentum of the preassembly, dark plus baryonic matter halo, J. The final result is the following expression for R_d (MMW98):

$$R_d = (2f_c)^{-0.5}f_R\frac{j_d\lambda}{m_d}r_{200}, \tag{13}$$

where $f_c(c)$ and $f_R(\lambda,\ c,\ m_d,\ j_d)$ are dimensionless factors of order unity, dependent on the exact shape of the initial and final mass distributions, respectively. Wherever directly comparable, our results are indistinguishable from those of MMW98 and Schaye (2004).

4.2. *The Vertical Mass Distribution of the Assembled Disk*

Once $\Sigma(R)$ is known, at a fixed R, the vertical density profile for the disk gas, $\rho_d(R, Z)$, is governed by two equations: the Poisson equation,

$$\frac{1}{R}\frac{\partial}{\partial R}(R\nabla\Phi_R) + \frac{\partial^2\Phi}{\partial Z^2} = 4\pi G(\varrho_h + \rho_d), \tag{14}$$

and the equation of hydrostatic equilibrium (in the Z-direction),

$$\sigma^2\frac{\partial\rho_d}{\partial Z} + \rho_d\frac{\partial\Phi}{\partial Z} = 0. \tag{15}$$

In these equations, $\Phi(R, Z)$ is the gravitational potential, $\Phi_R(R, Z)$ is the radial component of the potential, $\varrho_h(r)$ is the density of the dark matter halo (determined numerically as in Appendix B), and σ is the (thermal) gas velocity dispersion. Since the gas is assumed to be isothermal, σ is taken to be constant throughout the disk.

Using the equation of hydrostatic equilibrium is equivalent to using the first moment in $\overline{v_Z}$ of the collisionless Boltzmann equation, with the velocity dispersion tensor, σ_{ij}^2 (see, e.g., § 4 of Binney & Tremaine 1987), set to $\delta_{ij}\sigma^2$. In the context of noninteracting stellar dynamics (cf. an equilibrated gas), this assumption would introduce an error of order (ZZ_d/RR_d) (Bahcall 1984).

The disk is assumed to be in centrifugal balance in the R-direction, and particles to travel stable, circular orbits, viz., $\nabla\Phi_R = V^2/R$, where $V(R)$ is the orbital velocity of the gas at a

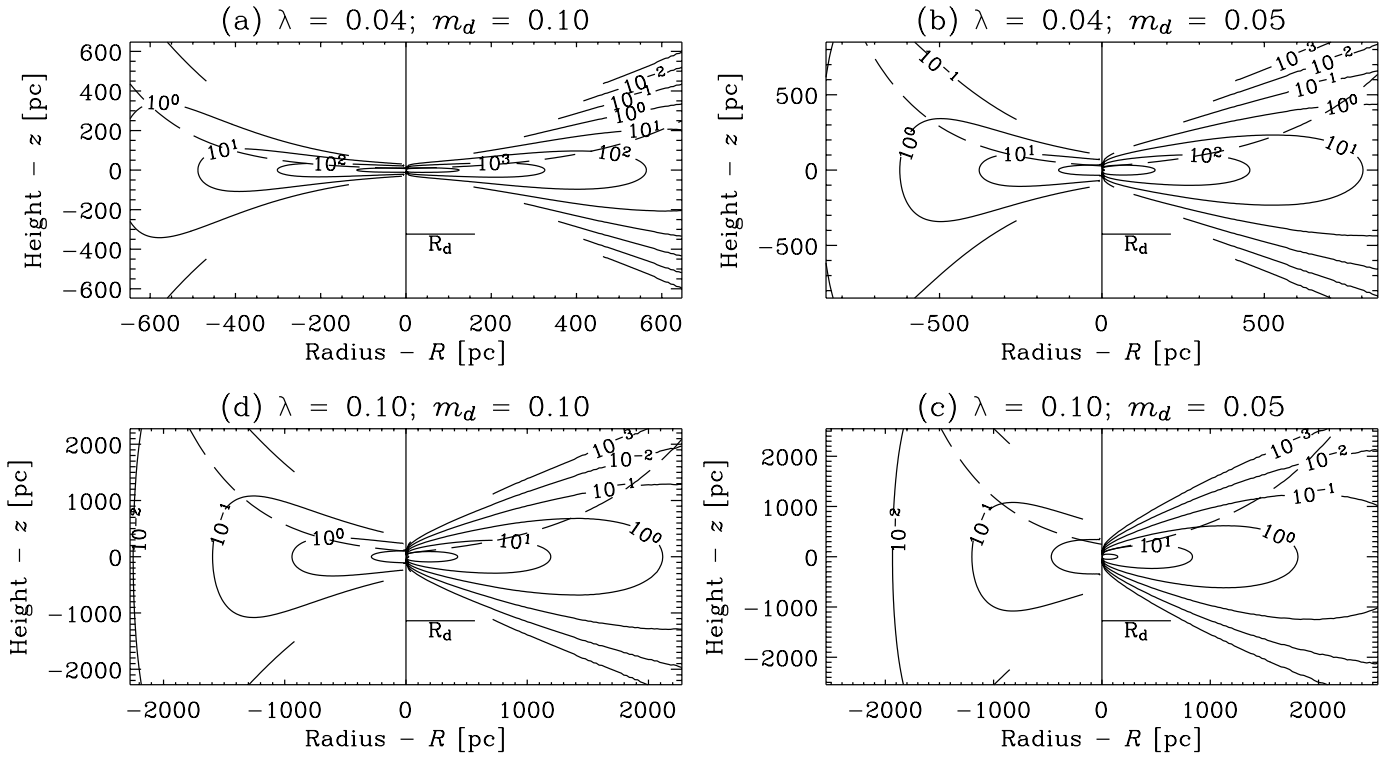

FIG. 2.—Mass distribution cross sections for $M_{200} = 10^9\ M_\odot$, illustrating the importance of the quantity ξ. Contours are of constant density, in units of m_p cm^{-3}. Each panel is $8R_d$ on a side. The disk scale height $Z_d(R)$ is shown by the long-dashed line; profiles are computed out to $4Z_d$. Models where ξ is computed as described in § 4.2 are shown on the right of each frame. For comparison, models with a Spitzer (1942) sech^2 profile are shown on the left.

radius R. With this substitution in equation (14), these two equations can be combined by differentiating equation (15) with respect to Z, to give (Bahcall 1984; Dove & Thronson 1993)

$$\left(\frac{1}{\rho_d}\frac{\partial\rho_d}{\partial Z}\right)^2 - \frac{1}{\rho_d}\frac{\partial^2\rho_d}{\partial Z^2} = \frac{4\pi G}{\sigma^2}\rho_d + \frac{4\pi G}{\sigma^2}\left(\varrho_h - \frac{1}{4\pi GR}\frac{\partial V^2}{\partial R}\right). \quad (16)$$

Following Bahcall (1984) and Dove & Thronson (1993), equation (16) is made dimensionless by defining a scale height for the disk $Z_d(R)$:

$$Z_d(R) \equiv \left[\frac{\sigma^2}{2\pi G\rho_0(R)}\right]^{1/2}, \quad (17)$$

where $\rho_0(R) = \rho_d(R,\ 0)$ is the volume density of the disk gas on the plane, at a distance R from the galactic center. (Note that Z_d is twice the exponential scale height; Bahcall 1984.) This leads (again, for a fixed R) to an ordinary differential equation for the dimensionless density $y(x) = \rho_d(R,\ Z)/\rho_0(R)$, where $x = Z/Z_d$:

$$\frac{d^2y}{dx^2} = \frac{1}{y}\left(\frac{dy}{dx}\right)^2 - 2y^2 - 2\xi(R,\ Z)y, \quad (18)$$

subject to the boundary conditions:

$$y(x=0) = 1, \quad \left.\frac{dy}{dx}\right|_{x=0} = 0. \quad (19)$$

It is worth making a few comments here about the quantity ξ in equation (18), which is defined as

$$\xi(R,\ Z) \equiv \frac{1}{4\pi G\rho_0}\left[4\pi G\varrho_h(r) - \frac{1}{R}\frac{\partial V^2(R)}{\partial R}\right], \quad (20)$$

but which can be rewritten in a more convenient form (Appendix B). If ξ is neglected altogether, then equation (18) is analytic, yielding $y(x) = \mathrm{sech}^2 x$ (Spitzer 1942). This formalism was first derived in the context of stellar dynamics close to the galactic plane and at modest distances from the galactic center, in which case both V and ϱ_h vary slowly across the volume of interest. Authors therefore typically neglect the $\partial V^2/\partial R$ term and treat ϱ_h as constant (e.g., Maloney 1993; Dove & Shull 1994).

This work, however, considers lower mass galaxies with manifestly nonflat rotation curves. Moreover, we are interested primarily in the outermost gas, since this is where the effects of the CBR are greatest; this is also where the effect of a nonzero ξ is most pronounced. It is therefore inappropriate to adopt these standard approximations here. The effect of a nonzero ξ is illustrated in Figure 2, which shows several model gas distributions, as well as the equivalent distribution assuming $y = \mathrm{sech}^2 x$.

In summary, the modeling procedure is as follows: $y(x)$ is determined by numerical integration of equation (18); ρ_0 is then fixed with reference to the definition of $\Sigma(R)$, viz.,

$$\rho_0 = \frac{2\pi G\Sigma^2(R)}{\sigma^2\left[2\int_0^\infty dx\, y(R,\ x)\right]^2}. \quad (21)$$

The presence of ρ_0 in the definition of ξ, as well as the implicit dependence of ξ on Z, makes it necessary to solve for $y(x)$

iteratively: a trial value of ρ_0 is used to determine the value of ξ in the integration of equation (18); the value of the integral is then used in equation (21) to obtain the next ρ_0. This process rarely requires more than several iterations for convergence.

5. THE MODEL. III. THE CHEMICAL COMPOSITION OF THE GAS AND THE CBR

Once the distribution of the disk gas is known, we determine its equilibrium chemical structure. Nine chemical species are identified within the gas, interacting via 21 collisional and nine photodestructive processes, as outlined in § 5.1. The gas is subject to a UV–X-ray CBR field; we describe the CBR parameterization, including gas self-shielding, in § 5.2. The solution for the equilibrium chemical composition of the gas is discussed in § 5.3.

5.1. *Chemical Species and Processes*

Following Haiman et al. (1996b) and Abel et al. (1997), we distinguish nine chemical species within the gas: H^0, H^-, H^+, H_2^0, H_2^+, He^0, He^+, He^{++}, and e^-, assuming a (primordial) helium mass fraction of 0.24. These chemical species are allowed to interact via the 21 collisional processes identified by Haiman et al. (1996b). In addition, the nine photoionization and photodissociation processes listed in Abel et al. (1997) are included. For ionization of He^0 and H_2^0, the more recent cross sections given by Yan et al. (1998) are employed. These expressions give much lower cross sections at high energies than those used by Abel et al. (1997), which reduces the rate of heating in marginally shielded areas but increases the penetration depth. The rate coefficients and cross sections for all of these processes are listed in full in Appendix A.

We completely neglect the effects of any dust or metals present in the gas, even though the IGM is expected to have seen considerable enrichment (Aguirre et al. 2001; Schaye et al. 2000). Metals can make a significant contribution to the cooling rate within the self-shielded region (Galli & Palla 1998; Katz et al. 1996); moreover, both dust and metals act as sites for H_2 production. These effects only become dominant, however, once two-body collisional production of H_2 becomes inefficient, well below the assumed temperature of 10^4 K (Buch & Zhang 1991; Galli & Palla 1998). Further, both dust and metals can absorb UV CBR radiation, thus contributing to the level of self-shielding within the cloud. The neglect of dust and metals therefore minimizes the predicted level of H_2 production and cooling.

For each individual chemical species, an expression can be written for the net rate of production/consumption in terms of the number densities of all nine species. It is most convenient and instructive to cast these expressions in terms of the relative chemical abundances, $x_i = n_i/n_{\rm H}$, where n_i is the number density of species i and $n_{\rm H}$ is the number density of hydrogen nuclei, viz.,

$$\frac{dx_i}{n_{\rm H}\,dt} = \sum_p \left[\sum_q \left(b_{pq}^{(i)} \alpha_{p,q} x_p x_q \right) + d_p^{(i)} \frac{\Gamma_p}{n_{\rm H}} x_p \right]. \quad (22)$$

Here $\alpha_{p,q}$ is the temperature-dependent rate coefficient for the reaction between p and q, Γ_p is the rate of photoionization/photodissociation of p, and $b_{pq}^{(i)}$, $d_p^{(i)} = 0$, ± 1, or ± 2, depending on the number of i particles produced/consumed in the reaction. Note that in the high-density/zero-radiation limit, the equilibrium abundances are independent of density; they are determined by the temperature-dependent α-values alone. Equation (22) encapsulates a particularly "stiff" set of nine highly nonlinear coupled differential equations for the chemical composition of the gas at a fixed point (Haiman et al. 1996b).

5.2. *The Photodestructive Effects of the CBR*

We consider the CBR in the range 13.6 eV $\leq h\nu \leq$ 40 keV (i.e., 912 Å $\geq \lambda \geq$ 0.31 Å; we find that photoionization of He and He^+ by soft X-rays is important in H_2 production, as well as being a significant photoheating mechanism). Within this range, the CBR spectrum, $J_{\rm CBR}(\nu)$, is described using a simple power law, characterized by its spectral index, η, and a normalization constant, J_{21}, chosen such that

$$J_{\rm CBR}(\nu,\ R,\ Z) = J_{21} \left(\frac{\nu}{\nu_{\rm H\,I}} \right)^{-\eta} e^{-\tau(\nu;\,R,\,Z)} \times 10^{-21}\ {\rm ergs^{-1}\ s^{-1}\ Hz^{-1}\ cm^{-2}\ sr^{-1}}. \quad (23)$$

Here $\nu_{\rm H\,I}$ is the frequency of the Lyα photon: $h\nu_{\rm H\,I} \approx 13.6$ eV. This parameterization is the simplest possible; to assess its relevance, we direct the reader to the compendium of CBR observations across the whole electromagnetic spectrum provided by Henry (1999). We test two extreme values of η (see § 7.3), in an attempt to bracket the real situation.

The frequency-dependent optical depth, τ, defined as

$$\tau(\nu;\ R,\ Z) = \sum_i \left[a_i(\nu) \int_Z^\infty dZ'\, n_i(R,\ Z') \right], \quad (24)$$

is used to modulate the CBR spectrum as it penetrates into the cloud, thus mimicking radiative transfer. Here $a_i(\nu)$ is the frequency-dependent cross section for photodestruction of species i, as given in Appendix A. Note that this treatment of radiative transfer is "one way" (no attempt has been made to account for radiation that crosses the midplane), leading to a very minor underestimation of the photoheating rate close to the plane.

We have made three substantial omissions in writing equations (23) and (24). First, any extinction of the CBR due to the presence of dust or metals has been neglected. Secondly, secondary ionizations by energetic electrons from X-ray ionizations are ignored. While these ionizations have no effect on the heating rate, the enhanced electron abundance would increase the H_2 production rate (and the total rate of cooling); the omission of secondary ionizations will therefore again lead to a lower limit on the shielded fraction. Thirdly, and most importantly, diffuse radiation produced by the gas itself has been ignored; that is, any photons (re)radiated by the gas are assumed to be free to escape unhindered. This assumption is justified for the optically thin outermost gas; its validity within the self-shielded region is certainly problematic, but a detailed treatment of radiative diffusion through the cloud would be dependent on both the thermal and dynamic evolution of the disk during its formation. As has been stated, this is beyond the scope of the present undertaking.

For each photoreaction except Solomon photodissociation of H_2 (see below), the photodissociation/photoionization rate, Γ_i, is then given by

$$\Gamma_i = 2\pi \int_{\nu_i}^\infty d\nu\, \frac{J(\nu)}{h\nu} a_i(\nu), \quad (25)$$

where ν_i is the threshold energy for dissociation of species i and the factor of 2π arises from integration over all angles of incidence onto the flat disk.

Photodissociation of H_2 occurs primarily in the Lyman-Werner bands (11.26–13.6 eV) via the two-step Solomon process (Stecher & Williams 1967):

$$H^0 + h\nu \rightarrow H_2^* \rightarrow 2H^0. \quad (26)$$

This process is harder to deal with than other photoprocesses, since the structure of the molecular energy levels (and so the cross section for photodestruction) is so much more complex. Moreover, Doppler broadening causes individual lines to overlap. Following Abel et al. (1997), absorption is therefore restricted to the Lyman band (12.24–13.51 eV); the CBR spectrum is treated as flat in this narrow range. Thus, rather than explicitly considering H_2 dissociating photons longward of 912 Å, the H_2 photodissociation rate is determined with reference to the mean number of photons near 1000 Å. The penetration probability formalism of de Jong et al. (1980), which modulates the unextinguished CBR field with a shielding factor, β, is used to simulate H_2 self-shielding, including the effects of Doppler broadening; the adopted expression for $\beta(\tau_{H_2})$ is given in Appendix A.

5.3. *Chemical Composition in Equilibrium*

Once the radiation density at a given point is known (and hence the rate coefficients for the various photochemical processes), it is possible to find the equilibrium chemical composition at that point by solving the system of equations represented by equation (22), with the time derivatives on the left-hand side set to zero. We do this using a modified Newton-Raphson algorithm (§ 9 of Press et al. 1992), where we compute the Jacobi matrix analytically. For $dx_i/dt = 0$, only six of the nine equations remain independent; it is also necessary to ensure that the total numbers of hydrogen nuclei, helium nuclei, and electrons are conserved separately (Katz et al. 1996).

As a diagnostic for the solution procedure *only*, Figure 3 shows the equilibrium chemical composition (in the absence of photodestruction) as a function of temperature. In this plot, the lines show our steady state solutions for the full set of equations encapsulated by equation (22); for comparison, the points show algebraic solutions for a reduced system of six equations obtained by omitting H^-, H_2^0, and H_2^+. Note that these results differ significantly from those in the directly comparable Figure 2 of Haiman et al. (1996b), which were obtained by direct integration of equation (22) until apparent convergence. In particular, our predicted H_2^0 abundance is 1–2 orders of magnitude lower than that of Haiman et al. (1996b), depending on the reaction rate expressions used (see Appendix A). The disagreement arises from the fact that the results shown in Haiman et al. (1996b) had not yet fully converged (Z. Haiman 2003, private communication; note that this does not affect the conclusions of Haiman et al. 1996b), as can be readily verified by direct substitution.

It should also be noted that, contrary to Haiman et al. (1996a), assuming chemical equilibrium is expected to lead to a lower limit on the actual H_2 abundance at $T = 10^4$ K and by extension on the H_2 cooling rate. The assumed rate coefficient for H_2 production is temperature independent, while that for consumption decreases with temperature; thus, the effect of a small decrease in temperature is an increase in the *net* rate of H_2 production (see, e.g., Fig. 4 of Abel et al. 1997). The net rate of H_2 production, therefore, can only go to zero if the H_2 abundance is dropping (or if the gas temperature is rising). The H_2 abundance in an equilibrating gas will therefore always reach equilibrium *from above*. (For a time-evolved demonstration of this point, see Fig. 4 of Haiman et al. 2000.)

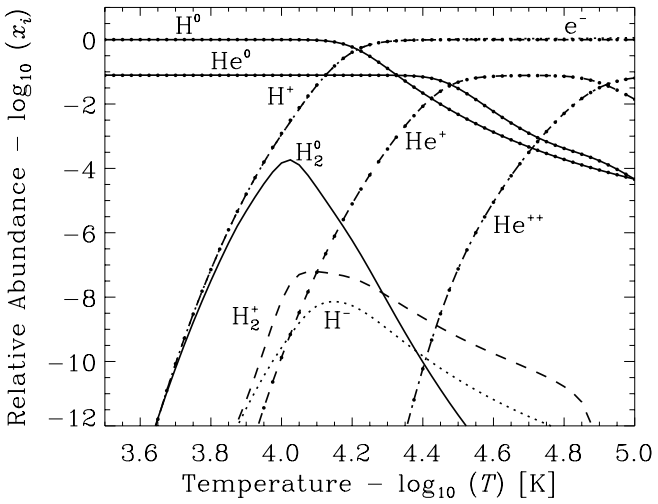

FIG. 3.—Equilibrium abundances of nine chemical species in the absence of any photodestruction. Chemical species are as labeled. The points show the equilibrium solutions for the reduced set of equations in which H^-, H_2^0, and H_2^+ are omitted. All abundances are scaled relative to the total hydrogen nucleus number density, n_H. Note that the electron abundance closely tracks the proton (H^+) abundance. This plot is provided only as a diagnostic for our solution procedure and should not be mistaken for actual evolutionary tracks (see § 5.3). [*See the electronic edition of the Journal for a color version of this figure*.]

Moreover, the situation of chemical equilibrium for H_2 is physically improbable for temperatures between $\sim 10^3$ and 10^4 K because in the absence of a source of photoheating, the gas will be thermally unstable; gas is far more likely to cool past these temperatures without equilibrating. In this sense, we emphasize the limited physical relevance of Figure 3, which is provided only as a diagnostic, and refer the reader to plots of nonequilibrium abundances in cooling and collapsing gas clouds like those found in Figures 3–5 of Oh & Haiman (2002), Figure 3 of Abel et al. (1997), or Figure 6 of Shapiro & Kang (1987).

6. THE MODEL. IV. THE MINIMUM SFR

At each point where H_2 cooling is found to be sufficient to initiate gravothermal instability, we introduce an additional internal radiation field and determine the minimum ISRF required to preserve thermal balance at 10^4 K; it is then possible to relate this ISRF to a minimum global SFR, $\mathcal{R}_{\rm min}$. In other words, we solve for the minimum ISRF-regulated SFR within the cloud.

This is done by including a 10th equation in the system of equations described in § 5.3, requiring that the collisional H_2 cooling rate (Λ_{H_2}, given in Appendix A) be balanced by the net rate of photoheating, viz.,

$$\sum_i \zeta_i - \Lambda_{H_2} = 0. \quad (27)$$

Here ζ_i is the rate of photoheating due to absorption by species i, which (for all processes except Solomon dissociation of H_2; see Appendix A) is given by

$$\zeta_i = n_i \int_{\nu_i}^{\infty} d\nu\,(h\nu - h\nu_i)\frac{J(\nu)}{h\nu}a_i(\nu), \quad (28)$$

where the radiation flux $J = J_{\rm CBR} + cu_d$ now has two components. The first is due to the CBR, where $J_{\rm CBR}$ is defined by equation (23); the second is the new diffuse ISRF, characterized

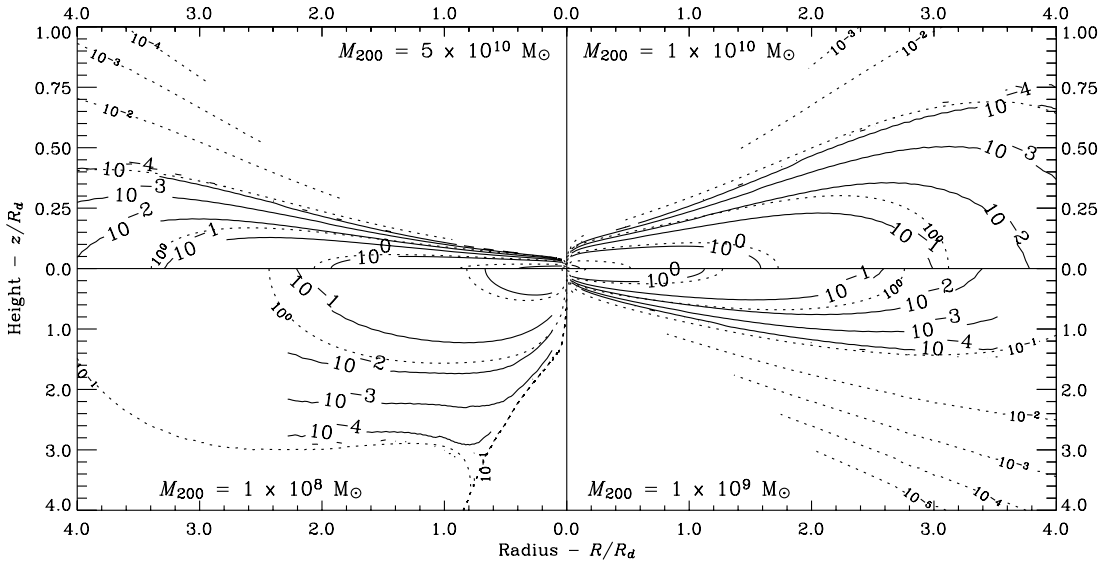

FIG. 4.—Contours of u_d, shown in cross section. Parameter u_d is the minimum ISRF energy density (at 1000 Å) required for thermal balance, determined as per § 6. Contours are shown (in units of the Habing flux) for example model $(\lambda,\ m_d) = (0.04,\ 0.05)$; clockwise from top left are models with $M_{200} = 5\times10^{10}$, 10^{10}, 10^{9}, and $10^{8}\ M_\odot$. Note the different ranges of the vertical axes in the upper (R_d) and lower ($4R_d$) panels. We only solve for u_d at radii where the gas is Toomre unstable at $T \geq 300$ K (see Fig. 1). Overlaid are contours of $n_{\rm H}$ (*dotted lines*), in units of cm^{-3}.

by the diffuse radiation energy density, u_d. Note that we use this same J in equation (25), so that the ISRF maintains thermal balance by some combination of photoheating and H_2 photodissociation.

The spectral shape of u_d is fixed with reference to the stellar population modeling utility Starburst99 (Leitherer et al. 1999),[2] which gives the spectral energy distribution (SED) for continuous star formation, $L_{99}(\nu)$, with units of ergs s^{-1} Hz^{-1} $(M_\odot\ {\rm yr}^{-1})^{-1}$. We adopt a standard SF scenario, which assumes a Salpeter IMF over a mass range of 1–100 $M_\odot$ and a metallicity of 0.020 ($Z_\odot$); our results are not strongly dependent on the choice of metallicity. Over the spectral range of interest, this SED converges after $\lesssim$10 Myr (a few times the lifetime of O–B stars); we use this converged SED. It should be noted that the Starburst99 SED only extends to 100 Å; since self-shielding against the diffuse ISRF is ignored, this will lead to only a minor underestimate of the rate of He I ionization and hence the rate of heating.

To convert this luminosity to a radiation density, it is necessary to make an estimate of the mean lifetime of each photon, $\tilde{t}_\nu$, which is done with reference to the mean free path of a photon with energy $h\nu$, $l_\nu \equiv (na_\nu)^{-1} = \tilde{t}_\nu c$. With these substitutions, u_d can be written as

$$u_d(\nu) = \mathcal{U} L_{99}(\nu)\left[c\sum_i n_i a_i(\nu)\right]^{-1}. \qquad (29)$$

While the spectral shape of u_d is given by L_{99}, its intensity is thus characterized by a scale factor, $\mathcal{U}$, which has the units of SFR per unit volume.

$\mathcal{U}$ is thus taken as the 10th unknown in a system of 10 equations, representing the minimum SFR (per unit volume, and assuming chemical equilibrium) required to produce a sufficient ISRF to maintain thermal balance at $T = 10^4$ K and so restore gravothermal balance. Note that $\mathcal{U}$ gives the SFR required to ensure the stability of the volume dV, not the SFR within dV. Figure 4 shows the resultant ISRF field in cross section for several exemplary models. A global SFR, $\mathcal{R}_{\rm min}$, is then found by integrating $\mathcal{U} = d\mathcal{R}_{\rm min}/dV$ over the whole disk.

Clearly, this is not a solution for the exact problem. That said, $\mathcal{R}_{\rm min}$ is expected to provide a reasonable estimate for the minimum SFR in the cloud, since "maximum damage" has been allowed; photons are delivered directly to wherever they are required. Under these conditions, so long as the actual SFR exceeds $\mathcal{R}_{\rm min}$, the non–star-forming remainder of the disk will remain stable; otherwise, some part of the disk will become unstable and the SFR will increase as a result.

7. THE MODEL. V. CHOICE OF PARAMETERS

The model now outlined consists of three components described by nine parameters. In this section, the fiducial values adopted for these parameters are discussed.

Cosmology enters our calculations via $\rho_{\rm crit}$, which is used to determine r_{200} for a given M_{200} (see eq. [9]; we adopt the now standard *WMAP* cosmology; Bennett et al. 2003; Spergel et al. 2003). Further, model disks are assumed to form at zero redshift. Since $\rho_{\rm crit}$ decreases with decreasing redshift, r_{200}, and hence R_d, will be greater for smaller disk formation redshifts (see eq. [9]); at a fixed R/R_d, this minimizes ρ_d and hence self-shielding and H_2 production, in keeping with the ethos of the model.

7.1. *The Progenitor Halo: M_{200}, c, λ*

The maximum virial mass considered is $M_{200} = 5\times10^{11}\ M_\odot$, typical of a nearby LSB galaxy. The minimum mass, $M_{200} \approx 10^8\ M_\odot$, is set by the form of the analytic model: below this limit, the solution process for Z_d and ρ_0 fails to converge at $R \sim R_d$. Mathematically, the long, slow rise of the rotation curve for lower mass galaxies makes ξ a large negative number at these radii, which eventually drives the vertical density profile, $y(x)$, to increase with distance off the plane (see eq. [20]). In this unphysical case, we truncate the integration and treat the stationary point as a sharp boundary. It may be that lower mass halos do not form exponential disks.

[2] Specifically, we used the data shown in Fig. 2*b*, which is available at http://www.stsci.edu/science/starburst99/.

The concentration parameter, c, is chosen using a simple scaling relation drawn from Figure 6 of Navarro et al. (1996):

$$c = 7.5\left(\frac{M_{200}}{M_*}\right)^{-1/9}, \quad (30)$$

where M_* is the (cosmology dependent) "nonlinear" mass, as defined in that work.

Numerical experiments show that the spin distribution in hierarchical clustering scenarios is well described by a "lognormal" distribution centered at $\lambda_0 = 0.04$ and with width $\delta\lambda = 0.5$ (Warren et al. 1992; Bullock et al. 2001a). Two values for the spin parameter, λ, are adopted to sample this distribution: a "typical" value of 0.04 and a "high" value of 0.10. These values approximately correspond to the 50% and 90% points of the expected distribution (see also MMW98).

7.2. *The Disk: m_d, j_d, σ, T*

Based solely on the *WMAP* cosmology (Spergel et al. 2003), baryons are expected to make up 16% (by mass) of all matter. However, not all of the baryonic material initially within an overdense region will ultimately settle into the disk. Moreover, of the mass that does collect at the center of the potential well, some fraction may be lost, e.g., through evaporation after reionization (Shaviv & Dekel 2003). Two fiducial values for the disk mass fraction are therefore adopted, $m_d = 0.05$ and 0.10 (see also MMW98).

Assuming that the halo's angular momentum originates from tidal torquing, it is natural to assume that both the dark and baryonic components obtain the same specific angular momentum (Fall & Efstathiou 1980); thus, at the time of virialization, $m_d = j_d$. While N-body and hydrodynamic simulations predict that j_d/m_d is significantly less than unity for assembled disks (Navarro & White 1994; Navarro & Steinmetz 1997), the MMW98 model requires that j_d/m_d be close to unity in order to produce disk sizes and properties that are in agreement with observations. Accordingly, we assume $j_d = m_d$ for the assembled disk.

The velocity dispersion, σ, is assumed to be constant, with the value of 8.2 km s^{-1}; this corresponds to the thermal velocity dispersion at $T = 10^4$ K. Our chosen value is larger than the canonical value of 6 km s^{-1}, which, using a recalibrated Toomre criterion ($Q < 2$), has been shown to predict the location of star formation thresholds in around $\frac{2}{3}$ of nearby spiral galaxies (Kennicutt 1989; Martin & Kennicutt 2001). However, Schaye (2004) has recently suggested that this agreement is largely coincidental: for typical high surface brightness disk galaxies only, the point where the disk is Toomre unstable assuming $c_s = \sigma =$ 6 km s^{-1} fortuitously coincides with where the transition to the cold phase becomes possible.

7.3. *The CBR Field: J_{21}, η*

The CBR parameters are fixed using the value of the low-redshift H I photoionization rate, $\Gamma_{\rm H\,I}$, obtained by Scott et al. (2002). They give a best-fit value of $\Gamma_{\rm H\,I} = 2.0^{+9.0}_{-1.4}\times10^{-13}$ s^{-1} for $z < 1$, where $\Gamma_{\rm H\,I}$ is related to the UV–X-ray CBR spectrum by

$$\Gamma_{\rm H\,I} = 4\pi\int_{\nu_{\rm H\,I}}^{\infty} d\nu\,\frac{J_{\rm CBR}(\nu)}{h\nu}a_{\rm H\,I}(\nu). \quad (31)$$

Here $a_{\rm H\,I}$ is the known cross section of H I (see Appendix A). For a power-law spectrum, this integral can be performed analytically, which makes it possible to fix the normalization constant J_{21} for a given spectral index η.

We consider two fiducial values for η. The first, $\eta = 2$, mimics a typical quasar spectrum (Scott et al. 2002); the second, $\eta = 0.7$, is chosen to match cosmological simulations of the Lyα forest and the observed X-ray CBR up to 40 keV (Haiman et al. 1996a). Using the Scott et al. (2002) value for $\Gamma_{\rm H\,I}$, these two choices correspond to values for J_{21} of 0.091 and 0.071, respectively.

8. RESULTS

We determine the amount and distribution of gas (assuming chemodynamical equilibrium at $T = 10^4$ K) that is subject to H_2 cooling in the presence of the CBR. Where the cooling rate due to H_2 processes exceeds the total rate of heating due to the CBR (i.e., where the gas is thermally unstable at 10^4 K), the gas is deemed "shielded." If this gas is also Toomre unstable at $T \geq 300$ K (i.e., if it will be gravothermally unstable in the cold phase), then it is deemed "unstable;" this gas is then prone to star formation. Additionally, by determining the ISRF intensity required to restore thermal balance at $T = 10^4$ K at each unstable point (thus restoring the situation of gravothermal stability), we estimate the minimum global SFR for each model.

A few trial models have been recalculated with greater precision to ensure that numerical effects do not significantly alter our results; our principal results, shown in Figures 6 and 7 below, are convergent to better than 1%.

8.1. *Localized Properties of Model Disks*

In Figure 5 we present some of the "local" scaling relations that arise within the models. Within each panel, the two distinct curves in regions of marginal shielding correspond to the two choices of CBR spectral index: the upper curve to $\eta = 2.0$, the lower one to $\eta = 0.7$.

First, in Figure 5*a*, the mass fraction of shielded gas is plotted as a function of surface density. These results are in good agreement with the first estimate made in § 2.2. A universal vertical density profile would ensure that all columns with the same Σ would be self-similar; the scatter in this plot (particularly in the $\eta = 2.0$ case, which has a higher proportional X-ray flux) therefore demonstrates the importance of detailed modeling of the vertical gas distribution. Within the scatter, the trend with increasing mass is right to left; more massive disks are better able to self-shield.

In Figure 5*b*, the minimum ISRF intensity required to maintain thermal balance at 10^4 K, u_d, is plotted as a function of gas density. The ISRF is presented in units of the Habing "flux" (the observed local ISRF energy density at 1000 Å), $g = 40\times10^{-18}$ ergs Å^{-1} cm^{-3} $= 1.33\times10^{-29}$ ergs Hz^{-1} cm^{-3} (Habing 1968). Within the self-shielded region, $u_d = 0.16 n_{\rm H}^{0.75} g$; this relation has been overplotted.

Finally, in Figure 5*c*, the minimum SFR (per unit projected area) required to preserve thermal balance, $\mathcal{R}_{\rm min}/dA$, is plotted as a function of H I column density. Again, within the self-shielded region, the curves converge toward a power law: $\mathcal{R}_{\rm min}/dA \approx 1.1\times10^{-14}(N_{\rm H\,I}/10^{21}\ {\rm cm}^{-2})^3 = 2.2\times10^{-17}\Sigma^3_{\rm H\,I}$. The results plotted in this panel agree very well with those of Corbelli et al. (1997). Note that surrounding the completely self-shielded region is one in which $\mathcal{R}_{\rm min}/dA$ exceeds this power law: in this region, the CBR actually *aids* SF, as the increased electron abundance due to X-ray ionization of He I enhances H_2 production.

Since thermal balance within the self-shielded region is achieved primarily through photodissociation of H_2 (rather than

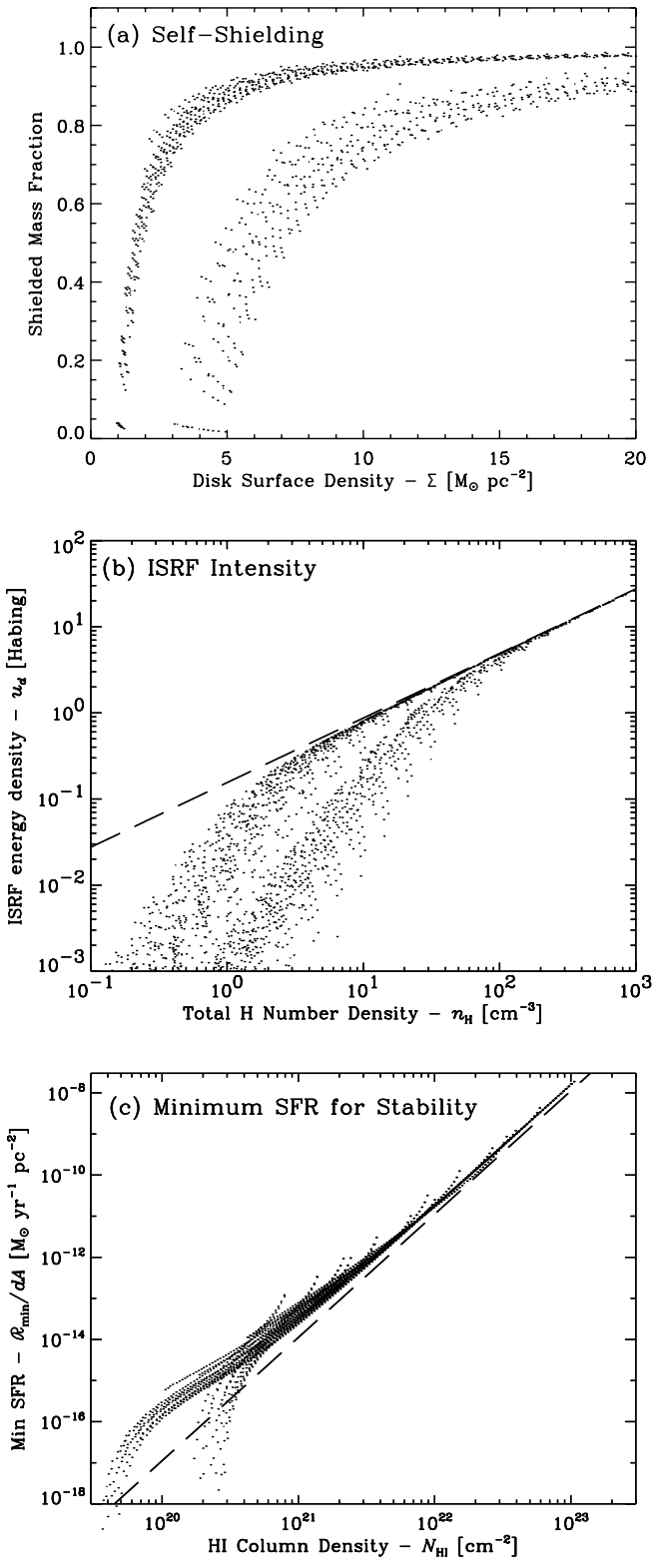

FIG. 5.—Local scaling relations arising in the model: (*a*) the fraction (by mass) of gas that is shielded from the CBR as a function of surface density; (*b*) the minimum ISRF intensity required to maintain thermal balance at 10^4 K as a function of density; (*c*) the minimum SFR required to maintain thermal balance as a function of H I column density. Representative points are drawn from each of the 32 parameter sets we have modeled. Within each panel, the two distinct curves correspond to $\eta = 2.0$ (*upper curves*) and 0.7 (*lower curves*). *Panel (c) should not be interpreted in terms of a Schmidt relation.* The SFR plotted in panel (*c*) is the SFR required elsewhere in the cloud to prevent SF in a column of density $N_{\rm H\,I}$; it is not the SFR expected in that column. [*See the electronic edition of the Journal for a color version of this figure.*]

heating by the CBR), the results of the ISRF calculation are sensitive to the H_2 photodissociative flux alone. The predictions for u_d are therefore robust, in that they should not be sensitive to the assumed ISRF spectral shape or to the exact distribution of the gas. However, the link between u_d and $\mathcal{R}_{\rm min}/dV$ is more tenuous, since the exact relation depends on the extent, location, and character of SF within the cloud.

8.2. *Integrated Properties of Model Disks*

Figure 6 shows the fractional mass of gas available for star formation for each of the 32 parameter sets that we have modeled. All else being equal, high-spin galaxies are less able to self-shield than those with lower spin. In the high-spin case, R_d is typically 8–10 times greater than for low spin, which reduces the surface density at a fixed R/R_d by a factor of order 50–100.

Also, a CBR spectrum with a shallower spectral index has a greater proportional X-ray flux and so is able to penetrate deeper into the cloud. At a fixed point, this increases the rate of photoheating; at the same time, the free electrons from helium ionizations lead to an increased H_2 production rate. Thus, in the $\eta = 0.7$ case, even though the CBR aids SF near the shielding boundary, this boundary is pushed deeper within the gas.

Recalling that the adopted values for λ represent the 50% and 90% points of the expected distribution, the majority of the disk is predicted to be unstable to SF in at least 50% of putative dark galaxies with baryonic masses greater than $5 \times 10^6\ M_\odot$. For baryonic masses greater than $10^8\ M_\odot$, the majority of mass is unstable in as much as 90% of all protogalaxies, depending on the distribution of m_d and the true CBR spectrum.

This result is directly contrary to that of Verde et al. (2002), who have predicted that all disks in halos with $M_{200} < 10^9\ M_\odot$ should remain dark. However, these authors have assumed a constant $c_s = \sigma = 6$ km s^{-1} when calculating the Toomre parameter, which implies $T \approx 8500$ K; in the context of this model, this is tantamount to ignoring H_2 cooling.

The predicted global minimum SFR within the putative dark galaxies, $\mathcal{R}_{\rm min}$, is plotted in Figure 7, as a function of the H I mass, $M_{\rm H\,I}$; in all cases, $M_{\rm H\,I}/m_d M_{200}$ is between 0.56 and 0.70. For comparison, some observed values of $M_{\rm H\,I}$ and SFR for actual dwarf galaxies have been overplotted. We have included a sample of six "relatively nearby" ($D \lesssim 15$ Mpc) Im and seven "more distant" (30 Mpc $\lesssim D \lesssim$ 100 Mpc) Im and Sm galaxies from Hunter et al. (1998); four H I–bright nearby galaxies, two each with "low" ($<10^{-4}\ M_\odot$ yr^{-1}) and "high" ($>10^{-4}\ M_\odot$ yr^{-1}) SFRs, from Young et al. (2003); seven LSB dwarf galaxies and four "normal" dwarfs from van Zee et al. (1997); 121 irregular galaxies from the sample of Hunter & Elmegreen (2004), which span more than 8 mag in absolute magnitude and surface brightness; and ESO 215-G?009, an extremely H I–rich dwarf galaxy, with the highest accurately measured H I mass-to-light ratio to date, $M_{\rm H\,I}/L_B \sim 20$ (Warren et al. 2004).

For each combination of λ and m_d, $\mathcal{R}_{\rm min} \propto M_{\rm H\,I}^{1.4}$. We note that this relation has the same power as the empirical Schmidt SF law, $\mathrm{SFR}/dA \propto \Sigma_{\rm H\,I}^{1.4}$ (Schmidt 1959; Kennicutt 1989, 1998; but see also Boissier et al. 2003). Moreover, even though these samples were selected to include widely disparate populations, these populations are not readily distinguished in Figure 7 for $M_{\rm H\,I} \gtrsim 10^8\ M_\odot$, nor is the exceptional ESO 215-G?009 clearly differentiated from the others.

These results are essentially independent of the CBR spectral index. For the H I fraction, which is determined principally by the outermost gas in the cloud, this is because the CBR spectrum has been normalized using $\Gamma_{\rm H\,I}$. On the other hand, $\mathcal{R}_{\rm min}$ is

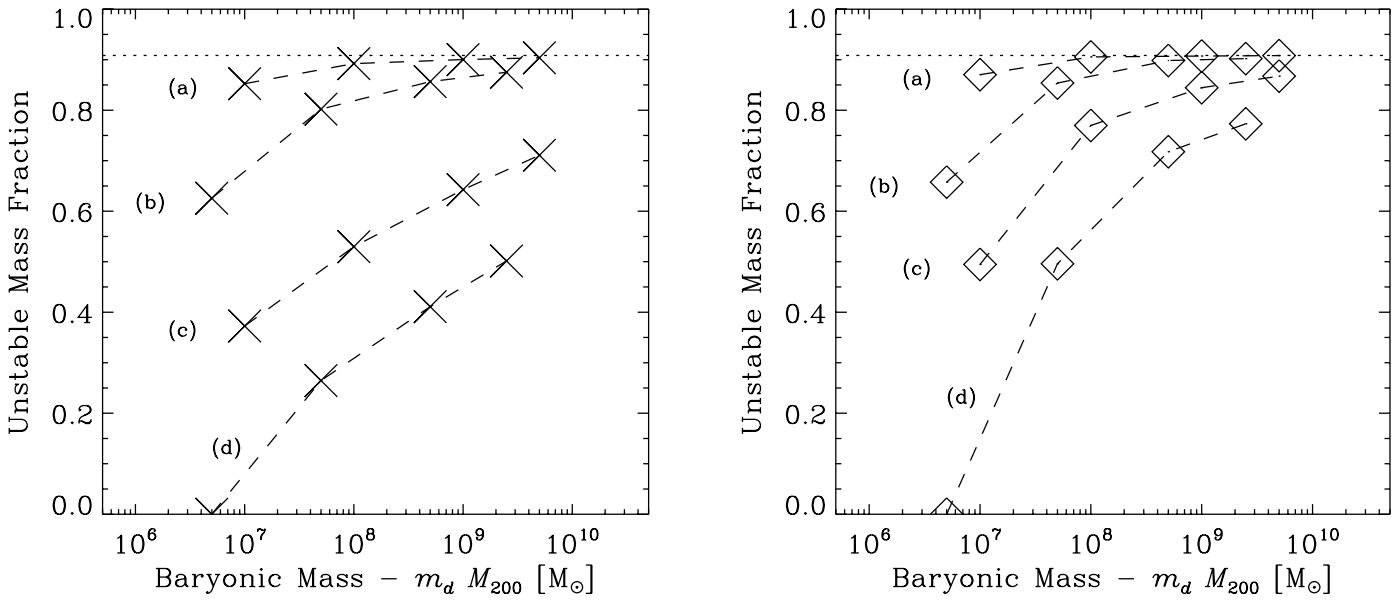

FIG. 6.—Fractional mass that is unstable to SF in the absence of an internal ISRF. The left panel shows models with $\eta = 0.7$; the right shows $\eta = 2.0$. Within each panel, from top to bottom, parameter sets are $(\lambda,\ m_d) = (0.04,\ 0.10)$, (0.04, 0.05), (0.10, 0.10), and (0.10, 0.05). In other words, lower spin galaxies are more unstable than higher spin; then for the same spin, a higher fractional disk mass reduces stability. Note that for an exponential disk, only ~90% of the mass resides within $4R_d$; this point is shown by the dotted line. Even for the "worst-case" CBR, the majority of baryonic matter is unstable to SF in at least 50% of all galaxies with baryonic masses greater than a few times $10^6\ M_\odot$; the same is true for all galaxies with baryonic masses greater than $\sim 10^9\ M_\odot$. [*See the electronic edition of the Journal for a color version of this figure.*]

determined principally by the dense gas closest to the plane of the disk and to the center of the galaxy, where the effect of the CBR is least. In combination with the results in Figure 6, this gives the surprising result that $\mathcal{R}_{\rm min}$ is essentially independent of the amount of gas that is deemed unstable; it depends on $M_{\rm H\,I}$ alone.

9. DISCUSSION

9.1. *The Effects of Substructure and Turbulence*

It is difficult to be quantitative about the effect of substructure, but, qualitatively, self-shielding is increased where the surface density is enhanced. Since, to first order, the CBR prevents a skin of fixed column depth from entering the cold phase, removing matter from a marginally shielded column and adding it to a well-shielded column will increase the overall shielded fraction. That is, the omission of substructure such as bars and turbulent cloudlets should again lead to an underestimate of the amount of gas prone to SF.

Wada et al. (2002) have shown that statistically stable turbulence, leading to a multiphase interstellar medium (ISM), can be produced and sustained with no stellar energy input. It is conceivable that this turbulence could dominate the velocity dispersion, hence providing a third mechanism for support against gravitational collapse. The turbulent velocity dispersion required to support gas at $T \sim 300$ K, however, is typically more than 15 km s^{-1}, whereas in the model these authors present, the turbulent velocity dispersion is only 1–5 km s^{-1}.

Moreover, Wada & Norman (2001) have shown that the timescale for cooling is far shorter than the several dynamical times required to establish this turbulent motion. In their model, the central gas cools within 0.1 Myr, whereas turbulent effects first become apparent only after ~2 Myr and stabilize after ~10 Myr. While turbulent effects are not unexpected in these galaxies, we therefore expect them to be insufficient to stabilize the gas and to emerge only after the first bout of star formation.

9.2. *An Important Sanity Check*

The crux of our argument is the assumption that, if and when H_2 cooling is initiated, the gas will rapidly evolve into a cold phase at $T \sim 300$ K. In § 2.3 we have argued that this transition will be fast in comparison to the dynamical timescale; we are now in a position to verify that this transition will, in fact, lead to Toomre instability.

Detailed integration of the coupled rate equations encapsulated by equation (22) has found that gas cooling from $T \gtrsim 10^4$ K to $T \lesssim 300$ K develops a nonequilibrium abundance of H_2, $x_{\rm H_2} \sim 10^{-3}$ (Shapiro & Kang 1987; Susa et al. 1998; Oh & Haiman 2002), nearly independently of the initial conditions. This occurs at $T \sim 4000$ K, when (two-body) collisional H_2 production and consumption processes become inefficient, leaving an unreplenishing freezeout abundance of H_2. In the presence of a photodestructive radiation field, these molecules can be quickly destroyed, preventing further cooling. Oh & Haiman (2002) have shown that, because of the scalings in the problem, the final temperature that a parcel of gas (at fixed n) is able to cool to depends only on the ratio J/n.

Oh & Haiman (2002) describe a method for estimating the minimum temperature achieved by a cooling parcel of gas, $T_{\rm min}$, by finding the temperature for which the timescale for cooling, t_c, becomes longer than that for photodissociation, $t_d = \Gamma_{\rm H_2}^{-1}$. With the approximation $\Lambda_{\rm H_2} \propto T^4$, and assuming the freezeout abundance, $x_{\rm H_2} = 10^{-3}$, this leads to the following expression for $T_{\rm min}$:

$$T_{\rm min} \approx 9000 \left(\frac{\beta J_{21}}{n}\right)^{1/3} \ {\rm K}. \tag{32}$$

(Recall that β is the H_2 self-shielding factor, discussed in § 5.2.) It should be noted that this analysis does not include dust-catalyzed H_2 production at $T \lesssim 3000$ K.

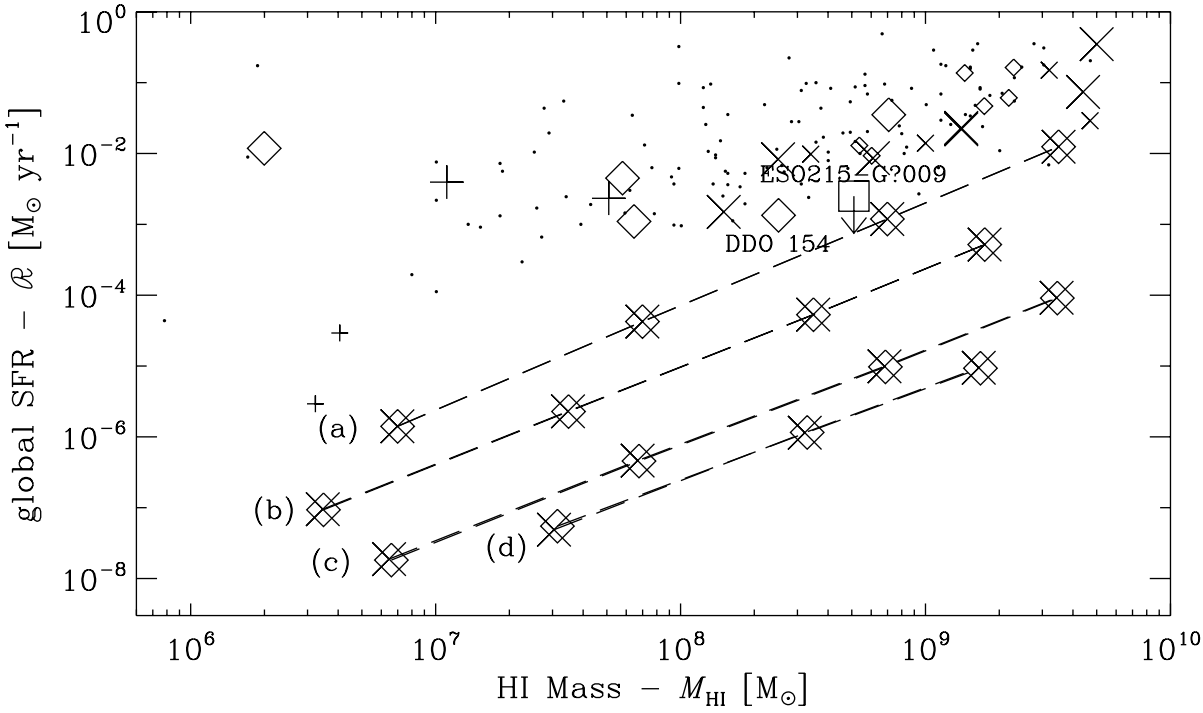

FIG. 7.—Minimum SFR required for stability, $\mathcal{R}_{\rm min}$, plotted against H I mass, $M_{\rm H\,I}$. As in Fig. 6, like models are connected with a dashed line to guide the eye: from top to bottom, models are (λ, m_d) = (0.04, 0.10), (0.04, 0.05), (0.10, 0.10), and (0.10, 0.05); $\eta = 0.7$ (2.0) models are plotted with crosses (diamonds). Note that the top and bottom pairs of lines represent the 50% and 90% points of the expected spin distribution, respectively, and that we have deliberately attempted to place a firm lower bound on $\mathcal{R}_{\rm min}$. For comparison, observed values of $M_{\rm H\,I}$ and SFR for a variety of dwarf galaxies are overplotted, including "nearby" (*large plus signs*) and "distant" (*small plus signs*) dwarfs from Hunter et al. (1998); "low" SFR (*small crosses*) and "high" SFR (*large crosses*) nearby H I–bright dwarfs from Young et al. (2003); "LSB" (*small diamonds*) and "normal" (*large diamonds*) dwarfs from van Zee et al. (1997); 121 members of a sample of irregular galaxies (*points*), spanning a large range in galactic parameters, from Hunter & Elmegreen (2004); and the extremely H I–rich dwarf ESO 215-G?009 (*large square*; Warren et al. 2004). See § 9.4 for a discussion of the significance of this relation. [*See the electronic edition of the Journal for a color version of this figure*.]

In Figure 8 we plot $T_{\rm min}$ (for gas in unstable regions) as a function of the critical temperature below which the gas becomes Toomre unstable, T_Q. Points that lie above the line $T_{\rm min} = T_Q$ represent regions where the gas will transit to a Toomre stable cold phase, in violation of the assumption on which we base our SF criterion. There do appear to be very rare violations of this assumption; however, these points are those closest to the shielding boundary (far from the midplane), in the most Toomre stable regions (far from the center), and make an insignificant contribution to the important integrated quantities.

9.3. *Implications for Star Formation*

Our model argues for a subtle but significant alteration to the (often unstated) conventional wisdom concerning SF: that H I is the raw material from which stars form. Within this paradigm, H I is commonly referred to as "unprocessed" gas, and it is common and reasonable to interpret the Schmidt "law" as a *causal* relation between an amount of gas and the level of SF.

The primary role of H_2 cooling in this picture lies in the fragmentation of a collapsing gas cloud into protostellar cloudlets. We have shown, however, that H_2 can also be important in initiating the first Toomre/Jeans instability. This is particularly important in the case of low-mass galaxies, which would remain Toomre stable if their thermal evolution were governed by H I alone.

The importance of this slight revision lies in the interpretation of H I observations. The picture outlined above interprets H I as, in some sense, the "original" state of hydrogen. In the alternate scenario, however, H_2 can be seen as a "default" state for hydrogen: wherever possible, H I will always develop into H_2. Moreover, since H_2 cooling virtually guarantees Toomre instability, some external source of UV radiation is required to prevent the production of stars from within the gas by dissociating H_2 into H I. In this new paradigm, H I is thus a *product* of SF rather than being its primary *fuel*, and the Schmidt relation is *symptomatic* of the mode of SF.

In other words, the Schmidt relation should not be interpreted as "given more H I, more stars will form," but instead as "more UV radiation (produced by short-lived O–B stars) is required to support a bigger H I cloud." (For a detailed exposition of this argument see Allen 2002.)

This revision is not so controversial: the importance of H_2 cooling in low-mass galaxy formation has been well established by works such as Peebles & Dicke (1968), Lepp & Shull (1984), Haiman et al. (1996b), and Kepner et al. (1997). Although these works are typically (but not exclusively) focused on the effect that radiative feedback from the first generation of stars has on galaxy formation, they clearly demonstrate the importance of H_2 in the global thermal and dynamic evolution of galaxies before protostellar fragmentation. Closer to home, the established physics of photodissociation regions (for a comprehensive review, see Hollenbach & Tielens 1999) demonstrate the importance of the UV ISRF in producing H I in our Galaxy.

9.4. *The $M_{\rm H\,I}$-SFR Relation and DDO 154–like Galaxies*

Our model predicts a hitherto undocumented relation between H I mass and SFR, which seems to hold over the range $M_{\rm H\,I} \sim 10^8$–10^{10} $M_\odot$: SFR $\propto M_{\rm H\,I}^{1.4}$. Within the model, this relation is a

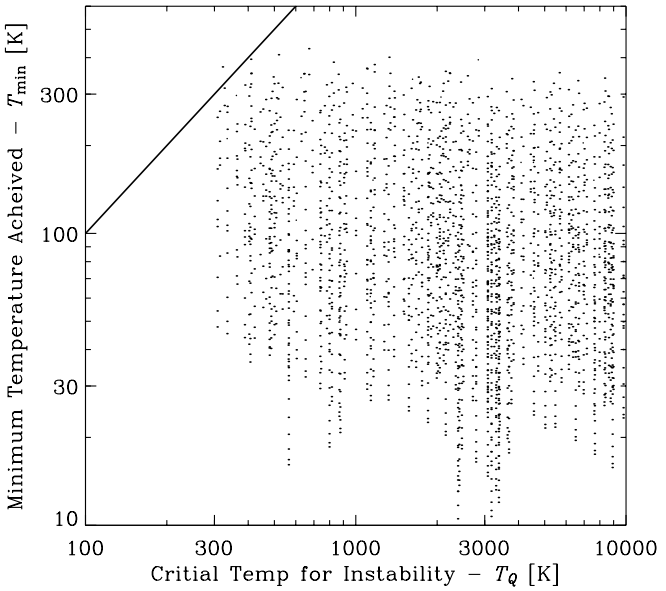

FIG. 8.—Minimum temperature achieved by cooling gas, T_{min}, plotted against the critical temperature below which the gas becomes Toomre unstable, T_Q, for representative points in regions deemed unstable. Where $T_Q > T_{min}$, gas is able to make an evolutionary transit to a Toomre unstable cold phase, in line with the argument presented in § 3. This is essentially always true.

manifestation of self-regulating SF (Lin & Murray 1992; Omukai & Nishi 1999), in which the ISM is kept warm and stable by a UV ISRF that is maintained by constant regeneration of O–B stars.

The slope of the relation is driven by the distribution of the densest gas in galaxy centers, since it is this gas that requires the most radiation to maintain thermal stability. The scatter is set primarily by the distribution of galactic spins, as well as variations in the relative mass of the disk in comparison to the halo, since, for a fixed mass, these determine the density near the center.

Because the model was designed to place a lower limit on the SFR, it cannot predict the zero point of such a relation. Within the self-shielded region, the H_2 cooling rate is typically less than 1%–10% of the net cooling rate; a significantly higher ISRF might be required to balance these other cooling mechanisms. Moreover, if there were, for example, a systematic run with mass in the relative contribution of metal line cooling in comparison to H_2 cooling, the inclusion of metals might alter the predicted slope of the relation. We stress that this would only be true, however, if metal line cooling were important in *initiating* the transition from warm to cold.

The general agreement between the character of the predicted and observed $M_{H\,I}$-SFR relations lends weight to the physical picture on which this simple model is based: the results in Figure 7 suggest that the well-observed population of dwarf galaxies represents the minimum rate of ISRF-regulated SF in galaxies.

Moreover, in the range $M_{H\,I} \sim 10^8$–$10^{10}\ M_\odot$, it seems that the majority of dwarfs are forming stars in the same way: "extreme" objects like DDO 154 and ESO 215-G?009 are not distinguished from "normal" dwarfs in the $M_{H\,I}$-SFR plane. In response to the question "why was so little gas processed into stars?," therefore, these galaxies' high H I mass-to-light ratios are not due to inefficiencies in the present mode of SF. Instead, they may be, for example, a reflection of late formation times or differing environmental effects.

9.5. *The Complete Prevention of SF*

Figures 6 and 7 show that SF is completely prevented in only one of the models: $(M_{200},\ \lambda,\ m_d) = (10^8\ M_\odot,\ 0.10,\ 0.05)$. As is apparent in Figure 1, this is due primarily to rotational support rather than an inability to self-shield: the gas is self-shielding at $R \approx 1R_d$ ($2.5R_d$) for $\eta = 0.7$ (2.0) but is never Toomre unstable at $T = 300$ K. However, if the gas were able to cool to ~30 K (i.e., if it could initiate metal line cooling), then up to 70% of the gas would become available for SF, and we would predict $\mathcal{R}_{min} \sim 10^{-9}\ M_\odot\ yr^{-1}$.

9.6. *The Detectability of Dark Galaxies*

Throughout this work, since the focus has been on gravothermal stability in the absence of internal SF, it has been implicitly assumed that the gas is predominantly H I (and not H II or H_2). Clearly, a very large stellar component would be necessary to completely ionize the gas, and it has been shown that the gas becomes gravothermally unstable in the presence of small amounts of H_2. It has also been argued that, once SF begins, the ISRF will regulate the SFR by dissociating H_2 into H I.

H I is detectable by the hyperfine transition at ~21 cm, provided that the excitation temperature, T_{spin}, exceeds the mean background radiation temperature, $T_{bg} \approx 3$ K. T_{spin} can be written as (Kulkarni & Heiles 1988)

$$T_{spin} = \frac{T + yT_{bg}}{1 + y}, \tag{33}$$

where, neglecting H I excitation by electron collisions, $y \sim (T/1000\ \mathrm{K})/(n_{H\,I}/0.2\ \mathrm{cm^{-3}})$. From Figure 5, it is apparent that $n_{H\,I}$ is essentially always in the range 10^{-1} to 10^2 cm^{-3} within the self-shielded region, in which case $T_{spin} = 800$–9800 K. Moreover, since H I is so optically thick to Lyα, even the low predicted SFRs act to thermalize T_{spin} via scattered Lyα photons.

Without exception, all models are thus predicted to have detectable H I emission, independently of their optical properties. Joint optical plus 21 cm surveys are therefore predicted to yield complete samples of galaxies.

10. THE EXISTENCE OF DARK GALAXIES: SUMMARY

We have developed a model for the long-time configuration of a baryonic dark galaxy, in order to determine whether such a galaxy, having formed, could conceivably remain dark. The model has the following features:

1. The gas is in a dynamically equilibrated disk (rotationally supported in the radial direction and in hydrostatic equilibrium vertically) at the end of the galaxy formation process.
2. The gas is isothermal at $T = 10^4$ K, which represents the end point of its thermal evolution in the absence of H_2 cooling. Together, the assumptions listed above limit the model to considering galaxies with $M_{200} \gtrsim 10^8\ M_\odot$.
3. The gas is in chemical equilibrium, in the presence of a UV–X-ray CBR. The CBR can prevent SF by preventing the gas from undergoing a transition (driven by H_2 cooling) from a warm, largely stable phase to a cold, mostly unstable phase.

We evaluate the stability of this situation of thermal, chemical, and dynamical equilibrium by determining whether or not H_2 cooling can induce gravothermal instability, and hence SF, in the face of the CBR. Additionally, wherever an ISRF is required (in addition to the CBR) to counterbalance H_2 cooling, we place an approximate lower bound on the expected SFR.

We find that, in the absence of an internal UV radiation field, gas that is sufficiently self-shielded to support an H_2 abundance as low as 10^{-4} is unable to remain gravothermally stable. This statement remains true independently of any other cooling processes, including metal line cooling: so as to place a firm lower bound on the amount of gas that will become gravothermally unstable, we have ignored the significant shielding and cooling effects of dust and metals when calculating the H_2 cooling rate at 10^4 K. Moreover, if the gas is even mildly enriched, then its final temperature drops by an order of magnitude from ~300 to ~30 K (Norman & Spaans 1997); particularly for the models most affected by the CBR, this could significantly reduce stability against SF (Fig. 1).

Our conclusions can be summarized as follows:

1. The model predicts a correlation between H I mass and SFR, which is observed in a wide variety of dwarf galaxies (see Fig. 7). This suggests that dwarf galaxies represent the minimum levels of ISRF-regulated SF in the universe. Further theoretical and observational studies of this relation should prove fruitful.

2. With only one exception, the modeled galaxies cannot avoid SF indefinitely and so will develop a stellar component. Even for baryonic masses as low $5\times10^6\ M_\odot$, the majority of gas in greater than 50% of all halos is unstable to SF in the absence of an internal ISRF (Fig. 6).

3. Without exception, all models would have detectable H I emission. That is, whatever dark or dim galaxies exist should be detectable in HIPASS-class surveys.

Above the detection limit of $M_{\rm H\,I}\sim7\times10^6\ M_\odot$ (Meyer et al. 2004), HIPASS did not detect any extragalactic H I clouds in the Local Group that did not also have a stellar component. Similar results are available for other groups and clusters (Zwaan 2001; Waugh et al. 2002). Within these limits, there is no reason to suspect a population of undetected baryonic dark galaxies.

We wish to thank Stuart Wyithe for many fruitful discussions in the course of this work and Joop Schaye for his helpful and insightful responses to an early draft. We also thank the anonymous referee, whose comments led to the calculation presented in § 9.2. E. N. T. was supported by a University of Melbourne–CSIRO collaborative grant while this work was being completed.

APPENDIX A

CHEMICAL PROCESSES IN THE DISK

The full list of reactions considered in this study is given in Table 1, along with their rate coefficients. All quantities in this section are listed in cgs units. The symbol T_n is used to denote T^{-n}. Note that the rate coefficient for reaction 21 has two terms, the second term arising from dielectric recombination (Cen 1992). The rates given for reactions 16 and 17 both are valid in the low-density ($n_{\rm H} < 10^4$ cm^{-3}) limit.

Also note that three of the expressions given in Table 3A of Haiman et al. (1996b) differ from those given in the works that those authors cite. They give the exponent of T_3 in the expression for the rate coefficient of reaction 2 as 0.2; it should read -0.2 (this typographical error is not present in Haiman et al. 1996a). More seriously, the multiplying factor in the expression for reaction rate 3,

TABLE 1
REACTION RATE COEFFICIENTS

Number	Reaction	Rate Coefficient, α (cm^3 s^{-1})	Reference
1	$H^0 + e^- \rightarrow H^+ + 2e^-$	$5.85\times10^{-11}T^{1/2}e^{-157{,}809.1/T}(1+T_5^{1/2})^{-1}$	1
2	$H^+ + e^- \rightarrow H^0 + h\nu$	$8.40\times10^{-11}T^{-1/2}T_3^{-0.2}(1+T_6^{0.7})^{-1}$	1
3	$H^0 + e^- \rightarrow H^- + h\nu$	$3.23\times10^{-17}T^{1/2}$	2
4	$H_2^0 + e^- \rightarrow H^- + H^0$	$2.7\times10^{-8}T^{-3/2}e^{-43{,}000/T}$	2
5	$H^0 + H^- \rightarrow H_2^0 + e^-$	1.3×10^{-9}	3
6	$H^0 + H^- \rightarrow 2H^0 + e^-$	$5.3\times10^{-20}T^{2.17}e^{-8750/T}$	4
7	$e^- + H^- \rightarrow H^0 + 2e^-$	$4\times10^{-12}Te^{-8750/T}$	4
8	$H^+ + H^- \rightarrow 2H^0$	$7\times10^{-7}T^{-1/2}$	5
9	$H^+ + H^- \rightarrow H_2^+ + e^-$	$4\times10^4T^{-1.4}e^{-15{,}100/T}$, $T > 10^4$ K; $10^{-8}T^{-0.4}$, $T \le 10^4$ K	4
10	$H^+ + H^0 \rightarrow H_2^+ + h\nu$	$3.255\times10^{-20}T - 4.152\times10^{-25}T^2 - 6.157e^{-17}$, $T > 4000$ K; $1.38\times10^{-23}T^{1.845}$, 200 K $\le T \le$ 4000 K	6
11	$H^+ + H_2^0 \rightarrow H_2^+ + H^0$	$2.4\times10^{-9}e^{-21{,}200/T}$	4
12	$H^0 + H_2^+ \rightarrow H_2^0 + H^+$	6.4×10^{-10}	2
13	$H^- + H_2^+ \rightarrow H_2^0 + H^0$	$5\times10^{-6}T^{-1/2}$	5
14	$e^- + H_2^+ \rightarrow 2H^0$	$1.68\times10^{-8}(T/300)^{-0.29}$	2
15	$H_2^0 + H^0 \rightarrow 3H^0$	$6.11\times10^{-14}e^{-29{,}300/T}$, $T > 7390$ K; $2.67\times10^{-15}e^{-(6750/T)^2}$, $T \le 7390$ K	7
16	$H_2^0 + H_2^0 \rightarrow H_2^0 + 2H^0$	$5.22\times10^{-14}e^{32{,}200/T}$, $T > 7291$ K; $3.17\times10^{-15}e^{-(4060/T)-(7500/T)^2}$, $T \le 7291$ K	4
17	$H_2^0 + e^- \rightarrow 2H^0 + e^-$	$4.38\times10^{-10}T^{0.35}e^{-102{,}000/T}$	4
18	$He^0 + e^- \rightarrow He^+ + 2e^-$	$2.38\times10^{-11}T^{1/2}e^{-285{,}335.4/T}(1+T_5^{1/2})^{-1}$	2
19	$He^+ + e^- \rightarrow He^{++} + 2e^-$	$5.68\times10^{-12}T^{1/2}e^{-631{,}515.0/T}(1+T_5^{1/2})^{-1}$	1
20	$He^+ + e^- \rightarrow He^0 + h\nu$	$1.50\times10^{-10}T^{-0.6353} + 1.9\times10^{-3}T^{-3/2}e^{-470{,}000/T}(1+0.3e^{-94{,}000/T})$	1
21	$He^{++} + e^- \rightarrow He^+ + h\nu$	$3.36\times10^{-10}T^{-1/2}T_3^{-0.2}(1+T_6^{0.7})^{-1}$	1

NOTES.—The expressions given for reactions 16 and 17 are valid for $n_{\rm H} < 10^4$ cm^{-3}. All coefficients are given in cgs units. Note that the rate coefficients for reactions 2, 3, and 6 given in Haiman et al. (1996b) disagree with those given in the works those authors cite (see text).

REFERENCES.—(1) Cen 1992; (2) Haiman et al. 1996b; (3) Rawlings 1988; (4) Shapiro & Kang 1987; (5) Dalgarno & Lepp 1987; (6) Rawlings et al. 1993; (7) Lepp & Shull 1983.

given as 5.57, should read 3.23. Also, the exponent of T in reaction rate 6, which should be 2.17, is omitted in Haiman et al. (1996b). These alterations reduce the equilibrium abundances of H_2 by several orders.

For photoionization and photodissociation reactions, we employ the reaction cross sections given by Abel et al. (1997), except for ionization of He^0 and H_2^0, for which we use the more recent expressions given by Yan et al. (1998). The different asymptotic behavior of the expression for He^0 ionization has a significant effect on the chemical structure of our models but does not drastically affect our conclusions. When calculating the cross section for H_2 ionization, we assume all H_2 to be in the para– configuration. This is appropriate only when reactions 5 and 12 are the dominant H_2 production mechanisms; within the model these reaction rates are typically 7 orders of magnitude higher than for reaction 13.

To account for self-shielding of H_2 against Lyman-Werner photons, including the effects of Doppler broadening and line overlap, we use the shielding factor of de Jong et al. (1980):

$$\beta(\tau) = \left\{ \left[\frac{\tau^2}{2} \ln\left(\frac{\tau^2}{\pi}\right) \right]^{-1/2} + \left(\frac{\tilde{a}}{\tau}\right)^{1/2} \right\} \mathrm{erfc} \left[\left(\frac{\tau \tilde{a}}{\pi a_{H_2^0}} \right)^{1/2} \right],$$

where erfc denotes the complimentary error function and $\tilde{a} = 9.2\times10^{-3}\sigma^{-1}$ cm km s^{-1} is called the Voigt parameter. Following de Jong (1977), we assume that each Solomon dissociation heats the gas by 2.5 eV.

Finally, we use the lengthy expression given by Lepp & Shull (1983) for the rate of H_2 cooling, Λ_{H_2}:

$$\Lambda_{H_2} = \left[\frac{L_{vH}}{1 + (L_{vH}/L_{vL})} + \frac{L_{rH}}{1 + (L_{rH}/L_{rL})} \right] n_{H_2}: \quad L_{vH} = 1.10\times10^{-18} e^{-6744/T},$$

$$L_{vL} = 8.18\times10^{-13} \left(\alpha_{16} n_{H^0} + \alpha_{17} n_{H_2^0} \right),$$

$$L_{rH} = 3.90\times10^{-19} e^{-6118/T,}$$

$$L_{rL} = \left(1.38\times10^{-22} e^{-9243/T} \right) Q_n,$$

$$Q_n = 1.2 (n_{H^0})^{0.77} + \left(n_{H_2^0} \right)^{0.77},$$

where each of the L-values has units of ergs s^{-1}. As given here, this expression is valid only for temperatures above 4031 K. For the calculation presented in § 9.2, we use $\Lambda_{H_2} \approx 4\times10^{-39} T^4$, which is a fit given by Omukai & Nishi (1999) and is a good approximation for 600 K $\lesssim T \lesssim$ 3000 K and $n \lesssim 10^4$ cm^{-3}. For all other cooling processes, we use the expressions given in Haiman et al. (1996b) and references therein.

APPENDIX B

CALCULATING THE QUANTITY ξ

The quantity $\xi(R, Z)$ in equation (16) is related to what Bahcall (1984) calls the effective halo potential, except that any vertical variation in ξ is neglected in that work. The effect of explicitly including ξ in determining the volume density of the disk gas is illustrated in Figure 2. In this section we motivate and describe the method by which we have calculated ξ in our integration of equation (16).

Throughout this section we use the symbols R and Z to represent the (cylindrical) galactocentric radius and the vertical distance from the midplane, respectively; we reserve the lowercase r for spherical coordinates. We also use the symbols M, ρ, and V here to refer to quantities that can be expressed analytically, as distinct from $\mathcal{M}$, ϱ, and $\mathcal{V}$, which must be computed numerically.

As described in § 4.2, ξ is defined as

$$\xi(R,\ Z) \equiv \frac{1}{4\pi G \rho_0} \left[4\pi G \varrho_h(R,\ Z) - \frac{1}{R} \frac{\partial \mathcal{V}^2(R,\ Z)}{\partial R} \right]. \tag{B1}$$

Here $\mathcal{V}^2(R,\ Z)$ is a sum in quadrature of contributions due to the (spherical) dark matter halo and the (cylindrical) baryonic disk, viz. (Freeman 1970; Binney & Tremaine 1987),

$$\mathcal{V}^2(R,\ Z) = \mathcal{V}_h^2(r) + V_d^2(R)$$

$$= \frac{G\mathcal{M}_h(r)}{r} + \frac{G m_d M_{200}}{R_d} 2y^2 [I_0(y)K_0(y) - I_1(y)K_1(y)], \qquad y \equiv \frac{R}{2R_d}, \tag{B2}$$

where I_n and K_n are nth-order modified Bessel functions of the first and second kinds, respectively. Note that there is an implicit thin disk assumption embedded in this expression, since it neglects any variation of V_d with Z.

The final halo mass distribution, $\mathcal{M}_h(r)$, is determined using the method described by MMW98, which assumes that the adiabatic invariant, $rM(r)$ (Barnes & White 1984), is conserved throughout the galaxy formation process. That is, a test particle initially located at a mean radius r_i moves through successive radii r in such a way that, at any given time,

$$r_i M_i(r_i) = r\mathcal{M}(r) = r[\mathcal{M}_h(r) + M_d(R)]. \tag{B3}$$

Here $M_i(r_i)$ is the (assumed) initial dark plus baryonic mass profile, $M_d(R)$ is the (trial) final disk mass profile, and $\mathcal{M}_h(r)$ is the (unknown) adiabatically contracted halo profile, viz.,

$$\mathcal{M}_h(r) = (1 - m_d)M_i(r_i). \tag{B4}$$

Combining equations (B3) and (B4), we obtain an expression that we solve numerically for r_i as a function of r:

$$M_d(r) + \left(1 - m_d - \frac{r_i}{r}\right)M_i(r_i) = 0. \tag{B5}$$

The resultant r_i can then substituted back into equation (B4) and thence equation (B2) to determine $\mathcal{V}$.

Of course, determining $\partial\mathcal{V}^2/\partial r$ is doubly expensive, since this numerical procedure must be repeated twice in order to estimate the derivative; the same is true of ϱ_h:

$$\varrho_h(r) = \frac{1}{4\pi r^2}\frac{d\mathcal{M}_h(r)}{dr}. \tag{B6}$$

However, we can rewrite the halo contribution to the $\partial\mathcal{V}^2/\partial r$ term in equation (B1) as (Kuijken & Gilmore 1989)

$$\begin{aligned}\frac{1}{R}\frac{\partial\mathcal{V}_h^2(r)}{\partial R} &= \frac{1}{R}\frac{\partial r}{\partial R}\frac{\partial}{\partial r}\left[\frac{G\mathcal{M}_h(r)}{r}\right] \\ &= \frac{1}{R}\frac{R}{r}\left[\frac{G}{r}\frac{\partial\mathcal{M}_h(r)}{\partial r} - \frac{G\mathcal{M}_h(r)}{r^2}\right] \\ &= \frac{1}{r}\left[\frac{G}{r}4\pi r^2\varrho_h(r) - (1 - m_d)\frac{GM_i(r_i)}{r^2}\right] \\ &= 4\pi G\varrho_h(r) - (1 - m_d)\frac{GM_i(r_i)}{r^3},\end{aligned} \tag{B7}$$

where we have used equations (B6) and (B4) to get from the second to the third line.

When this expression is substituted back into equation (B1), the ϱ_h terms cancel, and we are left with

$$\xi(R,\ Z) = \frac{1}{4\pi G\rho_0}\left[(1 - m_d)\frac{GM_h(r_i)}{r^3} - \frac{1}{R}\frac{\partial V_d^2(R)}{\partial R}\right]. \tag{B8}$$

In this new expression, both M_h and $\partial V_d^2/\partial R$ (Binney & Tremaine 1987) are known analytically.

Thus, by incorporating a lookup table for $r_i(r)$, we have developed a means of retaining all contributing terms in equation (16), without significantly increasing computation time. Moreover, with our rewritten definition of ξ, we have eliminated the need to compute $d\mathcal{M}_h/dr$ and $\partial\mathcal{V}^2/\partial r$ numerically. We found these modifications to decrease the computation time for the vertical distribution of the gas by a factor of roughly 100.

REFERENCES

Abel, T., Anninos, P., Zhang, Y., & Norman, M. L. 1997, NewA, 2, 181
Aguirre, A., Hernquist, L., Schaye, J., Katz, N., Weinberg, D. H., & Gardner, J. 2001, ApJ, 561, 521
Allen, R. J. 2002, in ASP Conf. Ser. 273, The Dynamics, Structure and History of Galaxies: A Workshop in Honour of Ken Freeman, ed. G. S. da Costa & H Jerjen (San Francisco: ASP), 183
Bahcall, J. N. 1984, ApJ, 276, 156
Barnes, J., & White, S. D. M. 1984, MNRAS, 211, 753
Bennett, C. L., et al. 2003, ApJS, 148, 1
Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
Boissier, S., Prantzos, N., Boselli, A., & Gavazzi, G. 2003, MNRAS, 346, 1215
Bothun, G. D., Impey, C. D., Malin, D. F., & Mould, J. R. 1987, AJ, 94, 23
Braine, J., Duc, P.-A., Lisenfeld, U., Charmandaris, V., Vallejo, O., Leon, S., & Brinks, E. 2001, A&A, 378, 51
Buch, V., & Zhang, Q. 1991, ApJ, 379, 647
Bullock, J. S., Dekel, A., Kolatt, T. S., Kravtsov, A. V., Klypin, A. A., Porciani, C., & Primack, J. R. 2001a, ApJ, 555, 240
Bullock, J. S., Kolatt, T. S., Sigad, Y., Somerville, R. S., Kravtsov, A. V., Klypin, A. A., Primack, J. R., & Dekel, A. 2001b, MNRAS, 321, 559
Carignan, C., & Beaulieu, S. 1989, ApJ, 347, 760
Carignan, C., & Freeman, K. C. 1988, ApJ, 332, L33
Cen, R. 1992, ApJS, 78, 341
Cole, S., & Lacey, C. 1996, MNRAS, 281, 716
Corbelli, E., Galli, D., & Palla, F. 1997, ApJ, 487, L53
Corbelli, E., & Salpeter, E. E. 1995, ApJ, 450, 32
Dalgarno, A., & Lepp, S. 1987, in IAU Symp. 120, Astrochemistry, ed. M. S. Varya & S. P. Tarafdar (Dordrecht: Reidel), 109
de Jong, T. 1977, A&A, 55, 137
de Jong, T., Dalgarno, A., & Boland, W. 1980, A&A, 91, 68

Disney, M., & Phillipps, S. 1987, Nature, 329, 203
Disney, M. J. 1976, Nature, 263, 573
Dove, J. B., & Shull, J. M. 1994, ApJ, 423, 196
Dove, J. B., & Thronson, H. A. 1993, ApJ, 411, 632
Doyle, M. T., et al. 2005, MNRAS, 361, 34
Elmegreen, B. G., & Parravano, A. 1994, ApJ, 435, L121
Fall, S. M., & Efstathiou, G. 1980, MNRAS, 193, 189
Fisher, J. R., & Tully, R. B. 1975, A&A, 44, 151
Freeman, K. C. 1970, ApJ, 160, 811
Galli, D., & Palla, F. 1998, A&A, 335, 403
Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421
Haiman, Z., Abel, T., & Rees, M. J. 2000, ApJ, 534, 11
Haiman, Z., Rees, M. J., & Loeb, A. 1996a, ApJ, 467, 522
Haiman, Z., Thoul, A. A., & Loeb, A. 1996b, ApJ, 464, 523
Henry, R. C. 1999, ApJ, 516, L49
Hollenbach, D. J., & Tielens, A. G. G. M. 1999, Rev. Mod. Phys., 71, 173
Hunter, D. A., & Elmegreen, B. G. 2004, AJ, 128, 2170
Hunter, D. A., Elmegreen, B. G., & Baker, A. L. 1998, ApJ, 493, 595
Impey, C. D., & Bothun, G. D. 1989, ApJ, 341, 89
Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Kennicutt, R. C., Jr. 1989, ApJ, 344, 685
———. 1998, ApJ, 498, 541
Kepner, J. V., Babul, A., & Spergel, D. N. 1997, ApJ, 487, 61
Kilborn, V. A., et al. 2000, AJ, 120, 1342
Krumm, N., & Burstein, D. 1984, AJ, 89, 1319
Kuijken, K., & Gilmore, G. 1989, MNRAS, 239, 571
Kulkarni, S. R., & Heiles, C. 1988, in Galactic and Extragalactic Radio Astronomy, ed. K. I. Kellerman & G. L. Verschuur (2nd ed.; Berlin: Springer), 95
Leitherer, C., et al. 1999, ApJS, 123, 3
Lepp, S., & Shull, J. M. 1983, ApJ, 270, 578
———. 1984, ApJ, 280, 465
Lin, D. N. C., & Murray, S. D. 1992, ApJ, 394, 523
Maloney, P. 1993, ApJ, 414, 41
Martin, C. L., & Kennicutt, R. C., Jr. 2001, ApJ, 555, 301
Meyer, M. J., et al. 2004, MNRAS, 350, 1195
Minchin, R., et al. 2005, ApJ, 622, L21
Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319 (MMW98)
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
———. 1997, ApJ, 490, 493
Navarro, J. F., & Steinmetz, M. 1997, ApJ, 478, 13
Navarro, J. F., & White, S. D. M. 1994, MNRAS, 267, 401
Norman, C. A., & Spaans, M. 1997, ApJ, 480, 145
Oh, S. P., & Haiman, Z. 2002, ApJ, 569, 558
Omukai, K., & Nishi, R. 1999, ApJ, 518, 64
Peebles, P. J. E., & Dicke, R. H. 1968, ApJ, 154, 891
Press, W., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C (2nd ed.; Cambridge: Cambridge Univ. Press)
Rawlings, J. M. C. 1988, MNRAS, 232, 507
Rawlings, J. M. C., Drew, J. E., & Barlow, M. J. 1993, MNRAS, 265, 968
Ryder, S. D., et al. 2001, ApJ, 555, 232
Schaye, J. 2004, ApJ, 609, 667
Schaye, J., Rauch, M., Sargent, W. L. W., & Kim, T. 2000, ApJ, 541, L1
Schmidt, M. 1959, ApJ, 129, 243
Scott, J., Bechtold, J., Morita, M., Dobrzycki, A., & Kulkarni, V. 2002, ApJ, 571, 665
Shapiro, P. R., & Kang, H. 1987, ApJ, 318, 32
Shaviv, N. J., & Dekel, A. 2003, preprint (astro-ph/0305527)
Spergel, D. N., et al. 2003, ApJS, 148, 175
Spitzer, L., Jr. 1942, ApJ, 95, 329
Stecher, T. P., & Williams, D. A. 1967, ApJ, 149, L29
Susa, H., Uehara, H., Nishi, R., & Yamada, M. 1998, Prog. Theor. Phys., 100, 63
Taylor, C. L., Brinks, E., Pogge, R. W., & Skillman, E. D. 1994, AJ, 107, 971
Toomre, A. 1964, ApJ, 139, 1217
van Zee, L., Haynes, M. P., Salzer, J. J., & Broeils, A. H. 1997, AJ, 113, 1618
Verde, L., Oh, S. P., & Jimenez, R. 2002, MNRAS, 336, 541
Wada, K., Meurer, G., & Norman, C. A. 2002, ApJ, 577, 197
Wada, K., & Norman C. A. 2001, ApJ, 547, 172
Warren, B. E., Jerjen, H., & Koribalski, B. S. 2004, AJ, 128, 1152
Warren, M. S., Quinn, P. J., Salmon, J. K., & Zurek, W. H. 1992, ApJ, 399, 405
Waugh, M., et al. 2002, MNRAS, 337, 641
Yan, M., Sadeghpour, H. R., & Dalgarno, A. 1998, ApJ, 496, 1044
Young, L. M., van Zee, L., Lo, K. Y., Dohm-Palmer, R. C., & Beierle, M. E. 2003, ApJ, 592, 111
Zwaan, M. A. 2001, MNRAS, 325, 1142